# AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High Throughput Screens

Kate A. Stafford[1], Brandon M. Anderson[1], Jon Sorenson[1], and Henry van den Bedem[1,2,*]

Atomwise, Inc., 717 Market Street, Suite 800, San Francisco, CA 94103, USA.
Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, USA.

## Abstract

Structure-based, virtual High Throughput Screening (vHTS) methods for predicting ligand activity in drug discovery are important when there are no or relatively few known compounds that interact with a therapeutic target of interest. State-of-the-art computational vHTS necessarily relies on effective methods for pose sampling and docking to generate an accurate affinity score from the docked poses. However, proteins are dynamic; in *vivo*, ligands bind to a conformational ensemble. *In silico* docking to the single conformation represented by a crystal structure can adversely affect the pose quality. Here we introduce AtomNet PoseRanker, a graph convolutional network trained to identify, and re-rank crystal-like ligand poses from a sampled ensemble of protein conformations and ligand poses. In contrast to conventional vHTS methods that incorporate receptor flexibility, a deep learning approach can internalize valid cognate and non-cognate binding modes corresponding to distinct receptor conformations. AtomNet PoseRanker significantly enriched pose quality in docking to cognate and non-cognate receptors of the PDBbind v2019 dataset. Improved pose rankings that better represent experimentally observed ligand binding modes improve hit rates in vHTS campaigns, and thereby advance computational drug discovery, especially for novel therapeutic targets or novel binding sites.

## Introduction

Successful drug discovery campaigns rely on identifying biologically active lead molecules that are chemically distinct from known compounds for the disease target. This is especially challenging when there is little or no nearby ligand data available, as is the case with novel targets, or when novel scaffolds are distant in chemical space. Structure-based, virtual High-Throughput Screening methods are designed to overcome this challenge, by identifying novel compounds with predicted activity from vast chemical libraries, for example, MCULE[1] or ENAMINE.[2] vHTS is routinely applied as a first step in the drug discovery process, with hit rates surpassing those of experimental screens.[3–5]

Conventional, structure-based vHTS approaches use an empirical or force-field based scoring function to dock distinct ligand poses to a mostly rigid receptor and predict affinity. Underlying structure-based vHTS approaches is the assumption that receptor-ligand binding poses correlate with experimentally observed affinities. While conventional approaches have led to several

exciting results, including a potent inhibitor for AmpV β-lactamase[6], they have several important drawbacks. First, identifying a scoring function that simultaneously gives high docking power (distinguishing correct docking poses from decoy poses) and high scoring power (generating an affinity score) has historically been challenging[7]. For example, the widely-used molecular docking scoring function AutoDock-Vina[8] excels at pose reproduction but is less competent at correlating poses to affinity, as assessed in the CASF benchmark[9]. Second, conventional methods bear substantial computational cost. Consequently, their ability to predict affinity is limited to small or medium-sized chemical libraries of tens to several hundred million compounds. Third, the dynamic nature of proteins is exploited in ligand binding[10]. Different receptor conformations can bind different ligand chemotypes, perhaps best exemplified by the 'DFG-in' and 'DFG-out' states occupied in different ratios by many protein kinases[11]. Docking a ligand to a rigid receptor conformational substate that deviates from its native bound state (e.g., an apo state) can result in inaccurate predictions of the bound complex that are not useful for further drug design applications. Unfortunately, incorporating receptor flexibility and representing binding-competent receptor conformations remains challenging in conventional methods[12,13] and substantially increases computational cost.

Machine learning (ML) and deep learning (DL) approaches can mitigate these limitations. ML can help determine features of the receptor-ligand complex that correlate with affinity to augment scoring functions and improve docking and screening power. For example, the $\Delta_{vina}RF_{20}$ scoring function combines twenty ligand, protein and pharmacophore features selected among a larger set of candidate features with random forest regression[14]. Post-scoring with $\Delta_{vina}RF_{20}$ improved docking and screening (ranking) power compared to the baseline AutoDock Vina scoring function[8]. A major advantage of learning approaches compared to conventional methods is that they can capitalize on the rapidly increasing availability of data to improve accuracy[15].

In contrast to ML-based methods, DL-based methods avoid the requirement to specify features, and instead learn relevant features directly from structural representations of the protein-ligand complex. In recent years structure-based DL architectures[16–21] have enabled high-throughput vHTS and contributed to the discovery of numerous new leads for drugs, often for challenging protein targets and diseases[22–24]. However, early structure-based DL approaches appeared to learn molecular features from a pose-free structure-based descriptor only, neglecting ligand binding modes and protein ligand interactions[25]. This is especially detrimental to predicting affinity for ligands docked to non-cognate receptors, which is a typical use-case in structure-based drug discovery (SBDD). To enforce sensitivity to protein-ligand interactions, more recent DL vHTS approaches include ligand binding mode information, either as a feature[26] or as a training label in a multi-task architecture[27,28]. Simultaneous learning on ligand binding modes substantially improved activity screening and led to better generalization beyond the training set. Importantly, predicting correct ligand binding modes has merit in its own right.[29] Precise protein-ligand interactions are vital for developing structure activity relationships in hit-to-lead and lead-optimization applications downstream from vHTS.[30]

The efficacy of structure-based virtual screening campaigns relies on an adequate representation of the protein conformational ensemble. Proteins and their ligands undergo conformational exchange under physiological conditions[31–33], and in many cases ligands bind through induced fit

(Fig 1A)[34,35] or bind short-lived intermediate states through conformational selection (Fig 1B)[36]. In those situations, the binding site of a crystal structure may be partially or even fully occluded in the absence of a ligand, hindering the discovery of potent binders. In cases where the protein's cognate ligand has small molecular weight, a holo crystal structure would limit opportunities for docking larger compounds even though those could be accommodated by the protein's full conformational ensemble (Fig 1C). Similarly, structurally uncharacterized disease mutations distal to the ligand binding site can shift the protein conformational ensemble, dramatically reducing or even depleting populations favorable for ligand binding in wild-type protein *in vivo* (Fig 1D,E)[37]. Such a situation would manifest itself by unfavorable binding kinetics in experimental assays despite highly ranked compounds in the virtual screen. Screening against multiple receptor conformations can fail to enrich experimentally validated active compounds if the conformations are higher-energy and unlikely to be accessible in solution. Careful selection of receptor states[38,39] can mitigate these effects of protein dynamics and increase virtual screening performance[40].

How to optimally represent protein conformational states computationally, and how to generalize docking scores to (unseen) conformational substates remain important open problems. Some, but not all, proteins with ligands in the Protein Data Bank (PDB) are structurally resolved in multiple conformations. For individual systems, sophisticated and resource intensive protocols including molecular dynamics simulations and Markov State Models can access biologically relevant conformations[41,42], but these methods scale poorly to large databases of thousands of receptors. Time-independent sampling is a less resource-intensive alternative to access conformational substates. A common limitation of screening with multiple conformations, of any origin, is the challenge in comparing docking scores for molecules docked to different conformations. This has limited the adoption of earlier ensemble approaches, but machine learning techniques can mitigate this problem[43]. Importantly, conventional docking and scoring protocols cannot internalize receptor conformational variability encoded in the thousands of structurally resolved receptor-ligand complexes in, for example, the PDBbind dataset.[44,45] While DL-approaches are relatively underexplored in ensemble docking, they can, in principle, learn and generalize how receptor flexibility accommodates distinct binding modes, both of cognate and non-cognate ligands.

Here, we introduce a DL-based method for binding mode prediction that exploits protein conformational ensembles instead of single structures in vHTS applications. Starting from a protein crystal structure or homology model, we use Rosetta's comparative modeling protocol (RosettaCM)[46] to sample low-energy conformations near the crystal structure to generate a six-member conformational ensemble, which includes the starting structure. We applied this protocol to each receptor-compound pair in PDBbind v2019 to generate FlexPDBbind v2019, a curated data set of conformational ensembles of drug targets. To enrich the quality of compound poses docked to these ensembles we adapted AtomNet GRAPHite, a new directional message passing graph convolutional network, for Pose Ranking applications, which we describe as AtomNet PoseRanker. To maximize learning on distinct receptor conformations, we additionally trained AtomNet PoseRanker on non-cognate poses, obtained from cross-docking ligands to distinct crystal structures of the same protein (uniprot identifier) in the PDBBbind v2019 dataset.

We first demonstrate that AtomNet PoseRanker enriches pose quality for compounds docked to single conformations of their cognate receptors as well as non-cognate receptors in the cross-docked PDBbind v2019 data set compared to conventional approaches. Our results indicate that training on a cross-docked data set teaches AtomNet PoseRanker to recognize distinct poses as valid in different receptor conformations. AtomNet PoseRanker trained on FlexPDBbind v2019 achieved nearly the same enrichment, suggesting that computationally sampled conformational ensembles can augment experimental ensembles when the latter are not available. Finally, we examined how conformationally diverse ensembles affect enrichment of active compounds in a virtual screen of Abl kinase. Ensemble-based procedures generally outperformed those using a single conformation. Strikingly, however, we found that using a *single* conformation with a cross-docked trained AtomNet PoseRanker achieved nearly identical enrichment as that using an ensemble, suggesting that DL approaches can infer flexibility from the training set.
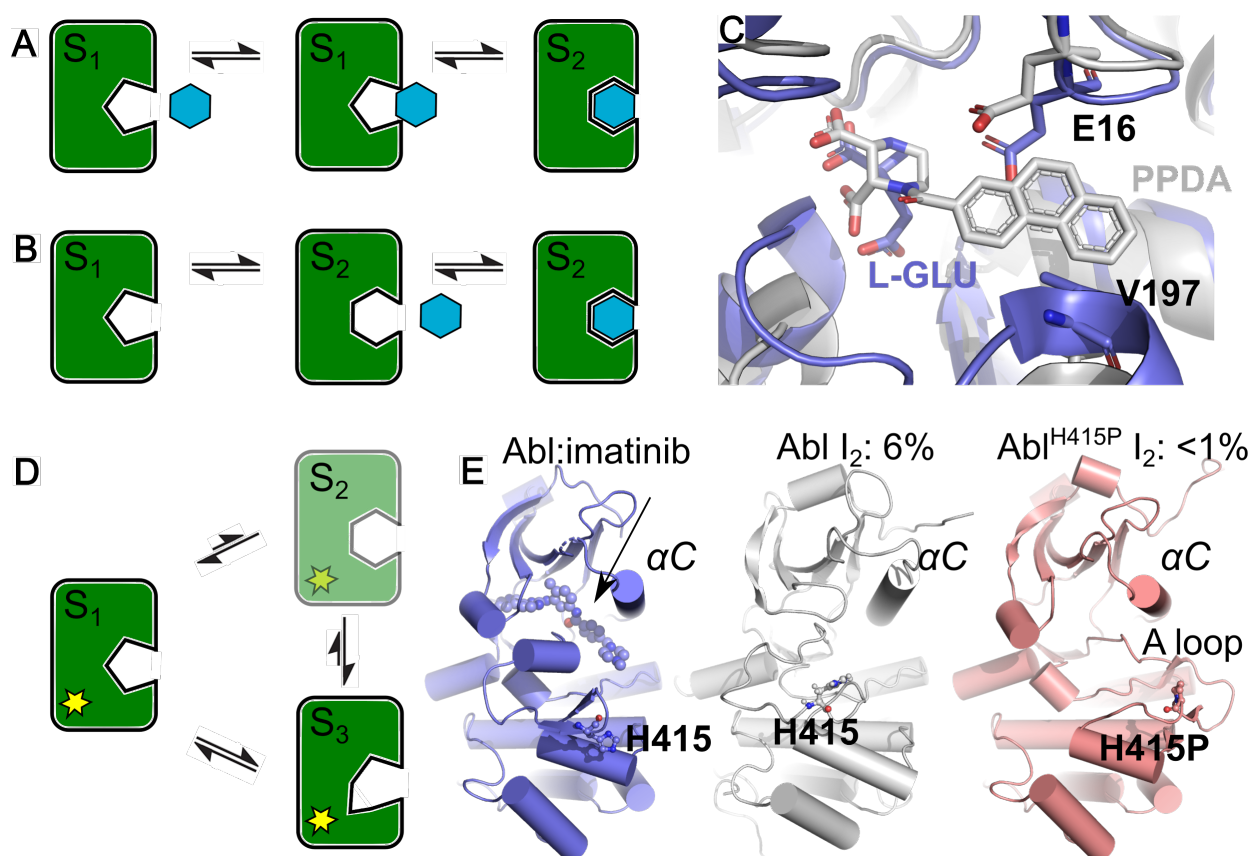


**Figure 1**: **Binding site plasticity challenges vHTS in structure-based drug design**. A) Induced fit. B) Conformational selection. C) The GluN2A NMDA receptor ligand-binding domain in complex with its cognate activating ligand L-glutamate (L-GLU, pdb ID 4nf8 (slate)) and antagonist 1-(phenanthrene-2-carbonyl)piperazine-2,3-dicarboxylic acid (PPDA, pdb ID 4nf6 (grey)). The L-Glu occupied site requires dramatic structural changes to accommodate the larger antagonist.  D) Mutations distal to the binding site (yellow) can shift the conformational equilibrium, thereby stabilizing non-binding substates or abrogating substates that are required for conformational selection. E) The compound imatinib in complex with Abl kinase (pdb ID 2hyy, slate). Imatinib binds Abl kinase in the inactive $I_2$ state, which occupies a population of ~6% in WT Abl (pdb ID 6xrg, grey). A distal H415P mutation in Abl kinase, at a distance of 18Å from the

active site, reduces affinity for the inhibitor imatinib five-fold. H415P destabilizes the $I_2$ state, reducing its population to below detectable levels (pdb ID 2f4j, salmon). Note the A-loop in the open conformation.

## Methods

**Preparing train and test data sets.** We downloaded the PDBbind data set v2019 (http://www.pdbbind.org.cn), and removed entries that contained cofactors, incomplete ligands, more than one ligand, incorrect valences, entries annotated as 'NMR', or entries for which the ligand was annotated as a peptide. This resulted in a data set containing 4,593 crystal structures from the 'refined' set, and 10,011 from the 'general' set. We split the data set into train (10,919) and test (5,206) sets by requiring that receptors share less than 70% or 50% sequence similarity between the sets ('seqsim70' and 'seqsim50' splits). For comparison, we also generated train and test sets identical in size to the seqsim70 and seqsim50 sets but split by enforcing that these sets do not share identical uniprot identifiers ('uniprot split').

**Preparing ligand structures.** We generated ligand structures for docking based on the ligands provided by PDBbind. To avoid biasing poses toward the crystal structures, and in contrast to previously reported studies, we discarded the native compound conformation and generated a UFF energy-minimized starting conformation from the ligand SMILES supplied by the PDB using RDKit[47]. This approach better reflects commercial vHTS campaigns, where a crystal pose or experimental structure of the ligand is typically lacking.

**Preparing receptor structures.** We prepared the receptors for docking by a three-step process. For consistency with our generated ensembles, we first read the crystal structure into Rosetta and removed any crystallographic buffer or water molecules, but retained metal ions of the following types: Na, Fe, Mg, K, Mn, Zn, and Ca. Using Rosetta, we filled in missing atoms for incomplete protein residues. We then define a bounding box surrounding each initial ligand used as the search space for subsequent docking. For this study, we accepted the protonation states of titratable residues found in PDBbind.
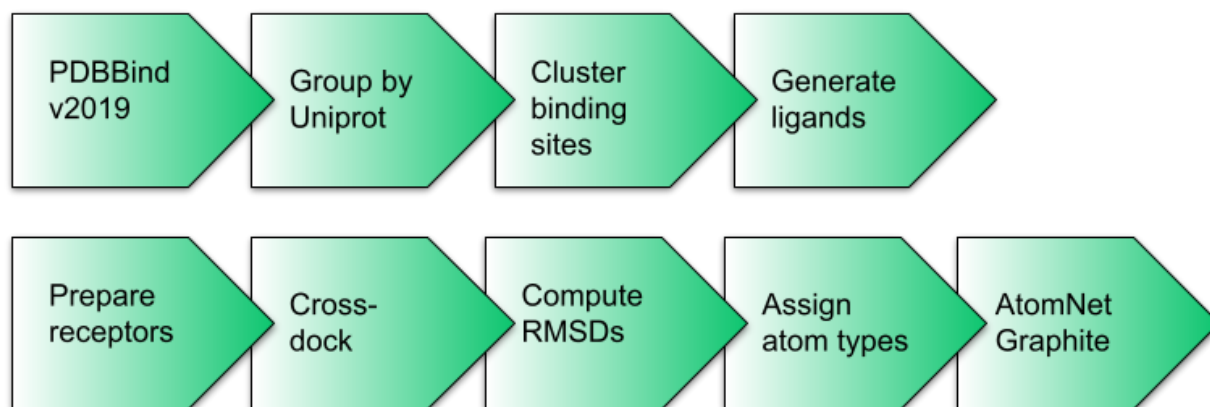


**Fig 2.** Flowchart illustrating data processing pipeline. Beginning from the PDBbind v2019 protein-ligand dataset, we clustered ligands for all structures of the same Uniprot into distinct binding sites to define cross-docking groups. We used Rosetta (version) to prepare the structures for docking,

including filling in missing sidechain atoms and adding protons. We prepared 3D conformations for each ligand using RDKit and used these conformations to dock to the crystal structures within each binding site cluster. We additionally prepared conformational ensembles using the Rosetta hybridization protocol and docked ligands within each binding site cluster.

**Cross-docking**. We superimposed receptors of the same uniprot ID using the chain identifiers present in the 'pocket' pdb file from PDBbind, keeping receptors that superimposed to within 5Å. We identified different binding sites within the same uniprot ID by clustering the center of mass of the superimposed ligands with DBSCAN[48] (parameters eps=5., min_samples=1). A small number of uniprot IDs ('targets') are represented by hundreds of receptor-compound pairs in PDBbind. To avoid those from dominating the cross-docked data set, we randomly selected five representatives from among the receptor-compound pairs to be included in the cross-docking procedure. Cross docking was performed on all members of each binding site cluster. The final training set consisted of targets that sampled at least one low-RMSD pose. This procedure gave a total of 27,166 target-compound pairs.

**Ensemble docking.** We generated five diverse receptor conformations using the rosetta hybridize protocol (Supplemental Material) with the `score3` and `score4_smooth_cart` scoring functions in stage 1 and 2 of the centroid stage, and the `ref2015_cart` scoring function with `metalbinding_constraint` in the full atom stage. Ligand parameterization relied on rosetta's templates whenever available or was generated with rosetta's `molfile_to_params.py` script using default settings. We used default weights for intra-ligand `hetatm_cst_weight=1.0` and receptor-ligand `hetatm_to_protein_cst_weight=1.0` constraints. Fragment insertions were disabled in stage 2: `fragprob_stage2=0.0`. We adjusted the stage 2 Monte Carlo temperature to 0.5: `stage2_temperature=0.5`. Next, we subjected each hybridized receptor conformation and the crystal structure to six rounds of energy relaxation with Rosetta's `FastRelax` protocol, retaining metal coordination with the `SetupMetalsMover` protocol combined with a `metalbinding_constraint = 1.0`. We observed that protein side-chain metal coordination was not preserved consistently throughout the hybridize protocol. In some cases that led to final models with unsatisfied coordinate-covalent bonds. We created a final ensemble of six receptor conformations by selecting the conformation with the lowest Rosetta energy from each of the six relaxed conformations starting from the hybridized models or crystal structure.

**Docking.** We used a slightly modified version of the smina docking software[49] with the vina scoring function to generate binding poses for the receptor-compound pairs with command line parameters `--exhaustiveness 384 --energy_range 99999 --num_modes 64 --mc_steps 3 --minimize_iters 40 --accurate_line --approximation linear --autobox_add 2.0 --seed 42`. Our modification introduces the additional parameter `mc_steps` which is used to tune the number of steps in the Monte Carlo search. For each receptor-compound pair we generated up to 64 poses with the default 1Å minimum difference between poses. We labeled poses within 2.5Å of the native pose a 'hit', and those greater than 4Å a 'miss'. Poses in between were discarded from the training set but retained in the testing sets.

We calculated the RMSD between corresponding heavy atoms of docked poses and native crystal structures of the corresponding compounds by first matching substructures using RDKit and accounting for symmetric substructures by using the minimum RMSD in the case of multiple matches. For the final dataset we required that each compound adopted at least one pose within 2.5Å of the native pose, evaluated after superposition of protein structures in the case of cross-docking.

**Descriptor generation.** From each docked complex, we generated descriptors for use in training and testing our models. We describe each heavy atom with its corresponding SYBYL atom type indicating its chemical environment, generated by converting to Mol2 format using OpenBabel[50]. We do not represent protons explicitly in our final descriptor sets, and we treat all metals as a single type. We treat the protein and ligand as distinct entities and encode their atom types separately.

**Markov State Models of Abl.** We obtained structures for 16 Abl kinase macrostates from the manuscript by Roux and coworkers (PDB ID 2HYY)[51]. We prepared a matching 16-member conformational ensemble using our rosetta-based ensemble-docking protocol. For each of those ensembles, we prepared the receptors for docking following our standard protocol described above. We prepared the single receptor conformation for this analysis based on the crystal structure 2HYY. We selected 8,946 compounds from our internal databases to dock to Abl kinase, including molecules with known activity (3,205), known non-binders (4,459) and random molecules (1,282). We ranked compounds using the AtomNet PoseRanker model trained on a dataset that excluded Abl kinase or receptors with more than 70% sequence similarity to Abl kinase. We calculated enrichment factors as $\mathrm{ef}(\nu) = ap^{-1}a(\nu)/n(\nu)$, where $ap$ is the total fraction of actives in the data set, $a(v)$ is the number of actives in the top $\nu$-percentile, and $n(\nu)$ is the total number of compounds in the top $\nu$-percentile.

**AtomNet PoseRanker.** Underlying AtomNet PoseRanker is our GRAPHite network architecture, a directional Message Passing Neural Network[52,53] in which nodes encode receptor or ligand atoms (Fig 2A). We do not impose a covalent structure on the graph. Instead, pairs of atoms within a distance threshold $D$ of each other can pass messages along a directed edge. To focus on the interface between ligand and receptor atoms, we use a flexible framework that allows us to configure distinct sets of 'source' and 'target' atoms for messages in each layer (Fig 2B). At the input layer $l = 0$, we construct separate embeddings for ligand $F^{(0,L)}$ and receptor $F^{(0,R)}$ atoms from their coordinates and SYBYL atom types.
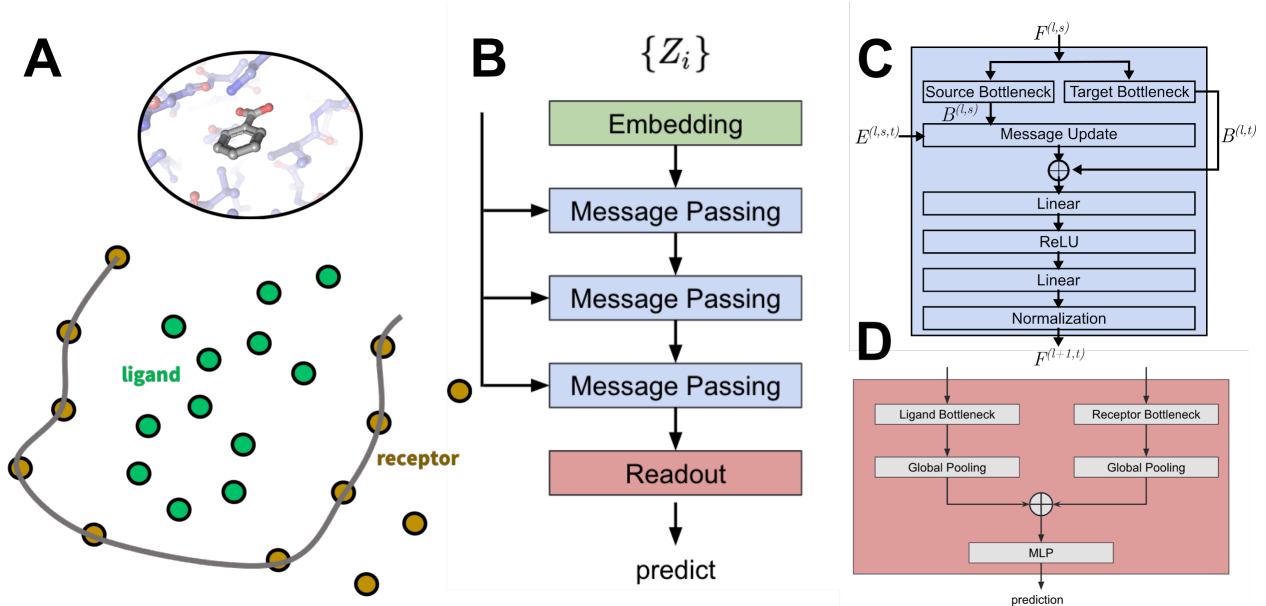
**Figure 3.** Architecture of AtomNet GRAPHite, a directional Message Passing Neural Network. A) Ligand and receptor atoms are represented as nodes in AtomNet GRAPHite, omitting their covalent structure. B) Overview of AtomNet GRAPHite's graph convolutional network architecture. After an initial embedding, any number of message passing layers (MPLs) can be stacked, optionally with skip connections, before final readout.

Each message passing layer $\ell = 1 \ldots \ell_{\max}$ in the network features both a source $F_i^{(\ell,s)}$ and a target $F_j^{(\ell+1,t)}$ set of atoms $\mathfrak{s}_\ell \rightarrow \mathfrak{s}_{\ell+1}$, where $s \in \mathfrak{s}_\ell$ and $t \in \mathfrak{s}_{\ell+1}$. The input to the message passing layers are updated features based upon dense skip connections and a linear bottleneck layer [63] (Fig 2C). We concatenate and mix the unique features for source atoms $s$ using a linear layer

$$B^{(\ell,s)} = \left( \bigoplus_\ell F^{(\ell,s)} \right) \cdot W^{(\ell,s)}.$$

The bottlenecked features are then used as the input to a message function, and updated using

$$M_j^{(\ell,t)} = \sum_{s \in \mathfrak{s}_\ell} \sum_j E_{ij}^{(t,s,\ell)} B_j^{(\ell,s)},$$

where $E_{ij}^{(t,s,\ell)}$ denotes an edge between source ($s$) atom $i$ and target ($t$) atom $j$. Accordingly, the message function $M_j^{(\ell,t)}$ for target atom $j$ is the sum of the messages from each source atom $s \in \mathfrak{s}_\ell$. The updated features for each layer are then constructed by concatenating with the source features of the atom species, and applying a $\mathrm{lin} - \mathrm{ReLU} - \mathrm{lin}$ multi-layer perceptron:

$$F^{(\ell+1,t)} = \mathrm{MLP} \left( M^{(\ell,t)} \oplus B^{(\ell,t)} \right).$$

Edge features

$$E^{(s,t,\ell)} = \sum_n c_n^{(s,t,\ell)} R_n \left( r_{ij} \right)$$

are encoded by linear combination of radial basis functions; we chose $R_n \left( r \right) = N_n j_0 \left( z_{0n} r / R_c \right)$, where $j_0 \left( x \right) = \sin x / x$ is a zeroth order spherical Bessel function, $N_n$ is a normalization constant, and $R_c$ is the maximum radius for passing messages between source atoms $s$ and target atoms $t$. Note that our implementation does not require a vanishing gradient for radial functions at the boundary.

Our readout layer is a global pooling operation to a final set of bottleneck features

$$P^{(s)} = \sum_i B_i^{(\ell_{\max}, s)}$$

.

These pooled features are then combined, and a $\text{lin} - \text{ReLU} - \text{lin}$ MLP generates the final predictions:

$$\hat{y} = \text{MLP} \left( P^{(L)} \oplus P^{(R)} \right).$$

**Data and software availability.** We provide input scripts and parameters for our Rosetta protocol in the Supplementary Information section. Our FlexPDBBind v2019 dataset consisting of sampled ensembles for PDBBind structures is freely accessible and available for download from http://atomwise.com/flexpdbbind2019.

# Results

**2.1 Ligand:receptor interfaces contribute to pose prediction accuracy.** Several studies suggest that structure-based vHTS CNN-based methods minimally rely on features of the receptor[54,55] but instead distinguish active from inactive compounds primarily by ligand-based features. Carefully designed data-augmentation can draw in receptor-based features in these models, thereby promoting generalizability.[56] By contrast, receptor features in graph-based models can affect binding mode prediction[57]. We designed AtomNet's configurable convolutional layers to deconvolve the role of the receptor and the ligand in determining pose quality, and to probe and optimize the role of ligand:receptor interactions. A hyperparameter controls the number and configurations of AtomNet's message passing layers. If 'l' denotes the set of ligand atoms, 'r' denotes the set of receptor atoms, and 'lr' denotes the combined set of atoms, each layer can pass messages from 'l' to 'r', 'r' to 'l', 'l' to 'l', 'r' to 'r', and each of 'l' or 'r' in these can be replaced by 'lr'.

**Table 1**. **Distinct configurations of convolutional layers respond differentially to mechanisms of molecular recognition.** Test AUC for distinct convolutional layer configurations in AtomNet PoseRanker. Test AUCs were computed from classifying poses obtained from self-docking on the PDBbind data set uniprot split.

| layer configuration | test ROC AUC |
|---|---|
| `l->l->l->l->l->l` | 0.67 |
| `lr->lr->lr->lr->lr->lr` | 0.91 |
| `l->l->r->l->r->l` | 0.91 |
| `l->l->r->l->r->l->r->l` | 0.92 |

To test the role of the ligand:receptor interface in classifying poses, we evaluated the effect of distinct layer configurations on distinguishing correctly docked poses from incorrectly docked poses for the uniprot split PDBbind data set. We tested layer configurations featuring transitions between ligand and receptor layers to focus on interface features. Table 1 reveals that, while the network demonstrates some capacity to identify correct poses based on ligand features alone, the test AUC (PDBbind self-docking, uniprot split) increased dramatically when we included the ligand:receptor interface in the layer configurations (test AUC > 0.9), compared to a layer configuration between ligand atoms only (test AUC = 0.67). When we probed the interface with additional convolutional layers, we observed a peak in test AUC at 0.92. Adding more layers did not result in further improvements. These findings suggest that our graph convolutional architecture is sensitive to mechanisms of molecular recognition in structure-based drug design. To minimize the potential of ligand- or receptor-based pattern memorization, for the remainder of the study we adopted the '`llrlrlrl`' layer architecture that includes edges only *between* ligand-receptor atom pairs.

**2.2 AtomNet PoseRanker improves (sm)vina and ML-based rankings.** Next, we compared AtomNet PoseRanker's ability to classify docked poses on the PDBbind dataset to similar classifiers. On the uniprot split, AtomNet PoseRanker's test AUC was 0.93. Other methods similarly reported AUCs in the 0.86-0.94 range[27] (Table 2). By training simultaneously on poses and activity, Lim et al[28] achieved an AUC = 0.94 on their data set. Direct comparison of AUCs with these methods is difficult owing to differences in the way the data is split which can lead to memorization effects[27], and slight random initializations of the methods.

**Table 2:** Pose classification compared to other methods across different train/test splits. *In contrast to other methods in this table, the model of Lim et al. was trained simultaneously on poses and activity. [†]Three-fold cross-validation on 90% sequence identity.

| | ROC AUC | | |
|---|---|---|---|
| **method** | uniprot | seqsim70 | seqsim50 |
| AtomNet PoseRanker | 0.93 | 0.90 | 0.89 |
| smina docking | 0.82 | 0.82 | 0.81 |

| | | | |
|---|---|---|---|
| Cornell et al.[57] | 0.86 | | |
| Ragoza et al.[58] | 0.815[†] | | |
| Lim et al.*,[28] | 0.94 | | |

We then tested AtomNet PoseRanker's ability to generalize out of the training data by evaluating its performance on the seqsim70 and seqsim50 data splits, which challenge memorization effects of the training data. We observed a small decrease in performance from AUC = 0.93 to 0.89.

In practical applications, pose ranking within a particular target class is more important than the pose quality across all targets as reported by AUC. AtomNet PoseRanker significantly enriched the fraction of poses within 2.5Å of the native pose among the top-n ranked poses compared to smina docking (Fig 3). When "gap" poses between 2.5Å and 4Å were excluded in training, 45% of the top poses in AtomNet PoseRanker were within 2.5Å of a crystal pose by target class (Fig 3A black line), compared to 39% of the top smina poses on the (most challenging) seqsim50 test set including "gap" poses. (Fig 3A, blue line).

**2.3 AtomNet PoseRanker enriches intra-target compound rankings in cross-docking.** While the re-docking performance of a deep learning model is often reported to evaluate its performance, correctly predicting the binding mode of a new ligand, possibly to a new target, is more important in SBDD. The new target may even lack a crystal structure, so vHTS will need to rely on a closely related structural model or a homology model. We therefore trained and evaluated AtomNet PoseRanker on a cross-docked data set (Methods). Table 3 reports the AUC for AtomNet PoseRanker compared to smina docking.

**Table 3**: cross docking pose ranking/classification compared to smina docking.

| | ROC AUC | | |
|---|---|---|---|
| **method** | uniprot | seqsim70 | seqsim50 |
| AtomNet PoseRanker | 0.91 | 0.90 | 0.86 |
| smina docking | 0.78 | 0.79 | 0.78 |

Unsurprisingly, distinguishing favorable docking poses from unfavorable poses is more challenging in cross-docking compared to self-docking, reflected by reduced AUCs for AtomNet PoseRanker and smina docking. There are numerous challenges in cross-docking: 1) the binding site often has a different size and/or shape owing to different (equilibrium) backbone and sidechain positions; 2) amino acid insertions, deletions, or substitutions can alter the steric and electrostatics characteristics of the site; 3) changes in solvation characteristics; and 4) experimental conditions like crystal packing, etc. Accordingly, test AUC were reduced slightly compared to self-docking across all data splits (Table 3).

We observed marked intra-target early enrichment among the top-ranked poses of the model trained on cross-docked crystal structures compared to the model trained on cognate receptors (Fig 3A, top-1 54% (pink) vs 45% (black). While the jump in performance could be attributable to a larger training set size, the reduced AUC suggests that the cross-docked model less easily distinguishes good poses from bad poses despite the larger data set. However, the model is more confident about good poses, likely because similar good poses occurred across multiple receptor conformations whereas invalid poses were reproduced less across receptors.
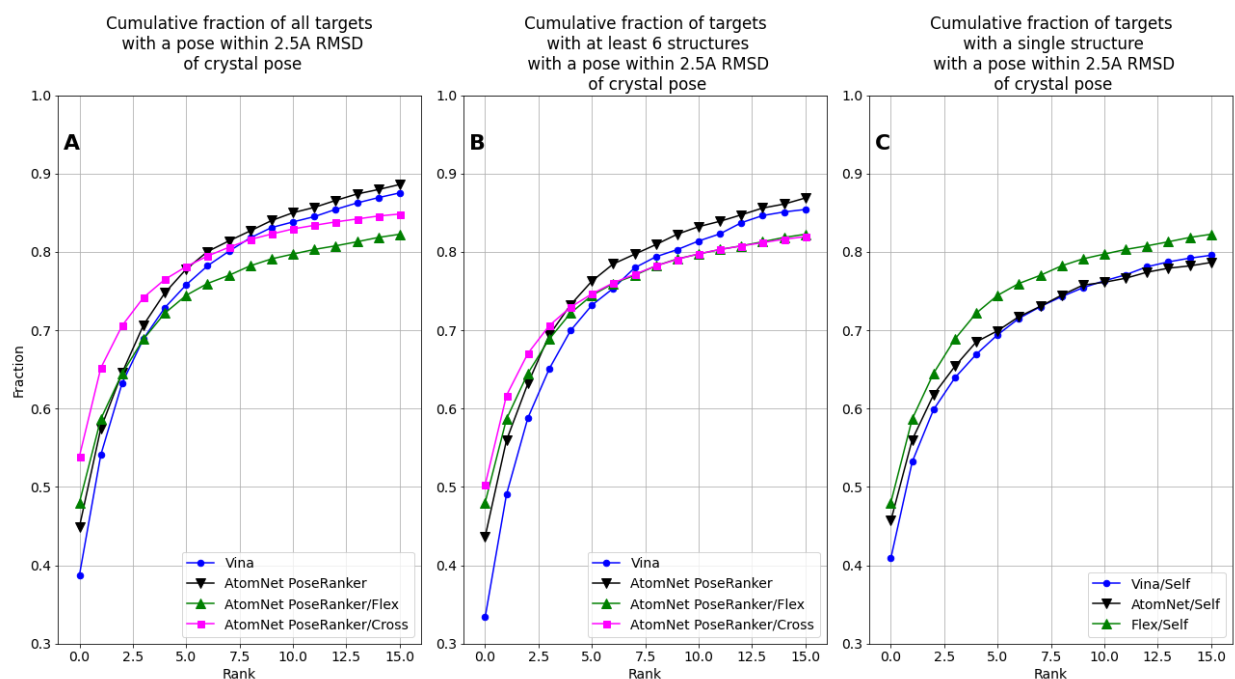


**Fig 3** Cumulative fraction of targets with docked poses within 2.5Å RMSD of the crystal pose across all receptors within a target class by pose rank for the seqsim50 split according to vina (blue), AtomNet PoseRanker trained on cross-docked crystal structures (black), and AtomNet PoseRanker trained on sampled conformational ensembles (green). Each figure shows the top sixteen ranked poses for each docking attempt in (A) the full dataset, (B) targets that include at least six PDB structures, for a direct comparison of experimental and sampled ensembles and (C) targets that include only one PDB structure, to illustrate the effects of sampling where additional experimental data is not available.

## 3. Ensemble-based docking

Next, to further examine how learning receptor conformational diversity can help AtomNet PoseRanker recognize valid binding modes, we developed a protocol that creates an ensemble of computationally sampled receptor conformations for docking based on Rosetta's comparative modeling protocol[46] (Methods). Aside from providing conformational variability when multiple experimental structures are lacking, computationally sampled ensembles offer advantages over

those constructed from experimental structures for a pose training task, including removing variations due to mutations or differences in construct; normalizing the numbers and sources of different conformations available for each protein in the dataset; and permitting the trained model to be used in practical applications to docked poses to homology models produced through Rosetta-based pipelines.

**3.1 FlexPDBbind: ensemble models for 13,000 biomolecular complexes.** For each biomolecular complex in the PDBbind 'refined' and 'general' sets we generated a six-member conformational ensemble. We applied Rosetta's `FastRelax` protocol to the crystal structures to generate the first member, and then repeatedly applied the `Hybridize` and `FastRelax` protocols to generate up to five additional conformations (Methods). This resulted in 13,063 biomolecular ensembles, for a total of 71,825 conformations. The mean Root Mean Square Deviation (RMSD) of the ensembles calculated over the CA atoms is 1.57 Å, and the Root Mean Square Fluctuations is 2.10 Å (Fig 4 A,B). This compares to a mean RMSD of 1.24 Å for the structurally aligned crystal structures. To evaluate the structural quality of the generated models, we calculated clash scores and overall Molprobity scores with Molprobity[59] (Fig 4 D,E). We discarded ensembles with outlying clash scores (scores over 10). We then docked ligands to each conformation of their cognate receptor using the protocol detailed in Methods. The RMSDs of docked ligand poses calculated to their starting pose in the source receptor conformation reveals a distribution sharply peaked around 3.5 Å with a long tail, illustrating enrichment of low-RMSD poses in sampled ensembles (Fig 4C). The distribution broadens when the RMSD is calculated with respect to the pose from the crystal structure, owing to random offsets in rigid body transformations of the generated structures (Fig 4C, red). Fig 4F shows two examples of conformations in FlexPDBbind.
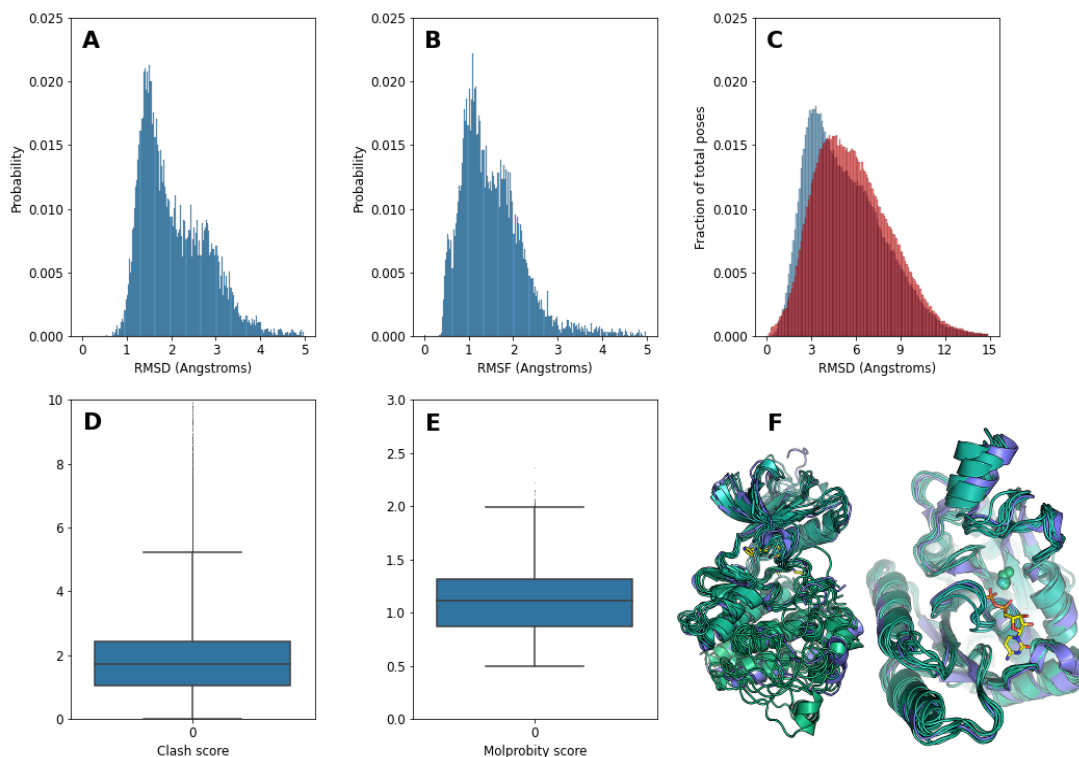
**Fig 4**: **Characteristics of the FlexPDBbind ensembles.** (A) RMSD and (B) RMSF distributions for 6-member sampled ensembles. (C) Distribution of RMSDs of ligand poses docked to sampled ensembles (blue) and crystal structures (red). (D) Molprobity clash scores and (E) overall scores for sampled ensembles. (F) Examples of a sampled conformational ensemble for ABL1 in complex with an inhibitor (top, PDB ID 5HU9) and dethiobiotin synthetase in complex with CDP (bottom, PDB ID 6CVF) with the crystal structure in purple, and ensemble members in green. The ligand corresponding to the crystal structure is shown in yellow.

## 3.2 Ensemble-docking mitigates limitations of induced-fit docking

AtomNet PoseRanker trained on the FlexPDBbind ensemble of cognate receptors slightly improved top-1 enrichment compared to the model trained on single conformations (Fig 3A; top poses: 48% (green) vs 45% (black) for single conformations). However, we did not observe improved enrichment beyond the top poses. Unsurprisingly, the model trained on cross-docked crystal structures outperformed the FlexPDBbind-trained model (Fig 3A pink; top poses: 54%), likely owing to receptor conformations that are closer to physiological sub-states and to a more diverse representation of binding modes for targets with a large number of receptor-compound pairs in PDBbind. Approximately 80% of targets in PDBbind have fewer than six structural representations. For a more direct comparison between the cross-docked set and FlexPDBbind, we repeated the analysis limited to targets that have at least six structural representations in the cross-docked set. (Fig 3B). While the top-1 enrichment on the FlexPDBbind subset was similar to that of the full FlexPDBbind set (Fig 3A), somewhat surprisingly the enrichment *gain* between the cross-docked and FlexPDBbind was substantially reduced on these subsets.

Strikingly, when we applied an AtomNet PoseRanker model trained on sampled ensembles to predict binding modes using a dataset of single receptor structures, we found that top poses were enriched over the model trained on single structures (Fig 3C, green vs. black line). While, again, the improved performance could in part perhaps be attributable to the larger training set for the sampled ensembles, the result does suggest that the model retains information about flexible binding sites, illustrating the value of performing additional conformational sampling where experimental data is limited.

## 3.3 Ensemble-trained AtomNet PoseRanker enriched Abl kinase actives on a conformational ensemble and a single conformation.

*3.1.1 Abl kinase*
As a practical application, we examined the effects of receptor conformational ensembles and compound pose reranking in identifying active compounds of Abl kinase (Abelson tyrosine kinase). Abl is of clinical significance due to its causal role in chronic myelogenous leukemia (CML); in about 90% of cases of CML, a chromosomal translocation forms a "Philadelphia chromosome", which creates a fusion between the Abl and break-point cluster (Bcr) genes to produce a protein with constitutive kinase activity[60,61]. Inhibitors of Abl kinase activity are in clinical use as treatments for CML and other cancers since the US approval of the targeted inhibitor imatinib in 2001 and have significantly improved clinical outcomes[61].

Kinase inhibitors are often divided into multiple classes or types depending on their binding mode[62]. Type I inhibitors are considered to bind to the "DFG-in" protein conformation while type II inhibitors bind the "DFG-out" conformation. The DFG-in conformation is sterically incompatible with the most common binding mode of type II inhibitors. Among known mutations conferring resistance to imatinib and related compounds, many exert their effect by changing the protein's conformational distribution and reducing occupancy of the favored binding state (Fig 1E). As a result, identifying both type I and type II inhibitors in a structure-based virtual screen of the ATP binding site is important for identification of new molecules of potential clinical utility, and requires the use of multiple protein conformations to adequately sample native-like ligand binding modes.

We evaluated the power of our ensemble-docking protocol to enrich known Abl inhibitors over an in-house curated library of negative examples including known non-binders and randomly selected molecules. We compared a conformational ensemble (n=16, RMSF=2.7Å) generated using our protocol to a conformational ensemble (n=16, RMSF=2.4Å) derived from a Markov-state model (MSM) based on extensive molecular dynamics simulations[42,51]. We also calculated the enrichment factor from the smina rank of docked poses to the (relaxed) crystal structure.

Docking compounds to an ensemble of MSM receptor conformations dramatically improved early enrichment of active compounds compared to docking to a single receptor conformation when we ranked compounds by smina score (Fig 6A, blue vs. yellow). We observed similar enrichment factors when we ranked compounds docked to an ensemble of rosetta-generated receptor conformations using AtomNet PoseRanker (Fig 6A, red). While these enrichment factors are near-indistinguishable, our Rosetta ensembles were generated at a fraction of the computational cost compared to MSM ensembles. Strikingly, early enrichment remained similarly elevated when we ranked compounds docked to a *single* receptor conformation using AtomNet PoseRanker (Fig 6A, magenta). Note that AtomNet PoseRanker model was trained on a cross-docked data set, which led the model to recognize distinct but valid poses of compounds corresponding to diverse receptor conformations. AtomNet PoseRanker in combination with the MSM ensemble did not achieve the same enrichment factors (Fig 6A, yellow), likely because the MSM receptor conformations are too distinct from the rosetta-generated conformations in the training set. Thus, docking against an ensemble of conformations can lead to early enrichment of active compounds, which can be retained using even a single receptor conformation with the AtomNet PoseRanker model.

We then compared the chemical diversity of active top compounds docked to single receptor conformations to those using the rosetta-ensemble. We selected the 1% top-ranked compounds for each method. These 90 compounds would approximately fill a 96-well plate used for experimental validation in a vHTS screen. We then selected all active compounds among the top 1% for the single receptor conformation/smina ranking (31 compounds) and the rosetta-ensemble/AtomNet PoseRanker (59 compounds) method. We excluded two compounds that ranked within the top 90 in both methods, since we are primarily interested in examining their differences. A t-SNE[63] projection of a principal components analysis computed from their ECFP4 fingerprints suggests that the rosetta ensemble is slightly more diverse compared to the single

conformation compounds (Fig 6B). Towards the lower left corner of the t-SNE projection, smina uncovers exceedingly fewer compounds. In the lower right corner, the first t-SNE coordinate separates a cluster of compounds at 't-SNE 1' > 2 that were enriched nearly exclusively in the ensemble. These compounds are all analogs of Ponatinib (Fig 6C), notably sharing the ethynyl linker to a imidazol-1,2-pyridazine-like hinge region of the compound. AtomNet PoseRanker identified the most hits in this cluster against receptor conformations 13 and 16. Altogether, two out of 25 hits in this cluster were identified from docking to a single receptor (blue circle, using smina, and another from AtomNet PoseRanker (Fig 6D)). Among all top 1% of compounds in both docking methods, only two Ponatinib analogs were identified using smina ranking.
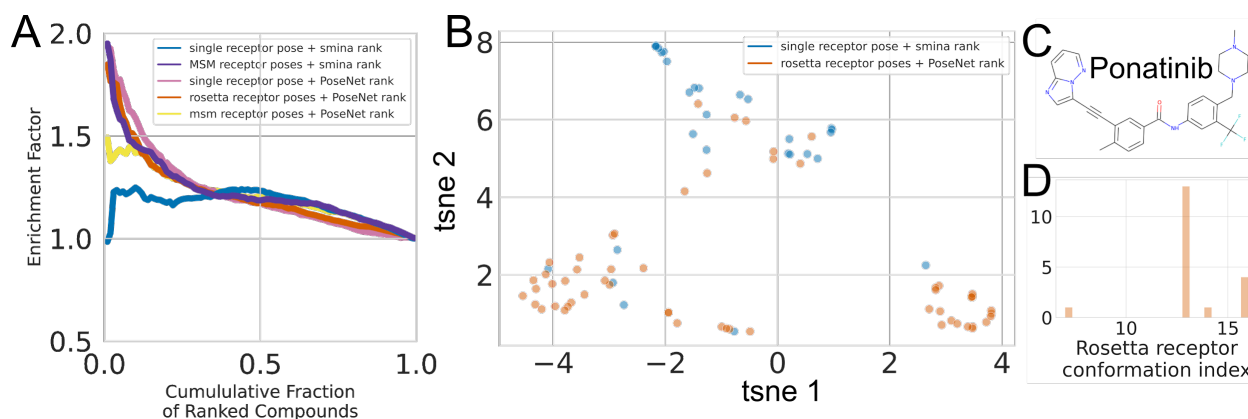


**Fig 6** Ensemble docking enriched known Abl inhibitors compared to single receptor conformations. A) Predicted probability of activity (top panel) and Enrichment Factors (bottom panel) for established Abl kinase inhibitors and confirmed non-binders and random compounds. Colors represent different combinations of ensembles and scoring algorithms. The dashed gray line represents the prevalence (35%) of active compounds in the set. Compounds are ranked from best to worst according to the score for each method, i.e., the value 0.5 on the x-axis represents the top-scoring 50% of compounds for each of the methods. D) Conformation 1 corresponds to the crystal structure.

## Conclusion

In this study we describe AtomNet PoseRanker, a graph-based convolutional neural network trained to identify high-quality crystal-like poses from docking. Importantly, and in contrast to previous work, we use ligand conformations prepared without knowledge of the ligand conformation in the PDB structure, which is representative of use cases encountered in virtual screening applications. Using the curated dataset PDBbind 2019, we demonstrate that our method can improve upon enrichment and ranking metrics for pose tasks compared to the baseline established by the physics-based scoring function vina, as implemented in the open-source docking software Smina. An important aspect of pose prediction in practical use cases for drug discovery campaigns is identifying a biologically relevant and binding-competent receptor conformation. Where there is limited crystallographic information available, such as only a single crystal structure or only a homology model for a target of interest, this can be particularly challenging. We introduce a simple Rosetta-based protocol for focused conformational sampling

and demonstrate that docking to small receptor ensembles can improve pose ranking metrics. The protocol is competitive in identification of active compounds for a given target compared to ensembles generated with much more computationally intensive sampling techniques, such as Markov-state models derived from molecular dynamics simulations.

## Author information

Kate Stafford; orcid.org/0000-0002-4144-4914
Brandon M. Anderson; orcid.org/0000-0001-6895-9902
Jon Sorenson; orcid.org/0000-0002-5301-8306
Henry van den Bedem; orcid.org/0000-0003-2358-841X

## Notes

AtomNet is a registered trademark of Atomwise Inc.

## References

(1)     Kiss, R.; Sandor, M.; Szalai, F. A. Http://Mcule.Com: A Public Web Service for Drug Discovery. *J. Cheminformatics* **2012**, *4* (1), P17. https://doi.org/10.1186/1758-2946-4-S1-P17.
(2)     Irwin, J. J.; Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59* (9), 4103–4120. https://doi.org/10.1021/acs.jmedchem.5b02008.
(3)     Damm-Ganamet, K. L.; Arora, N.; Becart, S.; Edwards, J. P.; Lebsack, A. D.; McAllister, H. M.; Nelen, M. I.; Rao, N. L.; Westover, L.; Wiener, J. J. M.; Mirzadegan, T. Accelerating Lead Identification by High Throughput Virtual Screening: Prospective Case Studies from the Pharmaceutical Industry. *J. Chem. Inf. Model.* **2019**, *59* (5), 2046–2062. https://doi.org/10.1021/acs.jcim.8b00941.
(4)     Schroedl, S. Current Methods and Challenges for Deep Learning in Drug Discovery. *Artif. Intell.* **2019**, *32–33*, 9–17. https://doi.org/10.1016/j.ddtec.2020.07.003.
(5)     Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432* (7019), 862–865. https://doi.org/10.1038/nature03197.
(6)     Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566* (7743), 224–229. https://doi.org/10.1038/s41586-019-0917-9.
(7)     Gabel, J.; Desaphy, J.; Rognan, D. Beware of Machine Learning-Based Scoring Functions—On the Danger of Developing Black Boxes. *J. Chem. Inf. Model.* **2014**, *54* (10), 2807–2815. https://doi.org/10.1021/ci500406k.
(8)     Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(9)     Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, *59* (2), 895–913. https://doi.org/10.1021/acs.jcim.8b00545.
(10)    Wankowicz, S. A.; de Oliveira, S. H. P.; Hogan, D. W.; van den Bedem, H.; Fraser, J. S. *Ligand Binding Remodels Protein Side Chain Conformational Heterogeneity*; preprint;

Biophysics, 2021. https://doi.org/10.1101/2021.09.21.461269.

(11) Vijayan, R. S. K.; He, P.; Modi, V.; Duong-Ly, K. C.; Ma, H.; Peterson, J. R.; Dunbrack, R. L.; Levy, R. M. Conformational Analysis of the DFG-Out Kinase Motif and Biochemical Profiling of Structurally Validated Type II Inhibitors. *J. Med. Chem.* **2015**, *58* (1), 466–479. https://doi.org/10.1021/jm501603h.

(12) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52* (5), 1262–1274. https://doi.org/10.1021/ci2005934.

(13) Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *J. Chem. Inf. Model.* **2010**, *50* (1), 186–193. https://doi.org/10.1021/ci9003943.

(14) Wang, C.; Zhang, Y. Improving Scoring-Docking-Screening Powers of Protein-Ligand Scoring Functions Using Random Forest. *J. Comput. Chem.* **2017**, *38* (3), 169–177. https://doi.org/10.1002/jcc.24667.

(15) Li, H.; Peng, J.; Sidorov, P.; Leung, Y.; Leung, K.-S.; Wong, M.-H.; Lu, G.; Ballester, P. J. Classical Scoring Functions for Docking Are Unable to Exploit Large Volumes of Structural and Interaction Data. *Bioinformatics* **2019**, *35* (20), 3989–3995. https://doi.org/10.1093/bioinformatics/btz183.

(16) Wallach, I.; Dzamba, M.; Heifets, A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-Based Drug Discovery. *ArXiv151002855 Cs Q-Bio Stat* **2015**.

(17) Feinberg, E. N.; Sur, D.; Wu, Z.; Husic, B. E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V. S. PotentialNet for Molecular Property Prediction. *ACS Cent. Sci.* **2018**, *4* (11), 1520–1530. https://doi.org/10.1021/acscentsci.8b00507.

(18) Cho, H.; Lee, E. K.; Choi, I. S. InteractionNet: Modeling and Explaining of Noncovalent Protein-Ligand Interactions with Noncovalent Graph Neural Network and Layer-Wise Relevance Propagation. *ArXiv200513438 Cs Q-Bio Stat* **2020**.

(19) Hassan-Harrirou, H.; Zhang, C.; Lemmin, T. RosENet: Improving Binding Affinity Prediction by Leveraging Molecular Mechanics Energies with a 3D Convolutional Neural Network. *bioRxiv* **2020**, 2020.05.12.090191. https://doi.org/10.1101/2020.05.12.090191.

(20) Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G. KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58* (2), 287–296. https://doi.org/10.1021/acs.jcim.7b00650.

(21) Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; Siedlecki, P. Development and Evaluation of a Deep Learning Model for Protein–Ligand Binding Affinity Prediction. *Bioinformatics* **2018**, *34* (21), 3666–3674. https://doi.org/10.1093/bioinformatics/bty374.

(22) Stecula, A.; Hussain, M. S.; Viola, R. E. Discovery of Novel Inhibitors of a Critical Brain Enzyme Using a Homology Model and a Deep Convolutional Neural Network. *J. Med. Chem.* **2020**, *63* (16), 8867–8875. https://doi.org/10.1021/acs.jmedchem.0c00473.

(23) Hsieh, C.-H.; Li, L.; Vanhauwaert, R.; Nguyen, K. T.; Davis, M. D.; Bu, G.; Wszolek, Z. K.; Wang, X. Miro1 Marks Parkinson's Disease Subset and Miro1 Reducer Rescues Neuron Loss in Parkinson's Models. *Cell Metab.* **2019**, *30* (6), 1131-1140.e7. https://doi.org/10.1016/j.cmet.2019.08.023.

(24) Huang, C.; Bernard, D.; Zhu, J.; Dash, R. C.; Chu, A.; Knupp, A.; Hakey, A.; Hadden, M. K.; Garmendia, A.; Tang, Y. Small Molecules Block the Interaction between Porcine Reproductive and Respiratory Syndrome Virus and CD163 Receptor and the Infection of Pig Cells. *Virol. J.* **2020**, *17* (1), 116. https://doi.org/10.1186/s12985-020-01361-7.

(25) Sieg, J.; Flachsenberg, F.; Rarey, M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2019**, *59* (3), 947–961. https://doi.org/10.1021/acs.jcim.8b00712.

(26) Morrone, J. A.; Weber, J. K.; Huynh, T.; Luo, H.; Cornell, W. D. Combining Docking Pose

Rank and Structure with Deep Learning Improves Protein–Ligand Binding Mode Prediction over a Baseline Docking Approach. *J. Chem. Inf. Model.* **2020**, *60* (9), 4170–4179. https://doi.org/10.1021/acs.jcim.9b00927.

(27) Francoeur, P. G.; Masuda, T.; Sunseri, J.; Jia, A.; Iovanisci, R. B.; Snyder, I.; Koes, D. R. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. *J. Chem. Inf. Model.* **2020**, *60* (9), 4200–4215. https://doi.org/10.1021/acs.jcim.0c00411.

(28) Lim, J.; Ryu, S.; Park, K.; Choe, Y. J.; Ham, J.; Kim, W. Y. Predicting Drug–Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation. *J. Chem. Inf. Model.* **2019**, *59* (9), 3981–3988. https://doi.org/10.1021/acs.jcim.9b00387.

(29) Bao, J.; He, X.; Zhang, J. Z. H. DeepBSP—a Machine Learning Method for Accurate Prediction of Protein–Ligand Docking Structures. *J. Chem. Inf. Model.* **2021**. https://doi.org/10.1021/acs.jcim.1c00334.

(30) Cappel, D.; Jerome, S.; Hessler, G.; Matter, H. Impact of Different Automated Binding Pose Generation Approaches on Relative Binding Free Energy Simulations. *J. Chem. Inf. Model.* **2020**, *60* (3), 1432–1444. https://doi.org/10.1021/acs.jcim.9b01118.

(31) van den Bedem, H.; Fraser, J. S. Integrative, Dynamic Structural Biology at Atomic Resolution—It's about Time. *Nat. Methods* **2015**, *12* (4), 307–318. https://doi.org/10.1038/nmeth.3324.

(32) van Zundert, G. C. P.; Hudson, B. M.; de Oliveira, S. H. P.; Keedy, D. A.; Fonseca, R.; Heliou, A.; Suresh, P.; Borrelli, K.; Day, T.; Fraser, J. S.; van den Bedem, H. *QFit-Ligand* Reveals Widespread Conformational Heterogeneity of Drug-Like Molecules in X-Ray Electron Density Maps. *J. Med. Chem.* **2018**, *61* (24), 11183–11198. https://doi.org/10.1021/acs.jmedchem.8b01292.

(33) Xie, T.; Saleh, T.; Rossi, P.; Kalodimos, C. G. Conformational States Dynamically Populated by a Kinase Determine Its Function. *Science* **2020**, *370* (6513). https://doi.org/10.1126/science.abc2754.

(34) Monod, J.; Wyman, J.; Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(35) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. Conformational Selection or Induced Fit: A Flux Description of Reaction Mechanism. *Proc. Natl. Acad. Sci.* **2009**, *106* (33), 13737–13741. https://doi.org/10.1073/pnas.0907195106.

(36) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.* **1958**, *44* (2), 98–104. https://doi.org/10.1073/pnas.44.2.98.

(37) Conformational states dynamically populated by a kinase determine its function | Science https://science.sciencemag.org/content/370/6513/eabc2754 (accessed 2021 -01 -11).

(38) Kamenik, A. S.; Singh, I.; Lak, P.; Balius, T. E.; Liedl, K. R.; Shoichet, B. K. Energy Penalties Enhance Flexible Receptor Docking in a Model Cavity. *bioRxiv* **2021**, 2021.04.20.440636. https://doi.org/10.1101/2021.04.20.440636.

(39) Fischer, M.; Coleman, R. G.; Fraser, J. S.; Shoichet, B. K. Incorporation of Protein Flexibility and Conformational Energy Penalties in Docking Screens to Improve Ligand Discovery. *Nat. Chem.* **2014**, *6* (7), 575–583. https://doi.org/10.1038/nchem.1954.

(40) Cleves, A. E.; Jain, A. N. Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket. *J. Chem. Inf. Model.* **2020**, *60* (9), 4296–4310. https://doi.org/10.1021/acs.jcim.0c00115.

(41) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114* (10), 2271–2278. https://doi.org/10.1016/j.bpj.2018.02.038.

(42) Paul, F.; Meng, Y.; Roux, B. Identification of Druggable Kinase Target Conformations

Using Markov Model Metastable States Analysis of Apo-Abl. *J. Chem. Theory Comput.* **2020**, *16* (3), 1896–1912. https://doi.org/10.1021/acs.jctc.9b01158.

(43) Chandak, T.; Mayginnes, J. P.; Mayes, H.; Wong, C. F. Using Machine Learning to Improve Ensemble Docking for Drug Discovery. *Proteins Struct. Funct. Bioinforma.* **2020**, *88* (10), 1263–1270. https://doi.org/10.1002/prot.25899.

(44) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47* (12), 2977–2980. https://doi.org/10.1021/jm030580l.

(45) Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-Wide Collection of Binding Data: Current Status of the PDBbind Database. *Bioinformatics* **2015**, *31* (3), 405–412. https://doi.org/10.1093/bioinformatics/btu626.

(46) Song, Y.; DiMaio, F.; Wang, R. Y.-R.; Kim, D.; Miles, C.; Brunette, T. J.; Thompson, J.; Baker, D. High-Resolution Comparative Modeling with RosettaCM. *Structure* **2013**, *21* (10), 1735–1742. https://doi.org/10.1016/j.str.2013.08.005.

(47) Landrum, Greg. RDKit: Open-Source Cheminformatics.

(48) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*; AAAI Press, 1996; pp 226–231.

(49) Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise | Journal of Chemical Information and Modeling https://pubs.acs.org/doi/abs/10.1021/ci300604z (accessed 2021 -01 -11).

(50) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; VanderMeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 33.

(51) Meng, Y.; Gao, C.; Clawson, D.; Atwell, S.; Russell, M.; Vieth, M.; Roux, B. Predicting the Conformational Variability of Abl Tyrosine Kinase Using Molecular Dynamics Simulations and Markov State Models. *J. Chem. Theory Comput.* **2018**, *14* (5), 2721–2732. https://doi.org/10.1021/acs.jctc.7b01170.

(52) Klicpera, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *ArXiv200303123 Phys. Stat* **2020**.

(53) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. *ArXiv170608566 Phys. Stat* **2017**.

(54) Wallach, I.; Heifets, A. Most Ligand-Based Classification Benchmarks Reward Memorization Rather than Generalization. *J. Chem. Inf. Model.* **2018**, *58* (5), 916–932. https://doi.org/10.1021/acs.jcim.7b00403.

(55) Sieg, J.; Flachsenberg, F.; Rarey, M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2019**, *59* (3), 947–961. https://doi.org/10.1021/acs.jcim.8b00712.

(56) Scantlebury, J.; Brown, N.; Von Delft, F.; Deane, C. M. Data Set Augmentation Allows Deep Learning-Based Virtual Screening to Better Generalize to Unseen Target Classes and Highlight Important Binding Interactions. *J. Chem. Inf. Model.* **2020**, *60* (8), 3722–3730. https://doi.org/10.1021/acs.jcim.0c00263.

(57) Morrone, J. A.; Weber, J. K.; Huynh, T.; Luo, H.; Cornell, W. D. Combining Docking Pose Rank and Structure with Deep Learning Improves Protein–Ligand Binding Mode Prediction over a Baseline Docking Approach. *J. Chem. Inf. Model.* **2020**, *60* (9), 4170–4179. https://doi.org/10.1021/acs.jcim.9b00927.

(58) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein–Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57* (4), 942–957. https://doi.org/10.1021/acs.jcim.6b00740.

(59)  Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *MolProbity* : All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (1), 12–21. https://doi.org/10.1107/S0907444909042073.

(60)  Hantschel, O.; Superti-Furga, G. Regulation of the C-Abl and Bcr–Abl Tyrosine Kinases. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (1), 33–44. https://doi.org/10.1038/nrm1280.

(61)  An, X.; Tiwari, A. K.; Sun, Y.; Ding, P.-R.; Ashby, C. R.; Chen, Z.-S. BCR-ABL Tyrosine Kinase Inhibitors in the Treatment of Philadelphia Chromosome Positive Chronic Myeloid Leukemia: A Review. *Leuk. Res.* **2010**, *34* (10), 1255–1268. https://doi.org/10.1016/j.leukres.2010.04.016.

(62)  Modi, V.; Dunbrack, R. L. Defining a New Nomenclature for the Structures of Active and Inactive Kinases. *Proc. Natl. Acad. Sci.* **2019**, *116* (14), 6818–6827. https://doi.org/10.1073/pnas.1814279116.

(63)  van den Maaten, L. J. P.; Hinton, G. E. Visualizing High-Dimensional Data Using t-SNE. *J Mach Learn Res* **2008**, *9*, 2579–2605.

(64)  Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger; *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700-4708

# AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High Throughput Screens

Kate A. Stafford[1], Brandon M. Anderson[1], Jon Sorenson[1], and Henry van den Bedem[1,2,*]

Atomwise, Inc., 717 Market Street, Suite 800, San Francisco, CA 94103, USA.
Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, USA.

## Rosetta Protocols

```
<ROSETTASCRIPTS>
      <SCOREFXNS>
          <ScoreFunction name="s3" weights="score3"/>
          <ScoreFunction name="s4" weights="score4_smooth_cart"/>
          <ScoreFunction name="fa" weights="ref2015_cart" >
              <Reweight scoretype="metalbinding_constraint" weight="1.0" />
          </ScoreFunction>
          <ScoreFunction name="r15" weights="ref2015" >
              <Reweight scoretype="pro_close" weight="0.0" />
          </ScoreFunction>
          <ScoreFunction name="cen" weights="cen_std"/>
      </SCOREFXNS>

      <MOVERS>
          <SetupMetalsMover name="metals" />
          <Hybridize name="hybridize" batch="1" stage1_increase_cycles="1.0"
stage2_increase_cycles="1.0" stage1_scorefxn="s3" stage2_scorefxn="s4"
fa_scorefxn="fa" fragprob_stage2="0.0" add_hetatm="1"
hetatm_to_protein_cst_weight="1.0" hetatm_cst_weight="1.0" stage2_temperature="0.5">
              <Template pdb="test_with_lig.pdb" cst_file="AUTO" weight="1.000" />

          </Hybridize>
          <FastRelax name="relax" scorefxn="fa" />
      </MOVERS>
      <PROTOCOLS>
          <Add mover="metals" />
          <Add mover="hybridize"/>
      </PROTOCOLS>
      <OUTPUT scorefxn="r15" />
   </ROSETTASCRIPTS>
```