EQUIBIND: GEOMETRIC DEEP LEARNING FOR DRUG BINDING STRUCTURE PREDICTION

Hannes Stärk*, Octavian-Eugen Ganea*, Lagnajit Pattanaik, Regina Barzilay & Tommi Jaakkola
Massachusetts Institute of Technology
{hstark, oct}@mit.edu

ABSTRACT

Predicting how a drug-like molecule binds to a specific protein target is a core problem in drug discovery. An extremely fast computational binding method would enable key applications such as fast virtual screening or drug engineering. Existing methods are computationally expensive as they rely on heavy candidate sampling coupled with scoring, ranking, and fine-tuning steps. We challenge this paradigm with EQUIBIND, an SE(3)-equivariant geometric deep learning model performing direct-shot prediction of both i) the receptor binding location (blind docking) and ii) the ligand's bound pose and orientation. EquiBind achieves significant speed-ups and better quality compared to traditional and recent baselines. Further, we show extra improvements when coupling it with existing fine-tuning techniques at the cost of increased running time. Finally, we propose a novel and fast fine-tuning model that adjusts torsion angles of a ligand's rotatable bonds based on closed-form global minima of the von Mises angular distance to a given input atomic point cloud, avoiding previous expensive differential evolution strategies for energy minimization.

1 INTRODUCTION

Drug discovery is an expensive process, e.g., a single drug costs around 1 billion dollars and takes 10 years of development and testing before potentially being FDA approved. Moreover, this process can fail at any point, e.g., due to unforeseen side effects or experimental disproof of the promised therapeutic efficacy. Worse, there are 10^{60} possible drug-like molecules (Reymond & Awale, 2012), going far beyond current experimental capabilities.

Accurate computational methods, e.g., deep learning (DL) based, can drastically reduce the molecular search space, but need to be extremely fast to scan the vast biological and chemical spaces for both desired and unexpected effects. For instance, a novel drug that inactivates an important cancer protein might negatively inhibit other essential proteins in the human body, potentially resulting in life-threatening side effects. Given that the human proteome contains up to $100\,000$ protein types, the current hope is to scan for these interactions in a computational manner before bringing a few promising candidates to *in vitro* and *in vivo* testing.

A core problem in drug discovery is understanding how drug-like molecules (ligands) interact and form complexes with target proteins (receptors) – *drug binding* – which is a prerequisite for virtual screening. This is a difficult problem with different facets and constraints: binding kinetics, conformational changes, and chemical and geometrical atomic interaction types are part of the domain knowledge describing ligand-protein binding mechanisms (Du et al., 2016). For instance, classical models for molecular complex formation are "lock-and-key", "induced fit", and "conformational selection", while hydrophobic, hydrogen-bonding, and π-stacking are the most frequent atomic binding interactions, but other types often occur during binding (de Freitas & Schapira, 2017).

Current *in silico* approaches for (3D) structural drug binding achieve high quality at a significant computational cost: in our experiments the GNINA method (McNutt et al., 2021) takes on average 146 seconds for a single ligand-receptor pair, while the popular commercial software Glide (Halgren et al., 2004) is up to 9 times slower. This is caused by the common strategy employed by all previous binding methods: first, a large set of candidate complexes (e.g., millions) is generated via thorough sampling of possible binding locations and poses (Hassan et al., 2017); then, scoring and ranking

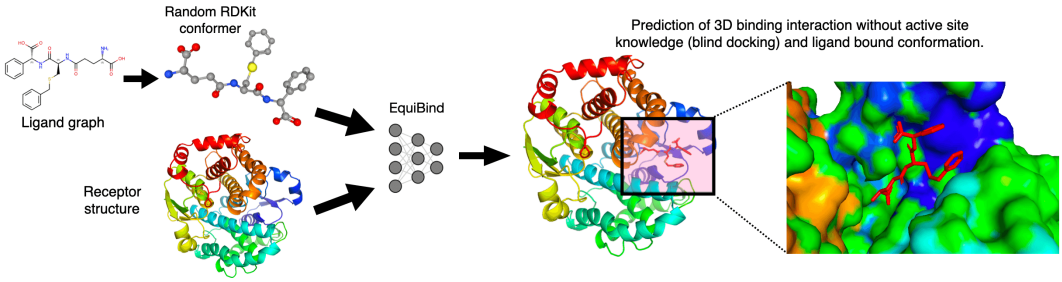

Figure 1: High-level overview of the structural drug binding problem tackled by EQUIBIND.

steps are used to retrieve the most promising instances; finally, an energy-based fine-tuning method is employed to best fit the ligand in the respective pocket locations.

Here, we introduce EQUIBIND, a novel geometric & graph deep learning model for structural drug binding – fig. 1. Inspired by Ganea et al. (2021a), we exploit graph matching networks (GMN) (Li et al., 2019) and E(3)-equivariant graph neural networks (E(3)-GNN) (Satorras et al., 2021) to perform a direct prediction of the ligand-receptor complex structure without relying on heavy sampling as prior work, thus achieving **significant inference time speed-ups**. Moreover, since 3D structural data suffers from scarcity (e.g., only around 19K experimental complexes are publicly available in the PDBbind database), it is crucial to inject the right physical, chemical, or biological *inductive biases* into DL models to avoid learning these priors from insufficient amounts of data and to create trustable models. Towards this goal, EQUIBIND:

- guarantees independence to the initial 3D placements and orientations of the two molecules, i.e., the exact same complex is always predicted for the same input unbound structures,
- incorporates an efficient mechanism for biologically plausible ligand flexibility by only altering torsion angles of rotatable bonds while keeping local structures (bond angles and lengths) fixed,

We focus on the *blind docking* scenario, i.e., zero knowledge of the protein's binding site or pocket. However, our method can easily be adapted to situations where the approximate binding location is known. Similar to (Zhang et al., 2020), we argue that errors in the ground truth binding pocket conformation heavily affect traditional docking methods that are conditioned on the receptor active site (Lang et al., 2009; Trott & Olson, 2010). In practice, the ground truth 3D locations of the binding atoms might be low-resolution, might not be known at all (e.g., for novel antigens), or we might be interested in discovering new druggable locations on a protein's surface that were previously thought to be undruggable (e.g. exploring allosteric binding locations rather than orthosteric sites).

Empirically, we investigate two settings: re-docking (i.e., taking the bound ligand structure out of a complex and asking the model to dock it) and flexible self-docking (i.e., ligands have no bound structure knowledge prior to docking). We assume a rigid receptor, but we model ligand flexibility by first predicting an atomic point cloud of the deformed molecule and then employing a fast algorithm to extract internal changes of rotatable bonds' torsion angles that would match the point cloud as well as possible. Instead of minimizing the root-mean-square deviation (RMSD) using expensive optimization strategies (e.g., differential evolution approaches (Méndez-Lucio et al., 2021)), we maximize the log-likelihood of a von Mises distribution that fits the torsion angles, proving closed-form expressions of the global optimum. Experimentally, we show improved quality in various metrics over popular and recent state-of-the-art baselines at a fraction of the running time. Finally, we show the power of combining EquiBind with existing energy-based methods to realize a hybrid DL approach. Indeed, we believe the future of computational drug discovery will follow the paradigm demonstrated here.

2 RELATED WORK

Protein and molecular structure prediction. Obtaining experimental 3D structural data of molecules and proteins is a highly expensive process. However, very recent DL models have produced a breakthrough in computational protein folding (Jumper et al., 2021b; Baek et al., 2021) and fast generation of small molecule low-energy conformation ensembles (Ganea et al., 2021b; Luo et al., 2021; Xu et al., 2020; Shi et al., 2021). These methods aim to accelerate the discovery of new structures and complement experimental data in various applications such as drug discovery.

Protein representations (for DL-based molecular interactions). To be useful for predicting molecular interactions, proteins must be modeled in specific ways to account for different views: backbone & side-chains, protein surface, atomic point cloud, or amino-acid sequence. Somnath et al. (2021) create a hierarchical representation of proteins and prove its utility in binding and function prediction. Gainza et al. (2020); Sverrisson et al. (2021) leverage geometric deep learning and mesh convolutional neural networks (CNN) to embed protein surface patches into fingerprints and allow for fast scanning and binding site identification, removing the need for handcrafted or expensive pre-computed features. However, these methods do not perform the full structural blind docking task that involves prediction of the binding site, of the orientations and placements of the two molecular structures, and of the internal conformational deformations during binding. Various other protein representations have been proposed for (graph) DL methods for individual structure prediction (Jing et al., 2020), protein-protein interactions (Dai & Bailey-Kellogg, 2021; Eismann et al., 2020; Townshend et al., 2019), or protein function prediction (Gligorijević et al., 2021).

Druggable binding site identification. Traditional computational methods for scanning proteins for their most "druggable" areas have leveraged various views such as utilizing the protein's 3D structure or/and residue sequence, extracting geometric features, building large template libraries, or relying on energy-based models (Macari et al., 2019). Recently, DL changed this paradigm, e.g., using 3D CNNs (Aggarwal et al., 2021; Jiménez et al., 2017; Torng & Altman, 2019b) or sequence models (Sankararaman et al., 2010).

Popular and more recent drug binding models. Representative docking software for drug-like molecules are AutoDock Vina (Trott & Olson, 2010) and its various extensions for improving speed (Trott & Olson, 2010), scoring (Koes et al., 2013) or for blind docking (Hassan et al., 2017). As mentioned in section 1, these methods employ a multi-stage strategy based on heavy candidate sampling, scoring, ranking, and fine-tuning. Various subsequent methods aimed to improve some parts of this pipeline (Zhang et al., 2020; Mohammad et al., 2021; McNutt et al., 2021; Francoeur et al., 2020), with a special emphasis on the scoring function – see below.

GNNs and CNNs for binding scoring functions and binding affinity prediction. Deep learning on 3D voxel images (via 3D CNNs) or interaction graphs (via GNNs) have improved the traditional hand-designed scoring function used in AutoDock Vina, enabling better fine-tuning of predicted docked poses, as well as direct binding affinity prediction from the 3D complex (McNutt et al., 2021; Francoeur et al., 2020; Ragoza et al., 2017; Wallach et al., 2015; Lim et al., 2019; Morrone et al., 2020; Jiang et al., 2021; Shen et al., 2021; Jastrzebski et al., 2020; Bao et al., 2021; Torng & Altman, 2019a; Li et al., 2021). However, some methods (Karimi et al., 2019; Gao et al., 2018) have found that using the protein sequence and the drug SMILES string already provide competitive predictions for binding affinity without the need for 3D structural data. Closer to our approach, Méndez-Lucio et al. (2021) has shown that optimizing the ligand's global 3D position and orientation and the torsion angles of rotatable bonds to minimize a GNN based scoring function improves fine-tuning of the ligand into the active site and its predicted bound pose.

Applications of drug binding and protein-protein docking methods. Computational docking methods are employed for various facets of drug discovery, e.g., fast virtual screening (Gniewek et al., 2021; Jastrzebski et al., 2020) or de novo binder generation (Masuda et al., 2020; Imrie et al., 2021; Drotár et al., 2021). A related problem is *protein-protein docking* in which recent methods have performed direct prediction of the complex structure from the two concatenated input sequences using evolutionary information (Evans et al., 2021), or have leveraged geometric deep learning to model rigid body docking (Ganea et al., 2021a) or side-chains structures (Jindal et al., 2021).

Incorporating Euclidean symmetries into GNNs. Injecting Euclidean 3D transformations into geometric DL models has become possible using equivariant message passing layers (Cohen & Welling, 2016; Thomas et al., 2018; Fuchs et al., 2020; Satorras et al., 2021; Brandstetter et al., 2021; Batzner et al., 2021). Our method follows Ganea et al. (2021a) to incorporate SE(3) pairwise equivariance into message passing neural networks for the drug binding problem. However, different from this method, we go beyond rigid docking and model ligand conformational flexibility.

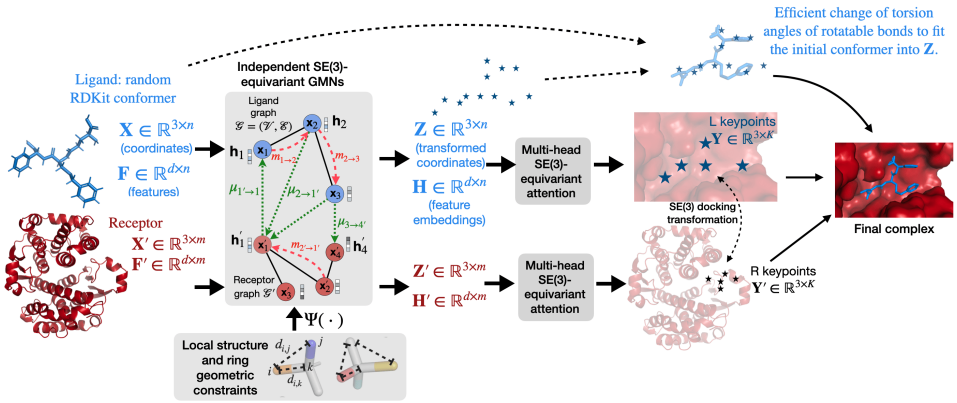

Figure 2: EQUIBIND model architecture.

3 EQUIBIND MODEL

We describe our EQUIBIND model, highlighted in fig. 1 and detailed in fig. 2. It takes as input a ligand molecular graph with a random associated unbound 3D conformer (e.g., generated using RDKit/ETKDG Riniker & Landrum (2015)), as well as a receptor-bound structure. As previously noted, we only model ligand flexibility in this work, assuming a rigid protein conformation.

K-NN graph representations. We represent both input molecules as spatial k-nearest neighbor (k-NN) graphs. The ligand graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ uses atoms as nodes with their respective 3D coordinates from the unbound conformer denoted as $\mathbf{X} \in \mathbb{R}^{3\times n}$, and initial features $\mathbf{F} \in \mathbb{R}^{d\times n}$ (e.g., atom type). Edges include all atom pairs within a distance cutoff of 4 Å. The receptor graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ has residues as nodes and their 3D coordinates $\mathbf{X}' \in \mathbb{R}^{3\times m}$ are given by the α-carbon locations. Each node is connected in the graph to the closest 10 other nodes at less than 30 Å distance. The receptor node features $\mathbf{F}' \in \mathbb{R}^{d\times m}$ and the ligand features are detailed in appendix C.

Independent E(3)-equivariant transformations. Similar to (Ganea et al., 2021a), an important geometric inductive bias is to predict the same binding complex no matter how the initial molecules are positioned and oriented in space. This is especially needed for data-scarce problems such as structural drug binding. Towards this goal, we use *Independent E(3)-Equivariant Graph Matching Network* (IEGMN) (Ganea et al., 2021a) which combines Graph Matching Networks (Li et al., 2019) and E(3)-Equivariant Graph Neural Networks (Satorras et al., 2021). This architecture jointly transforms both features and 3D coordinates to perform *intra* and *inter* neural graph message passing. Formally, $IEGMN(\mathbf{X}, \mathbf{F}, \mathbf{X}', \mathbf{F}') = \mathbf{Z} \in \mathbb{R}^{3\times n}, \mathbf{H} \in \mathbb{R}^{d\times n}, \mathbf{Z}' \in \mathbb{R}^{3\times m}, \mathbf{H}' \in \mathbb{R}^{d\times m}$, where $\mathbf{Z}, \mathbf{Z}'$ are coordinate transformations, while $\mathbf{H}, \mathbf{H}'$ are feature embeddings. The core property of IEGMNs is that stacking any number of such layers guarantees that any independent rotation and translation of the original input structures will be exactly reflected in the outputs, i.e., $IEGMN(\mathbf{UX} + \mathbf{b}, \mathbf{F}, \mathbf{U}'\mathbf{X}' + \mathbf{b}', \mathbf{F}') = \mathbf{UZ} + \mathbf{b}, \mathbf{H}, \mathbf{U}'\mathbf{Z}' + \mathbf{b}', \mathbf{H}'$ for any orthogonal matrices $\mathbf{U}, \mathbf{U}' \in \mathbf{SO}(3)$ and translation vectors $\mathbf{b}, \mathbf{b}' \in \mathbb{R}^3$. In practice, the $\mathbf{Z}, \mathbf{H}, \mathbf{Z}', \mathbf{H}'$ outputs shown in fig. 2 are obtained by stacking several IEGMN layers. The exact update equations are detailed in Appendix C.

The role of Z. The output of the coordinate E(3)-equivariant transformations denoted as $\mathbf{Z}$ and $\mathbf{Z}'$ will be used in different roles: to identify the rigid body transformation and the binding site, as well as to model ligand flexibility by training $\mathbf{Z}$ to represent the deformed atomic point cloud. We detail both steps below.

3.1 RIGID TRANSFORMATION THROUGH BINDING KEYPOINTS

To identify the rigid SE(3) transformation to dock the ligand in the right position and orientation, we follow (Ganea et al., 2021a) and compute ligand and receptor keypoints of size K using an SE(3)-equivariant multi-head attention mechanism defined as $\mathbf{y}_k := \sum_{i=1}^{n} \alpha_i^k \mathbf{z}_i, \quad \mathbf{y}'_k := \sum_{j=1}^{m} \beta_j^k \mathbf{z}'_j$ where $\alpha_i^k = softmax_i(\frac{1}{\sqrt{d}} \mathbf{h}_{1i}^\top \mathbf{U} \mu(\varphi(\mathbf{H}_2)))$ and similarly defined β_j^k are attention coefficients, with $\mathbf{U}$ a parametric learnable matrix. These keypoints are trained to match the ground truth binding pocket points using an optimal transport loss that recovers their alignment (detailed in (Ganea et al.,

2021a)). In our case, ground truth binding pocket points are defined as midpoints of segments connecting ligand atoms to receptor atoms (e.g., from side-chains) that are closer than 4 Å . For models incorporating ligand flexibility, these pocket points are defined as all ligand atoms that are closer than 4 Å to any receptor atom. When the ligand and receptor are separated, we seek to identify the corresponding binding sites, and their exact matching using the two predicted keypoint sets $\mathbf{Y}, \mathbf{Y}' \in \mathbb{R}^{3xK}$. If predicted perfectly, the SE(3) transformation to superimpose $\mathbf{Y}$ and $\mathbf{Y}'$ would precisely correspond to the binding SE(3) transformation to perform ligand docking.

3.2 Modeling Chemically Plausible Ligand Flexibility

It has previously been assumed that the most flexible parts of drug-like molecules are rotatable bonds, while *local atomic structures (LAS)* (bond lengths and adjacent bond angles) or small rings are mostly rigid (Trott & Olson, 2010; Zsoldos et al., 2007; Huang, 2018; Méndez-Lucio et al., 2021). We here follow this hypothesis in two different ways as below.

We model ligand flexibility through $\mathbf{Z}$, which will represent a good approximation of the deformed atomic point cloud of the original conformer (i.e., into its bound structure). We train our model with two extra loss function terms: ligand-RMSD (Root-mean-square deviation) and Kabsch-RMSD [1].

Distance Geometric Constraints. Our first goal is to enforce LAS distance constraints in the IEGMN layers after each coordinate transformation, i.e., through a tailored differentiable function Ψ , which we call *"LAS distance geometry (DG) projection."* While a hard constraint might be difficult to impose exactly, we find the following soft strategy to work well. Formally, the transformed coordinates $\mathbf{Z}$ satisfy the LAS DG constraints iff the following function is minimized w.r.t. $\mathbf{Z}$ for a fixed given (random low-energy unbound) conformer $\mathbf{X}$:

$$\mathcal{S}(\mathbf{Z},\mathbf{X}) = \sum_{\{(i,j)\in\mathcal{E}\}} (d^2_{\mathbf{X}}(i,j) - d^2_{\mathbf{Z}}(i,j))^2 + \sum_{\{i,j:\text{2-hops away in } \mathcal{G}\}} (d^2_{\mathbf{X}}(i,j) - d^2_{\mathbf{Z}}(i,j))^2 + \sum_{\{i,j:\text{i in aromatic ring with j}\}} (d^2_{\mathbf{X}}(i,j) - d^2_{\mathbf{Z}}(i,j))^2$$

Thus, our definition of Ψ is hard-coding a fixed number (T) of gradient descent layers that aim to minimize $\mathcal{S}$:

$$\Psi(\mathbf{Z}) = \Psi_T \circ \ldots \circ \Psi_1(\mathbf{Z}), \quad \Psi_t(\mathbf{Z}) = \mathbf{Z} - \eta \nabla_{\mathbf{Z}} \mathcal{S}(\mathbf{Z},\mathbf{X}), \forall t$$

which is easy since gradients of $\mathcal{S}$ have a simple closed-form expression. A similar approach can be employed for modeling various other rigid substructures such as aromatic rings. T and the correction step size η are model hyperparameters chosen as described in Appendix C.

Fast Point Cloud Ligand Fitting. However, while helpful for model training, the previous gradient descent-based projection is not guaranteed to enforce hard LAS DG constraints and, thus, might produce implausible conformers in practice as we show in fig. 12.

To address this issue, we only change the torsion angles of the initial (RDKit) unbound conformer $\mathbf{X}$ to match $\mathbf{Z}$ as well as possible while keeping LAS fixed and, thus, hard-guaranteeing chemically plausible output bound conformers. The output will be a new conformer $\mathbf{C} \in \mathbb{R}^{3xn}$ with $\mathcal{S}(\mathbf{C},\mathbf{X}) = 0$. First, $\mathbf{C}$ is initialized as $\mathbf{X}$, and only its rotatable bonds' torsion angles are changed.

A choice is to optimize $\mathbf{C}$ for minimizing Kabsch-RMSD($\mathbf{Z},\mathbf{C}$). However, such an approach requires an iterative optimization strategy of all torsion angles of rotatable bonds, which can be done using a differential evolution algorithm as in Méndez-Lucio et al. (2021), or other local random search strategies. This is computationally expensive (e.g., 51 minutes for a single 44 rotatable bond molecule). A gradient-based method that could better capture the various molecular interactions, but computing the gradients of a point cloud w.r.t. its bonds' torsion angles is non-trivial given the geometric dependencies between dihedral angles, i.e., eq. (2).

We present a much cheaper alternative for which a closed form solution exists (does not require optimization): we compute the dihedral angles of rotatable bonds of $\mathbf{C}$ as maximum likelihood estimates of von Mises distributions on dihedral angles of $\mathbf{Z}$. This reduces to the following maximization:

$$\max_{\{\angle(kij,ijl)\}} \sum_{(k,i),(i,j),(j,l)\in\mathcal{E}} \cos(\angle_{\mathbf{Z}}(kij,ijl) - \angle(kij,ijl)) \tag{1}$$

[1] RMSD after superimposition, or RMSD of atomic positions.

where $\angle_{\mathbf{Z}}(kij, ijl)$ are the clockwise dihedral angles of $\mathbf{Z}$, and $\angle(kij, ijl)$ are the dihedrals of $\mathbf{C}$ that we seek to optimize.

However, we have to explicitly take into account that *all the dihedral angles for the same rotatable bond (i,j) are coupled* by the following constraint (Ganea et al., 2021b):

$$\begin{aligned} &\angle(kij, ijl) =_{2\pi} \angle(k'ij, ijl') + \angle(kij, k'ij) + \angle(ijl', ijl), \\ &\forall (i,j) \in \mathcal{E}, \forall k, k' \in \mathcal{N}_i, \forall l, l' \in \mathcal{N}_j \end{aligned} \tag{2}$$

where $\angle(kij, k'ij)$ and $\angle(ijl', ijl)$ depend only on the local structures of nodes i and j, respectively, thus will not change if the torsion angle of bond (i,j) changes.

To minimize eq. (1), we can simply do it independently per each rotatable bond $(i,j) \in \mathcal{E}$. Let us fix one such bond (i,j) and use the notations: $\Delta_{kl} = \angle(kij, ijl)$ and $\beta_{kk'l'l} = \angle(kij, k'ij) + \angle(ijl', ijl)$. Additionally, for any angle α, we define: $\mathbf{A}_\alpha := \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$ and $\mathbf{s}_\alpha := \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix}$. Thus, we rewrite the constraint in eq. (2) as $\mathbf{s}_{\Delta_{kl}} = \mathbf{A}_{\beta_{kk'l'l}} \mathbf{s}_{\Delta_{k'l'}}, \forall k, k' \in \mathcal{N}_i, \forall l, l' \in \mathcal{N}_j$. eq. (1) is then rewritten for bond (i,j), up to a constant:

$$\max_{\{\Delta_{kl}\}} \sum_{k \in \mathcal{N}_i} \sum_{l \in \mathcal{N}_j} \langle \mathbf{s}_{\Delta_{kl}}, \mathbf{s}^*_{\Delta_{kl}} \rangle \tag{3}$$

Choosing any fixed $k_0 \in \mathcal{N}_i$, $l_0 \in \mathcal{N}_j$, the above becomes

$$\max_{\Delta_{k_0 l_0}} \sum_{k \in \mathcal{N}_i} \sum_{l \in \mathcal{N}_j} \langle \mathbf{A}_{\beta_{kk_0l_0l}} \mathbf{s}_{\Delta_{k_0l_0}}, \mathbf{s}^*_{\Delta_{kl}} \rangle = \mathbf{s}^\top_{\Delta_{k_0l_0}} \mathbf{v} \tag{4}$$

where $\mathbf{v} := \sum_{k \in \mathcal{N}_i} \sum_{l \in \mathcal{N}_j} \mathbf{A}^\top_{\beta_{kk_0l_0l}} \mathbf{s}^*_{\Delta_{kl}}$. This has the closed form solution $\mathbf{s}_{\Delta_{k_0l_0}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$, which finally gives all dihedral angles Δ_{kl} in closed form. One can easily verify that the choice of $k_0 \in \mathcal{N}_i$ and $l_0 \in \mathcal{N}_j$ will not affect the values of the predicted dihedrals $\Delta_{kl}, \forall k \in \mathcal{N}_i, \forall l \in \mathcal{N}_j$.

4 EXPERIMENTS

Data. We provide a new time-based dataset split and preprocessing pipeline for DL drug binding methods. We make this data and associated scripts available together with code to reproduce results or perform fast docking with the provided model weights at `https://github.com/HannesStark/EquiBind`. We use protein-ligand complexes from PDBBind Liu et al. (2017), which is a subset of the Protein Data Bank (PDB) Berman et al. (2003) that provides 3D structures of individual proteins and complexes. The newest version, PDBBind v2020, contains 19 443 protein-ligand complexes with 3890 unique receptors and 15 193 unique ligands. We describe our preprocessing to remove pathologies of the data, the motivation for the new time split, and the exact dataset split in appendix B.

Baselines. Quick Vina-W (QVina-W) is a classical docking program specifically developed for "wide" or blind docking. SMINA (Koes et al., 2013) builds on AutoDock Vina by designing an improved and empirical scoring function. GNINA (McNutt et al., 2021; Francoeur et al., 2020) further develops a DL scoring function using CNNs and a grid-based featurization scheme. GLIDE (Halgren et al., 2004) is a popular commercial docking software of which we use the 2021-4 release. We run GLIDE, GNINA, and SMINA with their default settings and for QVina-W we increase the exhaustiveness (parameter controlling the search time) to 64 with which it is still faster than the other baselines.

EquiBind models. Our model can be applied in various scenarios, see caption of table 1. First, the EQUIBIND-U model generates an uncorrected ligand point cloud $\mathbf{Z}$ that does not necessarily have valid bond angles and lengths. The standard EQUIBIND takes this output and applies our fast point cloud ligand fitting in section 3.2 to obtain a realistic molecular structure. The model EQUIBIND-R treats the ligand as a rigid body, being trained with no flexibility loss terms. The fine-tuning model EQUIBIND + Q builds on top of this output by searching refined conformations using Quick Vina 2 in a 5 Å bounding box around the ligand predicted by EQUIBIND-R. The instantiations EQUIBIND

+ Q2 does the same with two times as many sampled ligand positions, and EQUIBIND + S as well as EQUIBIND-R + S instead use SMINA for fine-tuning.

Evaluation Metrics. We use the ligand root mean square deviation (L-RMSD), the centroid distance, and the Kabsch-RMSD. We calculate all metrics after hydrogens are removed. The *centroid distance* measures the ability of the model to find the correct binding pocket (for a given ligand) via the distance between the averaged coordinates of the predicted and true bound ligand atoms. *Kabsch RMSD* is the lowest possible RMSD that can be obtained by SE(3) transformation of the ligand (i.e., RMSD after superimposition with the Kabsch algorithm). *L-RMSD* is the mean squared error between the atoms of the predicted and bound ligands. All RMSDs are calculated using OpenBabel's symmetry corrected and atom order invariant RMSD tool `obrms`. Next to the mean and cumulative distributions, we report the percent of predictions below a given error threshold.

Finally, we show the average number of seconds needed to make a prediction for a test complex, given the receptor and the initial ligand structure. The receptor preparation time is excluded (mainly an additional 393 sec for GLIDE) since this step is commonly only performed once before docking many ligands to the same receptor. We ran all runtime measurements on the same machine using 16 logical CPU cores (except for GLIDE, which does not support multithreading – detailed in Appendix C), once with and once without access to a 6GB GTX 1060 GPU.

Table 1: **Flexible blind self-docking.** All methods receive a random RDKit conformer of the ligand molecule as input and are tasked to find its binding site and the right orientation + conformation in which it binds. EQUIBIND-U refers to the model producing uncorrected atomic point clouds **Z** that are not necessarily chemically plausible ligands. EQUIBIND performs our fast conformer fitting corrections – see section 3.2. EQUIBIND+Q predicts an approximate ligand position and fine-tunes it using QuickVina 2. EQUIBIND+Q2 samples more candidate positions, and EQUIBINDS uses SMINA for fine-tuning. *GLIDE runtime details in Appendix C.

			LIGAND RMSD ↓						CENTROID DISTANCE ↓						KABSCH	
	AVG. SEC.	AVG. SEC.		PERCENTILES ↓			% BELOW THRESHOLD ↑			PERCENTILES ↓			% BELOW THRESHOLD ↑		RMSD ↓	
METHODS	CPU	GPU	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	MEDIAN
QVINA-W	49	-	13.6	2.5	7.7	23.7	40.2	20.9	11.9	0.9	3.7	22.9	54.6	41.0	**2.1**	1.9
GNINA	247	146	13.3	2.8	8.7	22.1	37.1	21.2	11.5	1.0	4.5	21.2	52.0	36.0	2.2	1.8
SMINA	146	-	12.1	3.8	8.1	17.9	33.9	13.5	9.8	1.3	3.7	16.2	55.9	38.0	2.2	1.9
GLIDE*	1405	-	16.2	2.6	9.3	28.1	33.6	21.8	14.4	**0.8**	5.6	26.9	48.7	36.1	2.2	1.9
EQUIBIND+Q	8	8	8.4	2.6	6.6	11.1	38.0	18.7	5.9	1.0	2.5	6.4	68.7	44.6	2.3	1.9
EQUIBIND+Q2	15	15	8.7	2.6	6.8	11.1	40.7	21.6	6.0	1.0	2.4	6.6	70.1	42.7	2.2	**1.6**
EQUIBIND+S	146	146	8.3	**2.1**	**5.6**	10.5	**46.4**	**24.6**	6.0	0.9	**2.0**	**6.2**	**71.0**	**50.6**	**2.1**	1.8
EQUIBIND	**0.16**	**0.04**	**8.2**	3.8	6.2	**10.3**	39.1	5.5	**5.6**	1.3	2.6	7.4	67.5	40.0	2.6	2.3
EQUIBIND-U	0.14	0.02	7.8	3.3	5.7	9.7	42.4	7.2	5.6	1.3	2.6	7.4	67.5	40.0	2.1	1.8

Blind self-docking. This set of experiments reflects the performance that can be expected in the most typical applications where the true ligand bond angles and distances (which are used in re-docking) are unknown. An initial approximate conformer has to be obtained from a 2D molecular graph for which we use a random RDKit conformer.

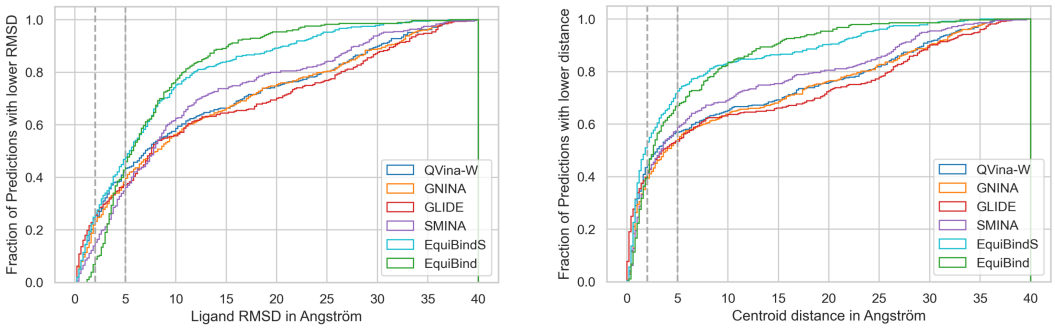

Figure 3: **Flexible blind self-docking.** Cumulative density histogram of the L-RMSD (top) and centroid distance (bottom) of EQUIBIND with and without SMINA for fine-tuning.

The results in Table 1 show that vanilla EQUIBIND performs well at identifying the approximate binding location and outperforms the baselines in metrics other than the 25th RMSD percentile and the fraction of predictions with and RMSD better than 2 Å. The fine-tuning extensions of EQUIBIND such as EQUIBIND + Q outperform or match the baselines in all metrics, while EQUIBIND + Q and EQUIBIND + Q2 also retain significant inference speed-ups, making our method *suitable for extremely high-throughput applications such as virtual screening over databases of hundred million molecules, e.g., ZINC*. Thus, practitioners can combine our method with previous fine-tuning baselines and trade quality over runtime depending on the downstream task of interest.

Figure 3 shows the same trend for the **RMSD**. EQUIBIND, which is three orders of magnitude faster than the fastest baseline, improves over the baselines for the predictions in the > 4 Å regime. EQUIBIND does better for complexes that are hard to predict (e.g., due to ligand size) and also outperforms the baselines in the low RMSD regime when using fine-tuning (EQUIBINDS). For the **centroid distances**, the exact conformer is less crucial, and the methods mainly have to find the correct binding pocket location. Here, EQUIBIND is already able to match the baselines in the low error regime without fine-tuning. Histograms for EQUIBIND + Q and EQUIBIND + Q2 are in Appendix Figure 5.

The **main observations** are that EQUIBIND is much faster than the baselines, has fewer predictions that are far off from the true conformer, and can use fast fine-tuning for very low-RMSD final predictions. The benefits through fine-tuning can be expected considering the difficulty of predicting the correct torsions jointly with the binding location and orientation in a single forward pass.

Table 2: **Blind re-docking.** The input is the bound ligand structure at a random location. The methods are tasked to find the binding site and the right binding location + orientation. EQUIBIND-R's and GLIDE's Kabsch RMSD is 0 since they treat the ligand as a rigid body while the other methods can only be run in a flexible mode where torsion angles are changed. *GLIDE runtime details in Appendix C.

			LIGAND RMSD ↓						CENTROID DISTANCE ↓						KABSCH	
	AVG. SEC.	AVG. SEC.		PERCENTILES ↓			% BELOW THRESHOLD ↑			PERCENTILES ↓			% BELOW THRESHOLD ↑		RMSD ↓	
METHODS	CPU	GPU	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	MEDIAN
QVINA-W	49	-	13.4	1.6	7.9	24.1	39.0	27.7	11.8	0.9	3.8	23.2	55.4	40.4	1.8	1.5
GNINA	247	146	12.2	1.3	6.1	22.9	46.8	32.2	10.9	0.7	2.8	22.1	58.4	43.8	1.7	1.4
SMINA	146	-	10.3	1.4	6.2	15.2	46.7	30.1	8.5	0.8	2.6	12.7	63.5	45.3	1.7	1.4
GLIDE*	1405	-	15.7	**0.5**	8.3	29.5	45.7	**43.4**	14.8	**0.3**	4.9	28.5	50.4	45.4	0	0
EQUIBIND-R	**0.14**	**0.02**	7.4	2.0	5.1	9.8	49.0	25.1	5.8	1.4	2.6	7.3	66.9	40.8	0	0
EQUIBIND-R+S	146	146	**7.0**	1.0	**3.4**	**9.6**	**57.0**	41.1	**5.3**	0.7	**1.4**	**4.7**	**76.0**	**59.2**	1.5	1.1

Blind re-docking. In these experiments, the bound ligand is extracted from the binding pocket, placed in a random location, and the methods have to re-dock it into the correct conformation. Thus the methods have access to the ground truth structure of the ligand, and all predictions will have the correct bond lengths and angles. EQUIBIND-R treats the ligand as a completely rigid body and only predicts a translation and rotation. Rigid re-docking results are of practical relevance for docking strategies where large amounts of conformations are generated for a single molecule and then rigidly docked to the receptor before using an additional scoring function to rank the predictions.

In table 2 we can observe that EQUIBIND-R can be particularly impactful for this strategy due to its much faster inference time. This is while outperforming the baselines in the metrics other than the 25th percentiles and the fraction of predictions with an error below 2 Å. For practical rigid re-docking applications, this could potentially be remedied by docking 10 times as many conformers while still retaining a 10 times speed-up over the fastest baseline.

Sensitivity to initial conformer. EQUIBIND's predictions depend on the initial conformer's torsion angles, bond angles, and bond lengths (the baselines only depend on initial bond angles and lengths). In fig. 4 we investigate the risk of an "unlucky" initial conformer leading to a high L-RMSD. For 363 complexes, we generate 30 different initial RDKit conformers. EQUIBIND predicts a binding structure using each of them, and we obtain 30 L-RMSD values of which we calculate the standard

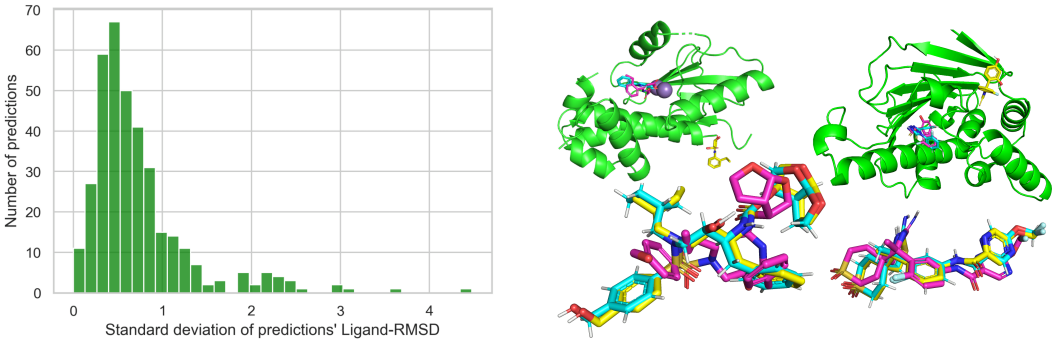

Figure 4: **Left:** Histogram of the standard deviations of the L-RMSDs of EQUIBIND's predictions when using different initial conformers. **Right:** Two cherry-picked example predictions where EQUIBIND has better RMSD than GNINA (top) and two where EQUIBIND performs worse than GNINA (bottom). The ground truth ligand is in cyan, EQUIBIND in pink, and GNINA in yellow.

deviation. We find a low sensitivity to the initial conformer, with the majority of predictions having a smaller standard deviation than 0.5

Visualizations. EQUIBIND's predictions are rarely far off from the true ligand, but there are cases where it struggles to find the exact torsion angles and, therefore, the right atom configurations in the ligand. Examples of this are in fig. 4 and show two cases where GNINA performs worse and produces a prediction that is far off while EQUIBIND is able to find the binding location. The other two cases, where GNINA is better, display how the baseline more exactly finds the true structure, but EQUIBIND still finds the correct approximate location. Further visualizations of predictions are in Appendix Figure 11.

5 CONCLUSION

The promising ability of deep neural networks to predict protein structures has sparked a large amount of research in computational drug discovery. Here, we proposed EQUIBIND, a deep neural model which relies on SE(3)-equivariant graph neural networks to predict bound protein-ligand conformations in a single shot. Our model shows strong empirical performance against state-of-the-art baselines, and we demonstrate its potential in a hybrid workflow by combining it with existing fine-tuning methods. We expect that EQUIBIND and similar models will progress the adoption of deep learning in drug discovery.

REFERENCES

R Aggarwal, A Gupta, V Chelur, C Jawahar, and UD Priyakumar. Deeppocket: Ligand binding site detection and segmentation using 3d convolutional neural networks, 2021.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

Jingxiao Bao, Xiao He, and John ZH Zhang. Deepbsp—a machine learning method for accurate prediction of protein–ligand docking structures. *Journal of Chemical Information and Modeling*, 2021.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. Se (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *arXiv preprint arXiv:2101.03164*, 2021.

H. Berman, K. Henrick, and H. Nakamura. Announcing the worldwide Protein Data Bank. *Nat Struct Biol*, 10(12):980, Dec 2003.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik Bekkers, and Max Welling. Geometric and physical quantities improve e (3) equivariant message passing. *arXiv preprint arXiv:2110.02905*, 2021.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Bowen Dai and Chris Bailey-Kellogg. Protein interaction interface region prediction by geometric deep learning. *Bioinformatics*, 2021.

Renato Ferreira de Freitas and Matthieu Schapira. A systematic analysis of atomic protein–ligand interactions in the pdb. *MedChemComm*, 8(10):1970–1981, 2017.

Pavol Drotár, Arian Rokkum Jamasb, Ben Day, Cătălina Cangea, and Pietro Liò. Structure-aware generation of drug-like molecules. *arXiv preprint arXiv:2111.04107*, 2021.

Xing Du, Yi Li, Yuan-Ling Xia, Shi-Meng Ai, Jing Liang, Peng Sang, Xing-Lai Ji, and Shu-Qun Liu. Insights into protein–ligand interactions: mechanisms, models, and methods. *International journal of molecular sciences*, 17(2):144, 2016.

Stephan Eismann, Raphael JL Townshend, Nathaniel Thomas, Milind Jagota, Bowen Jing, and Ron O Dror. Hierarchical, rotation-equivariant neural networks to select structural models of protein complexes. *Proteins: Structure, Function, and Bioinformatics*, 2020.

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew W Senior, Timothy Green, Augustin Žídek, Russell Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with alphafold-multimer. *bioRxiv*, 2021. doi: 10.1101/2021.10.04.463034.

Paul G Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B Iovanisci, Ian Snyder, and David R Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling*, 60(9):4200–4215, 2020.

Fabian B Fuchs, Daniel E Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. *arXiv preprint arXiv:2006.10503*, 2020.

Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, D Boscaini, MM Bronstein, and BE Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020.

Octavian-Eugen Ganea, Xinyuan Huang, Charlotte Bunne, Yatao Bian, Regina Barzilay, Tommi Jaakkola, and Andreas Krause. Independent se(3)-equivariant models for end-to-end rigid protein docking. *arXiv preprint arXiv:2111.07786*, 2021a.

Octavian-Eugen Ganea, Lagnajit Pattanaik, Connor W Coley, Regina Barzilay, Klavs F Jensen, William H Green, and Tommi S Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *arXiv preprint arXiv:2106.07802*, 2021b.

Kyle Yingkai Gao, Achille Fokoue, Heng Luo, Arun Iyengar, Sanjoy Dey, and Ping Zhang. Interpretable drug target prediction using deep neural representation. In *IJCAI*, volume 2018, pp. 3371–3377, 2018.

Vladimir Gligorijević, P Douglas Renfrew, Tomasz Kosciolek, Julia Koehler Leman, Daniel Berenberg, Tommi Vatanen, Chris Chandler, Bryn C Taylor, Ian M Fisk, Hera Vlamakis, et al. Structure-based protein function prediction using graph convolutional networks. *Nature communications*, 12(1):1–14, 2021.

Pawel Gniewek, Bradley Worley, Kate Stafford, Henry van den Bedem, and Brandon Anderson. Learning physics confers pose-sensitivity in structure-based virtual screening. *arXiv preprint arXiv:2110.15459*, 2021.

Thomas A Halgren, Robert B Murphy, Richard A Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–1759, 2004.

Nafisa M Hassan, Amr A Alhossary, Yuguang Mu, and Chee-Keong Kwoh. Protein-ligand blind docking using quickvina-w with inter-process spatio-temporal integration. *Scientific reports*, 7 (1):1–13, 2017.

Sheng-You Huang. Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges. *Briefings in bioinformatics*, 19(5):982–994, 2018.

Fergus Imrie, Thomas E Hadfield, Anthony R Bradley, and Charlotte M Deane. Deep generative design with 3d pharmacophoric constraints. *bioRxiv*, 2021.

Stanisław Jastrzebski, Maciej Szymczak, Agnieszka Pocha, Stefan Mordalski, Jacek Tabor, Andrzej J Bojarski, and Sabina Podlewska. Emulating docking results using a deep neural network: a new perspective for virtual screening. *Journal of Chemical Information and Modeling*, 60(9): 4246–4262, 2020.

Dejun Jiang, Chang-Yu Hsieh, Zhenxing Wu, Yu Kang, Jike Wang, Ercheng Wang, Ben Liao, Chao Shen, Lei Xu, Jian Wu, et al. Interactiongraphnet: A novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. *Journal of medicinal chemistry*, 2021.

José Jiménez, Stefan Doerr, Gerard Martínez-Rosell, Alexander S Rose, and Gianni De Fabritiis. Deepsite: protein-binding site predictor using 3d-convolutional neural networks. *Bioinformatics*, 33(19):3036–3042, 2017.

Akhil Jindal, Sergei Kotelnikov, Dzmitry Padhorny, Dima Kozakov, Yimin Zhu, Rezaul Chowdhury, and Sandor Vajda. Side-chain packing using se (3)-transformer. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*, pp. 46–55. World Scientific, 2021.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2020.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873), Aug 2021a.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021b.

Mostafa Karimi, Di Wu, Zhangyang Wang, and Yang Shen. Deepaffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, 35(18):3329–3338, 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

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, 2013.

Greg Landrum. Rdkit: Open-source cheminformatics software, 2016.

P Therese Lang, Scott R Brozell, Sudipto Mukherjee, Eric F Pettersen, Elaine C Meng, Veena Thomas, Robert C Rizzo, David A Case, Thomas L James, and Irwin D Kuntz. Dock 6: Combining techniques to model rna–small molecule complexes. *Rna*, 15(6):1219–1230, 2009.

Shuangli Li, Jingbo Zhou, Tong Xu, Liang Huang, Fan Wang, Haoyi Xiong, Weili Huang, Dejing Dou, and Hui Xiong. Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 975–985, 2021.

Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. Graph matching networks for learning the similarity of graph structured objects. In *International Conference on Machine Learning*, pp. 3835–3845. PMLR, 2019.

Jaechang Lim, Seongok Ryu, Kyubyong Park, Yo Joong Choe, Jiyeon Ham, and Woo Youn Kim. Predicting drug–target interaction using a novel graph neural network with 3d structure-embedded graph representation. *Journal of chemical information and modeling*, 59(9):3981–3988, 2019.

Zhihai Liu, Minyi Su, Li Han, Jie Liu, Qifan Yang, Yan Li, and Renxiao Wang. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of Chemical Research*, 50 (2):302–309, 2017.

Shitong Luo, Chence Shi, Minkai Xu, and Jian Tang. Predicting molecular conformation via dynamic graph score matching. *Advances in Neural Information Processing Systems*, 34, 2021.

Gabriele Macari, Daniele Toti, and Fabio Polticelli. Computational methods and tools for binding site recognition between proteins and small molecules: from classical geometrical approaches to modern machine learning strategies. *Journal of computer-aided molecular design*, 33(10): 887–903, 2019.

Tomohide Masuda, Matthew Ragoza, and David Ryan Koes. Generating 3d molecular structures conditional on a receptor binding site with deep generative models. *arXiv preprint arXiv:2010.14442*, 2020.

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):1–20, 2021.

Oscar Méndez-Lucio, Mazen Ahmad, Ehecatl Antonio del Rio-Chanona, and Jörg Kurt Wegner. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence*, 3(12):1033–1039, 2021.

Taj Mohammad, Yash Mathur, and Md Imtaiyaz Hassan. Instadock: A single-click graphical user interface for molecular docking-based virtual high-throughput screening. *Briefings in Bioinformatics*, 22(4):bbaa279, 2021.

Joseph A Morrone, Jeffrey K Weber, Tien Huynh, Heng Luo, and Wendy D Cornell. Combining docking pose rank and structure with deep learning improves protein–ligand binding mode prediction over a baseline docking approach. *Journal of chemical information and modeling*, 60(9): 4170–4179, 2020.

Open Babel development team. Open babel, 2005. URL `http://openbabel.org/wiki/Main_Page`.

Matthew Ragoza, Joshua Hochuli, Elisa Idrobo, Jocelyn Sunseri, and David Ryan Koes. Protein–ligand scoring with convolutional neural networks. *Journal of chemical information and modeling*, 57(4):942–957, 2017.

Jean-Louis Reymond and Mahendra Awale. Exploring chemical space for drug discovery using the chemical universe database. *ACS chemical neuroscience*, 3(9):649–657, 2012.

Sereina Riniker and Gregory A Landrum. Better informed distance geometry: using what we know to improve conformation generation. *Journal of chemical information and modeling*, 55(12): 2562–2574, 2015.

Sriram Sankararaman, Fei Sha, Jack F Kirsch, Michael I Jordan, and Kimmen Sjölander. Active site prediction using evolutionary and structural information. *Bioinformatics*, 26(5):617–624, 2010.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n)-equivariant graph neural networks. *arXiv preprint arXiv:2102.09844*, 2021.

Chao Shen, Xueping Hu, Junbo Gao, Xujun Zhang, Haiyang Zhong, Zhe Wang, Lei Xu, Yu Kang, Dongsheng Cao, and Tingjun Hou. The impact of cross-docked poses on performance of machine learning classifier for protein–ligand binding pose prediction. *Journal of cheminformatics*, 13(1): 1–18, 2021.

Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. *arXiv preprint arXiv:2105.03902*, 2021.

Vignesh Ram Somnath, Charlotte Bunne, and Andreas Krause. Multi-scale representation learning on proteins. *Advances in Neural Information Processing Systems*, 34, 2021.

Freyr Sverrisson, Jean Feydy, Bruno E Correia, and Michael M Bronstein. Fast end-to-end learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15272–15281, 2021.

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. *arXiv preprint arXiv:1802.08219*, 2018.

Wen Torng and Russ B Altman. Graph convolutional neural networks for predicting drug-target interactions. *Journal of chemical information and modeling*, 59(10):4131–4149, 2019a.

Wen Torng and Russ B Altman. High precision protein functional site detection using 3d convolutional neural networks. *Bioinformatics*, 35(9):1503–1512, 2019b.

Raphael Townshend, Rishi Bedi, Patricia Suriana, and Ron Dror. End-to-end learning on 3d protein structure for interface prediction. *Advances in Neural Information Processing Systems*, 32:15642–15651, 2019.

Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

Izhar Wallach, Michael Dzamba, and Abraham Heifets. Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*, 2015.

Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, and Jian Tang. Learning neural generative dynamics for molecular conformation generation. In *International Conference on Learning Representations*, 2020.

Wenyi Zhang, Eric W Bell, Minghao Yin, and Yang Zhang. Edock: blind protein–ligand docking by replica-exchange monte carlo simulation. *Journal of cheminformatics*, 12:1–17, 2020.

Zsolt Zsoldos, Darryl Reid, Aniko Simon, Sayyed Bashir Sadjad, and A Peter Johnson. ehits: a new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics and Modelling*, 26(1):198–212, 2007.

A ADDITIONAL RESULTS

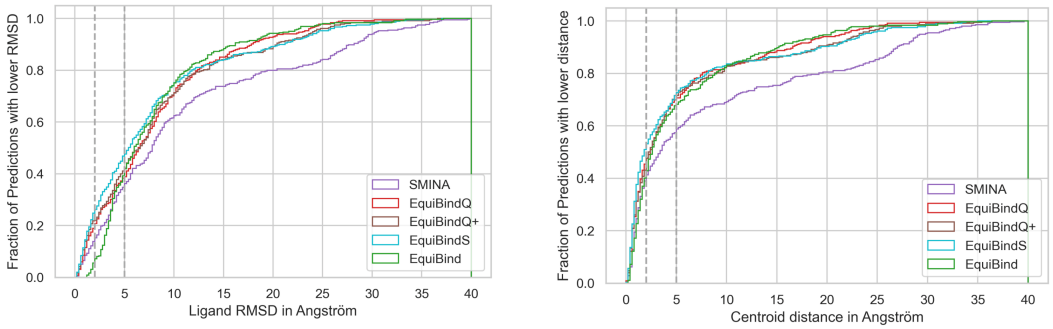

Figure 5: **Flexible blind self-docking.** Left: Cumulative histogram of the L-RMSD. Right: Cumulative histogram of the centroid distance.

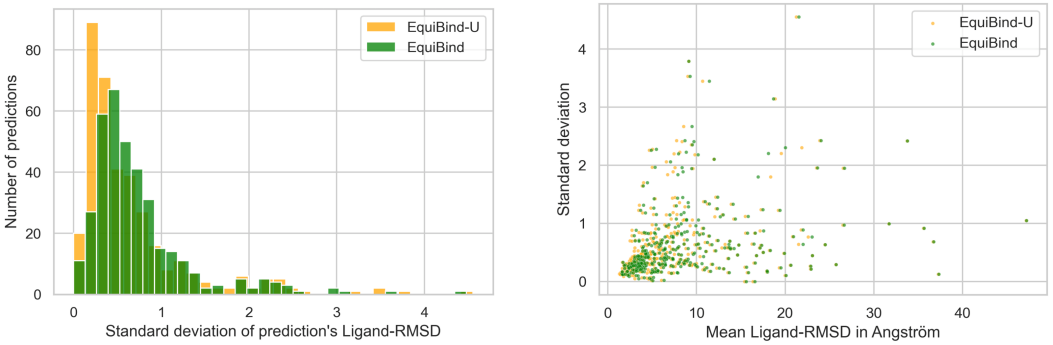

Figure 6: **Sensitivity to initial conformer.** Left: Histogram of the standard deviations of the L-RMSDs of EQUIBIND's and EQUIBIND-U's predictions when using 30 different initial conformers. Right: Scatter-plot with a point for each complex showing the mean L-RMSD and standard deviation of EQUIBIND's or EQUIBIND-U's 30 predictions from 30 different initial input conformers.

Table 3: **Flexible blind self-docking.** All methods receive a random RDKit conformer of the ligand as input and are tasked to find the binding site and the correct binding structure. Comparison of EQUIBIND-SA which additionally uses surface atoms of the receptor and EQUIBIND-A, which makes a prediction using EQUIBIND-R and refines it using an all-atom subgraph in a 10 Å radius around the predicted ligand. The methods have to be compared with EQUIBIND-U as corrections would still need to be applied.

			LIGAND RMSD ↓						CENTROID DISTANCE ↓						KABSCH	
	AVG. SEC.	AVG. SEC.		PERCENTILES ↓			% BELOW THRESHOLD ↑			PERCENTILES ↓			% BELOW THRESHOLD ↑		RMSD ↓	
METHODS	CPU	GPU	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	MEDIAN
EQUIBIND-SA	0.14	0.02	8.6	3.6	6.1	11.4	41.3	3.8	6.1	1.4	2.7	7.3	66.6	40.5	2.4	2.0
EQUIBIND-A	5.56	5.44	8.2	3.8	6.0	10.2	43.0	3.3	6.1	1.7	3.3	7.0	63.3	30.3	3.0	2.7
EQUIBIND-U	0.14	0.02	7.8	3.3	5.7	9.7	42.4	7.2	5.6	1.3	2.6	7.4	67.5	40.0	2.1	1.8

In table 3 we compare EQUIBIND which only implicitly uses atom level locations with approaches that explicitly use atoms as nodes in the processed graph. EQUIBIND-SA uses the surface atoms of the receptor, which are found using the MSMS tool. This additional step does not impact inference runtime, which we measure without preprocessing as explained in Section 4. EQUIBIND-A makes a prediction using EQUIBIND-R and uses an all-atom subgraph in a 10 Å radius around the predicted ligand to further refine the prediction. This additional step significantly impacts inference runtime. Both methods require around 2-3 times more GPU RAM than EQUIBIND.

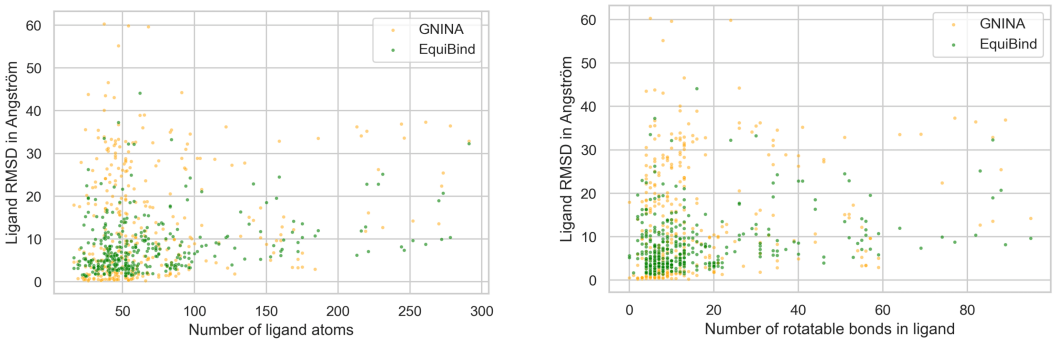

Figure 7: **Performance correlation with molecule size.** Left: Scatter-plot showing the L-RMSD and the number of ligand atoms for each prediction of EQUIBIND and GNINA. Right: The same for the number of rotatable bonds in the ligand.

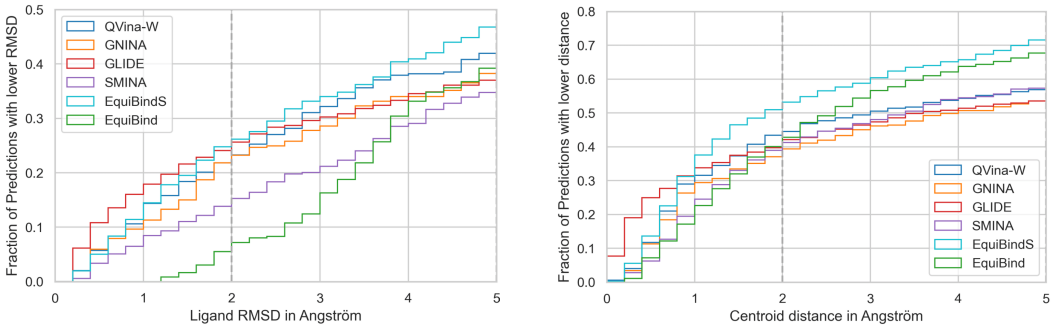

Figure 8: **Flexible blind self-docking.** Zoomed in versions of the histograms in Figure 3 where only the 0-5 Å range is shown. Left: Cumulative histogram of the L-RMSD. Right: Cumulative histogram of the centroid distance.

In fig. 8 we can again observe that EQUIBIND struggles to produce many predictions in the low L-RMSD range. When adding an additional fine-tuning step such as with EQUIBINDS, the model is able to match or outperform the baselines in all L-RMSD ranges. Thus predictions speed can be traded off for additional accuracy via fine-tuning EQUIBIND's predictions with classical physics-based methods. Above the 3.8 Å cutoff, the vanilla EQUIBIND starts outperforming the baselines even without fine-tuning.

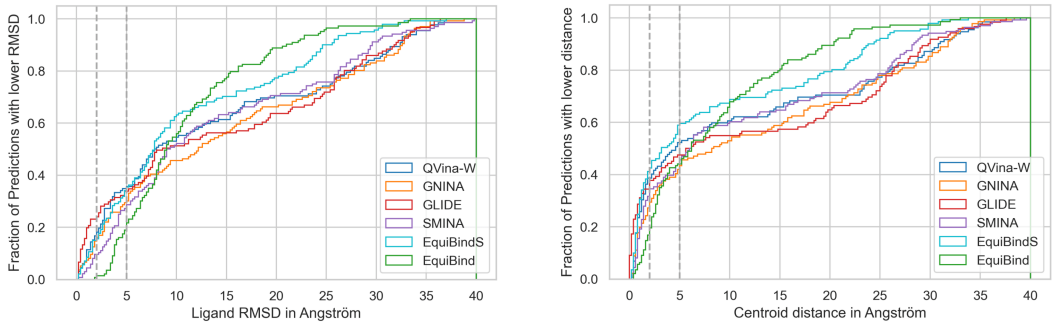

Figure 9: **Flexible blind self-docking new receptors.** Results when removing all complexes from the time split based test set whose receptor was present in a complex that is older than 2019. Cumulative density histogram of the L-RMSD (top) and centroid distance (bottom) of EQUIBIND with and without SMINA for fine-tuning.

Fast point cloud fitting. In fig. 12 we visualize our novel fast point cloud ligand fitting described in section 3.2. The point clouds produced by the uncorrected flexible EQUIBIND-U are not realistic

Table 4: **Flexible blind self-docking new receptors.** Results when removing all complexes from the time split based test set whose receptor was present in a complex that is older than 2019. 144 complexes remain. The run times are still averaged over all test complexes. All methods receive a random RDKit conformer of the ligand molecule as input and are tasked to find its binding site and the right orientation + conformation in which it binds. EQUIBIND-U refers to the model producing uncorrected atomic point clouds **Z** that are not necessarily chemically plausible ligands. EQUIBIND performs our fast conformer fitting corrections – see section 3.2. EQUIBIND+Q predicts an approximate ligand position and fine-tunes it using QuickVina 2. EQUIBIND+Q2 samples more candidate positions, and EQUIBINDS uses SMINA for fine-tuning. *GLIDE runtime details in Appendix C.

			LIGAND RMSD ↓						CENTROID DISTANCE ↓						KABSCH	
	AVG. SEC.	AVG. SEC.		PERCENTILES ↓			% BELOW THRESHOLD ↑			PERCENTILES ↓			% BELOW THRESHOLD ↑		RMSD ↓	
METHODS	CPU	GPU	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	MEDIAN
QVINA-W	49	-	16.9	3.4	10.3	28.1	31.9	15.3	15.2	1.3	6.5	26.8	47.9	35.4	**2.2**	1.9
GNINA	247	146	16.7	4.5	13.4	27.8	27.8	13.9	15.1	2.0	10.1	27.0	39.5	25.7	2.3	1.8
SMINA	146	-	15.7	4.8	10.9	26.0	25.7	9.0	13.6	1.6	6.5	25.7	41.7	29.9	2.3	1.9
GLIDE*	1405	-	19.6	**3.4**	18.0	31.4	28.7	**19.6**	18.1	1.1	17.6	29.1	40.6	29.4	2.3	**1.7**
EQUIBIND+Q	8	8	11.5	5.5	8.7	15.7	22.9	9.3	9.1	1.6	5.5	14.0	47.9	30.7	2.3	1.9
EQUIBIND+Q2	15	15	12.0	4.1	8.0	**19.7**	28.0	13.3	9.8	1.3	4.4	18.5	53.9	30.8	2.3	**1.7**
EQUIBIND+S	146	146	11.9	3.6	**7.9**	**19.7**	**33.3**	14.6	9.7	**1.0**	**4.0**	18.2	**57.6**	**40.3**	**2.2**	1.8
EQUIBIND	**0.16**	**0.02**	**11.3**	5.9	9.1	**14.3**	18.8	0.7	**8.9**	2.6	6.3	**12.9**	43.8	16.7	2.7	2.2
EQUIBIND-U	0.14	0.02	11.0	5.7	8.8	14.1	21.5	1.4	8.9	2.6	6.3	12.9	43.8	16.7	2.2	1.8

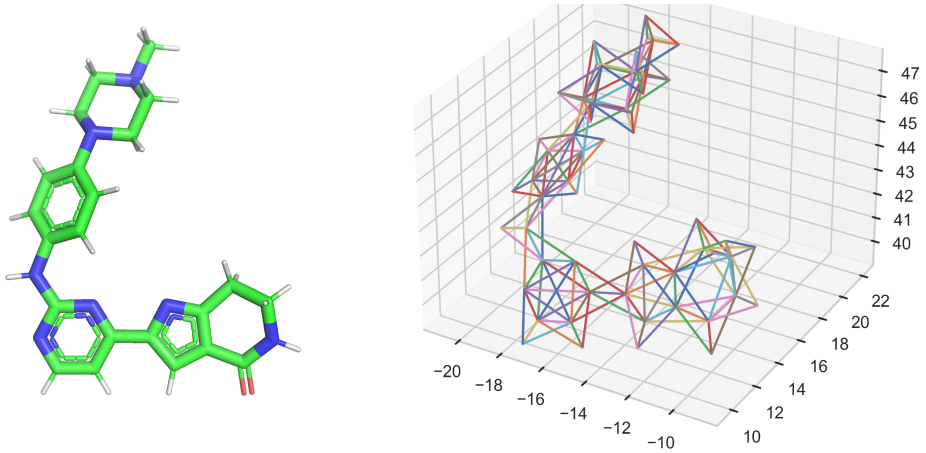

Figure 10: **LAS distance geometry constraint visualization.** Visualization of the interatomic distances which are included in the LAS distance geometry constraints in Equation 3.2. The pairwise distances in rings are only included if the ring is aromatic, like in the bottom left ring of the depicted molecule. The torsion angles in non-aromatic rings, such as the one in the bottom right, remain flexible.

molecules. The corrections use a conformer with valid bond lengths and angles and change its torsions to most closely match the point cloud.

Limitations. One drawback of EQUIBIND is that it only implicitly models the atom positions of side chains. This is done via the local frame encoding features of Jumper et al. (2021a) that we employ in the α-carbon graph of the receptor. Explicitly representing these atoms might improve precise docking. We experimented with surface atoms and fine-tuning approaches that use an atom subgraph of the receptor with results in Appendix A. However, this yielded only small or no improvements while adding considerable computational complexity. We leave further exploration of this strategy for future work. These results are in line with prior protein modeling techniques such

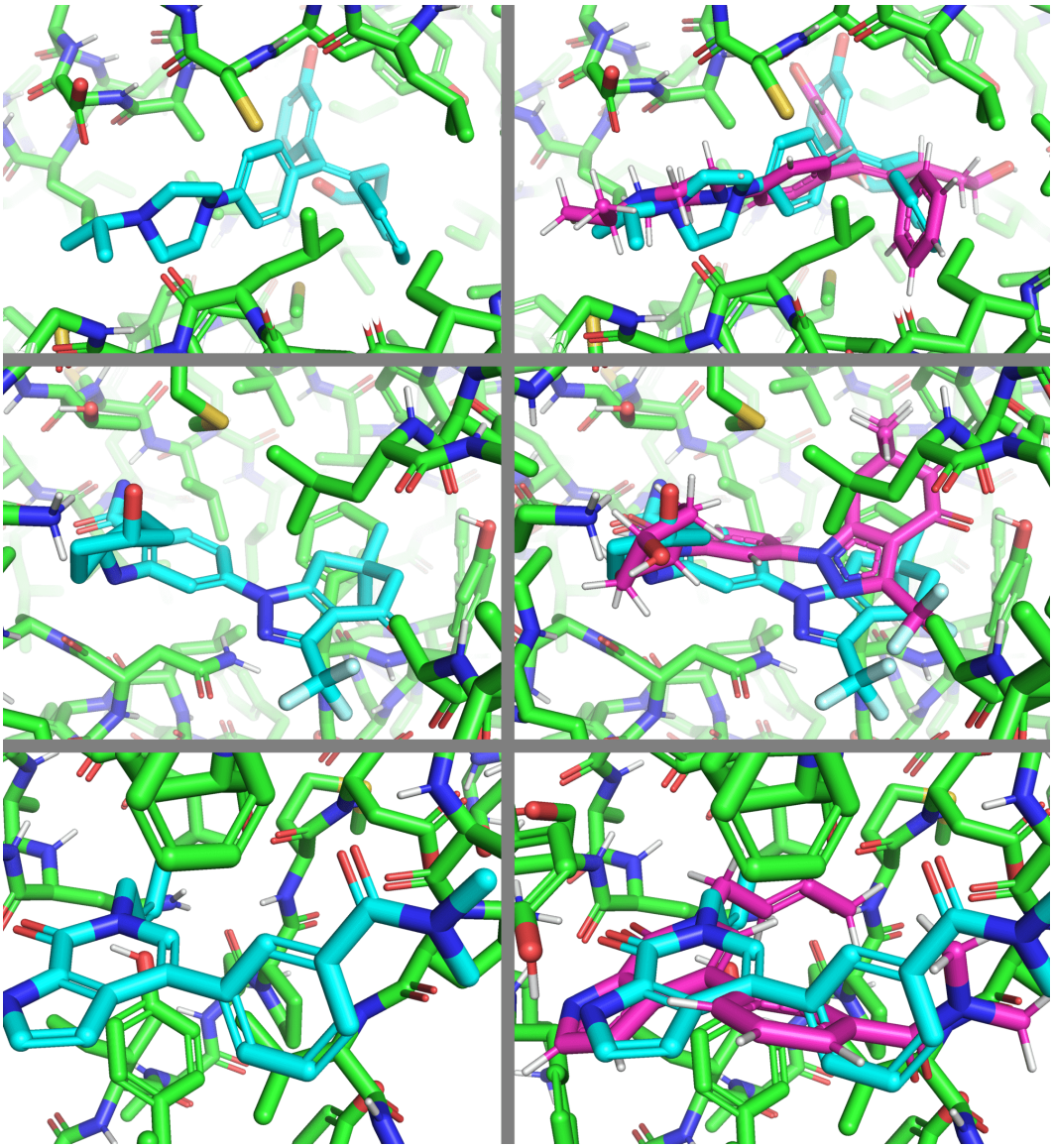

Figure 11: Three cherry-picked example predictions of EQUIBIND (cyan) on the left with the same image on the right, including the true bound conformer in pink.

as AlphaFold2, Jumper et al. (2021a) which successfully predicts side chains based on only residue level information.

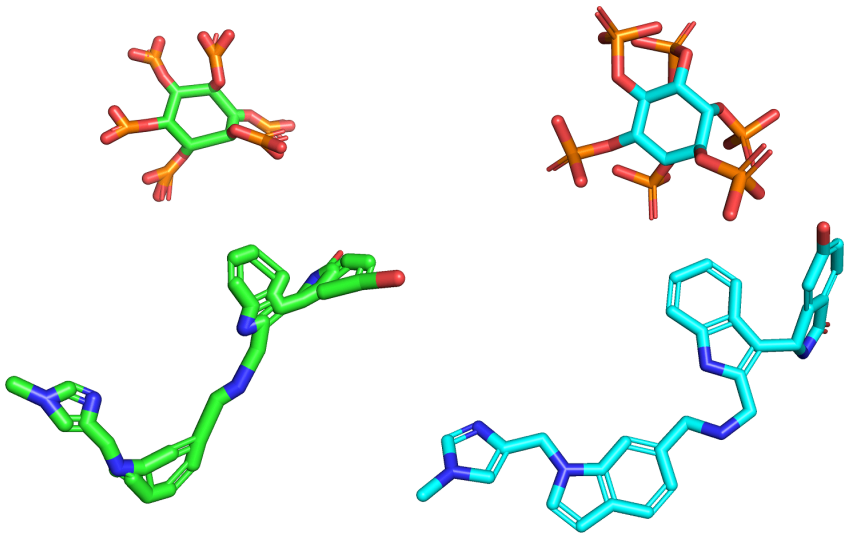

Figure 12: Left: in green are two uncorrected pseudo-molecules predicted by EQUIBIND-U. Right: the final output of EQUIBIND with corrections using our fast flexible conformer fitting applied to produce a conformer with realistic bond angles and lengths.

B DATASET

Motivation for new test set and time split. Docking methods are often evaluated using the PDBBind core set, which contains 285 hand-curated high-resolution complexes. However, this might not reflect the performance in real-world applications where data might not be of similar high quality. Due to the differences in resolution and the average ligand size (32 heavy atoms in PDBBind versus 24 in the core set), the complexes of the core set can be considered easier to predict than the average complex. Moreover, some of the previous methods might have been validated or trained on a subset of the core set and thus, report optimistic quality numbers. To better reflect the average complex encountered in applications, we employ a test set that only contains complexes that were discovered in 2019 or later, while the train and validation sets only use strictly older complexes.

Dataset split. Of the 19 119 preprocessed complexes, 1512 were discovered in 2019 or later. From these, we randomly sample 125 unique proteins and collect all new complexes containing them (363) to create the final test set. The low number of test samples is chosen to make it feasible to compare with time-consuming classical physics-based docking methods. From the remaining complexes that are older than 2019, we remove those with ligands contained in the test set, giving 17 347 complexes for training and validation. These are divided into 968 validation complexes, which share no ligands with the remaining 16 379 train complexes. Results when only testing on new receptors are in Appendix A.

Preprocessing. The time split is done after preprocessing the 19 443 complexes of PDBBind v2020 as follows. First, we drop all complexes that cannot be processed by the RDKit library (Landrum, 2016), leaving 19 119 complexes. We process each ligand and receptor with OpenBabel Open Babel development team (2005) and add all missing hydrogens to the ligands using RDKit. Next we correct all receptor hydrogens and add missing ones using `reduce`[2].

A significant remaining data issue is symmetric receptor structures comprised of the same protein repeated multiple times. In these cases, the ligand could equally likely bind to the pocket of each of the proteins, i.e., multiple ligand correct positions are possible. However, the ground truth ligand is only placed in one of those locations. Examples are in fig. 13. We address the majority of these cases by only keeping the connected components of the receptor, which have an atom within a 10 Å radius of any ligand atom.

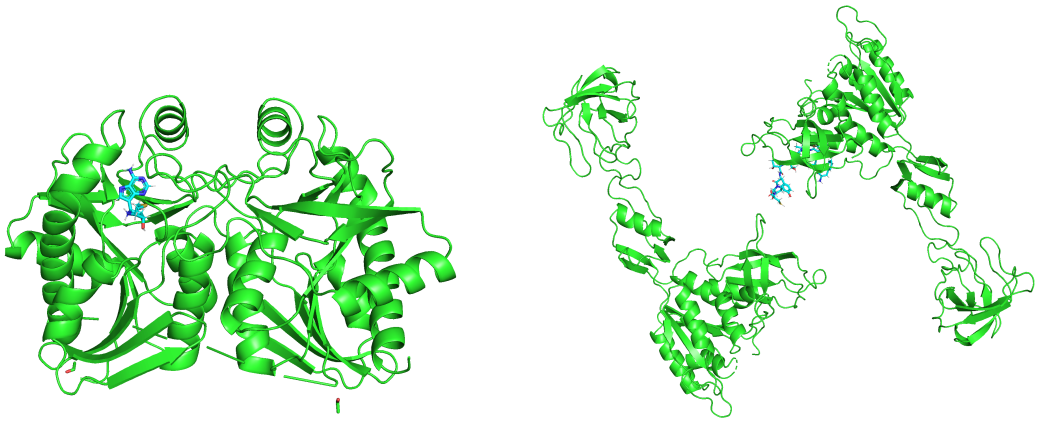

Figure 13: Examples of symmetric receptor complexes with multiple equally valid binding positions for the ligand.

C IMPLEMENTATION DETAILS

IEGMN Layer. Our choice for a single l-th layer is:

$$\mathbf{m}_{j\to i} = \varphi^e(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\|^2, \mathbf{f}_{j\to i}), \forall (i,j) \in \mathcal{E} \cup \mathcal{E}'$$

[2] `https://github.com/rlabduke/reduce`

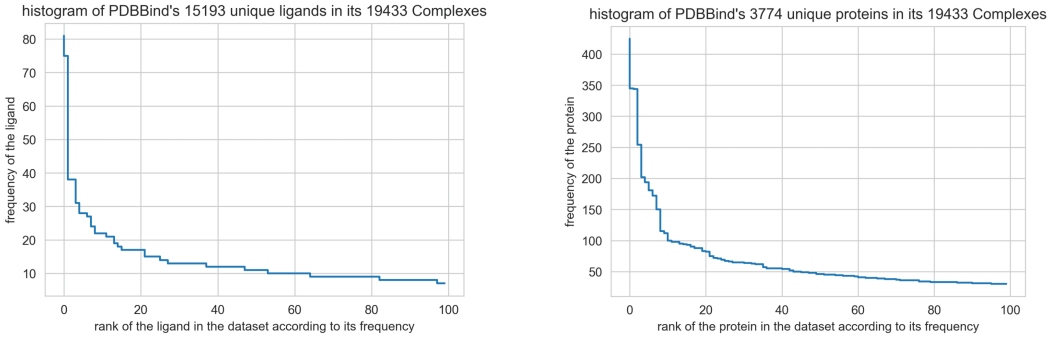

Figure 14: Histograms that show how often each unique ligand and receptor appears in the PDBBind dataset.

$$\mu_{j'\to i} = a_{j'\to i}\mathbf{W}\mathbf{h}_{j'}^{(l)}, \forall i \in \mathcal{V}, j' \in \mathcal{V}' \text{ or } i \in \mathcal{V}', j' \in \mathcal{V}$$

$$\mathbf{m}_i = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{m}_{j\to i}, \forall i \in \mathcal{V}\cup\mathcal{V}'$$

$$\mu_i = \sum_{j'\in\mathcal{V}'} \mu_{j'\to i}, \forall i \in \mathcal{V}, \quad \text{and} \quad \mu_i' = \sum_{j\in\mathcal{V}} \mu_{j\to i'}, \forall i \in \mathcal{V}'$$

$$\mathbf{x}_i^{(l+1)} = \Psi\left(\mathbf{x}_i^{(l)} + \sum_{j\in\mathcal{N}(i)} \frac{\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}}{\|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\|}\varphi^x(\mathbf{m}_{j\to i})\right)$$

$$\mathbf{h}_i^{(l+1)} = (1-\beta)\cdot\mathbf{h}_i^{(l)} + \beta\cdot\varphi^h(\mathbf{h}_i^{(l)}, \mathbf{m}_i, \mu_i, \mathbf{f}_i), \forall i \in \mathcal{V}\cup\mathcal{V}'$$

where $a_{j\to i}$ are SE(3)-invariant attention coefficients derived from $\mathbf{H}$ embeddings, $\mathcal{N}(i)$ are the graph neighbors of node i, $\mathbf{W}$ is a parameter matrix, and the various φ functions are modeled using shallow neural networks, with φ^x outputting a scalar and others a d-dimensional vector.

When modeling ligand flexibility, we found it useful to incorporate additional geometric constraints on transformed coordinates through Ψ models described in section 3.2.

Implementation Details. We optimize our model using Adam (Kingma & Ba, 2014) and do early stopping with patience of 150 epochs based on the percentage of predicted validation set complexes with an RMSD better than 2 Å. All hyperparameters and the employed ligand and node features are described in Appendix C. Code to reproduce results or perform fast docking with the provided model weights is available at `https://github.com/HannesStark/EquiBind`.

Featurization. For the α-carbons in the receptor graph, we use the residue type as a feature. The edges have two attributes. Firstly, the interatomic distances encoded with Gaussian basis functions with 15 different variances—secondly, the local frame orientation encodings as they are employed by Jumper et al. (2021a) and Ganea et al. (2021a).

In the ligand, the edges have features that are encoded in the same fashion as for the receptor. Meanwhile, the atoms have the following features: atomic number; chirality; degree; formal charge; implicit valence; the number of connected hydrogens; the number of radical electrons; hybridization type; whether or not it is in an aromatic ring; in how many rings it is; and finally, 6 features for whether or not it is in a ring of size 3, 4, 5, 6, 7, or 8.

GLIDE Runtime. While baselines like GNINA, SMINA, and QVina-W can leverage multiple CPU cores while making the predictions for a single complex, this is not the case for GLIDE, which only uses a single thread when processing a complex (this and the following information is described here `https://www.schrodinger.com/kb/1165`). GLIDE has an application that supports starting multiple processes with each processing a different complex in parallel and for distributing these processes across CPU cores. However, each process also requires a separate software license.

Further Hyperparameters. We use a learning rate of 10^{-4} for EQUIBIND and 3×10^{-4} for EQUIBIND-R. The learning rate is reduced by a factor of 0.6 after 60 epochs of no improvement in

our main validation criterion, which is the percentage of predicted validation set complexes with an RMSD better than 2 Å. The models with the best validation score are then tested on the time-based test set.

Table 5: Search space for all EQUIBIND models through which we searched to obtain a strong performance on the validation set. The final parameters for the standard EQUIBIND model are marked in **bold**.

PARAMETER	SEARCH SPACE
LAS DG STEP SIZE η	1, 0.01, **0.001**, 0.0001
LAS DG NUMBER OF STEPS T	**1**, 5, 10
OPTIMAL TRANSPORT LOSS WEIGHT	0, 0.1, 0.5, **1**, 2, 10
INTERSECTION LOSS WEIGHT	0, 0.1, 1, **3**, 10, 50, 100
PROPAGATION DEPTH	[5, 7, **8**]
INTERSECTION σ	8 (BASED ON LOSS ON VAL-SET)
INTERSECTION γ	8 (BASED ON LOSS ON VAL-SET)
KABSCH RMSD LOSS WEIGHT	0, **1**
HIDDEN DIMENSION	32, **64**, 100
NON LINEARITIES	**LEAKY-RELU**, RELU, SELU
LEARNING RATES	0.0009, 0.0003, **0.0001**, 0.00007
DROPOUT	0, 0.05, **0.1**, 0.2
NUM ATTENTION HEADS	10, 20, **30**, 50, 100
NORMALIZATION	**BATCHNORM**, LAYERNORM, GRAPHNORM