Advancing Ligand Binding Affinity Prediction with Cartesian Tensor-Based Deep Learning

Jie Yu,[◦,1,2,3] Xia Sheng,[◦,1,4] Zhehuan Fan,[◦,1,4] Zhaokun Wang,[◦,1,4] Duanhua Cao,[◦,1,5] Yongxin Hao,[1,6] Yingying Zhang,[1,6] JingXin Rao,[1,4] Panpan Shao,[1,3,7] Mingan Chen,[1,3,8] Kaixian Chen,[1,4] Dan Teng,[1,4] Xiaomin Luo,[1,4] Sulin Zhang[*,1,4], Mingyue Zheng[*,1,2,4]

[1]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

[2]School of Information Science and Technology, Shanghai Tech University, Shanghai, 201210, China

[3]Lingang Laboratory, Shanghai 200031, China

[4]University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

[5]Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

[6]Division of Life Science and Medicine, University of Science and Technology of China, Hefei, 230026, Anhui, China.

[7]School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing 210023, China

[8]School of Physical Science and Technology, Shanghai Tech University, Shanghai, 201210, China

Corresponding Author

[*](Mingyue Zheng) E-mail: myzheng@simm.ac.cn

[*](Sulin Zhang) E-mail: slzhang@simm.ac.cn

Author Contributions

[◦]J.Y., X.S., Z.H.F., Z.K.W., and D.H.C. contributed equally to this work. M.Y.Z. and J.Y. designed the research study. J.Y., X.S. and Z.H.F. developed the method and wrote the code. All authors contributed to the analysis of the results. J.Y., M.Y.Z. and X.S. wrote the paper. All authors read and approved the manuscript.

Note

The authors declare no competing financial interest.

0. Abstract

We present PBCNet2.0, a cartesian tensor-based Siamese Neural Network for protein-ligand relative binding affinity prediction. Trained on 8.6 million complex pairs, PBCNet2.0 achieves zero-shot performance comparable to computationally intensive physics-based simulations. Our prioritization experiments show that PBCNet2.0 speeds up binding affinity optimization by 718% while reducing resource use by 41%. Through extensive retrospective experiments, we demonstrate that PBCNet2.0 intrinsically comprehends protein-ligand interactions, showing high sensitivity to intermolecular interactions and exceptional perception of spatial geometric information. Strikingly, PBCNet2.0 exhibits an emergent capability to predict affinity changes induced by binding residue variations, highlighting its potential for identifying resistance mutation. We prospectively validated these capabilities on ENPP1 and ALDH1B1 targets, where PBCNet2.0 successfully identified affinity shifts arising from subtle molecular interactions and conformational differences, and pinpointed critical binding residues with an 83% hit rate. This combination of computational efficiency, spatial geometric perception of binding site, and generalizable affinity prediction establishes PBCNet2.0 as a transformative tool for developing pharmacological probes for all human proteins.

1. Introduction

The human genome harbors a vast number of proteins that remain underexplored despite their potential therapeutic value[1,2]. Many of these proteins are poorly understood and have not been adequately studied, yet they hold significant promise for drug development. A key factor contributing to this underutilization is the lack of high-quality, well-characterized molecular probes[1]. Recognizing this gap, the global initiative Target 2035 has been launched with the ambitious goal of identifying pharmacological probes for all human proteins[3,4].

Within this context, protein-ligand binding affinity prediction[5–7] plays crucial roles in accelerating the development of molecular probes. While physics-based methods such as free energy perturbation (FEP, e.g., Schrödinger's FEP+[8]) have achieved chemical accuracy (root-mean-square error $\sim 1.1\ \mathrm{kcal} \cdot \mathrm{mol}^{-1}$) in predicting relative binding affinities, their substantial computational cost, complexity in system setup, and dependency on expert intervention limit their broad applicability[9]. Consequently, artificial intelligence (AI), especially deep learning, presents a compelling opportunity for high-throughput and accurate affinity predictions by learning directly from structural data.

AI models explicitly incorporating three dimensional (3D) structural information have shown considerable promise[5,10] to this end. For instance, in 2024, Cao et al. proposed EquiScore—an equivariant graph neural network integrating key physical priors of ligand-protein interactions—which achieved robust performance in virtual screening[5]. Concurrently, Moon et al. introduced the PIGNet2.0 model, a physics-informed graph neural network that explicitly models various types of protein-ligand interactions to predict binding free energies, demonstrating efficacy in virtual screening and lead optimization[11]. Our previous work introduced the Pairwise Binding Comparison Network (PBCNet), a Siamese Neural Network (SNN) architecture enhanced by a physics-informed graph attention mechanism, specifically designed to rank binding affinities among congeneric ligands—a pivotal task in binding affinity optimization[6]. PBCNet demonstrated substantial advantages over existing high-throughput methods and, when fine-tuned with limited data, achieved performance comparable to the industry gold standard, Schrödinger's FEP+. However, despite these successes, a notable performance gap (in zero-shot settings) still persists between PBCNet and

Schrödinger's FEP+. We identified two key limitations that contribute to this disparity. First, PBCNet's predictive power is constrained by its training dataset. Although carefully curated, this dataset comprises only ~20,000 unique small molecules and 960 protein targets, covering a relatively narrow chemical and biological space. Recent insights from scaling laws[12–14]—empirically validated across diverse scientific domains, including protein structure prediction (e.g., AlphaFold3[15]) and large language models (e.g., GPT-4[16])—demonstrate that model performance scales positively with both data volume and model complexity. The limited scale of PBCNet's training set likely restricts its generalization capabilities across broader chemotypes and target families. Second, PBCNet's geometric encoding strategies introduce unintended biases. The model incorporates atom pairwise statistical potentials[17] (APSP) and angular encoding modules to refine attention weights during message passing. While these modules enforce physical priors (e.g., distance-dependent interactions, directional bonding), they rely on precomputed and predefined geometric features derived from known atom pairs and fixed angle bins. This approach not only restricts the model's capacity to learn context-dependent interactions from raw structural data but also exhibits limitations when encountering unseen atom types in the predefined APSP set[18]. Consequently, overly prescriptive geometric assumptions may hinder the discovery of noncanonical binding motifs, a limitation particularly relevant for systems with unconventional ligand chemistries.

To address these limitations and advance the field of protein-ligand binding affinity prediction, we present PBCNet2.0, a next-generation model leveraging equivariant graph neural networks (E-GNNs)[19]. Our approach is guided by two key insights from recent advances in artificial intelligence. First, inspired by scaling laws, we expanded our training dataset to encompass 8.6 million protein-ligand complex pairs from BindingDB[20], representing a 14-fold increase over PBCNet's original dataset. This expanded dataset covers a broader chemical and biological landscape, with 0.28 million unique small molecules and 1,122 distinct protein targets, enabling more comprehensive learning of diverse interaction patterns. Second, we fundamentally redesigned our model architecture by implementing a Cartesian-based E-GNN[28] framework for message passing. This novel approach naturally encodes three-dimensional structural relationships through equivariant transformations, moving beyond PBCNet's reliance on predefined potential functions to learn geometric constraints directly from data. This paper presents our investigation into these innovations through several key aspects:

First, we evaluate PBCNet2.0's performance in lead optimization tasks, comparing it with existing computational methods including physics-based approaches and AI models. We then analyze the model's ability to capture and interpret protein-ligand interactions through detailed ablation studies and geometric perturbation experiments. The discussion extends to an unexpected emergent capability: the prediction of binding affinity changes induced by pocket mutations, which has notable implications for understanding drug resistance[21–24]. Finally, we validate our findings through prospective experiments on two drug targets, ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1)[25] and aldehyde dehydrogenase 1B1 (ALDH1B1)[26], demonstrating the practical utility of our approach in real-world molecular probe development and lead optimization scenarios.

2. Results

2.1 PBCNet2.0: a structure-based $\Delta\mathrm{pAct}$ predictor

PBCNet2.0 has been structured to extract representations from a pair of pocket-ligand structures via an equivariant message passing to predict the binding affinity change. In this section, we delineate the training details, model framework, and the application scenarios of PBCNet2.0. In the following text, $\Delta\mathrm{pAct}$ is used as a general symbol representing negative log-converted relative binding affinity when no specific type of binding affinity (IC_{50}, EC_{50}, K_{i}, and K_{d}) is defined.

The recent studies on scaling law indicate a positive correlation between the model performance and the scale of training data as well as the complexity of the model, providing a principle for AI system performance optimization. Therefore, we constructed a large-scale training set encompassing 8.6 million protein-ligand complex pairs (Fig. 1a), representing a 14-fold expansion compared to the PBCNet's training dataset (0.6 million). The ***Training set*** section provides a more detailed description of the training data processing procedure.

The PBCNet2.0 framework, as visualized in Figure 1b, comprises three components: (1) the message-passing module, (2) the readout module, and (3) the prediction module. The message-passing module is designed to transmit information across the pocket-ligand interaction graphs and generate node-level representations. The major difference between PBCNet2.0 and PBCNet lies in this module. Specifically, an equivariant graph neural network[19,27] with Cartesian tensors[28] is incorporated in PBCNet2.0 to endow the model with the capability to fully consider distance and angular geometric information, thereby capturing intricate protein-ligand interaction details and molecular conformation information. In practical, each hidden channel of a node's representation is expressed as a Cartesian rank-2 tensor, corresponding to a 3×3 matrix. This matrix can further be decomposed into three independent 3×3 matrices with 1, 3, and 5 degree of freedom (Equation 1 in ***Methods*** section), representing the scalar, vector, and tensor features, respectively. These features are updated and combined in parallel using straightforward 3×3 matrix products, which maintains the model equivariance in a cost-effective way. Through the message-passing step, the representations of protein pocket and ligand atoms are fully interacted and updated. Moving forward, pocket atoms are removed from the interaction graphs, and ligand atoms are retained. The readout

module focuses on deriving the graph- and pair-level representations. Graph-level representations, $\boldsymbol{x}_{ref}$ and $\boldsymbol{x}_{query}$, are obtained by averaging the invariant node representations of the reference and query ligands, respectively. Then, the pair-level representation is got by $\Delta\boldsymbol{x} = \boldsymbol{x}_{ref} - \boldsymbol{x}_{query}$, differing from PBCNet, where we concatenated $\boldsymbol{x}_{ref}$, $\boldsymbol{x}_{query}$, and $\Delta\boldsymbol{x}$ to construct the pair-level representation. Our empirical observations reveal that employing $\Delta\boldsymbol{x}$ as the pair-level representation allows the model to devote more attention to protein-ligand interactions. This is important because if a model solely focuses on ligand molecular structures, $\boldsymbol{x}_{ref}$ and $\boldsymbol{x}_{query}$ would be similar as they are structural analogues, making $\Delta\boldsymbol{x}$ close to a null vector. This scenario would create an unsolvable optimization problem. Instead, employing $\Delta\boldsymbol{x}$ as the pair-level representation helps prevent the model from simply memorizing ligand features. In the prediction module, this pair-level representation feeds into a three-layer feedforward neural networks to predict the relative binding affinity, $\Delta\hat{y}$. To achieve anti-symmetry $f(\Delta\boldsymbol{x}) = -f(-\Delta\boldsymbol{x})$ during training, we also input $-\Delta\boldsymbol{x} = \boldsymbol{x}_{query} - \boldsymbol{x}_{ref}$ into the fully connected layer and calculate the combined loss, implementing a computationally efficient sign-flip augmentation strategy. The ***Methods*** section provides a more detailed description of the modeling process.

For inference, PBCNet2.0 only needs the binding poses of two structurally similar molecules bound to the same protein to predict their relative binding affinity (Fig. 1c). This capability allows the model to rank a set of compounds based on their predicted binding affinities, providing valuable guidance for lead optimization and probe development. Beyond its primary design, PBCNet2.0 has demonstrated an unexpected capability: when given structures where a same ligand molecule bound to protein pockets containing specific residue mutations, the model can predict changes in binding affinity caused by these mutations (Fig. 1c). We will explore this capability in detail in subsequent sections.

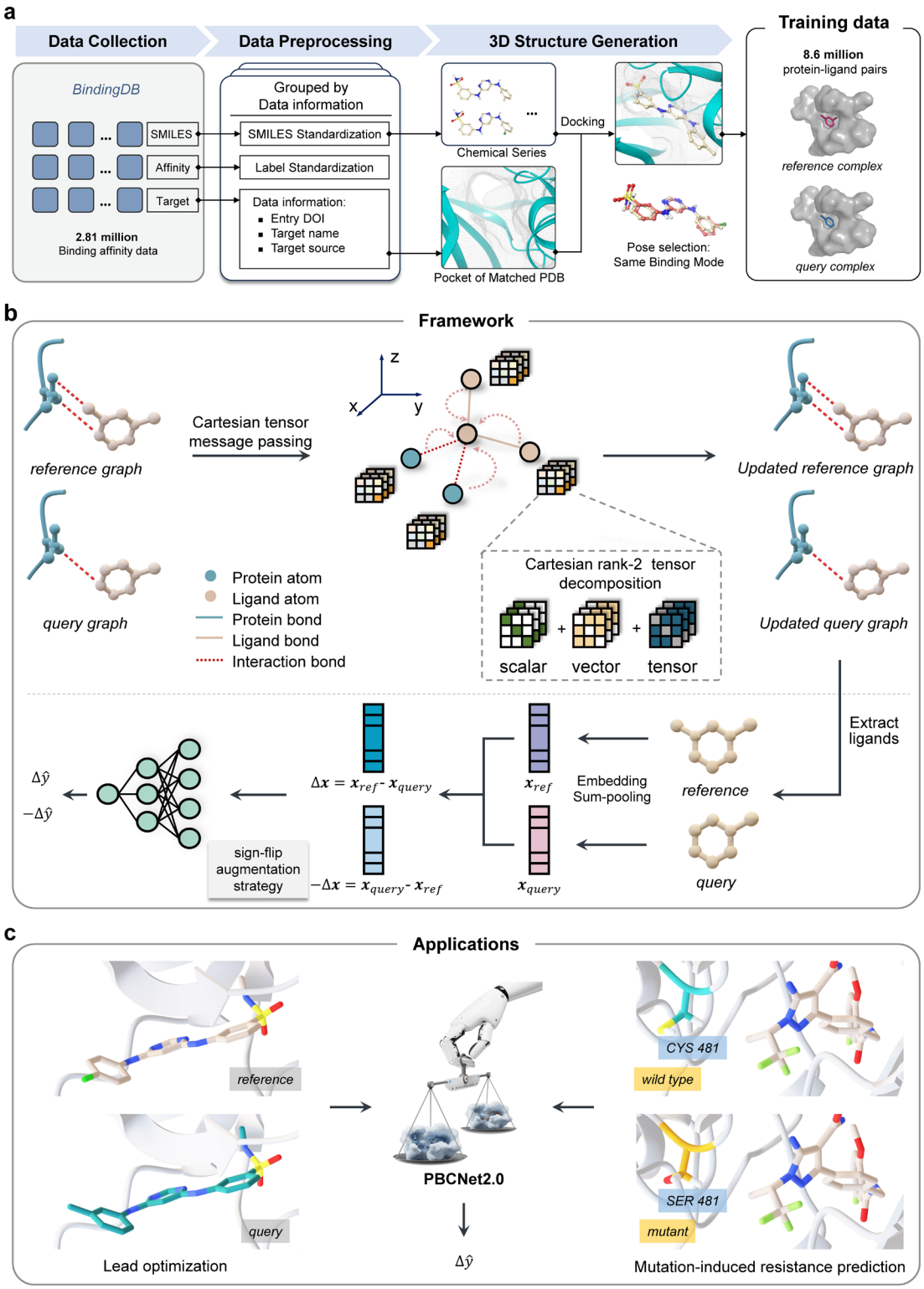

Figure 1 **a. Processing of training data for PBCNet2.0.** The raw data was collected from BindingDB (2023.12 version), and input poses were generated through Glide docking. PBCNet2.0 is trained on 8.6 million protein-ligand pairs. The training pairs share the same protein pocket but differ in ligand structures. **b. The framework of PBCNet2.0.** The complexes are processed through a Cartesian rank-2 Tensor-based message-passing module, designed to facilitate mutual information exchange between the ligands and the protein pocket. This module yields node-level representations. In this process,

each hidden dimension of each node-level representation is represented as a 3×3 matrix, which can be decomposed into three matrices with degrees of freedom 1, 3, and 5, respectively, using the irreducible representation of tensors. This decomposition enables efficient equivariant information propagation through matrix multiplication. In the Readout module, molecular representations (graph-level) are obtained by averaging the scalar node-level representations of the ligand atoms. The difference between the representations of the reference and query molecules ($\boldsymbol{x}^{ref}$ and $\boldsymbol{x}^{que}$) serves as the final pair-wise representation ($\Delta\boldsymbol{x}$), which is then fed into the Prediction module to output the predicted relative binding affinity $\Delta\hat{y}$. A computationally efficient sign-flip augmentation strategy is also implemented here. **c. The application of PBCNet2.0.** During inference, the PBCNet2.0 framework requires only the binding poses of two structurally similar small molecules bound to the same protein to predict their relative binding affinity. Additionally, when given structures with identical ligand molecules bound to protein pockets containing specific residue mutations, the model can predict the resulting changes in binding affinity.

2.2 PBCNet2.0 Achieves Substantial Improvement in Lead Optimization Tasks with Perception of Protein - Ligand Interaction Information

Following our previous work on PBCNet[6], we first systematically evaluated the performance of PBCNet2.0 in lead optimization tasks through three experimental paradigms: (1) **Zero-shot prediction**, which assess guidance capability during early-stage optimization when no structure-activity relationship (SAR) data is available; (2) **Few-shot prediction**, which evaluates performance in mid-to-late-stage optimization where a modest amount of SAR data enables model fine-tuning; and (3) **Prioritization experiment**, which validates the model's ability to identify high-activity compounds in real-world lead optimization scenarios. All metrics and methodologies matched our previous work, as summarized them in the Supplementary Section 1.

The zero-shot prediction results, presented in Figure 2a and Supplementary Figure 1, show the average performance of various methods across 16 chemical series referred to as the FEP test set (combining FEP1 and FEP2 sets). We assessed ranking performance using Pearson's correlation coefficient (R) and Spearman's rank correlation coefficient (ρ) (refer to ***Evaluation metrics*** section). We compared against eleven baseline methods (detailed in ***Baseline models*** section): three traditional structure-based approaches (Glide SP[29], MM-GB/SA[30], and Schrödinger's FEP+[8]), three sequence-based AI models (PSICHIC[7], BIND[31], and PLAPT[32]), and five structure-based AI models (PIGNet2[11], RTMScore[33], GenScore[34], OnionNet-2[35], and PBCNet[6]). The detailed prediction results of each method on each chemical series is summarized in the Supplementary Data 1.

Benchmarking results establish PBCNet2.0's superior predictive capability, achieving $\rho =$

0.67 and $R = 0.66$ on the FEP set, an 18% enhancement over its predecessor PBCNet ($\rho = 0.57$, $R = 0.56$) and significantly outperforming ($p \leq 0.05$) all other AI comparators. Notably, PBCNet2.0 approaches the performance of the gold-standard Schrödinger FEP+ ($\rho = 0.70$ and $R = 0.70$, $\Delta\rho = 0.03$ and $\Delta R = 0.04$), with statistically indistinguishable ranking accuracy ($p = 0.39 > 0.05$). This result makes PBCNet2.0 the first AI approach to achieve parity with industrial-strength FEP simulations in critical evaluation metrics.

To visualize PBCNet2.0's improvements, we compared predicted $\Delta pAct$ values from both PBCNet2.0 and PBCNet for test systems from the FEP set (Supplementary Fig. 2). PBCNet2.0's predictions exhibit a stronger correlation with experimental values and cover a broader value range. For example, in the pfkfb3 system, PBCNet2.0's predictions span [-2, 2], closely matching the experimental value range, while PBCNet's predictions remain limited to [-1, 1].

Figure 2b illustrates the few-shot prediction results, with detailed data summarized in Supplementary Data 2. The x-axis indicates the number of compounds used for fine-tuning, while the y-axis shows the ranking capability metrics. Cyan dots represent PBCNet's performance, red dots show PBCNet2.0's performance, and a blue dashed line indicates the Schrödinger's FEP+ baseline. Fine-tuning proves highly effective for PBCNet2.0: its performance improves steadily as more fine-tuning ligands are added, consistently outperforming both PBCNet and Schrödinger's FEP+. This data-efficient refinement capability makes PBCNet2.0 valuable when limited SAR data is available for a target of interest.

The real-world utility of PBCNet2.0 is demonstrated through retrospective prioritization experiments across 8 pharmaceutically relevant targets (Fig. 2c). The detailed data and evaluation process are summarized in Supplementary Data 3 and the Supplementary Section 1. Benchmarking results show that PBCNet2.0 outperforms both PBCNet and MM-GB/SA on all metrics. Specifically, PBCNet2.0 achieved an **advantage order** of 13.57, while compared to PBCNet's 11.18 and MM-GB/SA's 5.89. The **advantage ratio**, which measures theoretical resource savings in model-guided lead optimization, shows PBCNet2.0 reduced resource investment by 40.51%, surpassing PBCNet (30.18%) and MM-GB/SA (14.93%). The **efficiency improvement ratio**, reflecting enhanced efficiency in compound optimization, shows PBCNet2.0 accelerated lead-optimization projects by 717.81%, surpassing both PBCNet (473.00%) and MM-GB/SA (281.31%). Statistical analysis confirms that the performance of PBCNet2.0 in the prioritization experiment is significantly ($p \leq$

0.05) superior to MM-GB/SA.

The above results indicate that PBCNet2.0 has achieved substantial improvement in the task of predicting relative affinity, demonstrating its value and effectiveness in probe development and compound optimization projects. In the following, we will explore its prediction basis.

Mastropietro et al. reported that many machine learning-based scoring models for protein-ligand binding suffer from a bias where predictions are predominantly driven by ligand memorization rather than understanding protein-ligand interactions[36]. To rigorously test whether PBCNet2.0 overcomes this bias. First, we conducted a strategic ablation study using the FEP dataset. We systematically removed all protein-ligand interaction edges from the original molecular graphs (**Complete graphs**), generating the **Incomplete graphs** (Supplementary Fig. 3a). We then monitored the performance dynamics of PBCNet2.0 on both graphs over the course of training (Supplementary Fig. 3b). During the initial phase of training (approximately up to 12.5K steps), both curves exhibit a similar upward trend, suggesting that the model initially mainly relies on molecular structure information for predictions. However, as training progresses beyond 12.5K steps, a marked divergence emerges. The model performance on the **Incomplete graphs** declines steadily, whereas that on the **Complete graphs** continues to improve. This phenomenon indicates that protein-ligand interaction information has emerged as the primary determinant of PBCNet2.0's predictions, implying that PBCNet2.0 may progressively develop an understanding of the interaction information.

To further verify that PBCNet2.0 genuinely learns protein-ligand interaction patterns rather than memorizing molecular features or structural variations, the methods were benchmarked by SAR-Diff set which technically designed to decouple structural modifications from binding affinity changes (refer to ***SAR-Diff held-out test set*** section). Specifically, the SAR-Diff set comprises eight pairs of chemical series (16 series total) where each pair share identical R-group modification pathways but exhibit distinct structure-activity relationships[37–39]. For clarity, Figure 2d presents an example of the partial modification process for the first series pair of SAR-Diff set, and Supplementary Figure 3c illustrates the lack of correlation in activity between the two series. The critical design forces models to recognize interaction-driven activity changes rather than relying on structural memorization. Successful prediction on this benchmark therefore requires models to explicitly capture how modified chemical groups interact with protein binding pockets.

Figure 2e presents the comprehensive evaluation results (Supplementary Data 5), with heatmap organized to show performance of different methods (y-axis) across series pairs (x-axis). Spearman correlation coefficients (ρ) quantify prediction quality. The two numbers in each heatmap cell stand for the ρ of a model on a given series pair. Each column indicates the ρ of different models on the same chemical series, as illustrated by the arrows between Figure2d and Figure2e. When the $\rho \geq 0.4$, the corresponding model is considered to have predictive power on the series. The darkest green cells denote that models successfully predicted both series within a pair, lighter green indicate models only own predictive ability for one series, and pale cells represent models have no predictive ability for either series. Remarkably, PBCNet2.0 demonstrated solid predictive capability across all series pairs, a feat unmatched by other methods, with the second-best RTMScore succeeding on only three pairs. Crucially, PBCNet2.0 and PBCNet uniquely avoided cases where both series in a pair were incorrectly predicted. Figure 3f provides a detailed comparison of average model performance on these 16 series (ranking performance measured by ρ and prediction accuracy measured by root-mean-square error, r.m.s.e.). Our model outperformed all competitors in both measures, with PBCNet ranking second. These results highlight that PBCNet2.0 successfully masters the interactions between modified groups and protein pockets, confirming that protein-ligand interaction patterns serve as the primary determinant of the model's prediction accuracy. Moreover, comparing Figure 2f with Figure 2a reveals that most models (except RTMScore) performed worse on the SAR-Diff set than on the FEP set, indicating that the SAR-Diff set is a more challenging external test set. This underscores the robustness of PBCNet2.0, which remains effective even under more demanding predictive scenarios.

These results establish PBCNet2.0 as a robust and efficient tool for lead optimization and molecular probe development, demonstrating significant improvements over PBCNet in ranking accuracy, prediction capability, adaptability, and computational efficiency. Furthermore, the rationale behind PBCNet2.0's predictions is elucidated, indicating its effective assimilation of protein-ligand interaction information for precise relative binding affinity prediction.

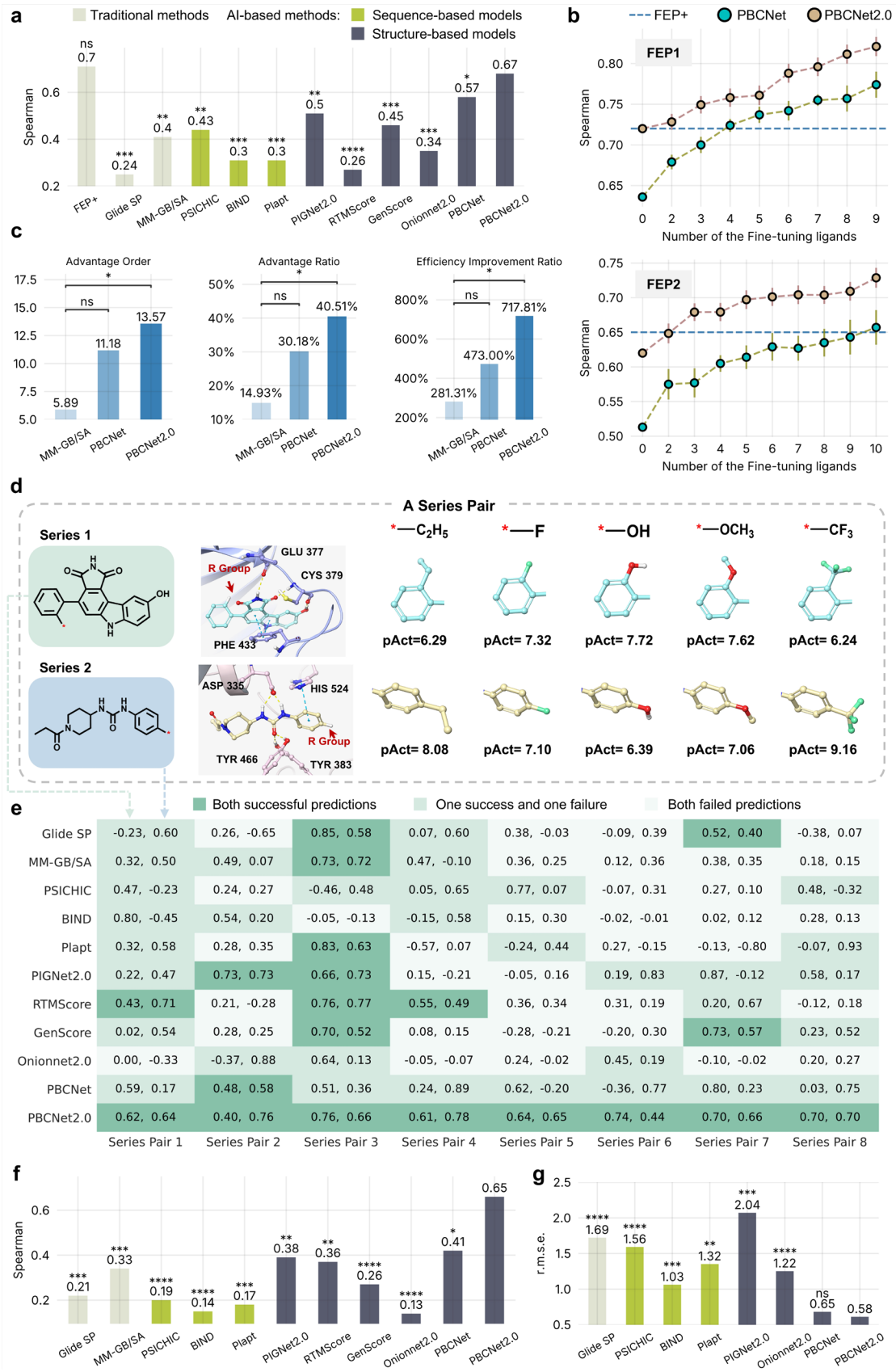

Figure 2 **a. Direct prediction results.** The x-axis represents various binding affinity prediction methods, while the y-axis denotes their average of their ranking capability (Spearman) on the FEP set. A two-sided Wilcoxon signed-rank test compared each model against PBCNet2.0, with significant levels:

*p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001, and ns (not significant) for p > 0.05. Different model types are shown in different colors. **b. Finetuned prediction results.** The x-axis represents the number of compounds used for fine-tuning, while the y-axis shows the ranking capability metrics. Cyan dots represent PBCNet performance, pink dots show PBCNet2.0 performance, and a blue dashed line indicates the Schrödinger's FEP+ baseline. **c. Selection experiment results.** The x-axis represents different methods, and the y-axis shows performance across eight systems. Three panels display distinct evaluation metrics. A one-sided Wilcoxon signed-rank test evaluated whether PBCNet2.0 and PBCNet significantly outperformed MM-GB/SA. **d. Illustration of the partial modification process for a chemical series pair from the SAR-Diff Set.** A schematic representation of structural changes within a chemical series pair. **e. The predictive performance of the models on the SAR-Diff set.** The x-axis lists 8 series pairs, and the y-axis represents 11 prediction methods. Each cell in the heatmap displays the Spearman correlation coefficient for a model on a given series pair. Dark green cells indicate models successfully predicted both series within a pair, light green cells show models with predictive ability for only one series, and pale cells represent models without predictive ability for either series. **f. Average performance of methods on SAR-Diff set.** The x-axis represents various binding affinity prediction methods, while the y-axis denotes the average of their ranking capability (Spearman) and prediction accuracy (r.m.s.e with ΔpIC_{50}) on the SAR-Diff set. The one-sided Wilcoxon signed-rank test was used to analyze whether the ranking performance of PBCNet2.0 was significantly superior to others and whether its prediction error was significantly lower. *p ⩽ 0.05, **p ⩽ 0.01, ***p ⩽ 0.001, ****p ⩽ 0.0001, and ns (not significant) for p > 0.5. Different types of models are distinguished using bars of different colors. For clarity, MM-GB/SA, RTMScore, and GenScore are not shown in the second figure due to large errors, see Supplementary Figure 4 for the full figure.

2.3 PBCNet2.0 Captures Subtle Intermolecular Interactions with Strict Geometric Constraints

Given that protein-ligand interaction information is a key factor for PBCNet2.0's predictions, we aimed to further investigate whether PBCNet2.0 can effectively identify critical protein-ligand interactions, and whether it has fully utilized the geometric information. In our previous work on PBCNet, the analysis primarily focused on the identification of hydrogen bond (H-bond) interactions[40]. Building on this, we first examined PBCNet2.0's ability to recognize H-bonds and revisited the two cases previously analyzed in the PBCNet study. The results, presented in Supplementary Figure 5, confirm that PBCNet2.0 retains the ability to effectively identify H-bond interactions, consistent with the findings from PBCNet.

In addition to H-bonds, we extended our analysis to fluorine orthogonal multipolar interactions, which is a crucial yet often overlooked protein-ligand interaction[41,42]. Many empirical expressions for intermolecular interactions, such as some general-purpose force fields, may fail to adequately account for this interaction[41,43]. Here, we aim to explore whether PBCNet2.0, a data-driven approach, can effectively learn and discern this interaction from an extensive dataset of structural information. As showed in Figure 3a, the strong electronegativity of the fluorine atom makes it a nucleophilic center, while the electron-withdrawing effect of the carbonyl oxygen renders the carbonyl carbon electrophilic. This complementary electron distribution facilitates orthogonal multipolar interactions between fluorine and the carbonyl carbon. This interaction is also featured by strict geometric constraints: the distance between the fluorine and carbon atoms should be less than 3.8 Å, and the $\mathrm{F} \cdots \mathrm{C} = \mathrm{O}$ angle should approximate $90°$. The accurate identification of this interaction demands explicit modeling of 3D spatial arrangements and angular dependencies. To rigorously evaluate this capability, we established the F-Opt benchmark test set (see ***F-Opt test set*** section), containing 10 carefully curated compound pairs where fluorine substitution induces fluorine orthogonal multipolar interactions and leads to improved binding affinity (Fig. 3b, Supplementary Fig. 6). Then, we systematically evaluated the binding affinity prediction methods on this benchmark. Figure 3c (Supplementary Data 7) provides a comparison of the absolute prediction errors for each model on each compound pair (represented by their corresponding identifiers in Protein Data Bank (PDB IDs)), and darker colors indicate larger errors. Red numeric annotations

represent sign discordances between predicted and experimental ΔpAct values ($\text{Sign}(\text{Predicted}\ \Delta\text{pAct}) \neq \text{Sign}(\text{Experimental}\ \Delta\text{pAct})$). Remarkably, PBCNet2.0 achieved superior accuracy with the lowest mean absolute error ($\text{m.a.e.} = 0.33$), outperforming all comparators. PBCNet ranked second in precision ($\text{m.a.e.} = 0.52$). Both PBCNet2.0 and PBCNet uniquely demonstrated 100% sign concordance, whereas other models exhibited error rates of $20\%-50\%$, demonstrating the exceptional robustness of the PBCNet2.0 and PBCNet.

We further performed interpretability analyses on two systems from F-Opt benchmark (5G4O and 4E3N, details provided in ***Interpretability analysis*** section) to elucidate the recognition mechanism of fluorine orthogonal multipolar interactions. As visualized in Figure 3d and Supplementary Figure 7, these interactions are explicitly depicted with purple dashed lines. Protein atoms within 5Å of fluorine atoms (central atoms) were annotated with corresponding weight values (range: $0.00 - 1.00$), where boldface numerals highlight interaction weights of the carbon atoms involved in the interactions. Both systems demonstrated consistently high weight assignments to these carbon atoms (range: $0.98 - 1.00$), confirming PBCNet2.0's capability to accurately identify geometrically constrained fluorine interactions. To further validate PBCNet2.0's angular perceptivity we performed a perturbation experiment on the 5G4O system. Specifically, we moved the central fluorine atom using Maestro to disrupt two key interaction angles: one reduced one angle from $79.2°$ to $52.9°$ and another increased from $95.6°$ to $133.7°$ (both $52.9°$ and $133.7°$ being unreasonable angles). Notably, during the moving process, the ligand's pose was automatically adjusted to maintain its structural rationality. As shown in Figure 3e, the weight values assigned to the interacting carbon atoms obviously decreased (from 0.98 to 0.5 and from 1.00 to 0.56), demonstrating the angular constraint recognition of the equivariant framework-equipped PBCNet2.0. This directional sensitivity, inherent to the equivariant framework design, enables precise geometric discrimination critical for modeling subtle intermolecular interactions.

While PBCNet demonstrates commendable predictive performance on the F-Opt test set, the rationale behind its predictions lacks support from interpretability analysis (Supplementary Figure 8a). Examination of PBCNet's built-in statistical atom-pair potential function, an empirical expression designed to capture distance-based intermolecular interactions, reveals that this function assigns low interaction probabilities to fluorine-carbonyl carbon atom pairs within reasonable distance range (Supplementary Figure 8c). This indicates that the empirical formulation fails to

account for fluorine orthogonal multipolar interactions. Moreover, we also did not observe PBCNet's ability to perceive angles (Supplementary Figure 8b). Coarse 3D modeling strategy prevents PBCNet from establishing a meaningful connection between predictions and physicochemical principles. In contrast, PBCNet2.0 discards empirically designed geometric functions and adopts a geometry-aware equivariant model framework. By integrating this architecture with large-scale training data, PBCNet2.0 achieves refined learning of intermolecular interaction information. This strategic improvement not only enables PBCNet2.0 to surpass other methods but also provides a valid paradigm for interaction information learning through data-driven approaches. The logical mapping between predictions and physicochemical principles makes PBCNet2.0 an even more compelling tool for probe development and lead optimization.

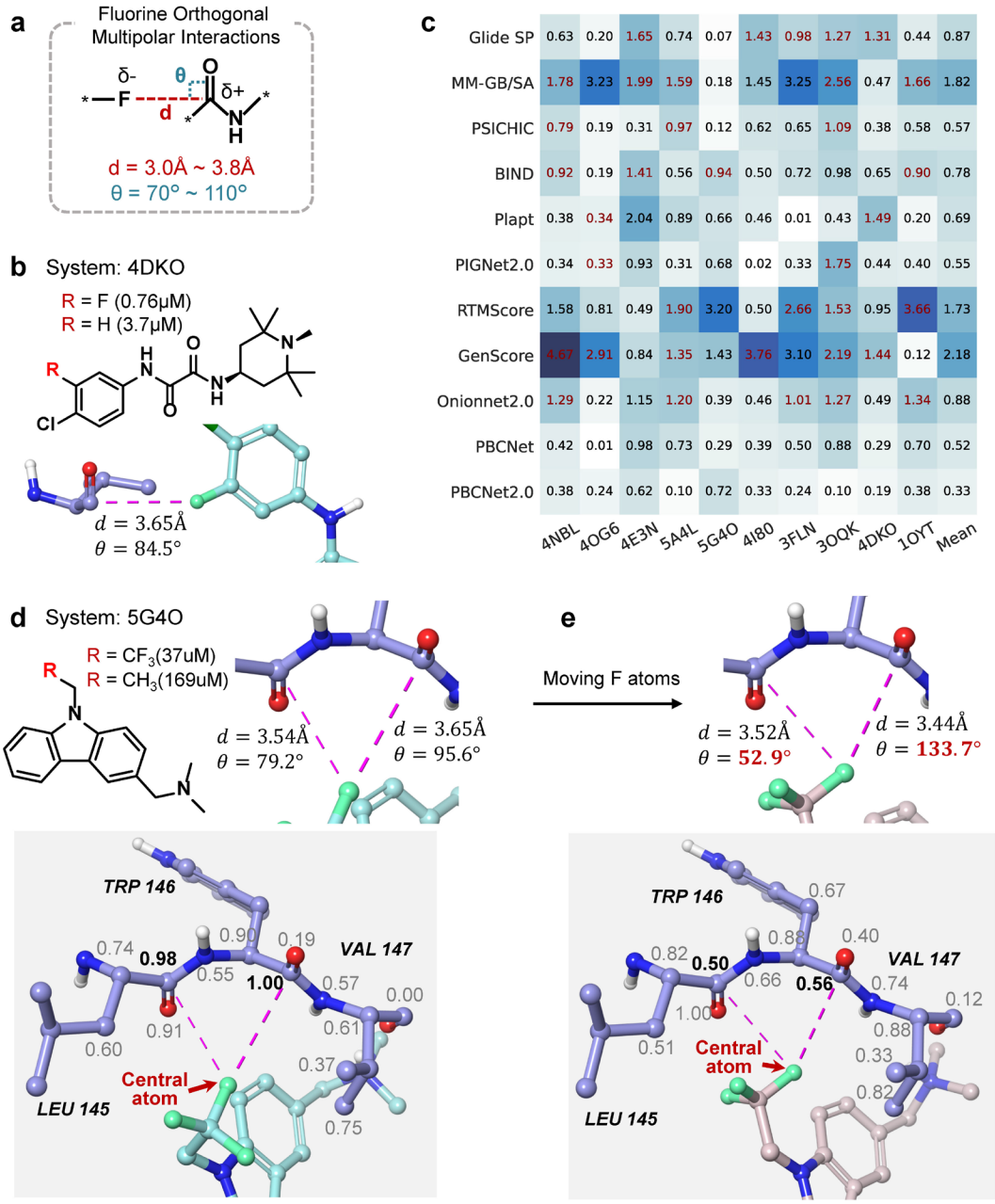

Figure 3 **a. Illustration of fluorine orthogonal multipolar interaction.** A schematic demonstrating its formation mechanism and geometric constraints. **b. Examples from the F-Opt Set.** Representative cases showcasing fluorine orthogonal multipolar interactions, including key distances and angles. **c. Absolute prediction errors of models on each compound pair:** The y-axis represents various methods, and the x-axis corresponds to compound pairs (denoted by PDB IDs). The numbers in the table indicate absolute errors, with red numbers highlighting incorrect predictions. **d. Interpretability analysis of the 5G4O system**. For a molecule with a trifluoromethyl R-group, two fluorine orthogonal multipolar interactions with the protein are identified. The binding mode and geometric constraints are illustrated, with fluorine orthogonal multipolar interactions represented by purple dashed lines. Protein atoms within 5 Å of the central fluorine atoms are annotated with weights, and bold numbers denote the weights of carbon atoms involved in the interactions. **e. Disruption of geometric constraints in fluorine orthogonal multipolar interaction**. The angles of fluorine orthogonal multipolar interactions were

disrupted, changing one angle from 79.2° to 52.9° and another from 95.6° to 133.7° (both being unreasonable angles). Consequently, the weight distribution of surrounding atoms was altered.

2.4 PBCNet2.0 Demonstrates Emergent Abilities for Identifying Drug-Resistant Mutations

Residue mutations can directly impact protein structure, stability, function, and molecular interactions[22–24] (Fig. 4a). Residue mutations may alter binding site properties and drug binding affinities, potentially leading to drug resistance[21]. For instance, in non-small cell lung cancer (NSCLC) treated with first-generation EGFR inhibitors (e.g., Gefitinib[44]), the acquired T790M mutation in EGFR kinase domain emerges in $> 50\%$ of resistant cases within 9-14 months[45,46], sterically hindering drug binding and causing therapeutic failure. This rapid resistance crisis necessitated urgent development of next-generation inhibitors, exemplified by Osimertinib[47]—a third-generation EGFR inhibitor specifically engineered to overcome T790M-mediated resistance. Critically, the stronger the drug selection pressure, the more rapidly resistance develops because drug target has been proved to be very plastic[48]. Hence, an early prediction and identification of mutations that may potentially impact ligand binding is important for lead optimization and probe development. Although PBCNet2.0 was not explicitly trained on data involving amino acid mutations (the training set only contained data for different ligands binding to the same protein site), we were curious whether PBCNet2.0 could have automatically learned to predict mutation-induced binding affinity changes. This hypothesis was motivated by the shared nature of both tasks—modeling changes in protein-ligand interactions—and by PBCNet2.0's ability to capture subtle molecular interactions as demonstrated above. To rigorously test this hypothesis, we constructed a Mutation Benchmark Set (see ***Mutation test set*** section) containing 8 pharmaceutically relevant targets with clinically observed binding pocket mutations. As shown in Fig. 4b (Supplementary Data 6), PBCNet2.0 demonstrated relative solid predictive performance, achieving an average Spearman correlation of 0.53 ($\rho = 0.53$). This is a stark contrast to scoring methods like Glide SP, BIND, PLAPT, RTMScore, and GenScore, which even exhibited negative correlations. A deeper target-level look into PBCNet2.0's predictions exposed varying performances across different targets. While PBCNet2.0 exhibited strong ranking performance for tACE[49] ($\rho = 0.76$) and HIV-1 PR[50]

($\rho = 0.75$) targets (Supplementary Figure 9a), it showed less satisfactory results for AKR1B1[51] ($\rho = 0.28$) and PfDFHR-TS[52] ($\rho = 0.26$) targets (Supplementary Figure 9b).

To further investigate the rationale behind PBCNet2.0's predictions, we conducted an interpretability analysis of **Lactobacillus casei Thymidylate Synthase** (LTS)[53] target. The N229C mutation abolished a key hydrogen bond between the ASN229 amide oxygen and ligand pyridine nitrogen (Fig. 4c). This mutation decreased the binding affinity ($\Delta pK_d = -0.62$)[53], closely matching PBCNet2.0's prediction ($\Delta pK_d = -0.79$). Additionally, we visualized the weights (details in ***Interpretability analysis*** section) assigned to neighboring atoms of the pyridine nitrogen during the model's message-passing phase (Fig. 4d). In Figure 4d, the hydrogen bond is depicted as green dashed lines, and protein atoms within 5 Å of the central nitrogen atom were annotated with weights, where boldface numeral indicates the weight of the oxygen atom involved in the interaction. We found that the involved oxygen atom was assigned a relative high weight (0.99) in wild-type simulations, with complete signal loss post-mutation. This phenomenon demonstrates the validity of PBCNet2.0's predictions. Through its acquired ability to capture protein-ligand interactions, the model has successfully generalized to mutation prediction tasks.

Thereafter, we retrospectively analyzed three cases of mutant-selective inhibitors that are either in clinical trials or already in the market: (1) Olutasidenib, approved by the Food and Drug Administration (FDA) in December 2022, targets the mutant **Isocitrate Dehydrogenase 1** (IDH1) (with ＞1000-fold selectivity) to treat acute myeloid leukemia (AML) [54]; (2) MRTX1133, a potent, noncovalent, and selective **Kirsten Rats Arcomaviral Oncogene Homolog** (KRAS) G12D inhibitor, entered clinical trials in March 2023 to address manifold tumors harboring $KRAS^{G12D}$ mutations[55]; and (3) Pirtobrutinib, the first-in-class noncovalent **Bruton's tyrosine kinase** (BTK) inhibitor approved by the FDA[56]. The structures, binding modes, and experimental and predicted relative binding affinities of these inhibitors are depicted in Figure 4e. As shown in Figure 4e, PBCNet2.0 accurately predicted relative binding affinities across these structurally diverse inhibitors, despite never being explicitly trained on mutation-specific data. This emergent capability[57,58] likely stems from the model's generalized understanding of protein-ligand recognition patterns, developed through training on a large-scale dataset containing 8.6 million entries.

The above experimental results demonstrate that PBCNet2.0 can identify critical mutations,

positioning it as a versatile tool for both probe development and molecular optimization.

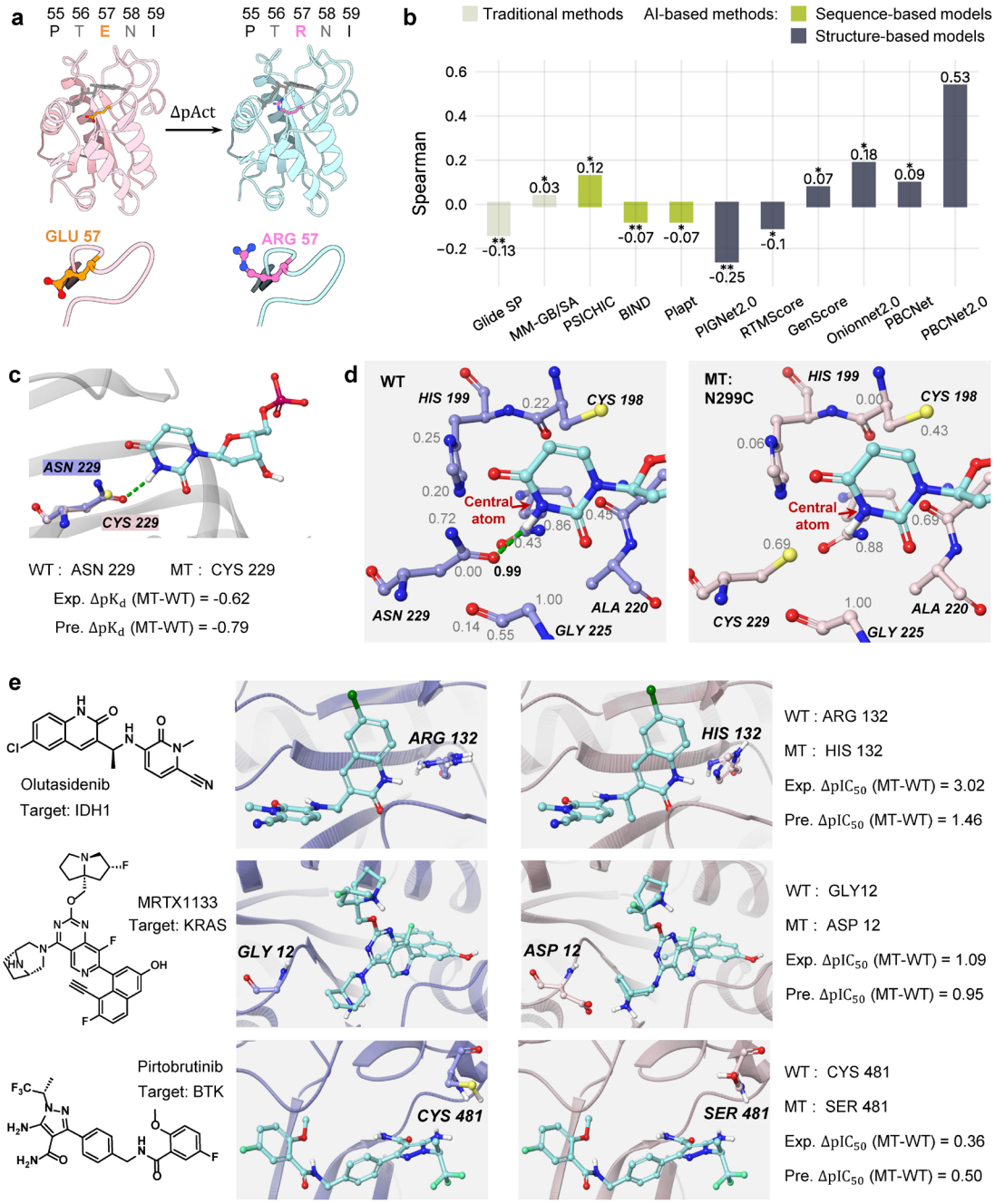

Figure 4 **a. Illustration of amino acid mutation.** A schematic representation of a mutation at position 113 in the protein pocket, where THR is mutated to TYR. This alters the protein pocket environment, affecting small molecule binding and resulting in changes to binding affinity. **b. Average performance of methods on the Mutation set.** The x-axis represents various binding affinity prediction methods, while the y-axis denotes the average of their ranking capability (Spearmen) on the Mutation set. The one-sided Wilcoxon signed-rank test was used to analyze whether the ranking ability of PBCNet2.0 was significantly superior to others. *$p \leq 0.05$ and **$p \leq 0.01$. Different types of models are distinguished using bars of different colors. **c. Binding Mode of the Mutation Case.** The mutation of ASN at position 229 to CYS causes the protein to lose a hydrogen bond with the small molecule, leading to a decrease in

pK_d by 0.62. PBCNet2.0 predicts a decrease of 0.79. **d Interpretability analysis.** The hydrogen bond is depicted as green dashed lines, and protein atoms within 5 Å of the central nitrogen atom were annotated with weights, where the bold number indicates the weights associated with the oxygen atom involved in the hydrogen bond. **e. Retrospective analysis of mutant-selective inhibitors.** Three mutant-selective inhibitors that are either in clinical trials or already in the market are shown: Olutasidenib (target: IDH1), MRTX1133 (target: KRAS), and Pirtobrutinib (target: BTK). For each compound, the figure shows chemical structures on the left, binding modes in wild-type and mutant proteins in the middle panels, and quantitative data on the right. The right section presents wild-type and mutant residues alongside experimental and predicted ΔpIC_{50} values for each protein-ligand pair.

2.5 Experimental Validation

Building upon our retrospective evaluation and analysis of PBCNet2.0, we operationalized it in a prospective experimental validation study to verify its capacity to capture subtle intermolecular interactions and predict resistance-associated mutations that govern ligand binding.

2.5.1 Experimental Validation of Subtle Molecular Interaction Capture in ENPP1 Inhibitors

In our previous work, we identified compound **A1** (Fig. 5a) as a promising ENPP1 inhibitor[25]. Docking-based binding mode analysis suggested it might form a fluorine orthogonal multipolar interaction with the target protein, a phenomenon briefly noted in our earlier publication. Here, we further investigate **A1** with detailed analysis.

First, we conducted an interpretability analysis on the fluorine of **A1** (Fig. 5b). Results showed a relative high weight (0.94) for the potential dipole interaction between this fluorine and the target protein's carbonyl oxygen. Then, based on **A1**, we designed an analog **A2** (Fig. 5a), replacing the fluorine with hydrogen to eliminate the potential dipole effect while minimizing other structural changes. Using PBCNet2.0, we predicted **A2**'s binding affinity for ENPP1 with **A1** as the reference. PBCNet2.0 predicted a substantial difference in binding affinity between **A1** and **A2**, specifically forecasting a 40-fold ($\Delta pAct = 1.61$) reduction in binding affinity upon fluorine-to-hydrogen substitution. This marked disparity in predicted binding affinities provides strong computational evidence supporting the existence of the fluorine orthogonal multipolar interaction.

To validate the accuracy of these computational results, we synthesized **A2** (Supplementary

Section 3) and conducted corresponding wet experiments (Supplementary Section 2). The protein thermal shift (PTS) assay revealed that both **A1** and **A2** increased the thermal stability of ENPP1 protein (Fig. 5c), suggesting direct binding of these compounds to ENPP1. However, **A1** exhibited a stronger thermostabilizing effect on ENPP1 (ΔT_m = 3.8°C) than **A2** (ΔT_m = 2.7°C). *In vitro* enzyme activity assays showed that **A1** inhibited ENPP1 enzymatic activity (IC_{50} = 13.3 nM) obviously more potently than **A2** (IC_{50} = 2284 nM) (Fig. 5d). These data suggested that **A1** likely possessed stronger binding affinity for ENPP1 protein. This was confirmed by subsequent surface plasmon resonance (SPR) studies (Fig. 5e-f), which showed K_D values of 35.3 nM and 1030 nM for **A1** and **A2** binding to ENPP1, respectively, indicating a substantially higher binding affinity of **A1**. The ΔpK_D measured by SPR assay was 1.46, closely aligning with PBCNet2.0's prediction of 1.61, with a minor error of just 0.15. These experimental observations collectively demonstrate the establishment of a robust fluorine-mediated orthogonal multipolar interaction between compound **A1** and ENPP1.

This prospective study provides compelling validation of PBCNet2.0's ability to accurately predict the impact of subtle intermolecular interactions on protein-ligand binding affinity. The strong correlation between computational predictions and experimental measurements supports the utility of PBCNet2.0 as a reliable tool for rational structure-based probe development and lead optimization.

2.5.2 Experimental Validation of Subtle Molecular Interaction Capture and Drug-Resistant Mutation Prediction in ALDH1B1 Inhibitors

Our preliminary study identified a racemic mixture exhibiting ALDH1B1 inhibitory activity. However, two unresolved questions persist regarding this compound: (1) Potential divergence in biological activity between stereoisomers, and (2) Identification of key amino acid residues that may lead to drug resistance. In this section, we attempt to solve these issues employing the PBCNet2.0.

To address the first question, we computationally analyze the compound's enantiomeric pair (**B1** and **B2**, Fig. 5h). Computational docking suggested that while **B1** and **B2** adopt largely similar binding modes, a critical difference emerged due to their distinct stereochemistry: the pyridine ring

of **B2** is unable to form the favorable $\pi - \pi$ stacking interaction[59,60] with PHE296 that is observed for **B1** (Fig. 5g). PBCNet2.0 predicted a notable difference in binding affinity, forecasting that **B1** would be approximately 10-fold more potent than **B2** (predicted $\Delta \mathrm{pAct} = 1.07$). Interpretability analysis corroborated this docking observation: the model explicitly highlighted the importance of the $\pi - \pi$ stacking interaction between the pyridine ring of **B1** and the PHE296 residue (Fig. 5i). To validate this computational prediction, wet experiments (Supplementary Section 2) were conducted. We synthesized **B1** and **B2** and determined their absolute configurations through X-ray diffraction analysis (Supplementary Section 4). PTS revealed that both **B1** and **B2** increased the thermal stability of ALDH1B1 protein (Fig. 5j). **B1** exhibited a stronger thermostabilizing effect on ALDH1B1 (ΔT_m = 8.8°C) than B2 (ΔT_m = 1.1°C). *In vitro* enzyme activity assays showed that **B1** inhibited ALDH1B1 enzymatic activity (IC_{50} = 20.0 nM) obviously more potently than **B2** (IC_{50} = 5574 nM) (Fig. 5k). These data suggested that **B1** likely possessed stronger binding affinity for ALDH1B1 protein. This was confirmed by subsequent SPR studies (Figure 5l-m), which showed K_D values of 15.2 nM and 2340 nM for **B1** and **B2** binding to ALDH1B1, respectively, indicating a substantially higher binding affinity of **B1**. The relative binding affinity measured by SPR assay was aligning with PBCNet2.0's prediction. While the precise fold-difference varied between prediction and experiment, PBCNet2.0 successfully issued a crucial warning regarding the stereochemistry-driven activity loss, demonstrating its capacity to capture vital protein-ligand interaction information and perceive subtle 3D geometric details.

Regarding the second question, we first conducted a pocket-directed *in silico* alanine scanning[61,62], computationally mutating each binding residue to alanine. Then, PBCNet2.0 was used to predict the binding affinity change for each mutation compared to the wild-type **B1**-ALDH1B1 complex. The prediction results have been showed in Figure 5n and Supplementary Figure 10. The six residues (PHE296, ASN457, PHE170, ILE458, LEU477, and GLU124), prioritized by PBCNet2.0, were selected and subsequently validated by the generation of point-mutated proteins and binding assays (Supplementary Section 2). PTS assays and SPR analyses revealed that alanine mutations at these sites (excluding GLU124) severely affected the binding of **B1** (Fig. 5o-t and Supplementary Figure 11). With a hit rate of ~83% (5/6), PBCNet2.0 demonstrated superior performance in identifying functionally critical residues, potentially implicated in drug resistance. Moreover, as a conventional method, MM-GB/SA only identified the PHE296 and PHE170 among

its prioritized residues (top 10 out of 43, Supplementary Fig. 7b). Other three critical residue (ASN457, ILE458, and LEU477) were not enriched by MM-GB/SA.

The employment of PBCNet2.0 yield three critical advancements: (1) Determination of binding mode through consensus analysis of molecular docking and AI-based interpretability analysis; (2) Identification of a stereo-chemically optimized ALDH1B1 inhibitor, **B1**, exhibiting 2-fold enhanced inhibitory potency compared to the racemic parent compound; (3) Discovery of potential resistance-conferring mutations through in silico mutagenesis scanning. Collectively, this AI-driven multimodal analysis provides critical structural insights and pharmacophoric blueprints for subsequent probe development and structure-guided lead optimization campaigns.

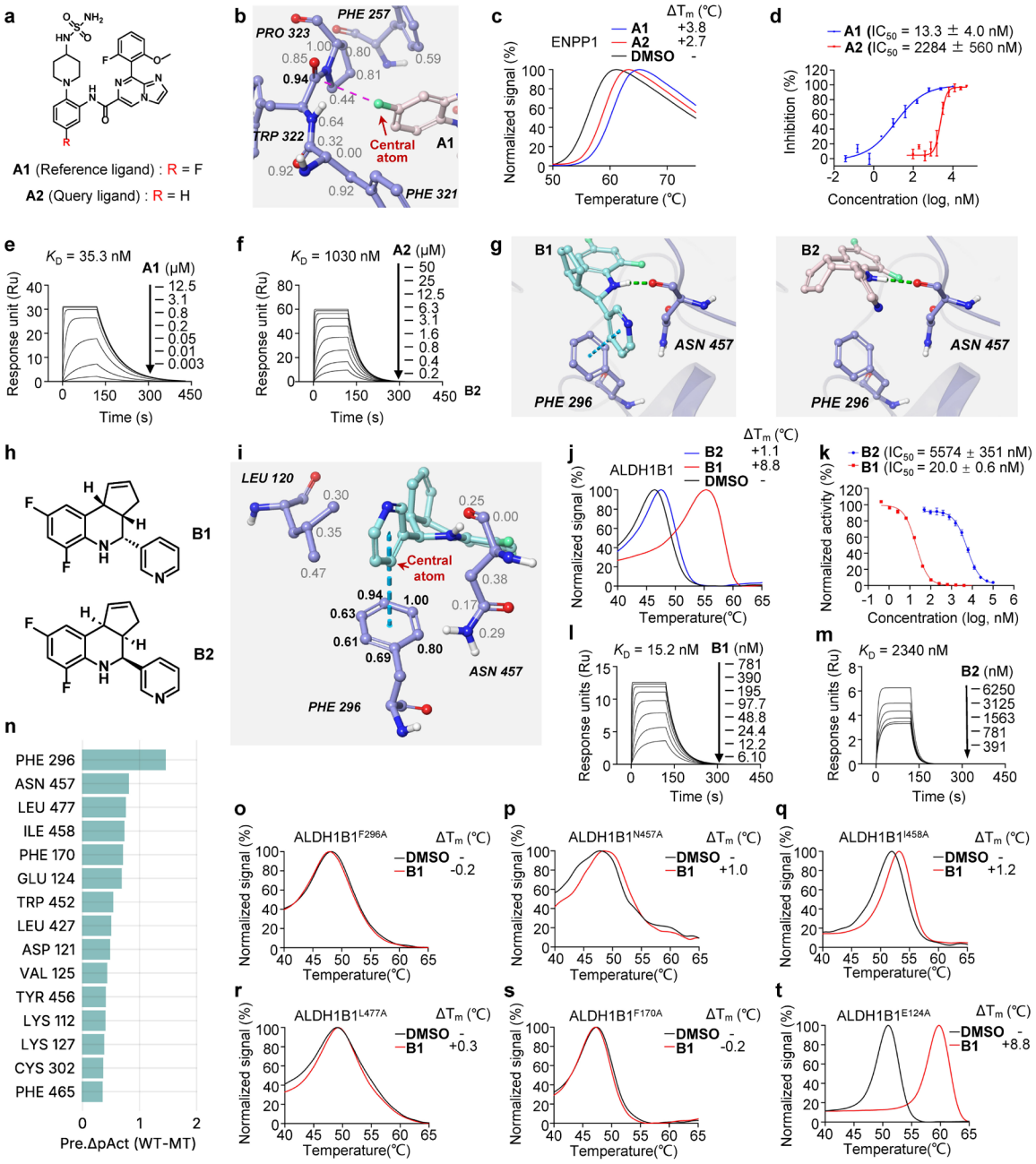

Figure 5. **a. Chemical structures of ENPP1 inhibitors.** Reference ligand **A1** (R = F) and query ligand **A2** (R = H) where the fluorine atom was replaced with hydrogen. **b. Interpretability analysis.** The fluorine orthogonal multipolar interaction between **A1** and TRP322 of ENPP1, with interaction distance (d = 3.22Å) and angle (θ = 109.2°). Fluorine orthogonal multipolar interaction represented by purple dashed lines. Protein atoms within 5 Å of the central fluorine atoms are annotated with weights, and bold numbers denote the weights of carbon atoms involved in the interactions. **c. Protein thermal shift assay.** Impact of **A1** (50 μM) and **A2** (50 μM) on the thermal stability of ENPP1 protein (2 μM) as determined by PTS assay. Temperature (°C) on the x-axis and normalized signal (%) on the y-axis. **d. *In vitro* enzyme activity assay.** Dose-dependent inhibition of **A1** and **A2** against ENPP1. The substrate for the ENPP1 enzymatic reaction is thymidine 5'-monophosphate p-nitrophenyl ester (TMP). Error bars represent the mean ± SEM of three independent experiments. **e-f. Surface plasmon resonance binding assay.** The binding kinetics measurement of **A1** and **A2** to ENPP1 protein were using SPR assay. Graphs

depicting equilibrium response units versus **A1** and **A2** concentrations were plotted. **g. Binding modes of B1 and B2.** While **B1** and **B2** adopt largely similar binding modes, a critical difference emerged due to their distinct stereochemistry: the pyridine ring of **B2** is unable to form the favorable $\pi - \pi$ stacking interaction with PHE296 that is observed for **B1**. **h. Chemical structures of ALDH1B1 inhibitors.** Three chiral centers of Reference ligand **B1** and query ligand **B2** were different. **i. Interpretability analysis.** The $\pi - \pi$ stacking interaction is depicted as blued dashed lines, and protein atoms within 5 Å of the central carbon atom were annotated with weights, where the bold number indicates the weights associated with the carbon atom on the benzene ring of PHE296. **j. Protein thermal shift assay.** Impact of **B1** (50 μM) and **B2** (50 μM) on the thermal stability of ALDH1B1 protein (5 μM) Temperature (°C) on the x-axis and normalized signal (%) on the y-axis. **k. *In vitro* enzyme activity assay.** Dose-dependent inhibition of **B1** and **B2** against ALDH1B1. Error bars represent the mean ± SEM of three independent experiments. **l-m. Surface plasmon resonance binding assay.** The binding kinetics measurement of **B1** and **B2** to ALDH1B1 protein were measured using SPR assay. Graphs depicting equilibrium response units versus **B1** and **B2** concentrations were plotted. **n. Mutation prediction results of PBCNet2.0.** The x-axis represents the value of the decrease in binding affinity predicted by the model after the mutation, and the y-axis represents the mutated residue. The 15residues prioritized by the model (out of 43) are shown. **o-t. Protein thermal shift assay.** Impact of **B1** on the thermal stability of various ALDH1B1 mutants.

3. Discussion

This work introduces PBCNet2.0, a next-generation deep learning framework that fundamentally advances the prediction of protein-ligand relative binding affinities. Integrating a vast 8.6 million-pair training dataset with a sophisticated Cartesian tensor-based equivariant network, PBCNet2.0 delivers unprecedented zero-shot prediction accuracy for relative binding affinities, achieving statistical parity with the gold-standard Schrödinger's FEP+ simulations while retaining high throughput. Critically, our investigations reveal that PBCNet2.0 develops a genuine understanding of protein-ligand interactions, demonstrating sensitivity to subtle geometric details and intermolecular interactions, rather than relying on ligand feature memorization. Furthermore, PBCNet2.0 yields an emergent capability to predict binding affinity changes arising from pocket residue mutations—a task for which it was not explicitly trained. The combined validation of retrospective and prospective experiments fully demonstrates PBCNet's potential as a powerful tool for both accelerating lead optimization and probe development.

Despite its strengths, PBCNet2.0 has certain limitations inherent in many current structure-based deep learning models. First, the model relies on static protein-ligand binding conformations as input, which prevents it from explicitly accounting for the entropic effects. This limitation restricts its ability to capture the conformational dynamics of proteins and ligands during the binding process. While PBCNet2.0's performance approaches that of methods like Schrödinger's FEP+ which account for entropic effects, further substantial improvements may necessitate incorporating structural flexibility, representing a key future challenge. Second, PBCNet2.0 struggles to predict negative data, i.e., cases where R-group modifications result in complete loss of ligand activity. This stems primarily from the scarcity of well-characterized negative data in public datasets. While some negative data are available, they often lack clarity regarding whether the ligand binds to the target protein or whether it adopts the same binding mode as positive ligands. As a result, the ambiguity of negative data risks adding noise that could mislead the model. Additionally, assigning affinity regression labels to such data remains challenging, as their binding affinity labels are often expressed as ">" values rather than precise activity measurements. The loss of activity caused by minor structural modifications is frequently due to steric clashes that prevent the modified ligand

from maintaining its original binding mode with the protein. Encouragingly, molecular docking techniques can provide early warnings in such cases, such as identifying obvious steric clashes in docking conformations or abrupt changes in binding modes caused by steric interference.

Future research holds several promising directions for further exploration. First, the integration of high-quality protein mutation datasets with the PBCNet2.0 training set might enable the model to focus on protein-ligand interaction information more directly. Training the model on such hybrid data might enhance its ability to generalize across diverse scenarios. Second, the training process of PBCNet2.0, which enables it to learn generalizable protein-ligand interaction knowledge, can potentially serve as a form of pretraining. By fine-tuning the pretrained model, it might be adapted to other downstream tasks, such as virtual screening, effectively mitigating the issue of ligand structure memorization often observed in non-pairwise training approaches. Finally, applying PBCNet2.0 to real-world compound optimization projects would provide a critical test of its practical utility. These applications not only offer opportunities to evaluate the model's performance but also help identify areas for improvement, ultimately facilitating its integration and adoption within drug discovery pipelines.

4. Methods and Data

4.1 Datasets

4.1.1 Training set

The training dataset for PBCNet2.0 comprises 8.6 million protein-ligand complexes pairs. The data processing pipeline is illustrated in Supplementary Figure 12 and can be divided into three main steps: (1) Data Collection, (2) Data Preprocessing, and (3) Three-Dimensional Structure Generation. First, during the data collection stage, we downloaded the BindingDB Database (2023.12 version)[20], obtaining ~2.81 million protein-ligand affinity data (IC_{50}, EC_{50}, K_i, and K_d).

Subsequently, we conducted data preprocessing. Ligands with SMILES[63] that failed during preparation with RDKit[64] were removed. Affinity data labeled as NaN, 0, or containing symbols such as '<' or '>' were also excluded. The affinity data were grouped based on three indices: 'Entry DOI', 'Target name', and 'Target source'. This grouping ensured that the binding affinity data within each group were associated with the same target and originated from the same publication. We then applied structural clustering to the molecules within each group, using a Tanimoto similarity threshold of 0.5 to define chemical series, thereby the molecules within each series were structural analogs. For cases where a molecule in a chemical series had multiple affinity labels, we followed a specific protocol: if the maximum affinity label of the molecule was less than three times its minimum affinity label, we used the average value as the final label; if the maximum label exceeded three times the minimum, we removed the molecule to ensure data quality. Then, we collected all available PDB entries for each compound series according to target information. To determine the final PBD used for docking, we extracted all co-crystallized ligands and calculated the Tanimoto similarities between the co-crystallized ligands and the compounds in each series. The co-crystallized ligands with maximum similarity and their corresponding PDBs are used for subsequent docking operation. To ensure data quality, if no co-crystallized ligands with a similarity greater than 0.5 were found for a chemical series, we excluded that series.

After confirming the ligands and the PDB structures to be used, we then used Glide[29] to generated 3D docking conformations. Initially, the ligands were preprocessed using the Schrödinger

LigPrep module with default parameters. On the protein side, the PDB files were prepared using the Protein Preparation Wizard in the Schrödinger suite, following the standard protocol. Water molecules that formed more than three hydrogen bonds with ligand or receptor atoms were retained, and the receptor grids were generated with the centroid of the co-crystallized ligand as the center. After preprocessing, docking was performed using the Glide module in Schrödinger with default parameters, allowing up to 200 poses per ligand to be output. It is well-established that similar ligands from the same chemical series should bind to a given protein in similar poses[6], which is also a main assumption in our modeling process. So, pose selection became a critical step here. For each series, the maximum common substructure (MCS) between each ligand and the co-crystallized ligand was first extracted. The root-mean-square deviation (r.m.s.d.) of each ligand pose and the experimentally determined pose of the co-crystallized ligand within the MCS was then calculated. Poses with an r.m.s.d. of 2.0 Å or less were considered acceptable, assuming they shared the same binding mode as the co-crystallized ligand. When multiple acceptable poses were found for a ligand, the pose with the highest Glide score was selected as the final pose. If no acceptable pose could be obtained through docking, the ligand was discarded to maintain data quality. All Schrödinger-related operations were performed using version 2022-3, utilizing the Schrödinger Python API.

After the preprocessing steps, we paired molecules from the same chemical series, with a maximum cutoff of 15,000 pairs generated per series. Priority was given to unpaired molecules to ensure comprehensive pairing within each series. Subsequently, we analyzed the label distribution of the generated data. We observed that the labels of the data were normally distributed, with the majority concentrated in the range of $[-1, 1]$ (Supplementary Fig. 13a, left). This concentration could potentially lead to overfitting, as a model might achieve low training error by simply predicting the mean value of the training labels (Supplementary Fig. 13b). To address this issue, we balanced the training data by under-sampling in the high-density regions and over-sampling in the low-density regions. The label distribution of the balanced training dataset is shown in Supplementary Figure 13a (right).

We also excluded the protein-ligand pairs that overlapped with the test sets. A comparison between the training set of PBCNet2.0 and that of PBCNet, as summarized in Supplementary Figure 13c, demonstrates that the training dataset outperforms the previous one in terms of both target and compound diversity.

Finally, we were concerned about the potential for the modification bias, which means that the activity changes of all compound pairs with the same R-group modification (e.g., from a methyl group to an ethyl group) might be consistently either increased or decreased. The risk of such a bias is that the model might simply memorize the structural differences between molecules and predict accordingly, leading to overfitting. To address this concern, we analyzed the distribution of binding affinity changes between compound pairs with the same R-group modifications. The results, shown in Supplementary Figure 13d, reveal that the activity changes broadly follow normal distributions with a mean of zero, indicating the absence of modification bias.

4.1.2 FEP held-out test set

The same as the work of PBCNet, we employed two well-established FEP benchmark datasets for performance assessment. The FEP1 set, obtained from Wang et al.[8], encompasses eight chemical series with experimentally validated binding free energies (ΔG) and corresponding FEP+ prediction results. The binding free energies were converted to pIC_{50} values assuming non-competitive binding conditions through the equation:

$$pIC_{50} \approx -\log_{10}(e^{\frac{\Delta G}{RT}})$$

where R represents the gas constant (1.987×10^{-3} $kcal \cdot K^{-1} \cdot mol^{-1}$), T is the thermodynamic temperature at 297 K, and e denotes Euler's number (2.718). Additionally, Schindler et al.[65] provided a complementary benchmark dataset (FEP2) containing eight congeneric series targeting pharmaceutically relevant proteins, with experimentally measured binding affinities (IC_{50} values). The FEP2 dataset introduces greater molecular diversity compared to FEP1, incorporating variations in net charge, charge distribution, ring systems, and core structures. For both datasets, we performed logarithmic transformation of the binding metrics and applied consistent ligand pairing strategies aligned with our training data protocols. The union of FEP1 and FEP2 sets is called FEP set. Specific data information is presented in Supplementary Table 1.

4.1.3 SAR-Diff held-out test set

The SAR-Diff set consists of eight pairs of chemical series, with each pair containing two chemical series. The molecules in these two series undergo the same R-group modification process,

but differ in their structure-activity relationships and scaffold structures. On this test set, the model must pay attention to the interactions between small molecule and protein pockets to make accurate predictions. Simply memorizing molecular structures is insufficient for prediction.

This dataset was obtained from Chembl database[66], with each series derived from the same assay (identified by the 'Chembl assay ID' index) to minimize the activity ranking error for each series. After downloading, we can obtain the structural information of the corresponding ligand molecules. Subsequently, we determined the PDB structures for docking in the same manner as the training set, and obtained 3D structural data. The data and related information are summarized in Supplementary Table 2.

4.1.4 Mutation test set

To assess the model's ability to capture protein-ligand interactions, we curated a protein mutation dataset comprising eight targets. This dataset was sourced from the mutation-induced drug resistance database (MdrDB)[67]. We filtered out mutation data where protein mutations changed the original protein-ligand interaction (including hydrogen bonds, hydrophobic interactions, salt bridges, π-stacking, π-cation interactions, and halogen bonds), ensuring that mutations occurred within the protein-ligand binding pocket. For each series of mutations within a target, the binding ligand is identical. Each series consists of a wild-type (WT) complex and a range of different mutant (MT) complexes with correspondingly relative binding energy labels ($\Delta\Delta G$). We then used Schrödinger Maestro to prepare the protein and ligand structures, perform docking, and select the most plausible conformations based on their binding modes to create the mutation test set. Finally, we collected mutation data for a total of 8 targets, and detailed targets information can be found in Supplementary Table 3.

4.1.5 F-Opt test set

We conducted an interpretability analysis of the model's performance, using protein-ligand fluorine orthogonal multipolar interactions as a case study. We selected a fluorine-involved lead optimization (F-Opt) dataset from different publications, which includes 10 pairs of complexes where ligand modifications from -H or -CH3 to -F or -CF3 resulted in binding affinity improvements

and the formations of fluorine orthogonal multipolar interactions. The selection criteria were: (1) the literature explicitly identifies fluorine orthogonal multipolar interactions as the cause of the activity improvements, and (2) co-crystal structures are available for visualizing the interactions. Detailed information about the proteins and ligands were summarized in the Supplementary Table 4.

4.2 PBCNet2.0 implement

PBCNet2.0 predicts the relative binding affinity for a pair of complexes. The core architecture (message-passing module) of PBCNet2.0 is an equivariant graph neural network with Cartesian tensor representations, which is a variant of TensorNet[28]. This architecture is unique in modeling atomic interactions in Cartesian coordinates, using Cartesian rank-2 tensors (represented as 3×3 matrices) as atom representations. The Cartesian rank-2 tensors can be cheaply decomposed into their irreducible components under the rotation group (Equation 1). So, in the message-passing phase, representation mixing could be simplified by straightforward 3×3 matrix product operations, which has contributed to the efficiency of equivariant PBCNet2.0. In the following, we first introduce the irreducible decomposition of the rank-2 Cartesian tensor, and then provide a detailed discussion to the construction of PBCNet2.0.

4.2.1 Irreducible Cartesian rank-2 tensor decomposition

Here, X denotes a rank-2 tensor (3×3 matrix) defined in Cartesian coordinates, with X_{ij} representing the value in the $i-\text{th}$ row and $j-\text{th}$ column. Any tensor X can be decomposed in the following manner:

$$X = \frac{1}{3}\mathrm{Tr}(X)\mathrm{Id} + \frac{1}{2}(X - X^T) + \frac{1}{2}\left(X + X^{\mathrm{T}} - \frac{2}{3}\mathrm{Tr}(X)\mathrm{Id}\right), \tag{1}$$

where $\mathrm{Tr}(X) = \sum_i X_{ii}$ is the trace operator, X^T denotes the transpose of X and Id is the identity matrix. The first term is proportional to the identity matrix, possessing only one degree of freedom. This term is invariant under rotations and represents a scalar. The second term is a skew-symmetric matrix and contains three independent components that transform as a vector under rotations. The third term is a symmetric and traceless, with five independent components that represent a higher-

order tensor. From a representation theory perspective, Equation 1 shows that the 9-dimensional Cartesian rank-2 tensor (X) can be decomposed into components of dimensions 1 (scalar), 3 (vector), and 5 (tensor), respectively. We refer to X as a "full tensor", while its components are referred to as scalar (I), vector (A), and tensor (S) features.

In neural networks, tensors can be constructed from relative position vectors. For example, given a vector $\boldsymbol{r} = (r_x, r_y, r_z)$ (e.g., the difference between the coordinates of two nodes in a graph), we can build a well-behaved tensor $X = I + A + S$ as follows:

$$I = f(\|\boldsymbol{r}\|)\mathrm{Id}, A = \begin{pmatrix} 0 & r_z & -r_y \\ -r_z & 0 & r_x \\ r_y & -r_x & 0 \end{pmatrix}, S = \boldsymbol{r}\boldsymbol{r}^{\mathrm{T}} - \frac{1}{3}\mathrm{Tr}(\boldsymbol{r}\boldsymbol{r}^{\mathrm{T}})\mathrm{Id}, \tag{2}$$

where $\|\boldsymbol{r}\|$ is the length of $\boldsymbol{r}$, $f(\cdot)$ is mapping function, and $\boldsymbol{r}\boldsymbol{r}^{\mathrm{T}}$ denotes the outer product of $\boldsymbol{r}$ with itself. We can also modify the full tensor by multiplying invariant quantities to each component, $X' = f_I I + f_A A + f_S S$, where$f_I$, f_A, and f_S can be constants or learnable invariant functions (e.g., functions of distances). These modifications preserve the transformation properties of the tensor under rotations.

4.2.2 Graph preprocessing and feature initialization

When constructing a protein–ligand interaction graph, we only considered residues in the range of 8 Å around the ligand, and these residues are defined as protein pocket. Each heavy atom in a pocket-ligand complex is regarded as a node in the graph, and each covalent bond is treated as an edge. Moreover, distance edges are also built between protein-ligand and ligand-ligand node pairs whose spatial distances are less than or equal to 5.0 Å. For protein–protein node pairs, the creation of distance edges followed a context-dependent approach: if any of the protein atoms in a protein–protein node pair was connected to a ligand atom via a distance edge, then a distance edge was created for the protein-protein node pair when their spatial distance was ≤ 5 Å. Otherwise, distance edges were only established if the protein-protein node pair distance was ≤ 3 Å. This graph construction strategy preserved the flow of protein-ligand interaction information while reducing the number of edges connected protein atoms far from the binding interface, thereby lowering computational overhead.

In terms of feature initialization, the original TensorNet architecture utilized only atom types

and edge lengths as node and edge initial features, respectively. Here, we enhanced the model by incorporating additional chemically relevant atom and edge features, enabling it to leverage more a priori chemical knowledge. Specifically, 9 types of atomic features (Supplementary Table 5 and bond types (Supplementary Table 6) were used to better encode atoms and their local chemical environments.

4.2.3 Initial tensor representations

PBCNet2.0 learns a Cartesian full tensor representation $X_i = \{X_i^0, X_i^1, \cdots X_i^d\}$ (d represents the number of hidden dimensions) for an atom a_i. In the following, we describe the computational process using X_i as an example. Atomic representation X_i can be decomposed into scalar, vector and tensor contributions I_i, A_i, S_i via Equation 1. Conversely, these components can be merged to form X_i.

We treat a_i as the central atom, $a_j \in \{a_{j_1}, a_{j_2}, \cdots, a_{j_n}\}$ (n is the number of neighbor atoms of a_i) represents a neighbor atom of a_i, and $e_{ij} \in \{e_{ij_1}, e_{ij_2}, \cdots, e_{ij_n}\}$ is an edge between a_i and a_j. The coordinates of a_i and a_j are presented as $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, the edge vector $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$ can be obtained, and $\|\boldsymbol{r}_{ij}\|$ presents the length of e_{ij}. We initialize each edge scalar features by the identity matrix $I_{ij}^{initial} = Id$, and each vector ($A_{ij}^{initial}$) and tensor ($S_{ij}^{initial}$) feature are initialized using the normalized edge vectors $\hat{\boldsymbol{r}}_{ij} = \frac{\boldsymbol{r}_{ij}}{\|\boldsymbol{r}_{ij}\|} = \left(\hat{r}_{ij}^x, \hat{r}_{ij}^y, \hat{r}_{ij}^z\right)$ through Equation 2. Thus, the initial full tensor representation of e_{ij} is $X_{ij}^{initial} = I_{ij}^{initial} + A_{ij}^{initial} + S_{ij}^{initial}$. Next, we aim to encode bond- and atom-invariant information, including edge distances, edge and atom chemical features (Supplementary Table 5 and 6), into the initial full tensor $X_{ij}^{initial}$. This information is encoded using linear layers, with the distance information extended by a radial basis function (RBF) before encoding:

$$\boldsymbol{Z}_i = Linear_1(\boldsymbol{z}_i), \boldsymbol{Z}_j = Linear_1(\boldsymbol{z}_j), \boldsymbol{Z}_{ij}^{atom} = \boldsymbol{Z}_i \oplus \boldsymbol{Z}_j,$$
$$\boldsymbol{Z}_{ij}^{chem} = Linear_2(\boldsymbol{z}_{ij}), \boldsymbol{Z}_{ij}^{dist} = \mathrm{RBF}(\|\boldsymbol{r}_{ij}\|),$$
$$\boldsymbol{Z}_{ij}^{bond} = \boldsymbol{Z}_{ij}^{chem} \oplus \boldsymbol{Z}_{ij}^{dist}, \quad (3)$$

where $\boldsymbol{z}_i$, $\boldsymbol{z}_j$ and $\boldsymbol{z}_{ij}$ are chemical features of a_i, a_j and e_{ij}, respectively. $\|\boldsymbol{r}_{ij}\|$ is the length of e_{ij}. RBF($\cdot$) and $\oplus$ are the radial basis function and concatenate operation. $\boldsymbol{Z}_i$, $\boldsymbol{Z}_j$ and $\boldsymbol{Z}_{ij}^{chem}$ are chemical embeddings of a_i, a_j and e_{ij}, while $\boldsymbol{Z}_{ij}^{dist}$ is distance embedding of e_{ij}.

$\boldsymbol{Z}_{ij}^{atom}$ and $\boldsymbol{Z}_{ij}^{bond}$ are atom pair-wise and bond-invariant representation of e_{ij}. Then $\boldsymbol{Z}_{ij}^{atom}$ and $\boldsymbol{Z}_{ij}^{bond}$ are combined with $I_{ij}^{initial}$, $A_{ij}^{initial}$ and $S_{ij}^{initial}$ according to:

$$f_I^{initial} = Linear_3(\boldsymbol{Z}_{ij}^{bond}), f_A^{initial} = Linear_4(\boldsymbol{Z}_{ij}^{bond}), f_s^{initial} = Linear_5(\boldsymbol{Z}_{ij}^{bond}),$$

$$X_{ij} = \boldsymbol{Z}_{ij}^{atom} \times \left(f_I^{initial} I_{ij}^{initial} + f_A^{initial} A_{ij}^{initial} + f_s^{initial} S_{ij}^{initial}\right), \quad (\mathbf{4})$$

where X_{ij} is the invariant information-merged initial tensor representation of e_{ij}. It should be noted that the original TensorNet[28] includes an additional distance decay function in Equation 4. The assumption of distance decay function is that atoms that are farther apart have less influence on each other, which obviously affects the identification of long-range interaction. Moreover, this function also inevitably introduces a bias, where the correlation weights between atom pairs become larger as their distance decreases, and vice versa. In PBCNet2.0, we have eliminated this function. We then get atom-wise initial tensor representation of a_i by aggregating all neighboring edge tensor representations:

$$X_i = \sum_{j \in \mathcal{N}(i)} X_{ij},$$

where $\mathcal{N}(i)$ means the neighbor set of the central atom a^i. At this point, Frobenius norm (an invariant norm) $\|X\|_{\mathrm{F}} = \mathrm{Tr}(X^T X)$ of X_i is computed and fed to a series of layers:

$$f_i^I, f_i^A, f_i^S = SiLU(Linear_7(SiLU\left(Linear_6\left(LN(\|X_i\|_{\mathrm{F}})\right)\right))),$$

where $LN(\cdot)$ and $SiLU(\cdot)$ are layer norm and SiLU activation. After the decomposition of this atom-wise tensor representation X_i into its irreducible representations via Equation 1, f_i^I, f_i^A and f_i^S are then used to modify component-wise linear combinations to obtain the final initial atomic tensor representations $X_i = f_i^I \times Linear_8(I_i) + f_i^A \times Linear_9(A_i) + f_i^S \times Linear_{10}(S_i)$.

4.2.4 Tensor representations updating

PBCNet2.0 consists of three message passing layers. We denote the input of the first layer as X_i^0 (i.e., the initial atomic tensor representation), the output of the first layer (which becomes the input of the second layer) as X_i^1, and so on. To introduce the message passing process, we focus on the transition from X_i^0 to X_i^1 as an example. Firstly, X_i^0 is normalized by

$$X_i^0 = \frac{X_i^0}{\|X_i^0\|_{\mathrm{F}} + 1},$$

Then X_i^0 is decomposed into scalar, vector and tensor features $X_i^0 = I_i^0 + A_i^0 + S_i^0$. The obtained

features are transformed by independent linear combinations to get an updated tensor embedding $X_i^{0_ud}$ according to

$$X_i^{0_ud} = Linear_{11}(I_i^0) + Linear_{12}(A_i^0) + Linear_{13}(A_i^0).$$

We then aim to merge the information from neighboring atoms used to further update $X_i^{0_ud}$. Firstly, the bond-invariant embeddings $\boldsymbol{Z}_{ij}^{bond}$ are concatenated with the $O(3)$-invariant scalar embedding of a_i and a_j ($\boldsymbol{x}_i^{0_scalar}$ and $\boldsymbol{x}_j^{0_scalar}$, see *Scalar embedding extracting* section for more details), and then are transformed into three invariant terms as follows:

$$\boldsymbol{x}_i^{0_scalar} = SiLU(Linear_{14}(LN\left(\|I_i^0\|_F \oplus \|A_i^0\|_F \oplus \|S_i^0\|_F\right))),$$

$$\boldsymbol{x}_j^{0_scalar} = SiLU(Linear_{14}(LN\left(\|I_j^0\|_F \oplus \|A_j^0\|_F \oplus \|S_j^0\|_F\right))),$$

$$\boldsymbol{x}_{ij}^{0_bond} = \boldsymbol{x}_i^{0_scalar} \oplus \boldsymbol{x}_j^{0_scalar} \oplus \boldsymbol{Z}_{ij}^{bond}$$

$$f_{ij}^I, f_{ij}^A, f_{ij}^S = SiLU(Linear_{17}(SiLU(Linear_{16}(SiLU\left(Linear_{15}(\boldsymbol{x}_{ij}^{0_bond})\right))))). \quad (5)$$

Note that we also eliminated the application of distance decay function compared to the original TensorNet here. Next, the messages M_{ij} sent from neighbors a_j to atom a_i are computed as:

$$X_j^{0_ud} = I_j^{0_ud} + A_j^{0_ud} + S_j^{0_ud},$$

$$M_{ij} = f_{ij}^I \times I_j^{0_ud} + f_{ij}^A \times A_j^{0_ud} + f_{ij}^S \times S_j^{0_ud}$$

These messages are then aggregated:

$$M_i = \sum_{j \in \mathcal{N}(i)} M_{ij}$$

where $\mathcal{N}(i)$ means the neighbor set of the central atom a^i. Then updated tensor embedding $X_i^{0_ud}$ is further updated by M_i with a parity-preserved form:

$$X_i^{0_ud} = X_i^{0_ud} \times M_i + M_i \times X_i^{0_ud}.$$

Next, $X_i^{0_ud}$ is further decomposed into scalar, vector, and tensor features ($I_i^{0_ud}$, $A_i^{0_ud}$, and $S_i^{0_ud}$) with Equation 1. These features are individually normalized according to:

$$I_i^{0_ud} = \frac{I_i^{0_ud}}{\|X_i^{0_ud}\|_F + 1}, A_i^0 = \frac{A_i^{0_ud}}{\|X_i^{0_ud}\|_F + 1}, S_i^0 = \frac{S_i^{0_ud}}{\|X_i^{0_ud}\|_F + 1}.$$

These normalized features are then used to compute independent linear combinations:

$$X_i^{0_ud} = Linear_{18}(I_i^{0_ud}) + Linear_{19}(A_i^{0_ud}) + Linear_{20}(S_i^{0_ud}).$$

Finally, a residual message ΔX_i^0 is computed in a parity-preserving manner and used to update X_i^0

to get X_i^1:

$$\Delta X_i^0 = X_i^{0_ud} + (X_i^{0_ud})^2,$$

$$X_i^1 = X_i^0 + \Delta X_i^0.$$

Repeating the above process, we can get X_i^2 and X_i^3 for the subsequent layers.

4.2.5 Scalar embedding extracting

Since the protein-ligand binding affinity is a $O(3)$-invariant property, the Frobenius norms of each component of X_i^3, which is also $O(3)$-invariant, are used to get final atomic scalar embedding $\boldsymbol{x}_i^{scalar}$,

$$\boldsymbol{x}_i^{scalar} = SiLU(Linear_{21}(LN\left(\left\|I_i^3\right\|_F \oplus \left\|A_i^3\right\|_F \oplus \left\|S_i^3\right\|_F\right))) \tag{6}$$

4.2.6 Readout

After the above process, we obtain the final scalar embedding $\boldsymbol{x}_i^{scalar}$ of each atom. We then summed the scalar embeddings from the ligand atoms as the final complex embedding $\boldsymbol{x}^{scalar}$,

$$\boldsymbol{x}^{scalar} = \sum_{i \in \mathcal{L}(lig)} \boldsymbol{x}_i^{scalar}$$

where $\mathcal{L}(lig)$ is the ligand atom set of a pocket-ligand complex. Finally, we utilize the embedding of the reference complex $\boldsymbol{x}_{ref}^{scalar}$ to subtract that of the query complex $\boldsymbol{x}_{query}^{scalar}$, thereby obtaining the difference embedding $\Delta\boldsymbol{x}^{scalar} = \boldsymbol{x}_{ref}^{scalar} - \boldsymbol{x}_{query}^{scalar}$, which then be fed into a 3-layer perceptron to obtain the final prediction $\Delta\hat{y}$. If the binding affinity of the reference complex y_{ref} is known, we can get the predicted binding affinity of tested complex $\hat{y}_{query} = y_{ref} - \Delta\hat{y}$. In the training process, to ensure, to the extent possible, that $Model(\Delta\boldsymbol{x}^{scalar}) = -Model(-\Delta\boldsymbol{x}^{scalar})$, we input $-\Delta\boldsymbol{x}^{scalar} = \boldsymbol{x}_{query}^{scalar} - \boldsymbol{x}_{ref}^{scalar}$ into the fully connected layer as well and calculate the loss together as a data augmentation strategy with a low computational overhead.

4.3 Baseline models

To thoroughly evaluate our method, we selected a range of representative baseline models. For traditional approaches, we chose Schrödinger's **Glide-SP**[29], **MM-GBSA**[30], and Schrödinger's

FEP+[8]. The Glide-SP and MM-GBSA modules were used with Schrödinger's default parameters, while the performance of FEP+ is based on the reported results[8,65].

For deep learning methods, we considered both sequence-based and structure-based approaches. The sequence-based methods include:

- **PSICHIC**[7]. It directly decodes protein-ligand interaction fingerprints from sequence data by incorporating physicochemical constraints. The weights and code of PSICHIC is from its original publication. The binding affinity values output by the model are pKd or pKi, and we directly evaluate the model using its outputs.
- **BIND**[31]. It is a graph neural network that uses Evolutionary Scale Modeling 2[68] (ESM-2) embeddings for proteins and graph-based representations for ligands to perform virtual screening. We ran the model with the default parameters provided in the paper and evaluated it using the IC_{50} values from the output.
- **PLAPT**[32]. It utilizes transfer learning from pretrained transformers like ProtBERT[69] and ChemBERTa[70], requiring only 1D protein and ligand string data for predictions. The model's weights are set to their default values, and the predicted IC_{50} values are directly used for performance calculation.

The structure-based methods differ in terms of input forms (ranging from 2D contact features to 3D graph structures), learning objectives (from direct binding affinity prediction to fine-grained interaction modeling), and model architectures, providing a comprehensive evaluation benchmark. These models comprise:

- **RTMScore**[33]**.** It predicts binding affinity by fitting a protein-ligand distance likelihood potential, using a Mixture Density Network (MDN) framework. In the pretrained models provided in the paper, we selected the best-performing model, rtmscore_model3. The paper defines a custom affinity score, where higher values indicate stronger binding affinity.
- **GenScore**[34]**.** Building on RTMScore, the model adds a ranking loss term to the training loss and incorporates a fine-tuning strategy. We evaluated the model using GatedGCN_ft_1.0, which showed the strongest ranking ability according to the results. The output affinity values are also custom-defined.
- **PIGNet2**[11]**.** The model focuses on modeling the interactions between protein-ligand pairs, refining the types of interactions into hydrogen bonds, metal bonds, van der Waals forces, and

hydrophobic interactions. During replication of the paper's results, we found that the model pda_3 had the performance most closely aligned with the reported results on the FEP test set, so we selected this model for comparison. The predicted binding free energy (ΔG) is converted into pIC_{50}.

- **OnionNet-2**[35]**.** It is a scoring model based on convolutional neural network to predict the binding affinity pKd, where the protein-ligand interactions are characterized by the number of contacts between protein residues and ligand atoms in multiple distance shells. We used the default parameter weights for the model.

4.4 Evaluation metrics

Model performance was evaluated using four metrics: the Pearson correlation coefficient (r), Spearman's rank correlation coefficient (ρ), root-mean-square error (r.m.s.e.), and mean-absolute error (m.a.e.). The correlation coefficients range from 0 to 1, where 1 indicates perfect correlation. The Pearson correlation measures the linear relationship between predicted and experimental binding affinities, while Spearman's correlation assesses the monotonic relationship between predicted and experimental rankings. Root-mean-square error measures the average magnitude of prediction errors, where lower values indicate better prediction accuracy. Mean absolute error represents the average absolute difference between predicted and experimental values, where lower values indicate better performance.

4.5 Interpretability analysis

In the message passing module of PBCNet2.0, the pocket-ligand complex is deemed a whole graph and information on atoms is transmitted across the molecule graph. This module provide atom–atom pairwise interpretation of key inter-molecular interactions, which is helpful for guiding the design a drug for a specific target. The invariant terms f_{ij}^{I}, f_{ij}^{A}, and f_{ij}^{S} obtained from the Equation 5 can be seen as a measure of importance between atoms a_i and a_j. We extracted these invariant terms from the first layer of the message-passing model and computed their average as relevance weights f_{ij} using the formula:

$$f_{ij} = \frac{f_{ij}^{I} + f_{ij}^{A} + f_{ij}^{S}}{3}.$$

Next, we performed min-max normalization on the weights of all neighboring atoms to obtain the final weights w_{ij} used in visualization:

$$f_{i_min} = \min_{j \in \mathcal{N}(i)} f_{ij},$$

$$f_{i_max} = \max_{j \in \mathcal{N}(i)} f_{ij},$$

$$w_{ij} = \frac{f_{ij} - f_{i_min}}{f_{i_max} - f_{i_min}}.$$

The normalized weights w_{ij} have a minimum value of 0 and a maximum value of 1, facilitating their use in visualization. An ideal model will tend to assign a higher weight to the pair of atoms which may form a key inter-molecular interaction because they usually have a larger influence on each other's chemical environment and on ligand binding affinity.

Notably, because the positions of the hydrogen atoms depended heavily on the corresponding program used to add hydrogens, and due to the sheer number of hydrogen atoms which would substantially increase the computational load, we did not take hydrogen atoms into account in the modeling process. Here, for hydrogen-bond donors, the heavy atoms covalently linked with hydrogen atoms were chosen for further analysis, consistent with previous practices in PBCNet.

5. Data availability

The unprocessed training data is from BindingDB source and can be found at The test data are available in GitHub .

6. Coda availability

The code of this work is private. The files used to support the results in this work are available in GitHub XXX.

7. Acknowledgments

This work was supported by XXX.

References

1. Oprea, T. I. *et al.* Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov* **17**, 317–332 (2018).

2. Dewson, G., Eichhorn, P. J. A. & Komander, D. Deubiquitinases in cancer. *Nat Rev Cancer* **23**, 842–862 (2023).

3. Ackloo, S. *et al.* Target 2035 – an update on private sector contributions. *RSC Med. Chem.* **14**, 1002–1011 (2023).

4. Ackloo, S. *et al.* CACHE (Critical Assessment of Computational Hit-finding Experiments): A public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding. *Nat Rev Chem* **6**, 287–295 (2022).

5. Cao, D. *et al.* Generic protein–ligand interaction scoring by integrating physical prior knowledge and data augmentation modelling. *Nat Mach Intell* **6**, 688–700 (2024).

6. Yu, J. *et al.* Computing the relative binding affinity of ligands based on a pairwise binding comparison network. *Nat Comput Sci* **3**, 860–872 (2023).

7. Koh, H. Y., Nguyen, A. T. N., Pan, S., May, L. T. & Webb, G. I. Physicochemical graph neural network for learning protein–ligand interaction fingerprints from sequence data. *Nat Mach Intell* **6**, 673–687 (2024).

8. Wang, L. *et al.* 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 (2015).

9. Cournia, Z., Allen, B. & Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **57**, 2911–2937 (2017).

10. Cao, D. *et al.* SurfDock is a Surface-Informed Diffusion Generative Model for Reliable and Accurate Protein-ligand Complex Prediction. Preprint at https://doi.org/10.1101/2023.12.13.571408 (2023).

11. Moon, S., Hwang, S.-Y., Lim, J. & Kim, W. Y. PIGNet2: a versatile deep learning-based protein–ligand interaction prediction model for binding affinity scoring and virtual screening. *Digital Discovery* **3**, 287–299 (2024).

12. Kaplan, J. *et al.* Scaling Laws for Neural Language Models. Preprint at https://doi.org/10.48550/arXiv.2001.08361 (2020).

13. Rosenfeld, J. S. Scaling Laws for Deep Learning. Preprint at https://doi.org/10.48550/arXiv.2108.07686 (2021).

14. Klug, T. & Heckel, R. Scaling Laws For Deep Learning Based Image Reconstruction. Preprint at https://doi.org/10.48550/arXiv.2209.13435 (2023).

15. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

16. OpenAI *et al.* GPT-4 Technical Report. Preprint at https://doi.org/10.48550/arXiv.2303.08774 (2024).

17. Muegge, I. & Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **42**, 791–804 (1999).

18. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**, 2507–2524 (2006).

19. Satorras, V. G., Hoogeboom, E. & Welling, M. E(n) Equivariant Graph Neural Networks. Preprint at https://doi.org/10.48550/arXiv.2102.09844 (2022).

20. Liu, T., Lin, Y., Wen, X., Jorissen, R. N. & Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res* **35**, D198–D201 (2007).

21. Cohen, M. L. Epidemiology of Drug Resistance: Implications for a Post—Antimicrobial Era. *Science* **257**, 1050–1055 (1992).

22. Alfalah, M., Keiser, M., Leeb, T., Zimmer, K. & Naim, H. Y. Compound Heterozygous Mutations Affect Protein Folding and Function in Patients With Congenital Sucrase-Isomaltase Deficiency. *Gastroenterology* **136**, 883–892 (2009).

23. Teng, S., Madej, T., Panchenko, A. & Alexov, E. Modeling Effects of Human Single Nucleotide Polymorphisms on Protein-Protein Interactions. *Biophys J* **96**, 2178–2188 (2009).

24. Hashimoto, K. & Panchenko, A. R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci U S A* **107**, 20352–20357 (2010).

25. Zhan, S. *et al.* Discovery of Imidazo[1,2-a]pyrazine Derivatives as Potent ENPP1 Inhibitors. *J. Med. Chem.* **67**, 18317–18333 (2024).

26. Chen, Y. *et al.* Aldehyde dehydrogenase 1B1 (ALDH1B1) is a potential biomarker

for human colon cancer. *Biochemical and Biophysical Research Communications* **405**, 173–179 (2011).

27. Nehil-Puleo, K., Quach, C. D., Craven, N. C., McCabe, C. & Cummings, P. T. E(n) Equivariant Graph Neural Network for Learning Interactional Properties of Molecules. *J. Phys. Chem. B* **128**, 1108–1117 (2024).

28. Simeon, G. & Fabritiis, G. de. TensorNet: Cartesian Tensor Representations for Efficient Learning of Molecular Potentials. Preprint at https://doi.org/10.48550/arXiv.2306.06482 (2023).

29. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).

30. Kollman, P. A. *et al.* Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33**, 889–897 (2000).

31. Lam, H. Y. I., Guan, J. S., Ong, X. E., Pincket, R. & Mu, Y. Protein language models are performant in structure-free virtual screening. *Briefings in Bioinformatics* **25**, bbae480 (2024).

32. Rose, T., Monti, N., Anand, N. & Shen, T. PLAPT: Protein-Ligand Binding Affinity Prediction Using Pretrained Transformers. 2024.02.08.575577 Preprint at https://doi.org/10.1101/2024.02.08.575577 (2024).

33. Shen, C. *et al.* 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 (2022).

34. Shen, C. *et al.* A generalized protein–ligand scoring framework with balanced

scoring, docking, ranking and screening powers. *Chem. Sci.* **14**, 8129–8146 (2023).

35. Wang, Z. *et al.* OnionNet-2: A Convolutional Neural Network Model for Predicting Protein-Ligand Binding Affinity Based on Residue-Atom Contacting Shells. *Front. Chem.* **9**, (2021).

36. Mastropietro, A., Pasculli, G. & Bajorath, J. Learning characteristics of graph neural networks predicting protein–ligand affinities. *Nat Mach Intell* **5**, 1427–1436 (2023).

37. Marshall, G. R. & Cramer, R. D. Three-dimensional structure-activity relationships. *Trends in Pharmacological Sciences* **9**, 285–289 (1988).

38. Peltason, L. & Bajorath, J. SAR Index: Quantifying the Nature of Structure−Activity Relationships. *J. Med. Chem.* **50**, 5571–5578 (2007).

39. Guha, R. On Exploring Structure–Activity Relationships. in *In Silico Models for Drug Discovery* (ed. Kortagere, S.) 81–94 (Humana Press, Totowa, NJ, 2013). doi:10.1007/978-1-62703-342-8_6.

40. Kenny, P. W. Hydrogen-Bond Donors in Drug Design. *J. Med. Chem.* **65**, 14261–14275 (2022).

41. Zhou, P., Zou, J., Tian, F. & Shang, Z. Fluorine Bonding — How Does It Work In Protein−Ligand Interactions? *J. Chem. Inf. Model.* **49**, 2344–2355 (2009).

42. Xing, L., Keefer, C. & Brown, M. F. Fluorine multipolar interaction: Toward elucidating its energetics in binding recognition. *Journal of Fluorine Chemistry* **198**, 47–53 (2017).

43. Pollock, J. *et al.* Rational Design of Orthogonal Multipolar Interactions with Fluorine in Protein–Ligand Complexes. *J. Med. Chem.* **58**, 7465–7474 (2015).

44. Herbst, R. S., Fukuoka, M. & Baselga, J. Gefitinib — a novel targeted approach to treating cancer. *Nat Rev Cancer* **4**, 956–965 (2004).

45. Yu, H. A. *et al.* Local therapy with continued EGFR tyrosine kinase inhibitor therapy as a treatment strategy in EGFR-mutant advanced lung cancers that have developed acquired resistance to EGFR tyrosine kinase inhibitors. *J Thorac Oncol* **8**, 346–351 (2013).

46. Yu, H. A. *et al.* Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res* **19**, 2240–2247 (2013).

47. Schmid, S., Li, J. J. N. & Leighl, N. B. Mechanisms of osimertinib resistance and emerging treatment options. *Lung Cancer* **147**, 123–129 (2020).

48. Structure-based methods for predicting target mutation-induced drug resistance and rational drug design to overcome the problem. *Drug Discovery Today* **17**, 1121–1126 (2012).

49. Watermeyer, J. M., Kröger, W. L., O'Neill, H. G., Sewell, B. T. & Sturrock, E. D. Probing the Basis of Domain-Dependent Inhibition Using Novel Ketone Inhibitors of Angiotensin-Converting Enzyme. *Biochemistry* **47**, 5942–5950 (2008).

50. Kim, E. E. *et al.* Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **117**, 1181–1182 (1995).

51. Steuber, H., Heine, A., Podjarny, A. & Klebe, G. Merging the Binding Sites of Aldose and Aldehyde Reductase for Detection of Inhibitor Selectivity-determining Features. *Journal of Molecular Biology* **379**, 991–1016 (2008).

52. Vanichtanankul, J. *et al.* Combined Spatial Limitation around Residues 16 and 108 of Plasmodium falciparum Dihydrofolate Reductase Explains Resistance to Cycloguanil.

Antimicrobial Agents and Chemotherapy **56**, 3928–3935 (2012).

53. Finer-Moore, J. S. *et al.* Partitioning Roles of Side Chains in Affinity, Orientation, and Catalysis with Structures for Mutant Complexes: Asparagine-229 in Thymidylate Synthase. *Biochemistry* **35**, 5125–5136 (1996).

54. Caravella, J. A. *et al.* Structure-Based Design and Identification of FT-2102 (Olutasidenib), a Potent Mutant-Selective IDH1 Inhibitor. *J. Med. Chem.* **63**, 1612–1623 (2020).

55. Hallin, J. *et al.* Anti-tumor efficacy of a potent and selective non-covalent KRASG12D inhibitor. *Nat Med* **28**, 2171–2182 (2022).

56. Wang, E. *et al.* Mechanisms of Resistance to Noncovalent Bruton's Tyrosine Kinase Inhibitors. *New England Journal of Medicine* **386**, 735–743 (2022).

57. Anderson, P. W. More Is Different. *Science* **177**, 393–396 (1972).

58. Wei, J. *et al.* Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research* (2022).

59. Chen, T., Li, M. & Liu, J. π–π Stacking Interaction: A Nondestructive and Facile Means in Material Engineering for Bioapplications. *Crystal Growth & Design* **18**, 2765–2783 (2018).

60. Xu, X., Jia, K., Qi, Q., Tian, G. & Xiang, D. Regulation of π–π interactions between single aromatic molecules by bias voltage. *Phys. Chem. Chem. Phys.* **26**, 14607–14612 (2024).

61. Heredia, V. V., Cooper, W. C., Kruger, R. G., Jin, Y. & Penning, T. M. Alanine Scanning Mutagenesis of the Testosterone Binding Site of Rat 3α-Hydroxysteroid Dehydrogenase Demonstrates Contact Residues Influence the Rate-Determining Step. *Biochemistry* **43**, 5832–5841 (2004).

62. Zou, H. *et al.* Molecular Insight into the Interaction between IFABP and PA by Using MM−PBSA and Alanine Scanning Methods. *J. Phys. Chem. B* **111**, 9104–9113 (2007).

63. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).

64. Landrum, G. RDKit: Open-source cheminformatics from machine learning to chemical registration. *RDKit* https://rdkit.org/docs/source/rdkit.Chem.Scaffolds.rdScaffoldNetwork.html (2019).

65. Schindler, C. E. M. *et al.* Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **60**, 5457–5474 (2020).

66. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* **40**, D1100–D1107 (2012).

67. Yang, Z. *et al.* A mutation-induced drug resistance database (MdrDB). *Commun Chem* **6**, 123 (2023).

68. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. (2023).

69. Elnaggar, A. *et al.* ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**, 7112–7127 (2022).

70. Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction. Preprint at https://doi.org/10.48550/arXiv.2010.09885 (2020).