SE(3) Equivariant Topologies for Structure-based Drug Discovery

Alvaro Prat†
aprat@ro5.ai

Hisham Abdel Aty†*
habdel-aty@ro5.ai

Aurimas Pabrinkis
apabrinkis@ro5.ai

Orestis Bastas
obastas@ro5.ai

Tanya Paquet
tpaquet@ro5.ai

Gintautas Kamuntavičius
gkamuntavicius@ro5.ai

Roy Tal
rtal@ro5.ai

AI Chemistry, Ro5
2801 Gateway Drive, Irving, 75063, TX, USA
† Equal contributions

Abstract

Modeling protein-ligand interactions is a challenging task that has been approached through an array of perspectives. From physics-based computational approaches to vast deep learning pipelines, *in silico* methods hold promise in reducing experimental overhead in the otherwise tedious and costly drug discovery campaigns. We introduce Protein-Ligand Equivariant Transformer (ProLET), a generalizable model built upon chemically inspired SE(3) equivariant geometric deep learning. We evaluate ProLET on a wide range of established standards, including the notoriously difficult PoseBusters and Merck's FEP benchmarks, consistently demonstrating superior performance in binding affinity prediction and pose estimation. We demonstrate its effectiveness across different stages in drug discovery, showing that ProLET can be used for lead optimization and hit identification as well as for prioritizing compounds that are selective towards a desired target. By bridging the gap between accuracy, efficiency, and generalizability, ProLET stands as a powerful and adaptive resource, signifying a step towards safe and reliable AI-driven drug discovery.

1 Introduction

The discovery of new drugs is a complex and expensive process involving multiple stages including target identification, lead discovery, lead optimization, and clinical trials. Typically, it is estimated that the average drug discovery campaign may take up to 15 years and cost billions of dollars [1]. In early-stage drug discovery, it is crucial to identify diverse molecular candidates that are potent, novel and selective, as this can help reduce the risk of side effects and increase the likelihood of success in clinical trials [2]. Across these stages, large-scale assays are typically run to empirically quantify the hit rates and dose-response curves of potential drug candidates towards a selected target protein. Known metabolic targets are also tested within the pipeline for collateral binding (selectivity) and

*Corresponding Author

Preprint. Under review.

to unmask the toxic and metabolic implications. However, running large-scale assays is laborious and oftentimes economically infeasible, which limits the number of compounds that can be tested and significantly slows down the drug discovery process. Therefore, given the rise of discovered diseases [3], there is an ever increasing need for accurate accelerated *in silico* solutions which can help identify novel promising drug candidates and reduce the need for high-throughput laboratories.

With the aim to address this issue, data-driven and computational structure-based drug discovery (SBDD) alternatives have gained significant attention over the last decade, particularly within the last few years [4, 5, 6, 7, 8]. Established physics-based SBDD methods such as molecular docking and data-driven quantitative structure activity relationship (QSAR) models have proven resourceful across drug discovery campaigns [9]. However, these often fail short from generalizing outside their domain [10, 11]. On the other hand, more accurate computational techniques such as molecular dynamics (MD) and free-energy perturbation (FEP) simulations have shown superior performance and robustness [12, 7]. However, whilst these methods have demonstrated resourcefulness in providing insights into the binding mechanism and ranking of potent binders, they are too computationally demanding and require significant domain knowledge. In turn, this limits their applicability to large-scale drug discovery campaigns.

Biologically, protein-ligand interactions are fundamental to explaining processes such as signal transduction [13], immune response [14], and enzymatic catalysis [15]. Understanding these interactions is crucial for drug discovery, as you can consequently design modulators (ligands) which alter the activity of a desired target [16]. In light of this, the estimation of binding modes and corresponding affinities are two key tasks in computational drug discovery. Recently, there has been a surge of deep learning models developed to address these tasks [17, 18, 19, 4, 6, 20, 21, 22, 23, 24, 25]. However, due to a lack of generalization, bridging these models in practice has been short from arduous [26].

Whilst some models achieve state-of-the-art results on some benchmarks, they often fail to generalize to others. One of the reasons for the lack of generalization is the use of simple feature reduction methods, such as pairwise atomic environments, to achieve translational and rotational invariances which are not built within the model architecture [25, 27]. In addition to rigorous and practical evaluation protocols, another well established issue in SBDD is the elusion of protein-specific information. Particularly, these models often fail to learn from protein-ligand interactions and instead learn to identify promiscuous binders that bind to multiple targets [28]. This is a major issue in drug discovery as it can lead to side effects and toxicity.

Geometric deep learning (GDL) has emerged as a powerful framework for learning from structured data, such as graphs and point clouds [29]. GDL atomistic models are designed to manipulate Euclidean geometries, exploiting the symmetries present in the data. This makes them particularly well-suited for learning invariant properties in multi-body physical systems. In this work, we propose Protein-Ligand Equivariant Transformer (ProLET), a novel chemically-inspired framework for protein-ligand modelling. Based on the principles of equivariant GDL, ProLET exploits irreducible representations to model symmetries present in multibody physical systems. We base our core architecture on Equiformer [30], an extension of the SE(3) Transformer [31], and perform significant modifications to adapt it to structural chemistry.

Our main contributions can be summarised as follows: (i) We introduce ProLET, a novel chemically inspired SE(3) equivariant transformer for protein-ligand binding modelling; (ii) We evidence state-of-the-art performance across various benchmarks, including CASF-2016, PoseBusters and the Merck FEP benchmark, consistently outperforming existing methods; (iii) We introduce two novel case studies to asses ProLET's ability to find selective compounds across competing protein target pairs; (iv) We utilize heterogeneous training data sources, including re-docked (RD) poses and molecular dynamics (MD) trajectories and highlight their independent significance.

2 Methods

2.1 ProLET

In structure-based drug discovery, small molecule binders (ligands) and their target proteins are often co-crystallized to discern their corresponding binding mode and elucidate the predominant interactions within the complex which may lead to a certain activity or binding outcome. However, this does not scale and is oftentimes impossible due to experimental difficulties [32]. In practice, a

virtual structure-based drug discovery pipeline should be able to generate probable poses *in silico*. Finding the correct binding mode is critical towards estimating the affinity of a ligand and is a key first step in SBDD (especially in later stages such as lead optimization). In this paper we deconstruct the task of pose estimation into two distinct tasks: (i) pose generation; (ii) pose discrimination. For the former, we use both physics-based and AI-based pose generation schemes to generate distributions of protein-ligand complexes, which we refer to as pose ensembles. For each conformation $c_i \in \mathcal{P}$ in the pose ensemble we model the pose probability p_i as well as its associated affinity a_i. Pose ensembles generated at the binding site are used to evaluate the overall affinity a_{ens} of a ligand. Here we leverage inference from sets via Boltzmann-weighted averaging to better capture entropic effects (Eq. 1), where k is the temperature of the operator.

$$a_{ens} = \sum_{i \in P} \frac{a_i e^{f_i}}{\sum_{i \in \mathcal{P}} e^{f_i}}, f_i = -\frac{p_i}{k} \tag{1}$$

2.2 Chemically Inspired Geometric Deep Learning

We model the protein-ligand complex as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes $v_i \in \mathcal{V}$, and edges $(e_i, e_j) \in \mathcal{E}, \forall i \in \mathcal{N}(j)$, where $\mathcal{N}(\cdot)$ represents a neighbourhood function. Nodes are assigned to all atoms, featurized with a vector defining its relative position from the ligand center of mass (CoM). Atomic nodes are furthermore described as one-hot vectors over the proton number, as well as a one-hot representation describing the owning species, *i.e.* if it belongs to the protein or the ligand.

We build 3 distinct families of edges to create our chemically-inspired Euclidean graph: bond ($\mathcal{E}_b$), interaction ($\mathcal{E}_i$) and structural ($\mathcal{E}_s$), where $\{\mathcal{E}_b \cup \mathcal{E}_i \cup \mathcal{E}_s\} = \mathcal{E}$. Correspondingly, each edge type has a unique neighbourhood lookup function: (i) Bond edges are defined by the chemical bond across neighbouring atoms. Additionally, these are augmented with the bond type (single, double, triple, aromatic) as well as the owning species; (ii) Interaction edges are defined by a radial distance function from nodes in the ligand graph to nodes in the protein graph. As the number of interactions scales cubically with distance, we set a maximum interaction threshold of 5Å. To reduce risk of oversquashing [33], we select at most 3 interaction edges per ligand node, prioritizing the closest heavy protein atoms; (iii) Structural edges are defined across all protein alpha carbons ($C\alpha$) residing within 15Å from the CoM, creating a convex hull over the complex. We add these edges to provide a more informed representation of the binding pocket, facilitating global message passing and improving the topological convergence towards the ligand nodes. In a way, these edges can be viewed as structure-informed skip connections [34, 35]. For all edges, a one-hot vector describing the edge family is also added as an additional feature. A schematic of the presented structure-based pipeline can be found in Fig. 1a. Further details on the architecture are included in A.2.

2.3 SE(3) Invariant Transformer for Protein-Ligand Modelling

As physical objects in space, protein-ligand complexes can be described using atomistic coordinate systems. Coordinate systems can be freely chosen and transformed using group actions from the Euclidean group (E(3)): translations, rotations and inversions. These can be separated into sub-groups: translations and rotations form SE(3), rotations alone form SO(3), and inversions form O(3). Since the pose or affinity of a molecular complex is invariant to the first two groups, yet not always to the last one due to enantiomeric specificity, we seek for a model that is SE(3) equivariant. In doing so, symmetries in the data can be inductively captured by the model, reducing the unnecessary complexity of learning these via data augmentation techniques. Formally, a function mapping between vector spaces X and Y is equivariant to a group of transformations G if, for any input $x \in X$, output $y \in Y$, and group element $g \in G$, the function satisfies $g \cdot f(x) = f(g \cdot x)$.

Group representations in the context of the 3D Euclidean group E(3) involve transformations acting on different quantities, such as scalars and Euclidean vectors, which may change under rotations (sign changes for vectors, scalars are invariant). Irreducible representations (irreps) of SO(3) are smaller decompositions of these group representations into Wiegner-D matrices, acting on independent vector spaces of different angular frequencies, denoted by degree L [36]. These irreps are formed by concatenating type-L vectors, which capture equivariant information under rotations in SO(3), enabling the analysis of geometric properties in space.

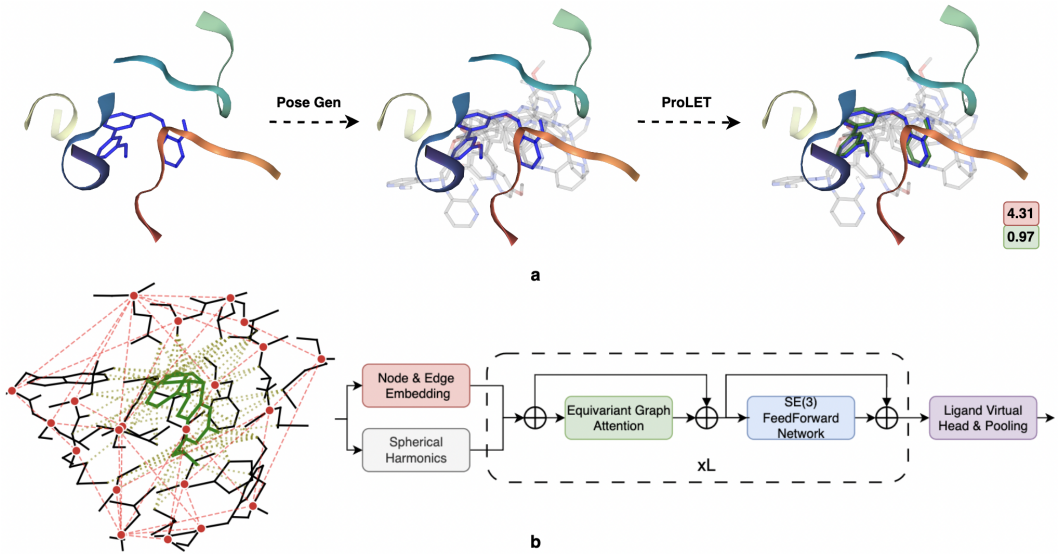

Figure 1: ProLET system framework. Figure **a** illustrates the defined structure-based pipeline: (i) A pose generation protocol generates conformations of a query ligand in the pocket of 2OHS, resulting in a pose ensemble (depicted as the set of faded-out poses); (ii) ProLET predicts the pose probability and corresponding affinity of each pose. A Boltzmann-weighted aggregate returns the predicted affinity of the pose ensemble (red square). The best predicted pose (green conformation) and its predicted probability (green square) are left for the reader. The original crystal structure of the query ligand is depicted in blue. Figure **b** presents an instance of the protein-ligand geometric graph generated via ProLET (left), where interaction edges (green dotted lines), structural edges (red dotted lines) and bond edges (black (protein) and green (ligand) lines) are displayed. A simplified version of ProLET's SE(3) Equivariant model architecture is shown on the right.

Positional Euclidean vectors $r_{ij} \in \mathrm{R}^3$ across pairwise nodes $(v_i, v_j) \in \mathcal{V}$ are transformed into irreducible representations via spherical harmonics. This generates the first layer of irreducible representations, which are concatenated at each node with their corresponding scalar irreps (atomic number embedding, node family embedding, etc.). Instead of linear matrix multiplication, separable tensor products across irreps u and v of degree l_1 and l_2 can be formally defined in a computationally efficient manner across each layer (Eq. 2), where $\mathcal{C}$ denotes the pre-computed Clebsch-Gordan coefficients, and m denotes the m-th component of the irrep. In our setup, we use a maximum degree of 2: $\forall l_i \in l, i \in \{0, 1, 2\}$ stemming from the spherical harmonic feature in the first layer.

$$(u \otimes v)^l_m = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} \mathcal{C}^{l,m}_{(l_1,m_1)(l_2,m_2)} u^{l_1}_{m_1} v^{l_2}_{m_2} \tag{2}$$

We create the SE(3) Invariant Transformer for molecular modelling by adapting the works from [37] and [31]. Specifically, we work on top of Equiformer [30] and use their separable tensor product attention with the following major modifications: (i) Instead of a radius graph built on distance neighbourhoods, we use the chemically inspired procedure outlined in section 2.2, drastically improving message passing efficiency and reducing over-squashing; (ii) We create a virtual node over ligand atoms only, removing the otherwise undesired dependency on local neighbourhood functions in the final aggregate operator. We believe this is imperative in this particular topology, as the proximity of a ligand to its binding site (strictly disjoint topologies) should not bias the final computational graph. For instance, as that the number of nodes scales cubically with the distance between a ligand conformation and a protein's binding site, this would lead to inconsistencies in scaling in the aggregate operator.

Note that in ProLET we maintain SE(3) equivariance across L stacked layers and achieve SE(3) invariance in the last layer by isolating the 0'th degree scalar contributions in the separable tensor product operator. Permutation invariance is ultimately achieved via a permutation invariant aggregate

over the ligand vector representations. Translational invariance is achieved by centering the coordinate system over the ligand CoM. See Figure 1b for an overview of the proposed model architecture.

2.4 Data

ReDocked. One of the most established structural datasets is PDBbind (version 2020) [38, 39], with its corresponding held-out benchmark subset, CASF-2016 [8]. PDBbind contains a curated set of $19k+$ protein-ligand crystal structures with empirically-obtained binding affinities: dissociation constants, K_i, K_d, or inhibition constant IC_{50} measurements. BindingMOAD [40] offers a more extensive list of ligand binders. Although (usually) devoid of affinity labels, BindingMOAD offers an additional $30k+$ structural complexes. For all datasets, we filter and pre-process the structures to remove any ligands/proteins with missing atoms or residues. We extract the pocket residues (any residue within 10Å from any ligand atom) from the protein structure and use these to define the protein graph for training.

Recent works demonstrate the benefit of including re-docked poses to augment the training data [6, 41]. Thus, we re-dock the crystal ligand in BindingMOAD and PDBbind using Smina [42]. Our docking pipeline generates poses for our pose ensemble both during training and inference. Pocket-defined docking is used with a bounding box of 5Å around the crystal ligand. We generate 100 poses per ligand, with a minimum root mean square deviation (RMSD) filter threshold of 0.25 Å from the crystal pose and an exhaustiveness of 8.

Molecular Dynamics. To further improve the generalization of the model and reduce potential memorization of static protein targets, we include protein-ligand poses generated through MD within the training set. The flexibility in both the protein and ligand aims to provide the model with a wider pool and a more representative distribution of the protein-ligand binding mode. We use the Misato dataset [43] which contains the MD simulations of a subset of 15k relaxed protein-ligand complexes from the PDBbind 2020 dataset. This constitutes our dynamic data sources used for training.

2.5 Training

We build our framework in Pytorch Geometric. We define our loss function by a weighted average of: (i) A smoothed cross-entropy loss between the predicted and true RMSD (Eq. 3), where $\phi : \mathrm{R} \rightarrow \mathrm{R}$ is an inverse softmax envelope over the RMSD offset by 2Å; (ii) A soft hinged mean square error between the predicted and true affinities (Eq. 4). We train our model using gradient descent with a batch size of 96 on an A10 GPU, 16 CPU machine until convergence ($\sim$ 20 epochs). We employ an AdamW optimizer with default parameters and an initial learning rate of $3e-4$ wrapped with a cosine annealing scheduler. See A.2 for more details.

$$\mathcal{L}_{ce}(\phi, \hat{\phi}) = -\big(\phi log(\hat{\phi}) + (1-\phi) log(1-\hat{\phi})\big) \tag{3}$$

$$\mathcal{L}_{aff}(a, \hat{a}, \phi) = ||\phi \cdot (a - \hat{a})||_2^2 \tag{4}$$

3 Results

ProLET is designed to be a versatile, generalizable tool applicable across various stages of drug discovery. In order to test the limits of our model, we assess its performance across a wide a range of established benchmarks for protein-ligand binding affinity and pose estimation. Namely, we evaluate ProLET against existing state-of-the-art (SOTA) methods leveraging both machine learning and physics-based propositions across three main benchmarks: CASF-2016 [8], PoseBusters [26] and the Merck FEP benchmark set [7]. We outline the importance of heterogeneous data for robust generalization and follow up with a case study on selectivity.

3.1 CASF-2016

The CASF-2016 benchmark consists of a subset of 285 held-out, carefully curated protein-ligand complexes from the PDBbind 2016 refined set. The affinities within this set consist of only high-quality K_i and K_d affinity values. Given the reportedly large similarity between protein targets in CASF-2016 and the rest of the PDBbind set [44], we use this benchmark as a preliminary indicator

of generalizability to unseen ligands. Primarily, we asses our performance by ranking scoring power with the benchmark's standard metrics: Pearson's correlation coefficient (r) and Root Mean Squared Error (RMSE) between the true and predicted affinity for each complex. We compare ProLET to prior art and present our results in Table 1.

Table 1: Comparative assessment of ProLET in the CASF-16 benchmark against prior art. For each model, inference is performed over the crystal (C), docked poses (D), or both. We use RMSE and Pearson's r to evaluate each entry.

Model	Year	Framework	Training	Inference	Pose Estimation	r	RMSE
Smina [42]	2013	Physics	—	C	T	0.55	—
MMGB-SA [45]	2015	Physics	—	C	T	0.65	—
GNINA [4]	2021	CNN	R	C	T	0.80	1.37
AEScore [27]	2021	MLP	R	C	F	0.83	1.22
PointTransformer [24]	2022	CNN + ATT	R	C	F	0.85	1.19
OnionNet-2 [23]	2021	CNN	R	C	F	**0.86**	1.16
ConBAP [46]	2024	GNN	R	C	T	**0.86**	1.13
Δ-AEScore [27]	2021	MLP	R	D & C	T	0.80	1.32
Hydrascreen [6]	2023	CNN	R	D & C	T	**0.86**	1.15
	2024		R + MD	D & C		**0.86**	**1.12**
ProLET	2024	SE(3) ATT	R + MD	C	T	**0.86**	**1.12**
	2024		R + MD	D		0.84	1.17

ProLET achieves top performance in both Pearson's r (0.86) and RMSE (1.12) across the 57 target clusters within this benchmark, with an average RMSE per cluster of 1.06 and a corresponding standard deviation of 0.41. In the more practical scenario where we employ docked poses (D) instead of crystal poses (C), ProLET still performs well, evidencing its suitability for virtual screening, where the true binding mode of a candidate ligand is unknown. Note that there are no reported values for inference via docked-only poses in literature, marking this an important result. However, although this suggests high generalization across protein targets, this result is a necessary yet insufficient indication to evaluate the practical implications of ProLET in finding potent drug candidates.

3.2 FEP Benchmark

The FEP benchmark is a collection of prospective FEP calculations [47] from Schrödinger's FEP+ workflow [48], and Molecular Mechanics Generalized Born Surface Area (MMGB-SA) calculations over multiple ligands across 8 different protein pharmaceutically relevant targets. As is common in FEP procedures, for each of the target proteins, the corresponding N ligands share a high degree of similarity (common scaffold), making this benchmark particularly difficult as the model needs to distinguish small changes in the ligand structure and its corresponding interactions. We compare ProLET to FEP+, MMGB-SA and GLIDE (Schrödinger's off-the-shelf docking tool) [49, 50]. FEP+ is a gold standard method for binding affinity prediction. However it is highly unscalable, requires significant domain knowledge, and is often prohibitively expensive.

As shown in Table 2, ProLET significantly outperforms industry-standard GLIDE and MMGB-SA scoring functions overall, with a ρ value increase of 63% and 42% respectively. Noteworthy is that ProLET is also superior to FEP+ on CDK8, EG5 and TNKS2 (R^2), all of which belong to different protein families. The robustness of our approach in this particularly difficult task suggests its suitability in accelerated and trustworthy lead optimization, where small changes in molecular structure often lead to large changes in biological activity.

Table 2: Performance comparison across 8 targets from the Merck FEP benchmark set [7]. The results are reported in terms of coefficient of determination (R^2), Spearman's correlation (ρ) and pairwise Root Mean Squared Error (RMSE). Averaged results contain 1σ standard deviations.

TARGET	N	FEP+			GLIDE			MMGB-SA			PROLET		
		R^2	ρ	RMSE	R^2	ρ	RMSE	R^2	ρ	RMSE	R^2	ρ	RMSE
CDK8	33	0.38	0.74	2.09	0.28	0.55	1.67	**0.60**	0.82	7.03	0.58	**0.85**	**1.29**
C-MET	24	**0.81**	**0.88**	**1.43**	0.34	0.56	2.16	0.36	0.64	5.96	0.32	0.59	1.86
EG5	28	0.50	**0.72**	1.23	0.01	-0.21	2.02	0.02	0.10	10.09	**0.57**	0.66	**0.68**
HIF-2α	42	**0.37**	**0.59**	1.60	0.15	0.41	1.56	0.29	0.48	11.69	0.21	0.52	**1.03**
PFKFB3	40	**0.63**	**0.79**	1.78	0.23	0.48	1.41	0.25	0.54	6.99	0.31	0.56	**1.94**
SHP-2	26	**0.50**	**0.78**	1.39	0.54	0.64	**1.05**	0.36	0.50	8.76	0.44	0.60	1.18
SYK	44	**0.25**	**0.42**	1.61	0.01	-0.02	1.49	0.00	-0.12	15.81	0.12	0.37	**0.78**
TNKS2	27	0.16	0.41	2.20	0.22	0.41	1.29	0.07	0.22	7.9	**0.32**	**0.54**	**0.90**
TOTAL	264	$0.44^{.64}_{.25}$	$0.65^{.79}_{.44}$	$1.68^{1.76}_{1.59}$	$0.20^{.41}_{.07}$	$0.33^{.54}_{.05}$	$1.57^{1.65}_{1.49}$	$0.24^{.41}_{.11}$	$0.38^{.59}_{.12}$	$9.72^{10.17}_{9.24}$	$0.34^{.49}_{.19}$	$0.58^{.71}_{.45}$	$1.21^{1.65}_{0.77}$

3.3 PoseBusters

The PoseBusters benchmark set acts as our temporal-split test set. It consists of 308 protein-ligand complexes with novel protein sequences from 2021 onwards. The task is to predict generated ligand pose that resembles the true binding mode of the ligand (RMSD ≤ 2). It is notoriously difficult for machine learning models to perform well on this benchmark due to it containing: (i) proteins with novel sequences which have not been trained on; (ii) a high proportion of co-binding structures where a substrate and the ligand co-exist within the binding pocket. Furthermore, each pose provided by the pose generation tool is first validated both sterically and energetically, ruling out a large portion of the suggested poses. Further details can be found in [26]. We compare ProLET to a wide range of SOTA deep-learning and physics-based approaches in Fig. 2.

In contrast to other DL methods, ProLET marks a significant improvement in pose estimation over physics-based approaches. Using docked poses generated from standard computational chemistry software, ProLET achieves a top-1 score of 74%, where the second best DL model (Diffdock) reports a 14% success rate over valid poses. Noteworthy is that ProLET is agnostic to the technique used to generate poses (Smina vs Physics), so long as these are chemically valid and are generated in the binding pocket. Nevertheless, generating poses from different protocols improves the diversity in the pose ensembles and in turn, ProLET's ability to identify the right binding mode.

We further explored the transferability of ProLET to AI-generated pose distributions extracted from Diffdock. Interestingly, although ProLET has not been trained on blind docked structures, we find that re-scoring via ProLET signifies an increase in top-1 performance of 180% with respect to Diffdock's original confidence scoring function. However, for the invalid poses, ProLET performs slightly worse than Diffdock. In addition to the frequent steric clashes and self-intersections present in Diffdock-generated poses [26], a large portion of false positives are picked up far away from the

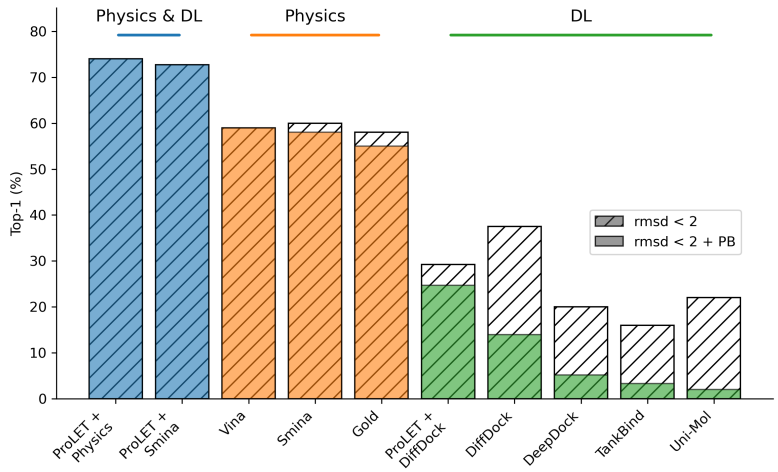

Figure 2: PoseBusters benchmark across state-of-the-art deep learning (green), physics-based (orange), and combined (blue) baselines. On the bars, "Physics" consists of stacked Vina, Smina and Gold docked conformations whilst "Smina" consists of Smina poses with added Vina poses where Smina failed to generate a single pose (13%).

binding pocket. We believe fine-tuning and/or training in blind docking scenarios would help boost ProLET's ability to find adequate binding modes across the whole protein structure.

It is noteworthy that our reported scores account for the limitations of Smina and other physics-based approaches in generating poses, where pose ensembles of up to 100 diverse poses do not include a single pose below 2Å RMSD ($\sim$ 7% of all complexes). Since a good pose does not exist, the average Top-1 penalizes ProLET unfairly. However, when excluding complexes in which there were no accurate and valid poses generated, the performance is thus increased from 74% to 82%.

3.4 Dataset Specificity

To better understand the significance of heterogeneous data in our training, we independently evaluate 10 model ensembles trained on MD data, RD data and both MD + RD data. We study the contributions of each ensemble towards the FEP, CASF and Posebusters benchmarks. As observed in Fig. 3, adding MD trajectories to the RD set consistently improves the performance in all 3 benchmarks. From an incremental perspective, the results clearly indicate that MD trajectories are especially important in ensuring generalisation in affinity estimation, as indicated by the $\sim 30\%$ improvement in average ρ across the 8 targets in FEP. This is likely the case due to the regularisation imparted in the structural changes present across molecular dynamic simulations, reducing the otherwise common overfitting nature of SBDD models. On the other hand, the lack of discriminative power behind MD-trained models in Posebusters highlights the importance of decoy poses in training. As shown in section A.1, the distributions of RMSDs in the MD set are substantially lower than in RD. This naturally leads to a substantial increment in false positives when evaluated in redocked data due to the distribution shift.

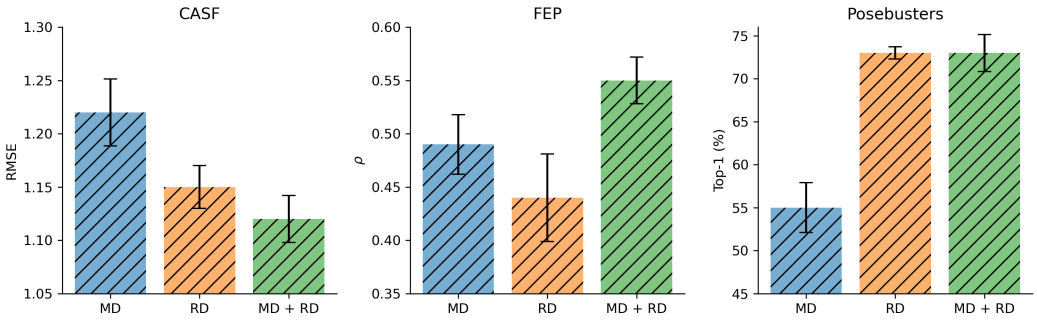

Figure 3: Comparative performance of ProLET trained using MD, RD or both datasets across the CASF, FEP and Posebusters benchmarks. Error bars denote a 1σ standard deviation bound computed from 10 separate model ensembles.

3.5 Selectivity

The ability to find ligands which are potent towards a desired protein target, whilst remaining inactive towards undesired target(s) is paramount in successful drug discovery. Proteins within and across different protein families often have a high degree of sequence and structural similarity, making the design of selective ligands notoriously difficult. This is often a common problem in drug discovery, particularly in kinases which share an ATP binding site, where off-target effects can lead to not only adverse side effects, but also determine the drug's metabolism and pharmacokinetics [51]. To that end, we design two novel case studies and assess ProLET's ability to prioritize selective binders for a given desired target over an undesired target. True experimental affinities are extracted from the Kinome dataset produced by Metz et al. [52]. We remove ligands for which the affinity measurements are unspecified (*i.e.* < 5.7). For each case study, we compute the difference between the experimental binding affinity of a set of ligands towards a desired and undesired target (Δ_t), and compare the differences to the predicted difference returned by ProLET (Δ_p). We calculate the model's ability to discern selective compounds by means of Spearman's ρ between Δ_t and Δ_p. Given the heteroscedastic nature of the limited affinity data available for this study, we opt to use ρ as our main selectivity metric. For all pairwise measurements, we set a minimum difference threshold of 0.69, which is the average $log_{10}(K_i)$ experimental error in binding affinity measurements [53], and only include ligand pairs which are above this threshold. This allows us to compute accurate metrics (ranking ligands that have very similar activity is likely to be experimental noise).

LCK/EGFR. The first case study involves kinases LCK and EGFR, both members of the tyrosine kinase family. We extract two PDB structures, 3KMM and 5XDK, which are the inhibitor-bound crystal structures of LCK and EGFR respectively. We dock the corresponding set of ligands to the binding site (keeping all residues 10Å away from the reference co-crystallized ligand) via Smina, using the same protocol described in section 2.4.

ProLET achieves a ρ value of 0.53 on the LCK/EGFR pair, indicating considerable ranking power in selective compounds towards LCK. Figure 4 illustrates this positive correlation, where ProLET successfully discriminates between ligands which are highly selective towards LCK over EGFR (LCK++) and ligands which are slightly more selective towards EGFR over LCK (EGFR+). When looking at the top 4 compounds, ProLET identifies 3 strong ($\Delta_t \geq log_{10}(50)$) and 1 moderately selective ($0 \geq \Delta_t < log_{10}(50)$) candidate. Furthermore, out of the 10 highest rank candidates, only 1 compound is not selective towards LCK ($\Delta_t < 0$), whilst 5 are highly selective and 4 are moderately selective.

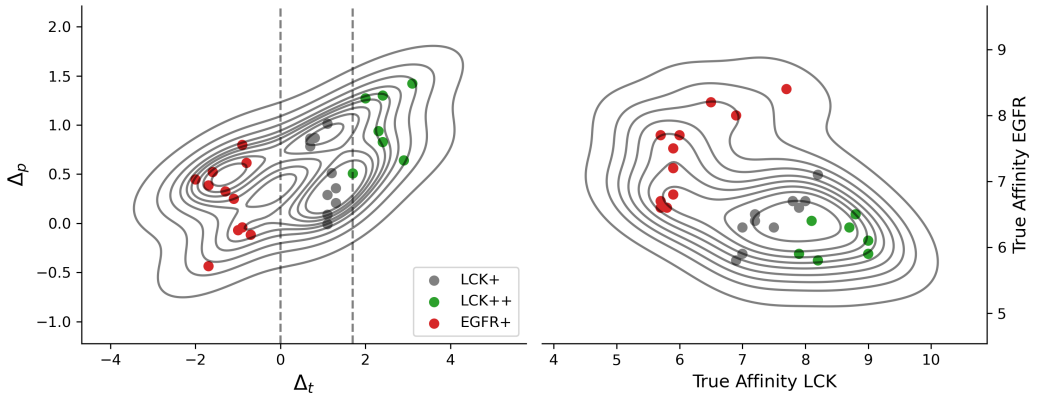

Figure 4: Overlay of Δ_t and Δ_p for active LCK/EGFR ligand pairs (left) and their corresponding affinities (right). Green, gray and red points correspond to highly selective, moderately selective and moderately unselective compounds for LCK.

CSF1R/PDGFRA. CSF1R and PDGFRA are two receptor kinases involved in cell growth and differentiation. We extract PDB structures 6T2W and 6JOL in a similar procedure as with the first case study. With this pair of proteins, ProLET achieves a similar ρ of 0.50. Given the highly correlated pairwise affinity between these two targets (narrow Δ_t), it is notably more difficult to find prominently selective ligands. As shown in Figure 6 in A.3, whilst the model is able to accurately rank selective ligands, it is less accurate than in the previous case study due to the lack of spread in the data.

4 Conclusion

In this study we introduce ProLET, a novel chemically inspired SE(3) Transformer for protein-ligand modelling. We propose alternatives to circumvent topological bottlenecks in conventional atomistic geometric deep learning and demonstrate our method's generalized ability to accurately predict protein-ligand binding affinities and identify promising binding modes.

We show that ProLET consistently outperforms existing physics-based and machine learning propositions across a range of established benchmarks. Particularly, our results signify a substantial boost in credibility for deep learning methods by surpassing prior art in the PoseBusters set, the otherwise "achiles heel" of ML-driven structure-based models. Moreover, in addition to leading the CASF-16 benchmark and demonstrating comparable performance without bound structures, we demonstrate impressive results in Merck's FEP collection, closely matching and occasionally improving upon FEP+, the industry leading method for lead optimization. Finally, we highlight the importance of combining docked poses and molecular dynamic trajectories in the training data and verify the suitability of ProLET in the complex and understudied task of target selectivity. Overall, these indications evidence ProLET's suitability in accelerating the otherwise costly and lengthy search of potent and diverse drug candidates.

Allowing faster development of new medications and a "fail fast" approach, ProLET could be employed to address unmet medical needs and improve public health outcomes. By reducing the reliance on extensive early experimental procedures which are often inconsistent, ProLET can not only lower the costs associated with drug development, but also reduce the environmental impact from running large scale experiments in a laboratory, making medications more affordable and accessible to a broader population either by allowing the development of cheaper drugs, or through re-purposing of current drugs. Furthermore, enhanced selectivity in compound prioritization can lead to more effective and tailored treatments, improving patient outcomes and reducing adverse effects.

References

[1] Steven M. Paul, Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg, and Aaron L. Schacht. How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews. Drug Discovery*, 9(3):203–214, March 2010.

[2] Shan-Liang Sun, Shi-Han Wu, Ji-Bo Kang, Yi-Yuan Ma, Lu Chen, Peng Cao, Liang Chang, Ning Ding, Xin Xue, Nian-Guang Li, and Zhi-Hao Shi. Medicinal Chemistry Strategies for the Development of Bruton's Tyrosine Kinase Inhibitors against Resistance. *Journal of Medicinal Chemistry*, 65(11):7415–7437, June 2022.

[3] Eric Ka-Wai Hui. Reasons for the increase in emerging and re-emerging viral infectious diseases. *Microbes and Infection*, 8(3):905–916, March 2006.

[4] Andrew T. McNutt, Paul Francoeur, Rishal Aggarwal, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. GNINA 1.0: Molecular docking with deep learning. *Journal of Cheminformatics*, 13(1):43, June 2021.

[5] Rocco Meli, Garrett M. Morris, and Philip C. Biggin. Scoring Functions for Protein-Ligand Binding Affinity Prediction Using Structure-based Deep Learning: A Review. *Frontiers in Bioinformatics*, 2, June 2022.

[6] Alvaro Prat, Hisham Abdel Aty, Gintautas Kamuntavičius, Tanya Paquet, Povilas Norvaišas, Piero Gasparotto, and Roy Tal. HydraScreen: A Generalizable Structure-Based Deep Learning Approach to Drug Discovery, September 2023.

[7] Christina E. M. Schindler, Hannah Baumann, Andreas Blum, Dietrich Böse, Hans-Peter Buchstaller, Lars Burgdorf, Daniel Cappel, Eugene Chekler, Paul Czodrowski, Dieter Dorsch, Merveille K. I. Eguida, Bruce Follows, Thomas Fuchß, Ulrich Grädler, Jakub Gunera, Theresa Johnson, Catherine Jorand Lebrun, Srinivasa Karra, Markus Klein, Tim Knehans, Lisa Koetzner, Mireille Krier, Matthias Leiendecker, Birgitta Leuthner, Liwei Li, Igor Mochalkin, Djordje Musil, Constantin Neagu, Friedrich Rippmann, Kai Schiemann, Robert Schulz, Thomas Steinbrecher, Eva-Maria Tanzer, Andrea Unzue Lopez, Ariele Viacava Follis, Ansgar Wegener, and Daniel Kuhn. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *Journal of Chemical Information and Modeling*, 60(11):5457–5474, November 2020.

[8] Minyi Su, Qifan Yang, Yu Du, Guoqin Feng, Zhihai Liu, Yan Li, and Renxiao Wang. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *Journal of Chemical Information and Modeling*, 59(2):895–913, February 2019.

[9] Gintautas Kamuntavičius, Alvaro Prat, Tanya Paquet, Orestis Bastas, Hisham Abdel Aty, Qing Sun, Carsten B. Andersen, John Harman, Marc Siladi, Dan Rines, Sarah J. L. Flatters, Roy Tal, and Povilas Norvaisas. Accelerated Hit Identification with Target Evaluation, Deep Learning and Automated Labs: Prospective Validation in IRAK1, January 2024.

[10] Jack Scantlebury, Lucy Vost, Anna Carbery, Thomas E. Hadfield, Oliver M. Turnbull, Nathan Brown, Vijil Chenthamarakshan, Payel Das, Harold Grosjean, Frank von Delft, and Charlotte M. Deane. A Small Step Toward Generalizability: Training a Machine Learning Scoring Function for Structure-Based Virtual Screening. *Journal of Chemical Information and Modeling*, 63(10):2960–2974, May 2023.

[11] Viet-Khoa Tran-Nguyen, Célien Jacquemard, and Didier Rognan. LIT-PCBA: An Unbiased Data Set for Machine Learning and Virtual Screening. *Journal of Chemical Information and Modeling*, 60(9):4263–4273, September 2020.

[12] Negin Forouzesh and Nikita Mishra. An Effective MM/GBSA Protocol for Absolute Binding Free Energy Calculations: A Case Study on SARS-CoV-2 Spike Protein and the Human ACE2 Receptor. *Molecules*, 26(8):2383, January 2021.

[13] Lester A. Rubenstein, Randy J. Zauhar, and Richard G. Lanzara. Molecular dynamics of a biophysical model for $B2$-adrenergic and G protein-coupled receptor activation. *Journal of Molecular Graphics and Modelling*, 25(4):396–409, December 2006.

[14] Erick Lu and Jason G. Cyster. G-protein coupled receptors and ligands that organize humoral immune responses. *Immunological Reviews*, 289(1):158–172, 2019.

[15] Pedro Ojeda-May, Ameeq UI Mushtaq, Per Rogne, Apoorv Verma, Victor Ovchinnikov, Christin Grundström, Beata Dulko-Smith, Uwe H. Sauer, Magnus Wolf-Watz, and Kwangho Nam. Dynamic Connection between Enzymatic Catalysis and Collective Protein Motions. *Biochemistry*, 60(28):2246–2258, July 2021.

[16] Alberto Schena, Rudolf Griss, and Kai Johnsson. Modulating protein activity using tethered ligands with mutually exclusive binding sites. *Nature Communications*, 6(1):7830, July 2015.

[17] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking, February 2023.

[18] Derek Jones, Hyojin Kim, Xiaohua Zhang, Adam Zemla, Garrett Stevenson, W. F. Drew Bennett, Daniel Kirshner, Sergio E. Wong, Felice C. Lightstone, and Jonathan E. Allen. Improved Protein–Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference. *Journal of Chemical Information and Modeling*, 61(4):1583–1592, April 2021.

[19] Zhirui Liao, Ronghui You, Xiaodi Huang, Xiaojun Yao, Tao Huang, and Shanfeng Zhu. DeepDock: Enhancing Ligand-protein Interaction Prediction by a Combination of Ligand and Structure Information. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 311–317, November 2019.

[20] Jeongtae Son and Dongsup Kim. Development of a graph convolutional neural network model for efficient prediction of protein-ligand binding affinities. *PLOS ONE*, 16(4):e0249404, April 2021.

[21] Hannes Stärk, Octavian-Eugen Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction, June 2022.

[22] Zichen Wang, Ryan Brand, Jared Adolf-Bryfogle, Jasleen Grewal, Yanjun Qi, Steven A. Combs, Nataliya Golovach, Rebecca Alford, Huzefa Rangwala, and Peter M. Clark. EGGNet, a Generalizable Geometric Deep Learning Framework for Protein Complex Pose Scoring. *ACS Omega*, 9(7):7471–7479, February 2024.

[23] Zechen Wang, Liangzhen Zheng, Yang Liu, Yuanyuan Qu, Yong-Qiang Li, Mingwen Zhao, Yuguang Mu, and Weifeng Li. OnionNet-2: A Convolutional Neural Network Model for Predicting Protein-Ligand Binding Affinity Based on Residue-Atom Contacting Shells. *Frontiers in Chemistry*, 9, October 2021.

[24] Yeji Wang, Shuo Wu, Yanwen Duan, and Yong Huang. A point cloud-based deep learning strategy for protein–ligand binding affinity prediction. *Briefings in Bioinformatics*, 23(1):bbab474, January 2022.

[25] Fangqiang Zhu, Xiaohua Zhang, Jonathan E. Allen, Derek Jones, and Felice C. Lightstone. Binding Affinity Prediction by Pairwise Function Based on Neural Network. *Journal of Chemical Information and Modeling*, 60(6):2766–2772, June 2020.

[26] Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences, November 2023.

[27] Rocco Meli, Andrew Anighoro, Mike J. Bodkin, Garrett M. Morris, and Philip C. Biggin. Learning protein-ligand binding affinity with atomic environment vectors. *Journal of Cheminformatics*, 13(1):59, August 2021.

[28] R. Özçelik, D. van Tilborg, J. Jiménez-Luna, and F. Grisoni. Structure-Based Drug Discovery with Deep Learning**. *ChemBioChem*, 24(13):e202200776, 2023.

[29] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017.

[30] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs, February 2023.

[31] Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks, November 2020.

[32] Jessica Lynn Grey and David H Thompson. Challenges and opportunities for new protein crystallization strategies in structure-based drug design. *Expert Opinion on Drug Discovery*, 5(11):1039–1045, November 2010.

[33] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature, November 2022.

[34] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA, June 2016. IEEE.

[35] Kexin Huang, Cao Xiao, Lucas M. Glass, Marinka Zitnik, and Jimeng Sun. SkipGNN: Predicting molecular interactions with skip-graph networks. *Scientific Reports*, 10(1):21092, December 2020.

[36] Mildred S. Dresselhaus, Gene Dresselhaus, and Ado Jorio. *Group Theory: Application to the Physics of Condensed Matter*. Springer Science & Business Media, December 2007.

[37] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, May 2018.

[38] Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. *Journal of Medicinal Chemistry*, 47(12):2977–2980, June 2004.

[39] Renxiao Wang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. The PDBbind Database: Methodologies and Updates. *Journal of Medicinal Chemistry*, 48(12):4111–4119, June 2005.

[40] Mark L. Benson, Richard D. Smith, Nickolay A. Khazanov, Brandon Dimcheff, John Beaver, Peter Dresslar, Jason Nerothin, and Heather A. Carlson. Binding MOAD, a high-quality protein–ligand database. *Nucleic Acids Research*, 36(Database issue):D674–D678, January 2008.

[41] Paul G. Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B. Iovanisci, Ian Snyder, and David R. Koes. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. *Journal of Chemical Information and Modeling*, 60(9):4200–4215, September 2020.

[42] David Ryan Koes, Matthew P. Baumgartner, and Carlos J. Camacho. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. *Journal of Chemical Information and Modeling*, 53(8):1893–1904, August 2013.

[43] Till Siebenmorgen, Filipe Menezes, Sabrina Benassou, Erinc Merdivan, Kieran Didi, André Santos Dias Mourão, Radosław Kitel, Pietro Liò, Stefan Kesselheim, Marie Piraud, Fabian J. Theis, Michael Sattler, and Grzegorz M. Popowicz. MISATO: Machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature Computational Science*, pages 1–12, May 2024.

[44] Minyi Su, Guoqin Feng, Zhihai Liu, Yan Li, and Renxiao Wang. Tapping on the Black Box: How Is the Scoring Power of a Machine-Learning Scoring Function Dependent on the Training Set? *Journal of Chemical Information and Modeling*, 60(3):1122–1136, March 2020.

[45] Samuel Genheden and Ulf Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5):449–461, May 2015.

[46] Ding Luo, Dandan Liu, Xiaoyang Qu, Lina Dong, and Binju Wang. Enhancing Generalizability in Protein–Ligand Binding Affinity Prediction with Multimodal Contrastive Learning. *Journal of Chemical Information and Modeling*, 64(6):1892–1906, March 2024.

[47] Lingle Wang, Jennifer Chambers, and Robert Abel. Protein–Ligand Binding Free Energy Calculations with FEP+. In Massimiliano Bonomi and Carlo Camilloni, editors, *Biomolecular Simulations: Methods and Protocols*, pages 201–232. Springer, New York, NY, 2019.

[48] Lingle Wang, Yujie Wu, Yuqing Deng, Byungchan Kim, Levi Pierce, Goran Krilov, Dmitry Lupyan, Shaughnessy Robinson, Markus K. Dahlgren, Jeremy Greenwood, Donna L. Romero, Craig Masse, Jennifer L. Knight, Thomas Steinbrecher, Thijs Beuming, Wolfgang Damm, Ed Harder, Woody Sherman, Mark Brewer, Ron Wester, Mark Murcko, Leah Frye, Ramy Farid, Teng Lin, David L. Mobley, William L. Jorgensen, Bruce J. Berne, Richard A. Friesner, and Robert Abel. 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. *Journal of the American Chemical Society*, 137(7):2695–2703, February 2015.

[49] Richard A. Friesner, Robert B. Murphy, Matthew P. Repasky, Leah L. Frye, Jeremy R. Greenwood, Thomas A. Halgren, Paul C. Sanschagrin, and Daniel T. Mainz. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry*, 49(21):6177–6196, October 2006.

[50] Richard A. Friesner, Jay L. Banks, Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, Eric H. Knoll, Mee Shelley, Jason K. Perry, David E. Shaw, Perry Francis, and Peter S. Shenkin. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47(7):1739–1749, March 2004.

[51] Danzhi Huang, Ting Zhou, Karine Lafleur, Cristina Nevado, and Amedeo Caflisch. Kinase selectivity potential for inhibitors targeting the ATP binding site: A network analysis. *Bioinformatics*, 26(2):198–204, January 2010.

[52] James T. Metz, Eric F. Johnson, Niru B. Soni, Philip J. Merta, Lemma Kifle, and Philip J. Hajduk. Navigating the kinome. *Nature Chemical Biology*, 7(4):200–202, April 2011.

[53] Carlos A. Hernández-Garrido and Norberto Sánchez-Cruz. Experimental Uncertainty in Training Data for Protein-Ligand Binding Affinity Prediction Models. *Artificial Intelligence in the Life Sciences*, 4:100087, December 2023.

A Appendix

A.1 RMSD Distributions

The Root Mean Square Deviation (RMSD) stands as a measure of spatial distance between two identical molecular conformations. In our work we use a symmetry corrected RMSD calculation via the `RDKit` python package. We compare the distributions of RMSDs for the MISATO dataset (MD trajectories) and the Redocked dataset (re-docked generated poses in the pocket). Note that for MD trajectories, an aligned version of the crystal ligand is used to calculat the RMSD, accounting for dynamic changes in the pocket.

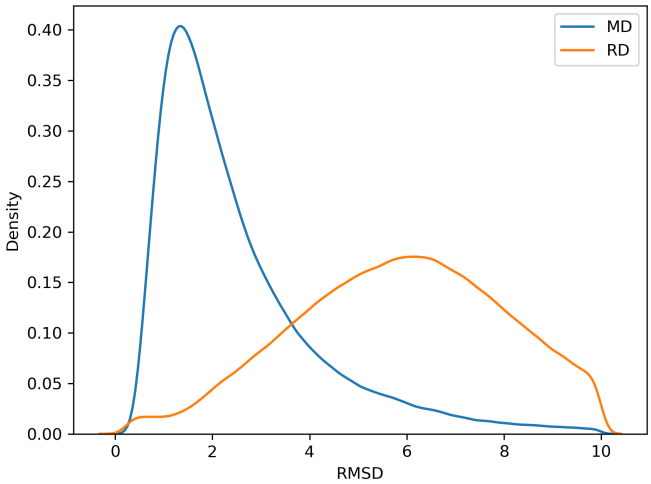

Figure 5: RMSD distributions for molecular dynamics (MD) trajectories vs. re-docked (RD) poses.

A.2 ProLET Architecture

We create an SE(3) transformer by stacking 6 layers of multi-head attention (4 attention heads). Each attention head contains a separable embedding composed of 3 irrep vectors of order different sizes `64x0e + 32x1o + 16x2e`, using the notation in [37].

In the first layer, node features consist of a concatenation of:

- L-0 with size $|n_{elem}|$
- L-1 with size $|n_{species}|$

where $n_{species}$ is a 2-dimensional one-hot vector defining the origin of the atom (ligand or protein) and n_{elem} is a a one-hot embedding of the element type, where $elem \in [H, C, N, O, P, S, F, Cl, Br, I]$. Note that separate embedding functions are used for each of these features.

Conversely, edges are initially featurized with scalar-only features:

- L-0 with size $|v_{bond}|$
- L-0 with size $|v_{species}|$

where v_{bond} is a 5-dimensional one-hot vector defining the bond type $bond \in$ [single, double, triple, aromatic, other], and $v_{species}$ is a 4-dimensional one-hot vector defining the species of the edge $species \in$ [ligand bond, protein bond, interaction, $a - C$ hull].

Spherical harmonics contributions (L-0: 1, L-1: 2, L-2: 1) calculated from separable radial (learnable) and angular (pre-computed) contributions are generated from the distance vector r_{ij} from nodes i and j and concatenated to the embedded edge features. With these, edge degree embeddings are

computed and concatenated to the 0-th degree node features, propagating the separable L-1 and L-2 irrep contributions. Via the separable tensor products, equivariant representations from the spherical harmonic features are propagated alongside the 0-th degree embeddings in the first layer.

In the last layer, learnable fully-connected separable tensor product (FCTsP) transform the aggregate SE(3) equivariant embedding into a 0-th degree SE(3) invariant irreps of dimension 256. These scalar logits are used as inputs to the loss functions during training using the following labels extracted from the training data.

- Pose (after a Sigmoid function)
- Affinity
- Energy (docking energy)
- RMSD

Note that both energy (Kcal/mol) and RMSD are only used as auxiliarly loss functions to help regularise the model during trianing.

A.3 Selectivity

As shown in Figure 6, the narrow and highly correlated distribution of pairwise affinities for the labelled ligands in the kinome dataset [52] makes it rather difficult to find selective compounds which are both potent towards CSF1R and not potent towards PDGFRA. However, ProLET still manages to rank these successfully: out of the 8 highest rank compounds, only 1 is moderately unselective whilst 2 are highly selective and 7 are moderately selective.

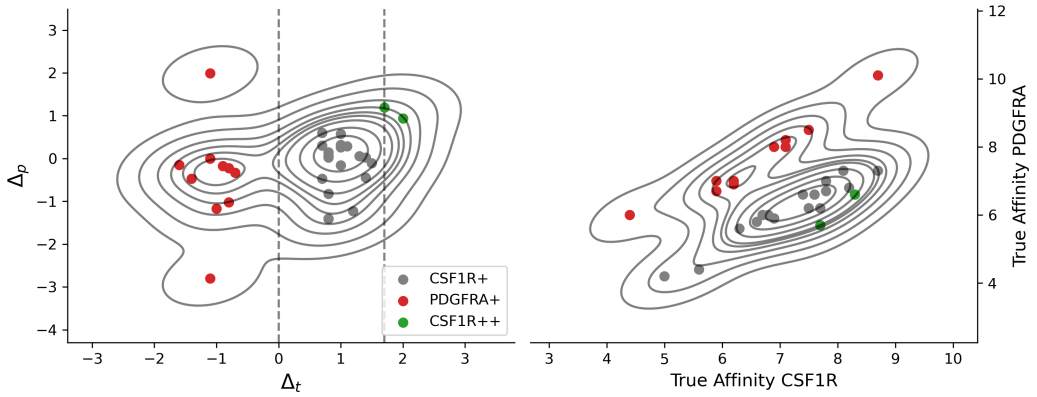

Figure 6: Overlay of Δ_t and Δ_p for active CSF1R/PDGFRA ligand pairs (left) and their corresponding affinities (right). Green, gray and red points correspond to highly selective, moderately selective and moderately unselective compounds for CSF1R.