# 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 binding is a complex challenge, with approaches ranging from physics-based simulations to advanced deep learning pipelines being actively explored. *In silico* methods are rapidly advancing, offering the potential to significantly reduce experimental overhead in the otherwise laborious 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. ProLET consistently outperforms existing methods in binding affinity prediction and pose estimation, excelling on challenging benchmarks such as PoseBusters and Merck's FEP. ProLET stands as a powerful and adaptive resource, addressing critical stages of drug discovery, including lead optimization and hit identification. Our approach marks a step forward towards targeted and accelerated development of novel therapeutics.

## 1 Introduction

Drug discovery is a lengthy and expensive endeavor, often taking up to 15 years and costing billions of dollars [1]. The high cost of large-scale assays limits the number of compounds that can be tested, highlighting the need for efficient *in silico* methods to accelerate and streamline the process. Data-driven and structure-based drug discovery (SBDD) methods, such as molecular docking and quantitative structure-activity relationship (QSAR) models [2, 3, 4, 5, 6, 7] have gained significant traction in the field. However, these methods tend to struggle with generalization [8, 9]. More accurate methods like molecular dynamics (MD) and free-energy perturbation (FEP) simulations [10, 5], though effective, are too computationally expensive for large-scale use. Recently, deep learning has shown promise in replacing traditional methods [11, 12, 13, 2, 4, 14, 15, 16, 17, 18, 19]. By learning and rationalizing over the key protein-ligand interactions which differentiate potent and weak drug candidates, these models aim to enable a more efficient design of modulators that affect

*Corresponding Author

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

target proteins [20, 7]. However, oftentimes these continue to struggle to generalize outside their training distribution and learn physics [21].

By leveraging data symmetries to better model these interactions [22], Geometric deep learning (GDL) offers a potential solution. To address the limitations of existing models, we introduce the Protein-Ligand Equivariant Transformer (ProLET), built on Equiformer [23], an extension of the SE(3) Transformer [24], with adaptations to structural chemistry. ProLET demonstrates superior performance on key benchmarks, consistently outperforming state-of-the-art methods in pose estimation and affinity prediction. It handles complex drug discovery tasks such as identifying selective binders across competing protein targets utilizing diverse training data from re-docked (RD) poses and molecular dynamics (MD) trajectories. This novel framework shows great potential in improving accuracy, efficiency, and generalizability of AI-driven drug discovery pipelines.

## 2 Methods

### 2.1 ProLET

Pose estimation is critical in estimating the affinity of a small molecule binder (ligand) and is a key component in SBDD. In this paper we distill pose estimation into two distinct tasks: (i) pose generation; (ii) pose discrimination. For the former, we use physics and AI-based generation schemes to generate distributions of protein-ligand complexes, which we refer to as pose ensembles. Following [4], 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 ensemble.

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

**Topological Deep Learning for SBDD.** 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. 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 $\{\mathcal{E}_b \cup \mathcal{E}_i \cup \mathcal{E}_s\} = \mathcal{E}$ and $\mathcal{E}_b \cap \mathcal{E}_i = \mathcal{E}_b \cap \mathcal{E}_s = \mathcal{E}_i \cap \mathcal{E}_s = \emptyset$. Correspondingly, each edge type has a unique neighbourhood lookup function: (i) In $\mathcal{E}_b$, bond edges are defined by the presence of a chemical bond across neighbouring atoms (single, double, triple, aromatic); (ii) Edges in $\mathcal{E}_i$ are defined by a radial distance function from nodes in the ligand graph to nodes in the protein graph. To reduce risk of over-squashing [25], we select at most 3 unique 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 12Å 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 topological convergence towards the ligand nodes, where global pooling is computed. This construction can be viewed as a proxy for efficient structure-informed skip-connections. The overall topology can be described as:

$$\mathcal{E} = \bigcup_{j} \bigcup_{i \in \mathcal{N}(j)} (\mathcal{E}_b(i,j) \cup \mathcal{E}_i(i,j) \cup \mathcal{E}_s(i,j)) \tag{2}$$

Finally, a one-hot scalar representation $\mathbf{e}(i,j)$ of the edges in $\mathcal{E}$ indicates the edge sub-type:

$$\mathbf{e}(i,j) = \left[\mathbb{I}\big((i,j) \in \mathcal{E}_b\big), \mathbb{I}\big((i,j) \in \mathcal{E}_i\big), \mathbb{I}\big((i,j) \in \mathcal{E}_s\big)\right] \tag{3}$$

A schematic of the presented structure-based pipeline can be found in Fig. 1a. Further details on the architecture, background an training are included in A.3 and A.2 and A.4.

**SE(3) Equivariant Transformer for Protein-Ligand Modeling.** Protein-ligand complexes can be described using atomistic coordinate systems, which can be transformed by translations, rotations, and inversions from the Euclidean group (E(3)). These transformations include SE(3) for translations and rotations, and O(3) for inversions. Since molecular pose or affinity remains unchanged under SE(3) but can vary with inversions, we focus on creating an SE(3)-equivariant model. This provides a strong inductive bias by allowing the model to naturally capture symmetries in the data, reducing the need for explicit data augmentation techniques. We create the SE(3) Invariant Transformer for molecular

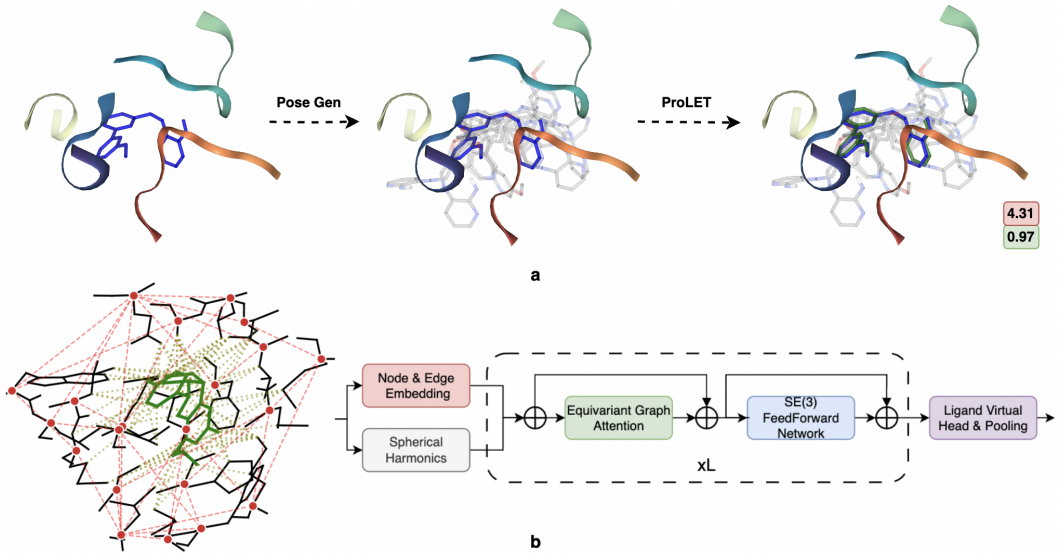


Figure 1: ProLET system framework. Fig. **a**: (i) generate pose ensembles in the 2OHS protein pocket; (ii) Predict pose probability and affinity; (iii) Compute affinity using a Boltzmann-weighted aggregate (red). The best predicted pose and its probability are highlighted (green), with the crystal ligand pose shown in blue. Fig. **b**: ProLET generates a protein-ligand geometric graph, displaying interaction (green), structural (red), and bond edges (black for protein, green for ligand). A simplified version of ProLET's SE(3) Equivariant model architecture is shown alongside.

modeling by adapting the works from [26] and [24]. Specifically, we work on top of Equiformer [23] and use their separable tensor product attention with the following chemically-inspired modifications: (i) Instead of a radius graph built on distance neighborhoods, we employ a topological alternative with separable radial basis functions, 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 neighborhood functions in the final aggregate operator (see 1b). 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.

### 2.2 Data

**ReDocked.** One of the most established structural datasets is PDBbind (version 2020) [27, 28], with its corresponding held-out benchmark subset, CASF-2016 [6]. 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 [29] offers additional $30k+$ ligand binders devoid of affinity labels. In both datasets we extract the binding pockets as in 2.1 and use these to define the coordinate space. In ProLET we create pose ensembles for training and inference by re-docking the crystal ligands using Smina [30]. Our docking pipeline generates up to 20 poses for our pose ensembles around the bounding box (5Å around the crystal ligand) with a minimum root mean square deviation (RMSD) filter threshold of 0.25Å from any pose.

**Molecular Dynamics.** To enhance model generalization and prevent memorization of static protein targets, we include protein-ligand poses from molecular dynamics (MD) simulations in the training set. This adds flexibility to both proteins and ligands, offering a broader, more representative distribution of binding modes. We use the Misato dataset [31], which contains MD simulations of 15k relaxed protein-ligand complexes from PDBbind 2020, as our dynamic data source for training.

## 3 Results

**CASF-2016.** The CASF-2016 benchmark consists of a subset of 285 held-out protein-ligand complexes from the PDBbind 2016 refined set. ProLET achieves top performance in both Pearson's

$r$ (0.86) and RMSE (1.12) across the 57 target clusters, with an average RMSE per cluster of 1.06 and standard deviation of 0.41. Although there is a high degree of protein similarity with the training set [4], a well established shortcoming of this benchmark, we demonstrate that ProLET is capable of maintaining a high level of accuracy when using docked poses only. In contrast, previous SBDD studies typically report affinity predictions obtained by additionally and/or solely considering the co-crystallized (true) pose. For more details, see section A.5.

**FEP Benchmark.** The FEP Benchmark is a collection of prospective FEP calculations [32] from Schrödinger's FEP+ workflow [33], and Molecular Mechanics Generalized Born Surface Area (MMGB-SA) calculations over multiple ligands across 8 pharmaceutically relevant protein targets. The high degree of similarity (common scaffold) across the $N$ ligands associates to each target makes this task particularly difficult, requiring high sensitivity to small structural changes. We compare ProLET to FEP+, MMGB-SA and GLIDE (Schrödinger's off-the-shelf docking tool) [34, 35]. FEP+ is a gold standard method for binding affinity prediction. However it is highly unscalable and requires significant domain knowledge. Table 1 shows ProLET significantly outperforming industry-standard GLIDE and MMGB-SA scoring functions overall, whilst being comparable FEP+ in pearson's $\rho$ and superior on 6 of the 8 targets (RMSE), all of which belong to different protein families.

Table 1: Performance comparison across 8 targets from the Merck FEP benchmark set [5].

| TARGET | N | FEP+ | | | GLIDE | | | MMGB-SA | | | PROLET | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $R^2$ | $\rho$ | RMSE | $R^2$ | $\rho$ | RMSE | $R^2$ | $\rho$ | RMSE | $R^2$ | $\rho$ | 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-2$\alpha$ | 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 | $0.44^{.64}_{.25}$ | $0.65^{.79}_{.44}$ | $1.68^{1.76}_{1.59}$ | $0.20^{.41}_{.07}$ | $0.33^{.54}_{.05}$ | $1.57^{1.65}_{1.49}$ | $0.24^{.41}_{.11}$ | $0.38^{.59}_{.12}$ | $9.72^{10.17}_{9.24}$ | $0.34^{.49}_{.19}$ | $0.58^{.71}_{.45}$ | $1.21^{1.65}_{0.77}$ |

**PoseBusters.** The PoseBusters benchmark set is our temporal-split test set, comprising of 308 protein-ligand complexes with novel protein sequences from 2021, assessing the prediction of true ligand binding poses (RMSD $\leq 2$). This benchmark is challenging for machine learning models due to the low degree of protein sequence similarity and high presence of (removed) co-binding structures [21]. We compare ProLET to a wide range of deep-learning and physics-based approaches in Fig. 2b. ProLET achieves a top-1 score of 74%. Whilst being relatively agnostic to the physics-based docking engine, generating poses from different protocols slightly improves ProLET's performance, likely due to an increase in coverage of the possible binding modes in the pocket. Noteworthy is that pose ensembles with up to 100 different poses do not include a single pose below 2Å RMSD ($\sim 7\%$ of all complexes) for the mixture of physics-based methods, highlighting their limitation in generating diverse candidates. Since a good pose does not exist, top-1 over-penalizes ProLET's pose estimation capacity. If we exclude complexes in which no valid poses are generated, the performance is further increased to 82%. In light of these results, we argue that coverage and diversity maximization are key desirables for future docking methods.

**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. We design two case studies to assess ProLET's ability to prioritize selective binders against an undesired target. We compute the difference between experimental binding affinities for a set of ligands towards a desired and undesired target ($\Delta_t$), and compare the differences to the predicted difference returned by ProLET ($\Delta_p$). The model's ability to discern selective compounds is measured through Spearman's $\rho$ between $\Delta_t$ and $\Delta_p$. We extract ground truth data from the Kinome dataset [36]. The first case study involves kinases LCK and EGFR, both members of the tyrosine kinase family. Using the protocol described in 2.2, we generate pose ensembles for each query ligand. ProLET achieves a $\rho$ value of 0.53 on the LCK/EGFR pairs, indicating considerable ranking power in selective compounds towards LCK (figure 2 c). 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. Similar results are observed over a second case study (CSF1R/PDGFRA). See more details in A.6.

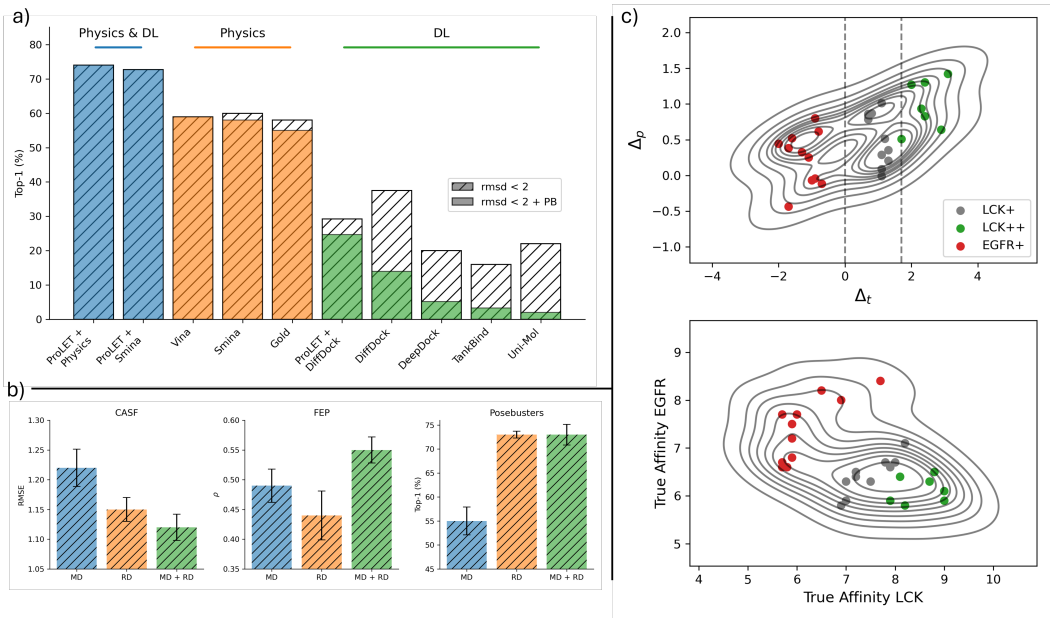


Figure 2: Summary of ProLET results: a) PoseBusters benchmark comparison with deep learning (green), physics-based (orange), and combined (blue) baselines; b) Performance using MD, RD, or both datasets across CASF, FEP, and PoseBusters, with $1\sigma$ error bars from a 10-model ensemble; c) Selectivity case study for LCK/EGFR ligands, showing $\Delta_t$, $\Delta_p$, and true $K_i$ experimental affinities.

**Data Heterogeneity.** We independently train ProLET on MD, RD, and both datasets and evaluate the differences in performance across the FEP, CASF, and PoseBusters benchmarks. As expected, adding MD trajectories to RD improved performance in all benchmarks (Fig. 2b), with a $\sim 30\%$ increase in average $\rho$ across 8 FEP targets. We suggest MD's structural variability likely reduces over-fitting in SBDD models, since the wider distribution in protein-ligand interactions prevents memorization. Nevertheless, MD-trained models under-perform in PoseBusters, highlighting the need for decoy poses during training (see A.1, where we illustrate how RMSD distributions from re-docked conformations and MD trajectories differ).

## 4 Conclusion

We introduce ProLET, an SE(3) Transformer for protein-ligand modeling that addresses topological limitations in traditional methods by constructing an efficient chemically-inspired convex hull over the protein-binding site. ProLET excels in predicting binding affinities and identifying the true binding modes of unseen ligands and proteins, outperforming prior art across PoseBusters and CASF-16. It also matches or surpasses FEP+ in Merck's FEP collection, showing its value for lead optimization and demonstrates utility in targeting selectivity. By integrating docked poses and molecular dynamics, ProLET enhances selectivity predictions, offering a faster, more cost-effective approach to drug development.

## 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] 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.

[3] 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.

[4] Alvaro Prat, Hisham Abdel Aty, Orestis Bastas, Gintautas Kamuntavičius, Tanya Paquet, Povilas Norvaišas, Piero Gasparotto, and Roy Tal. HydraScreen: A Generalizable Structure-Based Deep Learning Approach to Drug Discovery. *J. Chem. Inf. Model.*, 64(15):5817–5831, August 2024. Publisher: American Chemical Society.

[5] 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.

[6] 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.

[7] Gintautas Kamuntavičius, Alvaro Prat, Tanya Paquet, Orestis Bastas, Hisham Abdel Aty, Qing Sun, Carsten B. Andersen, John Harman, Marc E. Siladi, Daniel R. Rines, Sarah J. L. Flatters, Roy Tal, and Povilas Norvaišas. Accelerated hit identification with target evaluation, deep learning and automated labs: prospective validation in IRAK1. *Journal of Cheminformatics*, 16(1):127, November 2024.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

[38] 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.

[39] 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.

[40] 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.

[41] 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.

# 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) in Figure 3. Note that for MD trajectories, an aligned version of the crystal ligand is used to calculate the RMSD, accounting for dynamic changes in the pocket. In this paper we refer to "RMSD" as the RMSD between the true bound ligand conformation extracted from X-Ray crystallography and the generated (query) conformation.

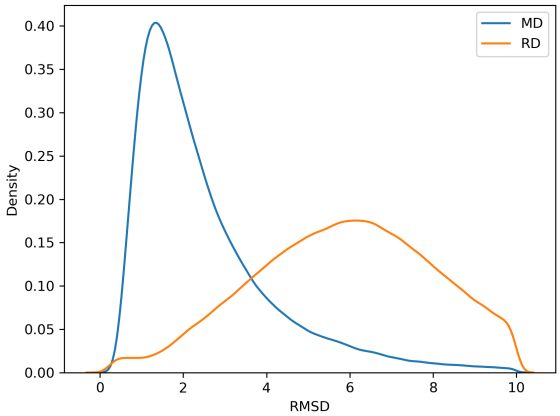


Figure 3: RMSD distributions for molecular dynamics (MD) trajectories vs. re-docked (RD) poses.

## A.2 SE(3) Equivariance

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 [37]. 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.

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. 4), 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.

$$(u \otimes v)^l_m = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} \mathcal{C}^{l,m}_{(l_1,m_1)(l_2,m_2)} u^{l_1}_{m_1} v^{l_2}_{m_2} \tag{4}$$

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.

### A.3 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 [26].

In the first layer, node features consist of a a separable fully connected tensor product mapping `(`$\mathrm{N}_e$ `+` $\mathrm{N}_s$`)x0e` into the node embedding dimension `128x0e + 64x1o + 32x2e`, where $N_s$ is a 2-dimensional one-hot vector defining the origin of the atom (ligand or protein) and $\mathrm{N}_e$ is a a one-hot embedding of the element type, where $e \in [H, C, N, O, P, S, F, Cl, Br, I]$. Note that separate embedding functions are used for each of these features.

Conversely, the edge features are initially represented by scalar-only features, denoted as:

$$\mathbf{X}_{\text{edge}} = \begin{bmatrix} v_{\text{bond}} \\ v_s \end{bmatrix}$$

where $v_{\text{bond}} \in \mathbb{R}^5$ is a one-hot encoded vector defining the bond type $b \in \{\text{single}, \text{double}, \text{triple}, \text{aromatic}, \text{other}\}$, and $v_s \in \mathbb{R}^4$ represents the species of the edge $s \in \{\text{ligand bond}, \text{protein bond}, \text{interaction}, a\text{-}C \text{ hull}\}$, and is also one-hot encoded.

The spherical harmonics contributions $\mathcal{Y}_\ell(\theta, \phi)$, where $\ell \in \{0, 1, 2\}$, are calculated using separable radial (learnable) and angular (pre-computed) components, generated from the distance vector $\mathbf{r}_{ij}$ between nodes $i$ and $j$. The spherical harmonics' components are concatenated with the edge features, as follows:

$$\mathbf{X}'_{\text{edge}} = \begin{bmatrix} v_{\text{bond}} \\ v_s \\ \mathcal{Y}_0(\mathbf{r}_{ij}) \\ \mathcal{Y}_1(\mathbf{r}_{ij}) \\ \mathcal{Y}_2(\mathbf{r}_{ij}) \end{bmatrix}$$

These concatenated features are then passed through the separable tensor products to propagate equivariant representations. The edge degree embeddings, calculated from these spherical harmonics contributions, are combined with the 0-th degree node features to form the updated node features:

$$\mathbf{E}'_{\text{node}} = \text{TensorProduct}(\mathbf{E}_{\text{node}}, \mathbf{X}'_{\text{edge}})$$

Finally, the separable representations of the $L_1$ and $L_2$ irreducible representations (irreps) from the spherical harmonics are propagated along with the 0-th degree embeddings, resulting in the updated node embeddings after the first layer.

In the last layer, a learnable fully-connected separable tensor product (FCTP) transforms the aggregate SE(3) equivariant embedding into a 0-th degree SE(3) invariant irreducible representation (irrep) of dimension 512. These scalar logits are used as inputs to the loss functions during training, with the following labels extracted from the training data. We maintain SE(3) equivariance across the $L$ stacked layers and achieve SE(3) invariance in the last layer by isolating the 0-th degree scalar contributions within the separable tensor product operator. Permutation invariance is ultimately achieved via a permutation invariant aggregation over the ligand vector representations, leading to 4 scalar outputs:

- Pose (after applying the Sigmoid function)
- Normalized Experimental Affinity
- Normalized Affinity (Vina docking energy)
- RMSD (after applying a weighted Tanh function)

Note that docking energy (Kcal/mol) and RMSD (Å) are used as auxiliary losses to help regularize the model during training.

### A.4 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. 5), where $\phi : \mathrm{R} \to \mathrm{R}$ 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. 6); A mean squared error between Vina affinity predictions (Eq. 7); The RMSD deviation (Eq. 8). The total loss is given in equation 9, where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are hyper-parameters that control the relative importance of each loss term. We use default values of 4, 2, 1 and 1 respectively.

$$\mathcal{L}_{ce}(\phi, \hat{\phi}) = -\big(\phi log(\hat{\phi}) + (1-\phi) log(1-\hat{\phi})\big) \tag{5}$$

$$\mathcal{L}_{aff}(a, \hat{a}, \phi) = ||\phi \cdot (a - \hat{a})||_2^2 \tag{6}$$

$$\mathcal{L}_{\text{vina}}(E_{\text{vina}}, \hat{E}_{\text{vina}}) = \|E_{\text{vina}} - \hat{E}_{\text{vina}}\|_2^2 \tag{7}$$

$$\mathcal{L}_{\text{RMSD}}(\text{RMSD}, \hat{\text{RMSD}}) = \|\text{RMSD} - \hat{\text{RMSD}}\|_2^2 \tag{8}$$

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{ce}(\cdot) + \lambda_2 \mathcal{L}_{aff}(\cdot) + \lambda_3 \mathcal{L}_{\text{vina}}(\cdot) + \lambda_4 \mathcal{L}_{\text{RMSD}}(\cdot) \tag{9}$$

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.

### A.5 CASF-2016

The CASF-2016 benchmark consists of a subset of 285 held-out 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 [38], we use this benchmark as a preliminary indicator of generalizability to unseen ligands. 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.

### A.6 Selectivity

**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 $\rho$ of $0.50$. Given the highly correlated pairwise affinity between these two targets (narrow $\Delta_t$), it is notably more difficult to find prominently selective ligands. As shown in Figure 4 in A.6, 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.

As shown in Figure 4, the narrow and highly correlated distribution of pairwise affinities for the labelled ligands in the kinome dataset makes it rather difficult to find selective compounds which are

Table 2: 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 [30] | 2013 | Physics | — | C | T | 0.55 | — |
| MMGB-SA [39] | 2015 | Physics | — | C | T | 0.65 | — |
| GNINA [2] | 2021 | CNN | R | C | T | 0.80 | 1.37 |
| AEScore [40] | 2021 | MLP | R | C | F | 0.83 | 1.22 |
| PointTransformer [18] | 2022 | CNN + ATT | R | C | F | 0.85 | 1.19 |
| OnionNet-2 [17] | 2021 | CNN | R | C | F | **0.86** | 1.16 |
| ConBAP [41] | 2024 | GNN | R | C | T | **0.86** | 1.13 |
| Δ-AEScore [40] | 2021 | MLP | R | D & C | T | 0.80 | 1.32 |
| Hydrascreen [4] | 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 |

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.

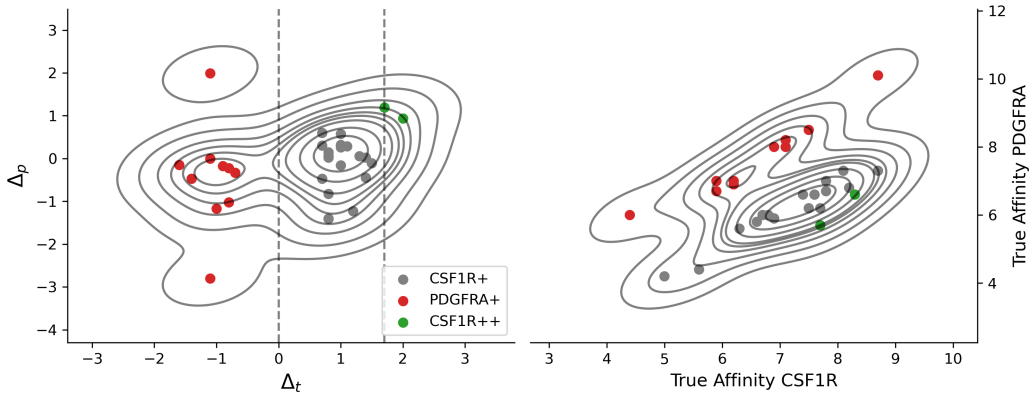


Figure 4: Overlay of $\Delta_t$ and $\Delta_p$ for active CSF1R/PDGFRA ligand pairs (left) and their corresponding affinities (right). Green, gray and red points correspond to highly selective, moderately selective and moderately unselective compounds for CSF1R.