Pre-training of Equivariant Graph Matching Networks with Conformation Flexibility for Drug Binding

Fang Wu[1,2†], Shuting Jin[2,3†], Yinghui Jiang[2†], Xurui Jin[2], Bowen Tang[2], Zhangming Niu[2], Xiangrong Liu[3], Qiang Zhang[4,5], Xiangxiang Zeng[6] and Stan Z. Li[1*]

[1]School of Engineering, Westlake University, Hangzhou, 310024, China.
[2]MindRank AI Ltd., Hangzhou, 310000, China.
[3]School of Informatics, Xiamen University, Xiamen, 361005, China.
[4]ZJU-Hangzhou Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, 311200, China.
[5]College of Computer Science and Technology, Zhejiang University, Hangzhou, 310013, China.
[6]School of Information Science and Engineering, Hunan University, Hunan, 410082, China.

*Corresponding author(s). E-mail(s): stan.zq.li@westlake.edu.cn;
Contributing authors: fw2359@columbia.edu; stjin@stu.xmu.edu.cn; yinghui@mindrank.ai; xurui@mindrank.ai; bowen@mindrank.ai; zhangming@mindrank.ai; xrliu@xmu.edu.cn; qiang.zhang.cs@zju.edu.cn; xzeng@foxmail.com;
†These authors contributed equally to this work.

Abstract

The latest biological findings observe that the motionless "lock-and-key" theory is not generally applicable and that changes in atomic sites and binding pose can provide important information for understanding drug binding. However, the computational expenditure limits the growth of protein trajectory-related studies, thus hindering the possibility of supervised learning. We present a novel spatial-temporal pre-training method based on the modified Equivariant Graph Matching Networks (EGMN), dubbed ProtMD which has two specially designed self-supervised learning tasks: atom-level prompt-based denoising generative task and conformation-level snapshot ordering task to seize the flexibility information inside MD trajectories with very fine temporal resolutions. The ProtMD can grant the encoder network the capacity to capture the time-dependent geometric mobility of conformations along MD trajectories. Two downstream tasks are chosen to verify the effectiveness of ProtMD through linear detection and task-specific fine-tuning. We observe a huge improvement from current state-of-the-art methods, with a decrease of 4.3% in RMSE for the binding affinity problem and an average increase of 13.8% in AUROC and AUPRC for the ligand efficacy problem. The results demonstrate a strong correlation between the magnitude of conformation's motion in the 3D space and the strength with which the ligand binds with its receptor.

Keywords: Deep Learning, Pre-training, Drug Binding, Molecular Dynamics

1 Introductions

The development of a new drug is well known to be very expensive (91; 100). Accurate drug binding prediction is a prerequisite for fast virtual screening (89; 105), which is to understand how drug-like molecules (ligands) interact with the target proteins (receptors). Recently, deep learning (DL)-based methods have emerged to drastically reduce the molecular search space and help accelerate the drug discovery process (9).

The process of a receptor accommodating a small molecule has been shown to be highly dynamic and time-dependent, thus the initial motionless 'lock-and-key' theory of ligand binding (32) has been abandoned. Currently, it is in favor of binding models that account for not only conformational changes, but random dynamic interaction (15; 19; 77; 103; 110). This is because the receptor and ligand flexibility are crucial for correctly predicting drug binding and other related thermodynamic and kinetic properties (1; 33). However, prior DL-based studies concentrate merely on a single, stable and static conformation (99) without considering the time-dependent mobility.

While crystallographic studies have convincingly demonstrated that protein flexibility matters in drug binding, the process demands expensive labor (30). Alternatively, molecular dynamics (MD) simulations (79) seek to approximate atomic motions by Newtonian physics (98) to reduce needed human labor; also MD can be used to incorporate flexibility into docking calculations. Some MD-based techniques allow a thorough sampling of the conformational space for large biomolecules, and can include the complete description of the pathway of the ligand binding to its target protein (25; 39; 95). However, the cost of MD prevents is prohibitively high with the growing size of reported protein trajectory-related data. Consequently, a completely supervised paradigm of training and inference on MD trajectories of protein-ligand pairs is infeasible.

Motivated by the above-mentioned reasons, in this paper, we aim to explore the mechanism behind the binding prediction from both spatial and temporal perspective , and land on self-supervised learning (20; 26; 116) to empower pre-trained models with the ability to learn temporal dependencies. We propose a simple yet effective self-supervised pre-training framework with the full employment of temporal sequences of **Prot**ein structures on **MD** trajectories termed as ProtMD as shown in Figure 1. Specifically, two types of self-supervised learning tasks are constructed, one for atom-level and the other for conformation-level, to better capture the internal and global information of MD trajectories. The former is a prompt-based denoising generative prediction , working at the atom-level. It asks the model to produce future conformations based on the current one. Unlike the naive generative self-supervised learning, a time-series prompt is added to regulate and control the time interval between the source and target conformations. This enables the molecular encoder to capture both short-term and long-term dependencies inside the MD trajectories. Apart from that, extra noise is injected into the input conformation to increase the task difficulty to prevent the model from overfitting. This setting conforms to the principles of enhanced sampling mechanism in MD simulations (2; 80; 112). The latter is through a conformation-level snapshot ordering task, which requires the model to identify the temporal order of a set of consecutive snapshots.

To fully unleash the potential of our proposed self-supervised method, we refine an *E(3)-Equivariant Graph Matching Network* (EGMN) (38) to cope with ligand binding modeling and use it as the backbone of the ProtMD. The EGMN as geometric network can jointly transforms both the features and 3D coordinates to perform message passing on *intra* and *inter* graphs. In the experiments, we train the ProtMD on the MD trajectories of sixty-four protein-ligand pairs with a total of 62.8K snapshots and then is leveraged for two downstream drug binding-related tasks, i.e., the binding affinity prediction and the ligand efficacy prediction. Our model leads to state-of-the-art results. Clear visualization strongly demonstrate that our pre-training approach can significantly and consistently improve the model performance and learn effective protein representations using MD data. It also proves the extraordinary ability of our ProtMD, pre-trained with a limited number of samples, to generalize to a diversity of downstream tasks. More importantly, we investigate

the underlying mechanism behind the success of ProtMD, and further demonstrate a tight correlation between the magnitude of spatial motion of conformation and the extent to which the ligand and the receptor bind with each other. This provides solid evidences that our ProtMD efficiently captures conformations' flexibility during the drug binding process [1]

2 The Framework of ProtMD

This section give a brief overview of our ProtMD, highlighted in Figure 1 and details in Figure 2. It consists of two parts: various spatial-temporal self-supervised learning tasks and an equivariant graph matching network. See the Method Section 6 for more descriptions on ProtMD.

2.1 Spatial-temporal Self-supervised Learning Tasks

To capture the temporal information of MD trajectory and boost the generalization ability of the model, we construct a spatial-temporal conformation sequences to represent MD trajectories of protein-ligand pairs and propose two self-supervised learning tasks to pre-train ProtMD model: the prompt-based denoising conformation generative task (atom-level) and the snapshot ordering task (conformation-level). Notably, these two levels of modeling is designed with different purposes. The atom-level task specializes in capturing the local context of each particle, while the conformation-level task is particularly to capture the long-term context within the complexity.

2.1.1 Spatial-temporal Protein Sequences

We consider MD trajectories of each protein-ligand pair with T timesteps. At each time $t \in [T]$, the ligand graph $\mathcal{G}_L^{(t)} = (\mathcal{V}_L^{(t)}, \mathcal{E}_L^{(t)})$ and the receptor graph $\mathcal{G}_R^{(t)} = (\mathcal{V}_R^{(t)}, \mathcal{E}_R^{(t)})$ use atoms as nodes with their respective 3D coordinates as $\mathbf{x}_L^{(t)} \in \mathbb{R}^{N\times 3}$ and $\mathbf{x}_R^{(t)} \in \mathbb{R}^{M\times 3}$, as well as the initial ψ_h-dimension roto-translational invariant features $\mathbf{h}_L^{(t)} \in \mathbb{R}^{N\times \psi_h}$ and $\mathbf{h}_R^{(t)} \in \mathbb{R}^{M\times \psi_h}$ (e.g. atom types, electronegativity). Edges include all atom pairs within a distance cutoff of 4Å. Then the spatial temporal protein sequence is represented as $\left\{\left(\mathcal{G}_L^{(t)}, \mathcal{G}_R^{(t)}\right)\right\}_{t=1}^{T}$. We denote a vector norm by $x = \|\mathbf{x}\|_2$, and the relative position by $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$.

2.1.2 Prompt-based Denoising Conformation Generative Task

Generative self-supervised learning is a classic track for unsupervised pre-training (7; 51; 63; 64; 67). It expects to learn effective representations by reconstructing each data point itself. Specifically to drug binding, a good self-supervised learning task should satisfy the following three essential properties. (1) The prediction target is reliable and easy to get. (2) The prediction target should reflect the temporal information within MD trajectories and is relevant to the drug binding. (3) Learned presentations should be diverse and distinguishable.

Guided by these criteria, we present a novel generative prediction task. First, We use a prompt-based pre-training approach to take into account both short-term and long-term temporal dependencies. We set a the prompt embedding $\mathbf{h}_{\text{prompt}} \in \mathbb{R}^{\psi_{\text{prompt}}}$, and append the prompting embedding $\mathbf{h}_{\text{prompt}}^{\Delta t_i}$ to the atomic feature at present timestep t as the model input to achieve the task of predicting the conformation of the next $(t+i)$-th step ($i \in \mathbb{Z}^+$), i.e., $\left(\mathbf{x}_L^{(t+i)}, \mathbf{x}_R^{(t+i)}\right)$. This enables our ProtMD approach to not only capture conformational changes between adjacent time steps, but also mine longer-term trajectory information. Then, we perturb the input conformation $\left(\mathcal{G}_L^{(t)}, \mathcal{G}_R^{(t)}\right)$ with a little random noise at each timestep. It is worth noting that random distortions of the geometry of ligands and receptors at a local energy minimum are almost certainly higher energy configurations (42). This denoising procedure that maps from a noised molecule to a local energy minimum endows the model with eligibility to learn a map from high energy to low energy, which reveals the binding strength to some extent and maintains exactly the hidden information we expect to encode from MD trajectories. Moreover, since the difference between adjacent conformations or conformations of close steps can

[1] The source code of this study providing the ProtMD is freely available in https://github.com/smiles724/ProtMD.

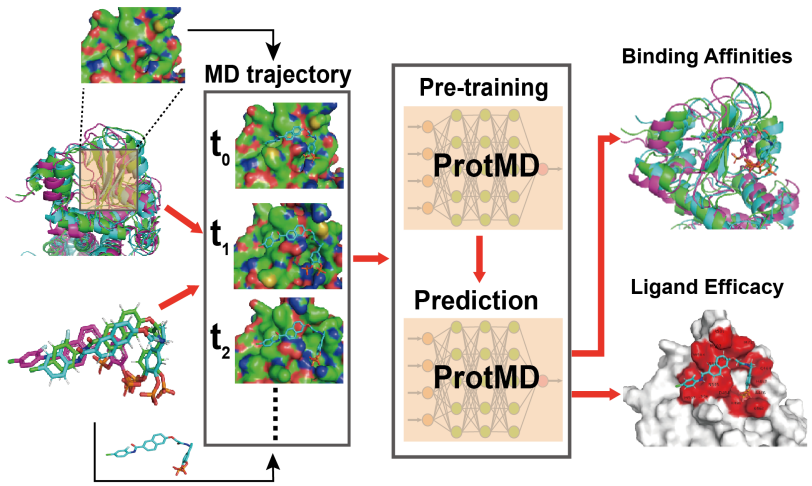


Fig. 1 High-level overview of our ProtMD based on the MD trajectories. The input during pre-training is a large amount of conformations at different timeframes. The output during inference is a wide variety of drug binding-related properties.

be tiny, noise serves to prevent overfitting and makes the pre-training more robust.

2.1.3 Snapshot Ordering Task

It has been widely proven that the shape and surface of each conformation carry crucial information for understanding potential molecular interactions (99). However, previous prompt-based denoising generative task is founded on the atom level and may fail to achieve that purpose. It is of necessity for us formulate a conformation-level self-supervised learning task to catch the global geometric information of conformations. Inspired by the classic sentence ranking task from NLP (53; 75), we design a snapshot ordering task. To be specific, we requires the model to order a set of closely-related conformations as a coherent sub-trajectory, which teaches the model to understand their dependencies from a global perspective (71).

2.2 Equivariant Graph Matching Network

To distinguish the ligand and receptor and also concern the geometric relationship between them in conformational, we refine *E(3)-Equivariant Graph Matching Network* (EGMN) (38) as the molecule encoder network for the ProtMD model. We strictly distinguish the intersections inside and across two graphs $\mathcal{G}_L^{(t)}$ and $\mathcal{G}_R^{(t)}$ respectively as $\mathcal{E}_L^{(t)} \cup \mathcal{E}_R^{(t)}$ and $\mathcal{E}_{LR}^{(t)}$ based on their spatial correlations. It avoids the underutilization of cross-graph edges information (e.g., interatomic distances) due to implicit positional relationships between ligands and receptors. For the protein-ligand at the t-th step, we input the set of atom embeddings $\left\{\mathbf{h}_L^{(t)}, \mathbf{h}_R^{(t)}\right\}$, and 3D coordinates $\left\{\mathbf{x}_L^{(t)}, \mathbf{x}_R^{(t)}\right\}$. Then it outputs a transformation on $\left\{\mathbf{h}_L^{(t+1)}, \mathbf{h}_R^{(t+1)}\right\}$ and $\left\{\mathbf{x}_L^{(t+1)}, \mathbf{x}_R^{(t+1)}\right\}$, where the latter is exactly the coordinates of the next timeframe. Concisely, $\mathbf{h}_L^{(t+1)}, \mathbf{x}_L^{(t+1)}, \mathbf{h}_R^{(t+1)}, \mathbf{x}_R^{(t+1)} = \text{EGMN}\left(\mathbf{h}_L^{(t)}, \mathbf{x}_L^{(t)}, \mathbf{h}_R^{(t)}, \mathbf{x}_R^{(t)}\right)$.

3 Experiments and Analysis

To thoroughly evaluate the efficiency of the representations learned by our ProtMD, we test its performance in two downstream tasks on the benchmark drug binding dataset with both linear-probing and fine-tuning, and compare it with multiple state-of-the-art methods. The linear-probing updates all model parameters while fine-tuning only updates the last linear layers (e.g, the prediction head).

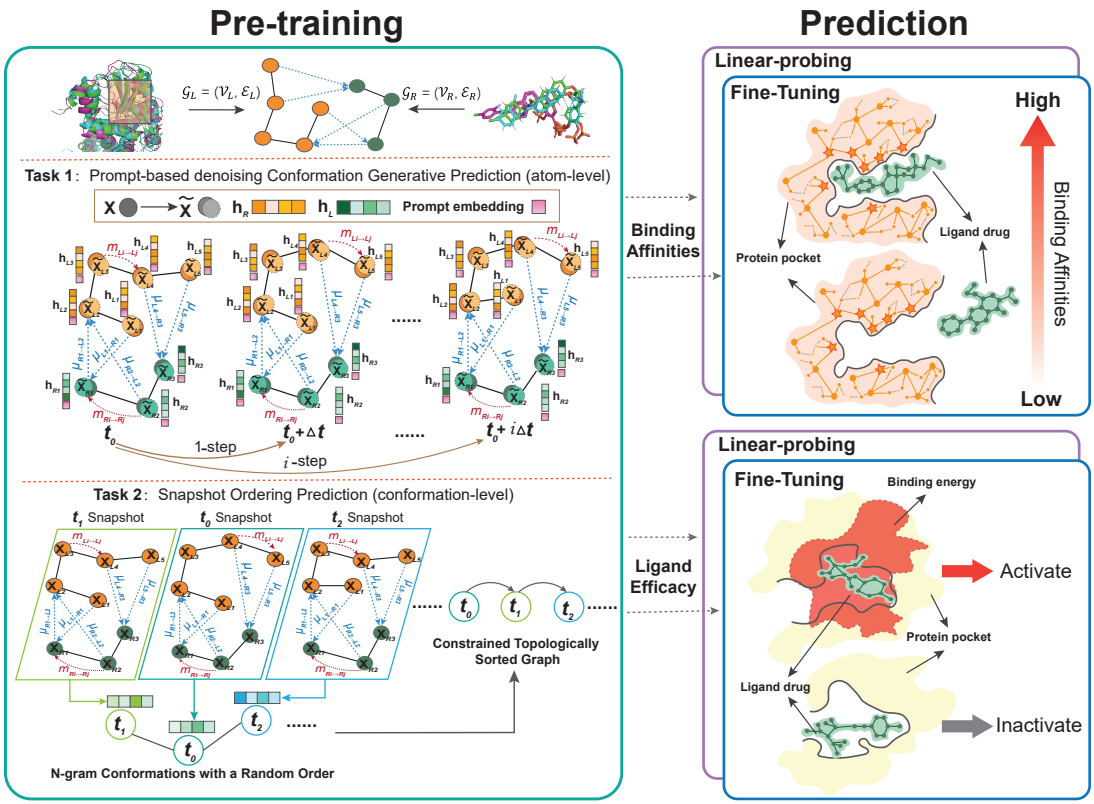


Fig. 2 Framework of our ProtMD pipelines. In the pre-training stage, two types of self-supervised learning tasks are introduced there to capture drug binding-related information hidden inside the MD trajectories. One is the prompt-based denoising generative prediction task, and the other is the snapshot ordering task. In the fine-tuning and linear-probing stage, the pre-trained model is utilized in two downstream problems, where the binding affinity prediction is a regression problem, and the ligand efficacy prediction is a classification problem.

3.1 Dataset and Setup

3.1.1 pre-training Data

In regards to the pre-training data collection, we selected sixty-four protein-ligand pairs in PDBbind and run their MD simulations. Conformations of each protein-ligand pair at a series of time intervals are generated by Amber (18), a high-performance toolkit widely accepted for molecular simulation. In addition, though the MD simulations are implemented with the whole protein-ligand pair as well as the solvents, we only use the pocket part as the model input instead of the entire protein in ProtMD for the following two major reasons. First, the pocket is the most crucial region that the protein interacts with the ligand, which undergoes the most violent spatial change during the interaction process and can reveal enough information about drug binding. This also explains why we adopt the pocket fraction as the model input for subsequent downstream tasks. Second, the pocket is much smaller and contains far less atoms than the integral protein, so the training speed is significantly faster. To be specific, we locate the pocket as atoms in proteins whose minimum distance to the ligand is shorter than a threshold of 6Å. More details regarding the experiments and conformation generations are elaborated in Appendix B

3.1.2 Downstream Data

Concerning binding affinity prediction, we adopt the PDBbind database (74; 113), a curated database containing protein-ligand complexes from the Protein Data Bank (12) and their corresponding binding strengths. The binding affinity

Table 1 Comparison of RMSE, R_p, and R_s on PDBbind. The best performance is marked bold and the second best is underlined for clear comparison. Results are reported with the mean and the standard deviation values for 3 experimental runs.

Model	# Params	Pre.	Sequence Identity (30 %)		
			RMSE	R_p	R_s
Sequence-based Methods					
DeepDTA (83)	1.93M	No	1.565 ± 0.080	$\underline{0.573 \pm 0.022}$	$\underline{0.574 \pm 0.024}$
LSTM (10)	48.8M	No	1.985 ± 0.016	0.165 ± 0.006	0.152 ± 0.024
TAPE (90)	93M	No	1.890 ± 0.035	0.338 ± 0.044	0.286 ± 0.124
ProtTrans (31)	2.4M	No	1.544 ± 0.015	0.438 ± 0.053	0.434 ± 0.058
Surface-based Method					
MaSIF (37)	0.62M	No	1.484 ± 0.018	0.467 ± 0.020	0.455 ± 0.014
Multi-scale Methods					
HoloProt (99)	1.44M	No	1.464 ± 0.006	0.509 ± 0.002	0.500 ± 0.005
Structure-based Methods					
3DCNN (105)	2.1M	No	1.429 ± 0.042	0.541 ± 0.029	0.532 ± 0.033
IEConv (47)	5.8M	No	1.554 ± 0.016	0.414 ± 0.053	0.428 ± 0.032
ProtMD (Linear-probing)	0.01M[1]	Yes	$\underline{1.413 \pm 0.032}$	0.572 ± 0.047	0.569 ± 0.051
ProtMD (Fine-tuning)	5.22M	Yes	$\mathbf{1.367 \pm 0.014}$	$\mathbf{0.601 \pm 0.036}$	$\mathbf{0.587 \pm 0.042}$

provided by PDBbind is experimentally determined and expressed in molar units of the inhibition constant (K_i) or dissociation constant (K_d). Accordingly, it is a regression problem. Similar to prior work (83; 99; 105), we do not distinguish these constants and predict the negative log-transformed affinity as $pK = -log(K)$. Besides, we select a 30% sequence identity threshold to limit homologous ligands or proteins and split those complexes into training, test and validation.

As for the ligand efficacy prediction problem (105), the dataset is also created from PDB (12). It contains a curated set of proteins from several families with both "active" and "inactive" state structures. Therefore, it is a traditional binary classification task. There are 527 small molecules with known activating or inactivating function modeled in using the program Glide (35).

3.2 Baselines

We choose wide-ranging popular or state-of-the-art baselines for comparison. Among them, four methods are based on sequences including LSTM (10), TAPE (90), ProtTrans (31), and DeepDTA (83). They take in pairs of ligand and protein SMILES as the input. While, two other approaches are established on molecular geometric structures. IEConv (47) designs a convolution operator that considers the primary, secondary, and tertiary structure of proteins and a set of hierarchical pooling operators for multi-scale modeling. 3DCNN (105) is also a competitive 3D method via convolution operations. Additionally, MaSIF (37) takes advantage of protein surfaces. HoloProt (99) introduces a multi-scale construction of protein representations, which connects surface to structure and sequence.

3.3 Binding Affinities Predictions

Table 1 reports the root-mean-squared error (RMSE), the Pearson correlation (R_p), and the Spearman correlation (R_s) of all baselines. We can observe that our ProtMD not only achieves the lowest RMSE, but also attains the highest Pearson and Spearman correlations compared to these state-of-the-art approaches. This indicates the strong capability and superiority of our pre-training method to learn efficacious representations for the estimation of drug binding.

It is worth noting that ProtMD with linear-probing can realize a RMSE of 1.413, which outperform all baselines. This phenomenon demonstrates the firm correlation between the

pre-training and downstream tasks, and deeper insights and exploration are offered in Section 3.6. More importantly, our model is pre-trained only in the trajectories of only sixty-four proteins, but examined in more than 3K proteins. This big gap strongly shows that our ProtMD has great generalization capability. Thus, the need for higher computational expenditure to model the trajectories of a large number of protein-ligand pairs is avoided. On the contrary, pre-training on a small group of binding pairs is adequate, and a more numerical analysis to verify this claim is provided in Section 3.7.

Besides, it can be found that ProtMD with fine-tuning yields a lower RMSE and higher correlations than ProtMD with linear-probing. This proves the necessity of tuning all parameters rather than only the parameters of the output predictor. It can also be demonstrated that structured-based methods generally surpass sequence-based and surface-based approaches.

3.4 Ligand Efficacy Prediction

Many proteins switch on or off their function by changing shape. Predicting which shape a drug will favor is thus an important task in drug design. To further demonstrate the validity of our ProtMD, we exam it in the ligand efficacy prediction task. To be explicit, it is formulated as a binary classification task where we predict whether a molecule bound to the structures will be an activator of the protein's function or not. Following Townshend et al. (105), we only use the regions within a radius of 5.5Å around the ligand as the model input and adopt a binary cross entropy loss as the loss function. Two metrics are used there: AUROC is the area under the receiver operating characteristic curves, and AUPRC is area under the precision-recall curve. Table 2 documents the results, which show that our ProtMD can realize the highest values of both AUPRC and AUROC simultaneously.

3.5 Generalization Ability to Unknown Structure Pairs

In Table 1, all previous methods are merely evaluated on experimentally known structure pairs. However, in real-world applications, the experimental cocrystal structures are not always accessible. As a remedy, *in silico* docking algorithms are adopted to predict the binding pose. Therefore, it is of necessity to validate our method on those predicted structures. Here we employ Equibind (100), a state-of-the-art paradigm, to quickly locate the binding site and the ligand's bound pose and orientation for all test samples in PDBbind. We delete 9 unsuccessfully docking pairs, where the distance between the ligand and the receptor is farther than 6 Å (i.e., no pocket exists). Table 3 reports the results and it can be found that ProtMD still achieves an RMSE as low as 1.474. This shows the generalization ability of our mechanism for unknown structures. Moreover, the decrease of the performance compared to Table 1 majorly comes from the inaccuracy of structure prediction method (i.e., Equibind). It is worth noting that Equibind has a very high ligand RMSE of 8.2 on average, whose predicted structures can significantly mislead our algorithm to forecast the binding properties.

3.6 Why Does ProtMD Work?

Ligands and receptors explore binding sites during the dynamic interaction process. Our experiments also firmly demonstrate the effectiveness of ProtMD capable of capturing such time-dependent dynamic binding information, but why self-supervised learning on MD trajectories is beneficial for downstream tasks still requires further analysis. In other words, we desire to quantitatively understand the correlation between MD simulations and drug binding-related properties such as the binding affinity. Notably, our design of ProtMD leads to transformations on both features and 3D coordinates, while the pre-training stage and the fine-tuning stage utilize different outputs to calculate the losses. On the one hand, the 3D coordinates is used in the self-supervised pre-training stage and correspond to the coordinates of future timeframe. On the other hand, the updated features is used in the fine-tuning stage and participate in acquiring the properties of the ligand-receptor pair. Therefore, it is motivating for us to explore the relationship between the outcome 3D coordinates and properties like binding affinities for our ProtMD before fine-tuning or linear-probing. To be specific, we compute the

Table 2 Comparison of AUROC and AUPRC on the ligand efficacy prediction task. Results are reported for 3 experimental runs.

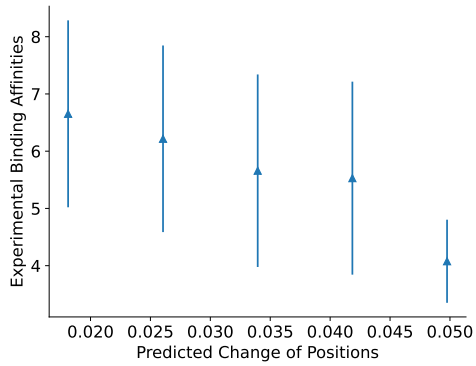
Metric	3DCNN (105)	3DGCN (105)	Cormorant (4)	DeepDTA (83)	ProtMD (Linear-probing)	ProtMD (Fine-tuning)
AUROC	0.589 ± 0.020	0.681 ± 0.062	0.663 ± 0.100	$\underline{0.696 \pm 0.021}$	0.548 ± 0.112	$\mathbf{0.742 \pm 0.039}$
AUPRC	0.483 ± 0.037	$\underline{0.598 \pm 0.135}$	0.551 ± 0.121	0.550 ± 0.024	0.516 ± 0.138	$\mathbf{0.724 \pm 0.041}$

Table 3 Comparison of RMSE, R_p, and R_s on predicted structures.

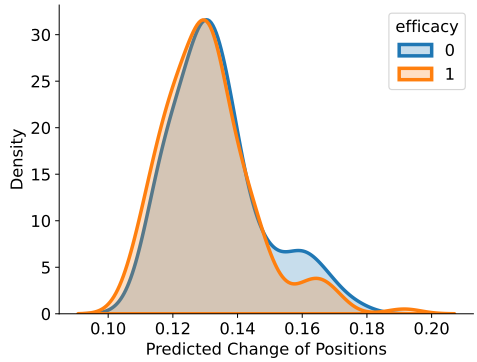
Model	Sequence Identity (30 %)		
	RMSE	R_p	R_s
No Pretrain	1.668 ± 0.048	0.447 ± 0.039	0.499 ± 0.041
Fine-tuning	$\mathbf{1.474 \pm 0.030}$	$\mathbf{0.517 \pm 0.038}$	$\mathbf{0.508 \pm 0.044}$

average predicted space shift between the input and output 3D coordinates for the ligand-receptor pair as $\Delta x_{LR} = \frac{\|\mathbf{x}_{LR}^{\text{out}} - \mathbf{x}_{LR}^{\text{in}}\|_2^2}{N+M}$. Δx_{LR} corresponds to the magnitude of the movement in the 3D space predicted by our ProtMD. Then we investigate the relation between this predicted spatial motion and the ground truth properties.

Figure 3 exhibits the results. We perform a linear least-squares regression and attain that $y = -65.90\Delta x_{LR} + 7.91$ with $R^2 = -0.2601$. It can be observed that for ligand binding affinity task, the change of spatial positions are highly associated with the binding affinities. For instance, if ProtMD forecasts a smaller change in the coordinates of the binding site, then this pair is more likely to have better binding interactions. This claim aligns with the discovery in Guterres and Im (44) that ligands with good initial binding modes tend to stay stable during MD simulations. Similarly, in the ligand efficacy prediction task, the average change is also concerned with the activity of protein-ligand pairs. To be concrete, for ligands with efficacy=0, the predicted change of positions is 0.1335 ± 0.0144 (mean $\pm$ std). For ligands with efficacy=1, the predicted change of positions is 0.1309 ± 0.0141. This indicates that more efficacious ligands have a slightly smaller positional change. We also calculate the Davis Bouldin (DB) index (108) to measure the separation of the two efficacy clusters, and it achieves a DB index as low as 8.09. All above analysis confirms that even though the pre-training stage and the fine-tuning stage employ different parts of the output of our ProtMD for backpropagation to update the model parameters (i.e., 3D coordinates for the pre-training stage and atomic features for the fine-tuning stage), ProtMD can properly capture the inherent link between 3D positions and their corresponding drug binding-related properties.



(a) Ligand Binding Affinity Prediction



(b) Ligand Efficacy Prediction

Fig. 3 Visualization of predicted spatial change and the corresponding drug binding-related properties. Sub-figure (a) is the error plot of the relationship between the predicted positional change and the binding affinities. Sub-figure (b) is the density plot of the predicted positional change and the corresponding efficacy.

3.7 How Many Trajectories Do ProtMD Need?

As previously announced, even though our pre-training dataset only contains the trajectories

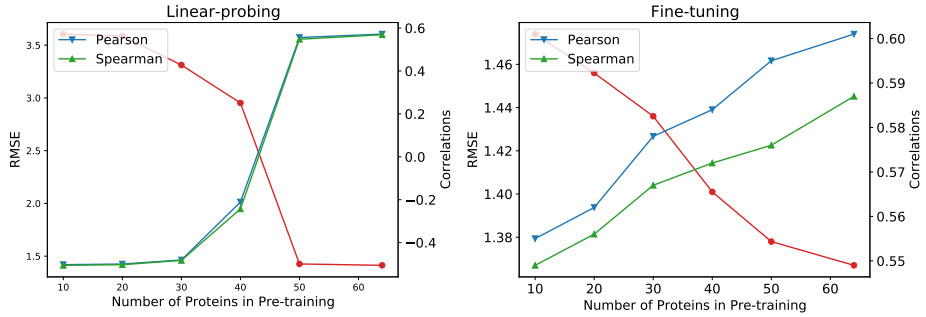


Fig. 4 Ablation study on the number of protein-ligand pairs used in the pre-training stage, where the red line denotes the RMSE, and the blue and green lines denote the Pearson and Spearman correlations, respectively. The left and right figures correspond to different strategies of linear-probing and fine-tuning, respectively.

of approximately sixty protein-ligand pairs, our model realizes unexpected generalization to all 3K samples in PDBbind. Thus, we take a further step to investigate the influence of the number of pre-training samples over the performance in downstream tasks. As shown in Figure 4, when there are only few proteins (e.g., 10 or 20 proteins), ProtMD with linear-probing performs badly. Its Pearson and Spearman correlations are negative. Nevertheless, when the number of proteins exceeds 50, the benefit for linear-probing is negligible. On the other hand, the improvement for fine-tuning persistently augments along with the increase of the number of pre-training samples. However, the increments of RMSE, Pearson and Spearman correlations also become smaller when the number of proteins used in the pre-training increases. Thus, it would be a contribution of providing a more beneficial pre-training database with a larger size and we leave it for future work to produce trajectories of more protein-ligand pairs.

3.8 Ablation Studies and Visualization

We also conduct extensive experiments to exam the effects of each component in our ProtMD in the ligand binding affinity prediction. First, we analyze if the performance of ProtMD with two self-supervised learning tasks outperforms its isolated components, i.e. when using only one task for pre-training. The second ablation axis analyzes the benefits of the noise injection and the prompt for the generative self-supervised learning.

As displayed in Table 4, the results clearly show that both self-supervised learning tasks contribute to the efficacy of learned representations. We further observe that both the noise trick and the prompt are more useful for linear-probing than fine-tuning, leading to a decrease of 0.068 in RMSE. an increase of 0.034 in the Pearson correlation and an increase of 0.039 in the Spearman correlation. This phenomenon supports our statement that noise can effectively increase the difficulty of pre-training task and prevent overfitting. Interestingly, fine-tuning the pre-trained model with a noiseless and non-prompt generative task can already realize very outstanding performance, which is even better than any sort of model with linear-probing.

To intuitively observe the representations that our self-supervised learning tasks have learned, we envision the representations by mapping them to the two-dimensional space by the Principal Component Analysis (PCA) and TSNE (109) algorithms in Figure 5. Remarkably, even without any label information, the representations learned from self-supervised learning follow some kind of pattern that is strongly related to the drug binding-related properties. For the efficacy binary classification problem, it has a DB index of 4.05. This demonstrates that our designed self-supervised learning tasks are an appropriate way to excavate structural and dynamical properties of molecular systems and comprehend the mechanism of physiochemical processes within the MD trajectory. It also aligns with the previous analysis that our ProtMD can achieve extraordinary performance in binding affinity prediction via linear-probing.

4 Related Work

4.1 Protein-ligand Modeling

With increasing availability of sequence and structure data, the area of protein representation learning has developed rapidly (57). Free energy-based

Table 4 Ablation study on ProtMD. The DG (naive) stand for the generative task without noise and prompt, and the SO stand for the snapshot ordering task. Results are documented with the mean and standard deviation for 3 runs.

	DG (naive)	Noise	Prompt	SO	RMSE	R_p	R_s
No Pre-train	-	-	-	-	1.541 ± 0.030	0.542 ± 0.048	0.527 ± 0.041
Linear-probing	✓	-	-	-	1.498 ± 0.030	0.533 ± 0.047	0.519 ± 0.039
	✓	✓	-	-	1.449 ± 0.028	0.536 ± 0.041	0.523 ± 0.035
	✓	✓	✓	-	1.430 ± 0.021	0.567 ± 0.036	0.558 ± 0.042
	✓	✓	✓	✓	$\mathbf{1.413 \pm 0.020}$	$\mathbf{0.572 \pm 0.035}$	$\mathbf{0.569 \pm 0.048}$
Fine-tuning	✓	-	-	-	1.403 ± 0.034	0.582 ± 0.038	0.561 ± 0.039
	✓	✓	-	-	1.386 ± 0.049	0.592 ± 0.044	0.578 ± 0.032
	✓	✓	✓	-	1.372 ± 0.027	0.600 ± 0.045	0.584 ± 0.033
	✓	✓	✓	✓	$\mathbf{1.367 \pm 0.024}$	$\mathbf{0.601 \pm 0.042}$	$\mathbf{0.587 \pm 0.030}$

simulations and DL-based scoring functions are two major computational methods for the binding affinity prediction (111). The latter is completely data-driven and can fast screen a vast number of compounds, attaching increasing interests.

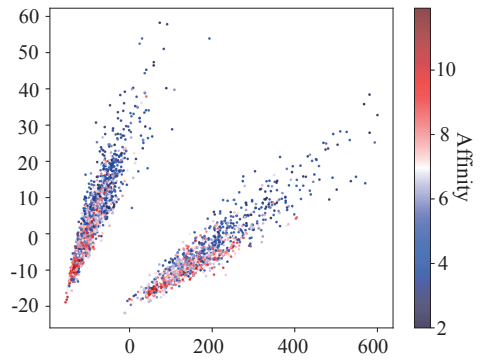
One-dimensional amino acid sequences continue to be the simplest and most abundant source of protein data, resulting in various methods (23; 50; 83) that borrow ideas from the area of NLP. Beyond that, previous methods ignore the spatial complexity of proteins and it has been proven that the exploitation of their 3D structures leads to improved performance (6; 105), which is further supported by the revolution in protein structure prediction (57). Some utilize 3D grids to capture the spatial distribution of the properties within molecular conformers, where 3DCNN (54; 89) have been the method of choice. Other studies use 3D voxel-based surface representations as inputs to 3DCNN (73; 82) for the prediction of protein-ligand binding sites (59). Apart from them, a protein structure can be naturally represented as a proximity graph over amino acid nodes, and a number of graph neural networks (GNNs) are been proposed (41). For instance, Structured Transformer (52) is designed for protein design, which encodes 3D geometry through relative orientations, and GraphQA (8) is introduced to solve model quality assessment. IEConv (47) introduces a graph convolution layer to incorporate both intrinsic and extrinsic distances between nodes. GVP (55; 56) extends standard dense layers to operate on collections of Euclidean vectors. However, none of those DL approaches consider a temporal perspective and take advantage of molecular dynamics simulations to describe the joint flexibility of proteins and ligands (49).

4.2 Protein Self-supervised Learning

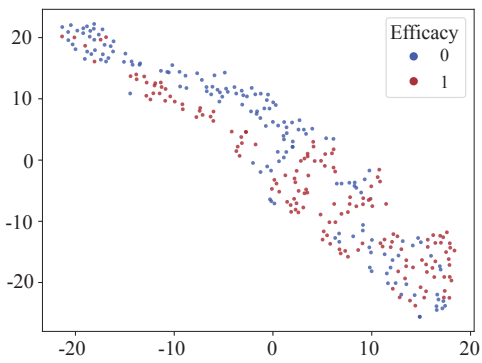
The self-supervised learning gains great progress on NLP tasks (88), and inspired by that, many methods have been expanded to the biological area (94; 117). Regarding representing proteins, most preceding studies focus on pre-training on unlabeled amino acid sequences because of their abundance (10; 11; 31; 90; 92). For instance, TAPE (90) uses the masked-token mechanism to pre-train the model and achieves good performance on several sequence-based prediction tasks. However, due to the fact that that protein functions are heavily governed by their folded structures, more and more attention has been draw to leverage the full spatial complexity of proteins. Hermosilla and Ropinski (46) uses contrastive learning for representation learning of 3D protein structures from the perspective of substructures. Apart from that, Z. Zhang et al. (118) combines a multi-view contrastive learning and a self-prediction learning to encode geometric features of proteins. Then these semantic representations learned from self-supervised learning are utilized for downstream tasks including structure classification (47), and function prediction (41). Nevertheless, no preceding research excavate the potential of pre-training on this sort of spatial-temporal data and enable the model to understand the flexibility of proteins, partly because of the high expenditure to run MD simulations.

5 Conclusion

Biological discovery demonstrates that the flexibility of both the receptor and ligand is deterministic in deciding strength of drug binding,



(a) Ligand Binding Affinity Prediction



(b) Ligand Efficacy Prediction

Fig. 5 Dimensionality reduction of the protein-ligand representations learned from our self-supervised learning tasks. The sub-figure **a** is drawn by PCA and the color is based on the strength of corresponding binding affinities. The sub-figure **b** is drawn by TSNE (109) and the color corresponds to different types of efficacy.

and MD simulations conventionally shoulder the responsibility to depict this dynamical process. In this work, in order to employ the time-dependent information of flexibility inside MD trajectories, we introduce a simple yet effective pre-training paradigm from both spatial and temporal perspectives for protein representation learning called ProtMD. It consists of two categories of self-supervised learning tasks: one is the atom-level prompt-based denoising generative task and the other is the conformation-level snapshot ordering task. And the improved EGMN as the backbone of protMD, to joint transformation of features and three-dimensional coordinates to achieve information transfer within and between graphs. We then linear probe and fine-tune the pre-trained models in the downstream drug binding prediction. Extensive experiments verify its effectiveness and ablation studies demonstrate the necessity of each component of our proposed ProtMD.

6 Methods

6.1 Prompt-based Denoising Generative Pre-training

6.1.1 Future Conformation Prediction

We use the conformation $\left(\mathcal{G}_L^{(t+1)}, \mathcal{G}_R^{(t+1)}\right)$ of the next timeframe as the target and models are required to forecast this prospective position. The objective is to maximize the likelihood as

$$L_1 = \sum_{t=k}^{T} \log P\Bigg(\left(\mathbf{x}_L^{(t+1)}, \mathbf{x}_R^{(t+1)}\right) \mid \left\{\left(\mathcal{G}_L^{(i)}, \mathcal{G}_R^{(i)}\right)\right\}_{i=1}^{t}; \theta \Bigg) \tag{1}$$

where k is the size of context window, and the conditional probability P is modeled via the encoder f_θ. Conventionally, several frameworks assume the Markov property on biomolecular conformational dynamics (21; 78) for ease of representation, i.e., $P\left(\left(\mathbf{x}_L^{(t+1)}, \mathbf{x}_R^{(t+1)}\right) \mid \left\{\left(\mathcal{G}_L^{(i)}, \mathcal{G}_R^{(i)}\right)\right\}_{i=1}^{t}\right) = P\left(\left(\mathbf{x}_L^{(t+1)}, \mathbf{x}_R^{(t+1)}\right) \mid \left(\mathcal{G}_L^{(t)}, \mathcal{G}_R^{(t)}\right)\right)$. We obey this rule and therefore set the length of context window as $k = 1$. As a consequence, our goal becomes:

$$L_1 = \log P\Bigg(\left(\mathbf{x}_L^{(t+1)}, \mathbf{x}_R^{(t+1)}\right) \mid \left\{\left(\mathcal{G}_L^{(t)}, \mathcal{G}_R^{(t)}\right)\right\}; \theta \Bigg) \tag{2}$$

6.1.2 Time-series Prompting for Motion Prediction

The previous formation of generative pre-training only guarantees the model to capture the change of conformations between adjacent timesteps. There exist other sorts of information that can only be excavated from long-term trajectories. In other words, in addition to the prediction of $\left(\mathbf{x}_L^{(t+1)}, \mathbf{x}_R^{(t+1)}\right)$, it is more reasonable to include $\left(\mathbf{x}_L^{(t+i)}, \mathbf{x}_R^{(t+i)}\right)$ $(i > 1)$ as the prediction target. A potential solution is to rely on the multi-task

learning, which allows the model to simultaneously output conformations of different time intervals. Nevertheless, with multi-task learning, the model intends to use a generalized representation and neglect the nuance between long-term and short-term trajectories.

Prompt tuning, with the emerging of GPT-3 (16), has gradually become a standard genre for pre-trained model tuning. By designing, generating and searching discrete or continuous prompts (40; 86), the gap between pre-training and fine-tuning is bridged, and the computational cost on fine-tuning the tremendous amounts of parameters is reduced. Motivated by these attractive benefits of prompts, we propose a novel prompt-based pre-training approach to concern both short-term and long-term temporal dependencies. By searching Explicitly, the prompt embedding $\mathbf{h}_{\text{prompt}} \in \mathbb{R}^{\psi_{\text{prompt}}}$ is concatenated to the atomic features of both ligand and receptor $\mathbf{h}_L^{(t)}$ and $\mathbf{h}_R^{(t)}$, respectively. This prompt serves as an indicator for the model to predict the new conformation after a certain period. To be specific, suppose there is a set of pre-defined time intervals $\{\Delta t_1, \Delta t_2, ...\}$, each time interval $\Delta t_i \in [T-1]$ has a corresponding learnable prompt embedding $\mathbf{h}_{\text{prompt}}^{\Delta t_i}$. Then if we expect the model to forecast the conformation after Δt_i, i.e., $\left(\mathbf{x}_L^{(t+\Delta t_i)}, \mathbf{x}_R^{(t+\Delta t_i)}\right)$, we append the prompting embedding $\mathbf{h}_{\text{prompt}}^{\Delta t_i}$ to the atomic feature at present timestep t as the model input. Afterwards, the dimension of initial ψ_h-dimension roto-translational invariant feature becomes $\psi_h + \psi_{\text{prompt}}$. During fine-tuning, a new prompt embedding $\mathbf{h}_{\text{prompt}}^{*} \in \mathbb{R}^{\psi_{\text{prompt}}}$ is assigned to help model differentiate the self-supervised learning pre-training and the specific downstream task.

6.1.3 Noise Perturbation

Moreover, we perturb the input conformation $\left(\mathcal{G}_L^{(t)}, \mathcal{G}_R^{(t)}\right)$ with a little noise at each timestep. This operation of perturbation has both theoretical and empirical supports. As demonstrated by Wu, Zhang, Jin, Jiang, and Li (115), the denoising diffusion architecture (5; 24; 48) has a strong connectivity with the enhanced sampling method in MD (13; 70; 81; 93), where energy is injected into the microscopic system to smooth biomolecular potential energy surface and decrease energy barriers. Besides, it has been shown in Godwin et al. (42) that the simple noise regularisation can be an effective way to address oversmoothing (17). A noise correction target can be added to prevent oversmoothing by enforcing diversity in the last few layers of our GNN, which is implemented with an auxiliary denoising autoencoder loss. In addition, we argue that as the difference between neighboring snapshots is small, the perturbation plays a critical part in preventing overfitting and improve generalization (3).

Towards this goal, the corrupted conformation is defined as $\left(\tilde{\mathcal{G}}_L^{(t)}, \tilde{\mathcal{G}}_R^{(t)}\right)$. The coordinates $\tilde{\mathbf{x}}_L^{(t)} = \mathbf{x}_L^{(t)} + \sigma^{(t)}$ and $\tilde{\mathbf{x}}_R^{(t)} = \mathbf{x}_R^{(t)} + \sigma^{(t)}$ is constructed by adding a noise, which is drawn from a normal distribution as $\sigma^{(t)} \sim \mathcal{N}\left(0, \sigma^2\right)$ and σ is a pre-defined hyperparamter to control the magnitude of perturbation. Remarkably, the translational invariant features of the ligand and the receptor $\mathbf{h}_L^{(t)}$ and $\mathbf{h}_R^{(t)}$ can be either perturbed or not.

6.2 Snapshot Ordering Pre-training

A set of n conformations with the order $\mathbf{t} = \{t_i\}_{i=1}^n$ can be described as $\left\{\left(\mathcal{G}_L^{(t_i)}, \mathcal{G}_R^{(t_i)}\right)\right\}_{i=1}^n$. The goal is to find the correct order $\mathbf{t}^*$ for them, with which the whole sub-trajectory has the greatest coherence probability as:

$$P\left(\mathbf{t}^* \mid \left\{\left(\mathcal{G}_L^{(t_i)}, \mathcal{G}_R^{(t_i)}\right)\right\}_{i=1}^n ; \theta\right) \geq P\left(\mathbf{t} \mid \left\{\left(\mathcal{G}_L^{(t_i)}, \mathcal{G}_R^{(t_i)}\right)\right\}_{i=1}^n ; \theta\right), \forall \mathbf{t} \in \Upsilon, \tag{3}$$

where $\mathbf{t}$ indicates any order of these conformations and Υ denotes the set of all possible orders. A series of approaches have been invented to tackle this rearranging problem (22; 43; 76; 114). There we select the topological sort method (58; 87), a standard algorithm for linear ordering of the vertices of a directed graph. Precisely, we have $\mathcal{C}_n$ set of constraints for this sub-trajectory. These constraints $\mathcal{C}_n$ represent the relative ordering between every pair of conformations in $\left\{\left(\mathcal{G}_L^{(t_i)}, \mathcal{G}_R^{(t_i)}\right)\right\}_{i=1}^n$. Hence, we have $|\mathcal{C}_n| = \binom{v_i}{2}$, and constraints $\mathcal{C}_n$ are learned using a multi-perceptron-layer (MLP) classifier. Notably, if we make $n = 2$, the snapshot

ordering task is deformed to the next sentence prediction (NSP) (26), which is criticized as a weak task for its comparison of similarity (101).

6.3 Equivariant Graph Matching Neural Network

Equivariance is ubiquitous in deep learning for microscopic systems. This is because the physical law controlling the dynamics of atoms stays the same regardless of the rotation and translation of biomolecules (45). Thus, it is essential to incorporate such inductive bias symmetry into model parameterization for modeling 3D geometry and achieving better generalization capacity (65; 66; 96; 104). Moreover, there are two distinct graphs in our circumstance, where the ligand graph $\mathcal{G}_L^{(t)}$ is much smaller than the receptor graph $\mathcal{G}_R^{(t)}$. To simply combine these two graphs together would confound the network and better representation capability is prohibited due to the nondiscrimination of small molecules and proteins. Therefore, it is of great need to make our model aware of the distinction between these two components.

To satisfy these requirements, we draw inspirations from recent models (36; 96; 104) and employ a variant of *E(3)-Equivariant Graph Matching Network* (EGMN) (38) as the molecule encoder. Notably, our refined EGMN has several key distinctions from prior work. On the one hand, receptor structures in Ganea et al. (38) are rigid and do not move during the whole binding process. Apart from that, since the positional relationship between ligand and receptor in their setting is implicit, they are unable to fully exploit the information within cross-graph edges (e.g., inter-atomic distances). As an alternative, they aggregate *intra*-messages assuming fully-connected edges $\mathcal{E}_{LR}^{(t)}$ between $\mathcal{G}_L^{(t)}$ and $\mathcal{G}_R^{(t)}$. On contrast, we strictly distinguish the intersections inside and across two graphs $\mathcal{G}_L^{(t)}$ and $\mathcal{G}_R^{(t)}$ respectively as $\mathcal{E}_L^{(t)} \cup \mathcal{E}_R^{(t)}$ and $\mathcal{E}_{LR}^{(t)}$ based on their spatial correlations.

Concisely, we construct the cross-graph edges $\mathcal{E}_{LR}^{(t)}$ based on their atomic pairwise distances in addition to the internal edges of $\mathcal{G}_L^{(t)}$ and $\mathcal{G}_R^{(t)}$. Then the layer of EGMN is formally defined as the following:

$$\mathbf{m}_{j\to i} = \phi_e\left(\mathbf{h}_i^{(t),l}, \mathbf{h}_j^{(t),l}, x_{ij}^{(t),l}\right), \forall e_{ij} \in \mathcal{E}_L^{(t)} \cup \mathcal{E}_R^{(t)}, \tag{4}$$

$$\boldsymbol{\mu}_{j\to i} = a_{j\to i}\mathbf{h}_j^{(t),l} \cdot \phi_d\left(x_{ij}^{(t),l}\right), \forall e_{ij} \in \mathcal{E}_{LR}^{(t)}, \tag{5}$$

$$\mathbf{x}_i^{(t),l+1} = \mathbf{x}_i^{(t),l} + \left(\mathbf{x}_i^{(t),l} - \mathbf{x}_j^{(t),l}\right)\phi_x(i, j^*) \tag{6}$$

$$\mathbf{h}_i^{(t),l+1} = \phi_h\left(\mathbf{h}_i^{(t),l}, \sum_j \mathbf{m}_{j\to i}, \sum_{j'} \boldsymbol{\mu}_{j'\to i}\right), \tag{7}$$

where ϕ_e is the edge operation, and ϕ_h denotes the node operation that aggregates the *intra*-graph messages $\mathbf{m}_i = \sum_j \mathbf{m}_{j\to i}$ and cross-graph message $\boldsymbol{\mu}_i = \sum_{j'} \boldsymbol{\mu}_{j'\to i}$ as well as the node embeddings $\mathbf{h}_i^{(t),l}$ to acquire the updated node embedding $\mathbf{h}_i^{(t),l+1}$. ϕ_x varies according to whether the edge e_{ij} is *intra*-graph or cross-graph. Particularly, $\phi_x = \phi_m\left(\mathbf{m}_{i\to j}\right)$ if $e_{ij} \in \mathcal{E}_L^{(t)} \cup \mathcal{E}_R^{(t)}$. Otherwise, $\phi_x = \phi_\mu\left(\boldsymbol{\mu}_{i\to j}\right)$ when $e_{ij} \in \mathcal{E}_{LR}^{(t)}$, where ϕ_m and ϕ_μ are two different functions to cope with different kinds of messages. It takes as input the edge embedding $\mathbf{m}_{i\to j}$ or $\boldsymbol{\mu}_{i\to j}$ as the weight to sum all relative distance $\mathbf{x}_i^{(t),l} - \mathbf{x}_j^{(t),l}$ and output the renewed coordinates $\mathbf{x}_i^{(t),l+1}$. ϕ_d operates on the inter-atomic distances $x_{ij}^{(t),l}$. $a_{j\to i}$ is an attention weight with trainable MLPs ϕ^q and ϕ^k, and takes the following form as:

$$a_{j\to i} = \frac{\exp\left(\left\langle \phi^q\left(\mathbf{h}_i^{(t),l}\right), \phi^k\left(\mathbf{h}_j^{(t),l}\right)\right\rangle\right)}{\sum_{j'}\exp\left(\left\langle \phi^q\left(\mathbf{h}_i^{(t),l}\right), \phi^k\left(\mathbf{h}_{j'}^{(t),l}\right)\right\rangle\right)}. \tag{8}$$

Specifically, the l-th layer of our encoder ($l \in [L]$) takes as input the set of atom embeddings $\left\{\mathbf{h}_L^{(t),l}, \mathbf{h}_R^{(t),l}\right\}$, and 3D coordinates $\left\{\mathbf{x}_L^{(t),l}, \mathbf{x}_R^{(t),l}\right\}$. Then it outputs a transformation on $\left\{\mathbf{h}_L^{(t+1),l}, \mathbf{h}_R^{(t+1),l}\right\}$ and $\left\{\mathbf{x}_L^{(t+1),l}, \mathbf{x}_R^{(t+1),l}\right\}$, where the latter is exactly the coordinates of the next timeframe. Concisely, $\mathbf{h}_L^{(t+1),l+1}, \mathbf{x}_L^{(t+1),l+1}, \mathbf{h}_R^{(t),l+1}, \mathbf{x}_R^{(t),l+1} = \text{EGMN}\left(\mathbf{h}_L^{(t),l}, \mathbf{x}_L^{(t),l}, \mathbf{h}_R^{(t),l}, \mathbf{x}_R^{(t),l}\right)$.

6.4 Fine-tuning and Linear-probing

When fine-tuning, since there is only one snapshot for each protein-ligand pair, we omit the temporal superscript. Then we average pool $\mathbf{h}_i$ across all atoms in both protein and ligand to extract a ψ_h-dimensional vector of features per example as $\mathbf{H} = \text{Pool}\left(\{\mathbf{h}_i\}_{i=1}^{N}\right) \in \mathbb{R}^{\psi_h}$. We expect to learn a projection from $\mathbf{H}$ to the target property such as the binding affinities and the ligand efficacy. We use a root-mean-squared-error (RMSE) loss and a binary cross-entropy loss to supervise the training for them, respectively.

Extracting features for linear probing follows a similar procedure to fine-tuning, except that those features are fixed and the encoder does not participate in the backpropagation.

Acknowledgments

This work is supported in part by the National key research and development program (Grant No. 2021YFA1301603) and National Natural Science Foundation of China (No. U21A20427).

Authors' contributions

F.W., S.Z.L., and S.J led the research. F.W., S.Z.L., S.J. and Y.J. contributed technical ideas. F.W., S.J., X.J. and Q.Z. developed the proposed method. F.W., S.J., Z.N., X.Z and X.L. performed analysis. S.Z.L., X.L., X.Z., Q.Z. and Z.N. provided evaluation and suggestions. All authors contributed to the manuscript.

Competing interests

Y.J., X.J., B.T and Z.N. are the employees of the MindRank AI Itd.. The other authors have no conflicts of interest.

Appendix A Molecular Dynamics Simulations

A.1 More Introductions about MD Simulations

Ab initio MD techniques have long and extensively been used to investigate structural and dynamical properties of a wide variety of molecular systems and understand the mechanism of physiochemical processes (34; 106; 107). It substantially accelerates the studies to observe biomolecular process in action, particularly important functional processes such as ligand binding (97), ligand- or voltage-induced conformational change (27), protein folding (72), or membrane transport (69; 102).

The most basic and intuitive application of MD is to assess the mobility or flexibility of various regions of a biomolecule. Instead of yielding an average structure by experimental structure determination methods including X-ray crystallography and cryo-EM, MD allows researchers to quantify how much various regions of the molecule move at equilibrium and what types of structural fluctuations they undergo, which is critical for protein function and ligand binding (14; 61; 68). To be explicit, on the one hand, simulations of the full ligand-binding process can reveal the binding site and pose of a ligand (28; 29; 60; 97). On the other hand, at a quantitative level, simulation-based methods provide essentially more accurate estimates of ligand binding affinities (free energies) than other computational approaches such as docking (85).

A.2 MD Simulations in Experiments

We use the pdb4amber program (18) to prepare the original PDB file downloads from the RCSB. The tags of all non-standard residues are sorted and the ones with the fewest occurrences are selected to be merged with the protein, and obtain the complex files. These files are manually inspected by a pharmaceutical expert to determine whether they are suitable as protein-ligand complex models.

It takes an RTX3080 GPU approximately 20 hours to run 100 nanoseconds (ns) per protein-ligand complex with the periodic boundary condition in the NPT ensemble. The detailed steps are described as the following:

(1) The solvated system is conducted 5,000 steps of minimization by specifying **MAXCYC=5000**, sander will use the steepest descent algorithm for the first **NCYC=2500** steps before switching to the conjugate gradient algorithm for the remaining (MAXCYC - NCYC).

(2) The NVT simulation is heated gradually from 0 to **303.15**K in the NVT ensemble during a period of **500 ps**. The heated system is equilibrated in the NPT ensemble during a period of 1 ns.

(3) Production simulation at the temperature of **303.15K** and the pressure of 1 atmospheric pressure (atm). The SHAKE algorithm is used to constrain all covalent bonds involving hydrogens and does not calculate the forces of bonds containing hydrogen. Finally, a 100 ns production simulation is performed and the structure snapshots are collected every 1 ps.

The names of pairs used in our ProtMD are listed as follows: 1TBF, 1TXI, 1ZKL, 1ZP5, 2E1Q, 2GH5, 2I0E, 2JED, 2JIF, 2NO6, 2Z5X, 3B6H, 3D4S, 3DPK, 3FVO, 3I8V, 3IAR, 3IW4, 3JZB, 3LW0, 3OLL, 3QXM, 3ROD, 3TKM, 3W2T, 4DJH, 4IAQ, 4IB4, 4MUW, 4NQD, 4PXZ, 4QTB, 4RWD, 4UDA, 4UXQ, 5AFJ, 5AX3, 5C37, 5CGD, 5DIQ, 5DSG, 5DYY, 5EDU, 6R4V, 6SSQ, 6WV3, 6X40, 7AOS, 7AYM, 7B0V, 7BR3, 7BVQ, 7BW1, 7C7S, 7CMV, 7CX3, 7DFW, 7DHL, 7EO4, 7JVP, 7JVR, 7LRC, 7VNR. The data will be realised once our paper get accepted.

Appendix B Experimental Details

B.1 EGMN Architecture

For each ligand-receptor pair, there are 10000 timestep (i.e., $T = 10000$). The dimension of roto-translational invariant features ψ_h and the prompt embeddings ψ_{prompt} are both 128. The pre-defined time intervals are set as $[1, 5, 10]$.

We use a 6-layer EGMN with a hidden dimensionality of 256, the twice of the input feature dimension. We adopt the coordinate normalization and a sum pooling method. The coordinate clamping value is set

Table B1 The training hyper-parameters.

Hyper-parameter	Description	Range
lr	The initial learning rate of ReduceLROnPlateau learning rate scheduler.	$[1e-4, 1e-5]$
min_lr	The minimum learning rate of ReduceLROnPlateau learning rate scheduler.	$[5e-6.5e-7]$
noise	The magnitude of noise injected into the input conformation during the pre-training stage.	$[1e-5, 1e-3, 1e-1, 1]$

as 2. A dropout rate of 0.15 is used for any layer. The maximum number of nodes for the input graph is set as 10000.

B.2 Downstream Dataset Splits

According to Townshend et al. (105) the split based on a 30% sequence identity threshold leads to training, validation, and test sets of size 3507. 466, and 490, respectively. In regards to the ligand efficacy prediction, complex pairs is splitted by protein target, the training, validation, and test sets have 608, 208, and 208 samples, respectively.

B.3 Training Details

We use Pytorch (84) to implement EGMN and a default random seed of 1234. On the pre-training stage, utilize the distributed training with 4 V100 GPUs and a batch size of 32 for each GPU. An Adam (62) optimizer is used and a ReduceLROnPlateau scheduler is enforced to adjust it with a factor of 0.6 and a patience of 10. The initial learning rate is 1e-4, and we apply no weight decay there. Each model is trained with 200 epochs. We split the trajectory of each protein-ligand pair into training and validation with a ratio of 9:1, and save the best model based on its performance on the validation set.

On the downstream fine-tuning and linear-probing stage, the number of GPUs, the configurations of optimizer and scheduler keep the same. The batch size is 64. Following Townshend et al. (105), we also only use the pocket position and the ligand as the model input. Only atoms within a distance of 6 Å from the ligand are used and the number of atoms in total (ligand + protein) is limited to no more than 600. We perform a hyperparamter sweep in Table B1 for different pre-training models and different strategies of linear-probing and fine-tuning.

B.4 Baselines

For protein-ligand binding affinity prediction, we use the reported values from Somnath et al. (99) and Townshend et al. (105). As for the model size, we use all available reported numbers from Somnath et al. (99). For 3DCNN, we download the code from the official repository [2] and compute the model parameters. For ligand efficacy prediciton, we use the baseline values from Townshend et al. (105).

B.5 Visualization Details

For the visualization in the main text, we try several different approaches for dimension reduction including PCA, TSNE, Uniform Manifold Approximation and Projection (UMAP), and Linear Discriminant Analysis (LDA). The results turn out that PCA and TSNE perform the best for our two downstream tasks, respectively. Since the efficiency of dimension reduction technique is out of the scope of this paper, we believe it will not cause any problem if different dimension reduction methods are applied to visualize the high-dimension representations.

To be concise, we apply the following setting of the TSNE algorithm to the presentations of protein-ligand pairs $\mathbf{H}$ for the dimension reduction in the ligand efficacy prediction. Concretely, the maximal iteration is 10000. The perplexity is 30.0. The learning rate is 200. The early exaggeration is 12.0. The angle is 0.5. Finally, the t-SNE reduces the outputs of EGMN into the 2-dimensional representations,

[2] 3DCNN: https://github.com/drorlab/atom3d/blob/master/examples/lba/cnn3d/train.py#L179

which then are plotted as 2D images. Additionally, we adopt PCA to reduce the dimension of EGMN's outcome using the popular *sklearn.decomposition* package. We use the default setting with an automatic Singular Value Decomposition (SVD) solver. Thus, it uses the LAPACK implementation of the full SVD or a randomized truncated SVD, depending on the shape of the input data and the number of components to extract.

References

Abagyan, R., & Totrov, M. (2001). High-throughput docking for lead generation. *Current opinion in chemical biology*, *5*(4), 375–382.

Abrams, C., & Bussi, G. (2014). Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*, *16*(1), 163–199.

An, G. (1996). The effects of adding noise during backpropagation training on a generalization performance. *Neural computation*, *8*(3), 643–674.

Anderson, B., Hy, T.S., Kondor, R. (2019). Cormorant: Covariant molecular neural networks. *Advances in neural information processing systems*, *32*.

Anderson, B.D. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, *12*(3), 313–326.

Atz, K., Grisoni, F., Schneider, G. (2021). Geometric deep learning on molecular representations. *Nature Machine Intelligence*, 1–10.

Bajaj, P., Xiong, C., Ke, G., Liu, X., He, D., Tiwary, S., ... Gao, J. (2022). Metro: Efficient denoising pretraining of large scale autoencoding language models with model generated signals. *arXiv preprint arXiv:2204.06644*.

Baldassarre, F., Menéndez Hurtado, D., Elofsson, A., Azizpour, H. (2021). Graphqa: protein model quality assessment using graph convolutional networks. *Bioinformatics*, *37*(3), 360–366.

Ballester, P.J., & Mitchell, J.B. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, *26*(9), 1169–1175.

Bepler, T., & Berger, B. (2019). Learning protein sequence embeddings using information from structure. *arXiv preprint arXiv:1902.08661*.

Bepler, T., & Berger, B. (2021). Learning the protein language: Evolution, structure, and function. *Cell systems*, *12*(6), 654–669.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., ... Bourne, P.E. (2000). The protein data bank. *Nucleic acids research*, *28*(1), 235–242.

Bernardi, R.C., Melo, M.C., Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1850*(5),

872–877.

Berneche, S., & Roux, B. (2001). Energetics of ion conduction through the k+ channel. *Nature*, *414*(6859), 73–77.

Boehr, D.D., Nussinov, R., Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, *5*(11), 789–796.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., ... others (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877–1901.

Cai, C., & Wang, Y. (2020). A note on over-smoothing for graph neural networks. *arXiv preprint arXiv:2006.13318*.

Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K.M., ... Woods, R.J. (2005). The amber biomolecular simulation programs. *Journal of computational chemistry*, *26*(16), 1668–1688.

Changeux, J.-P., & Edelstein, S. (2011). Conformational selection or induced fit? 50 years of debate resolved. *F1000 biology reports*, *3*.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *International conference on machine learning* (pp. 1597–1607).

Chodera, J.D., & Noé, F. (2014). Markov state models of biomolecular conformational dynamics. *Current opinion in structural biology*, *25*, 135–144.

Cui, B., Li, Y., Chen, M., Zhang, Z. (2018). Deep attentive sentence ordering network. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4340–4349).

Dalkiran, A., Rifaioglu, A.S., Martin, M.J., Cetin-Atalay, R., Atalay, V., Doğan, T. (2018). Ecpred: a tool for the prediction of the enzymatic functions of protein sequences based on the ec nomenclature. *BMC bioinformatics*, *19*(1), 1–13.

De Groot, S.R., & Mazur, P. (2013). *Non-equilibrium thermodynamics*. Courier Corporation.

De Vivo, M., Masetti, M., Bottegoni, G., Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *Journal of medicinal chemistry*, *59*(9), 4035–4061.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dror, R.O., Arlow, D.H., Maragakis, P., Mildorf, T.J., Pan, A.C., Xu, H., ... Shaw, D.E. (2011). Activation mechanism of the β2-adrenergic receptor. *Proceedings of the National Academy of Sciences*, *108*(46), 18684–18689.

Dror, R.O., Green, H.F., Valant, C., Borhani, D.W., Valcourt, J.R., Pan, A.C., ... others (2013). Structural basis for modulation of a g-protein-coupled receptor by allosteric drugs. *Nature*, *503*(7475), 295–299.

Dror, R.O., Pan, A.C., Arlow, D.H., Borhani, D.W., Maragakis, P., Shan, Y., ... Shaw, D.E. (2011). Pathway and mechanism of drug binding to g-protein-coupled receptors. *Proceedings of the National Academy of Sciences*, *108*(32), 13118–13123.

Durrant, J.D., & McCammon, J.A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, *9*(1), 1–9.

Elnaggar, A., Heinzinger, M., Dallago, C., Rihawi, G., Wang, Y., Jones, L., ... others (2020). Prottrans: towards cracking the language of life's code through self-supervised deep learning and high performance computing. *arXiv preprint arXiv:2007.06225*.

Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Berichte der deutschen chemischen Gesellschaft*, *27*(3), 2985–2993.

Fischer, M., Coleman, R.G., Fraser, J.S., Shoichet, B.K. (2014). Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nature chemistry*, *6*(7), 575–583.

Frenkel, D., & Smit, B. (2001). *Understanding molecular simulation: from algorithms to applications* (Vol. 1). Elsevier.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., ... others (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, *47*(7), 1739–1749.

Fuchs, F., Worrall, D., Fischer, V., Welling, M. (2020). Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, *33*, 1970–1981.

Gainza, P., Sverrisson, F., Monti, F., Rodola, E., Boscaini, D., Bronstein, M., Correia, B. (2020). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, *17*(2), 184–192.

Ganea, O.-E., Huang, X., Bunne, C., Bian, Y., Barzilay, R., Jaakkola, T., Krause, A. (2021). Independent se (3)-equivariant models for end-to-end rigid protein docking. *arXiv preprint arXiv:2111.07786*.

Ganesan, A., Coote, M.L., Barakat, K. (2017). Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug discovery today*, *22*(2), 249–269.

Gao, T., Fisch, A., Chen, D. (2020). Making pre-trained language models better few-shot learners. *arXiv preprint arXiv:2012.15723*.

Gligorijević, V., Renfrew, P.D., Kosciolek, T., Leman, J.K., Berenberg, D., Vatanen, T., ... others (2021). Structure-based protein function prediction using graph convolutional networks. *Nature communications*, *12*(1), 1–14.

Godwin, J., Schaarschmidt, M., Gaunt, A.L., Sanchez-Gonzalez, A., Rubanova, Y., Veličković, P., ... Battaglia, P. (2021). Simple gnn regularisation for 3d molecular property prediction and beyond. *International conference on learning representations.*

Gong, J., Chen, X., Qiu, X., Huang, X. (2016). End-to-end neural sentence ordering using pointer network. *arXiv preprint arXiv:1611.04953*.

Guterres, H., & Im, W. (2020). Improving protein-ligand docking results with high-throughput molecular dynamics simulations. *Journal of Chemical Information and Modeling*, *60*(4), 2189–2198.

Han, J., Rong, Y., Xu, T., Huang, W. (2022). Geometrically equivariant graph neural networks: A survey. *arXiv preprint arXiv:2202.07230*.

Hermosilla, P., & Ropinski, T. (2022). Contrastive representation learning for 3d protein structures. *arXiv preprint arXiv:2205.15675*.

Hermosilla, P., Schäfer, M., Lang, M., Fackelmann, G., Vázquez, P.P., Kozlíková, B., ... Ropinski, T. (2020). Intrinsic-extrinsic convolution and pooling for learning on 3d protein structures. *arXiv preprint arXiv:2007.06252*.

Ho, J., Jain, A., Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, *33*, 6840–6851.

Hollingsworth, S.A., & Dror, R.O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*(6), 1129–1143.

Hou, J., Adhikari, B., Cheng, J. (2018). Deepsf: deep convolutional neural network for mapping protein sequences to folds. *Bioinformatics*, *34*(8), 1295–1303.

Hu, Z., Dong, Y., Wang, K., Chang, K.-W., Sun, Y. (2020). Gpt-gnn: Generative pre-training of graph neural networks. *Proceedings of the 26th acm sigkdd international conference on knowledge discovery & data mining* (pp. 1857–1867).

Ingraham, J., Garg, V., Barzilay, R., Jaakkola, T. (2019). Generative models for graph-based protein design. *Advances in neural information processing systems*, *32*.

Jernite, Y., Bowman, S.R., Sontag, D. (2017). Discourse-based objectives for fast unsupervised sentence representation learning. *arXiv preprint arXiv:1705.00557*.

Jiménez, J., Skalic, M., Martinez-Rosell, G., De Fabritiis, G. (2018). K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of chemical information and modeling*, *58*(2), 287–296.

Jing, B., Eismann, S., Soni, P.N., Dror, R.O. (2021). Equivariant graph neural networks for 3d macromolecular structure. *arXiv preprint arXiv:2106.03843*.

Jing, B., Eismann, S., Suriana, P., Townshend, R.J., Dror, R. (2020). Learning from protein structure with geometric vector perceptrons. *arXiv preprint arXiv:2009.01411*.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... others (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*(7873), 583–589.

Kahn, A.B. (1962). Topological sorting of large networks. *Communications of the ACM*, *5*(11), 558–562.

Kandel, J., Tayara, H., Chong, K.T. (2021). Puresnet: prediction of protein-ligand binding sites using deep residual neural network. *Journal of cheminformatics*, *13*(1), 1–14.

Kappel, K., Miao, Y., McCammon, J.A. (2015). Accelerated molecular dynamics simulations of ligand binding to a muscarinic g-protein-coupled receptor. *Quarterly reviews of biophysics*, *48*(4), 479–487.

Khafizov, K., Perez, C., Koshy, C., Quick, M., Fendler, K., Ziegler, C., Forrest, L.R. (2012). Investigation of the sodium-binding sites in the sodium-coupled betaine transporter betp. *Proceedings of the National Academy of Sciences*, *109*(44), E3035–E3044.

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D.P., & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, *31*.

Kingma, D.P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Köhler, J., Klein, L., Noé, F. (2019). Equivariant flows: sampling configurations for multi-body systems with symmetric energies. *arXiv preprint arXiv:1910.00753*.

Köhler, J., Klein, L., Noé, F. (2020). Equivariant flows: exact likelihood generative learning for symmetric densities. *International conference on machine learning* (pp. 5361–5370).

Larsson, G., Maire, M., Shakhnarovich, G. (2016). Learning representations for automatic colorization. *European conference on computer vision* (pp. 577–593).

Li, J., Shaikh, S.A., Enkavi, G., Wen, P.-C., Huang, Z., Tajkhorshid, E. (2013). Transient formation of water-conducting states in membrane transporters. *Proceedings of the National Academy of Sciences*, *110*(19), 7696–7701.

Liang, R., Swanson, J.M., Madsen, J.J., Hong, M., DeGrado, W.F., Voth, G.A. (2016). Acid activation mechanism of the influenza a m2 proton channel. *Proceedings of the National Academy of Sciences*, *113*(45), E6955–E6964.

Liao, Q. (2020). Enhanced sampling and free energy calculations for protein simulations. *Progress in molecular biology and translational science* (Vol. 170, pp. 177–213). Elsevier.

Lin, J., Nogueira, R., Yates, A. (2021). Pretrained transformers for text ranking: Bert and beyond. *Synthesis Lectures on Human Language Technologies*, *14*(4), 1–325.

Lindorff-Larsen, K., Piana, S., Dror, R.O., Shaw, D.E. (2011). How fast-folding proteins fold. *Science*, *334*(6055), 517–520.

Liu, Q., Wang, P.-S., Zhu, C., Gaines, B.B., Zhu, T., Bi, J., Song, M. (2021). Octsurf: Efficient hierarchical voxel-based molecular surface representation for protein-ligand affinity prediction. *Journal of Molecular Graphics and Modelling*, *105*, 107865.

Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., ... Wang, R. (2015). Pdb-wide collection of binding data: current status of the pdbbind database. *Bioinformatics*, *31*(3), 405–412.

Logeswaran, L., & Lee, H. (2018). An efficient framework for learning sentence representations. *arXiv preprint arXiv:1803.02893*.

Logeswaran, L., Lee, H., Radev, D. (2018). Sentence ordering and coherence modeling using recurrent neural networks. *Thirty-second aaai conference on artificial intelligence.*

Ma, B., Shatsky, M., Wolfson, H.J., Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein science*, *11*(2), 184–197.

Malmstrom, R.D., Lee, C.T., Van Wart, A.T., Amaro, R.E. (2014). Application of molecular-dynamics based markov state models to functional proteins. *Journal of chemical theory and computation*, *10*(7),

2648–2657.

McCammon, J.A., Gelin, B.R., Karplus, M. (1977). Dynamics of folded proteins. *Nature*, *267*(5612), 585–590.

Miao, Y., Feher, V.A., McCammon, J.A. (2015). Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation*, *11*(8), 3584–3595.

Miao, Y., & McCammon, J.A. (2017). Gaussian accelerated molecular dynamics: Theory, implementation, and applications. *Annual reports in computational chemistry* (Vol. 13, pp. 231–278). Elsevier.

Mylonas, S.K., Axenopoulos, A., Daras, P. (2021). Deepsurf: a surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics*, *37*(12), 1681–1690.

Öztürk, H., Özgür, A., Ozkirimli, E. (2018). Deepdta: deep drug–target binding affinity prediction. *Bioinformatics*, *34*(17), i821–i829.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... others (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Perez, A., Morrone, J.A., Simmerling, C., Dill, K.A. (2016). Advances in free-energy-based simulations of protein folding and ligand binding. *Current opinion in structural biology*, *36*, 25–31.

Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A.H., Riedel, S. (2019). Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*.

Prabhumoye, S., Salakhutdinov, R., Black, A.W. (2020). Topological sort for sentence ordering. *arXiv preprint arXiv:2005.00432*.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. (2018). Improving language understanding by generative pre-training.

Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J., Koes, D.R. (2017). Protein–ligand scoring with convolutional neural networks. *Journal of chemical information and modeling*, *57*(4), 942–957.

Rao, R., Bhattacharya, N., Thomas, N., Duan, Y., Chen, X., Canny, J., ... Song, Y.S. (2019). Evaluating protein transfer learning with tape. *Advances in neural information processing systems*, *32*, 9689.

Reymond, J.-L., & Awale, M. (2012). Exploring chemical space for drug discovery using the chemical universe database. *ACS chemical neuroscience*, *3*(9), 649–657.

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., ... others (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, *118*(15).

Rocchia, W., Masetti, M., Cavalli, A. (2012). Enhanced sampling methods in drug design. *Physico-Chemical and Computational Approaches to Drug Discovery. The Royal Society of Chemistry*, 273–301.

Rong, Y., Bian, Y., Xu, T., Xie, W., Wei, Y., Huang, W., Huang, J. (2020). Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, *33*, 12559–12571.

Salmaso, V., & Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Frontiers in pharmacology*, *9*, 923.

Satorras, V.G., Hoogeboom, E., Welling, M. (2021). E (n) equivariant graph neural networks. *International conference on machine learning* (pp. 9323–9332).

Shan, Y., Kim, E.T., Eastwood, M.P., Dror, R.O., Seeliger, M.A., Shaw, D.E. (2011). How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, *133*(24), 9181–9183.

Śledź, P., & Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Current opinion in structural biology*, *48*, 93–102.

Somnath, V.R., Bunne, C., Krause, A. (2021). Multi-scale representation learning on proteins. *Thirty-fifth conference on neural information processing systems.*

Stärk, H., Ganea, O.-E., Pattanaik, L., Barzilay, R., Jaakkola, T. (2022). Equibind: Geometric deep learning for drug binding structure prediction. *arXiv preprint arXiv:2202.05146*.

Sun, Y., Zheng, Y., Hao, C., Qiu, H. (2021). Nsp-bert: A prompt-based zero-shot learner through an original pre-training task–next sentence prediction. *arXiv preprint arXiv:2109.03564*.

Suomivuori, C.-M., Gamiz-Hernandez, A.P., Sundholm, D., Kaila, V.R. (2017). Energetics and dynamics of a light-driven sodium-pumping rhodopsin. *Proceedings of the National Academy of Sciences*, *114*(27), 7043–7048.

Teague, S.J. (2003). Implications of protein flexibility for drug discovery. *Nature reviews Drug discovery*, *2*(7), 527–541.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., Riley, P. (2018). Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*.

Townshend, R.J., Vögele, M., Suriana, P., Derry, A., Powers, A., Laloudakis, Y., ... others (2020). Atom3d: Tasks on molecules in three dimensions. *arXiv preprint arXiv:2012.04035*.

Tuckerman, M. (2010). *Statistical mechanics: theory and molecular simulation*. Oxford university press.

Tuckerman, M.E., & Martyna, G.J. (2000). *Understanding modern molecular dynamics: Techniques and applications* (Vol. 104) (No. 2). ACS Publications.

Van Der Maaten, L. (2014). Accelerating t-sne using tree-based algorithms. *The journal of machine learning research*, *15*(1), 3221–3245.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*(11).

Vogt, A.D., & Di Cera, E. (2012). Conformational selection or induced fit? a critical appraisal of the kinetic mechanism. *Biochemistry*, *51*(30), 5894–5902.

Wang, D.D., Zhu, M., Yan, H. (2021). Computationally predicting binding affinity in protein–ligand complexes: free energy-based simulations and machine learning-based scoring functions. *Briefings in Bioinformatics*, *22*(3), bbaa107.

Wang, J., Arantes, P.R., Bhattarai, A., Hsu, R.V., Pawnikar, S., Huang, Y.-m.M., ... Miao, Y. (2021). Gaussian accelerated molecular dynamics: Principles and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *11*(5), e1521.

Wang, R., Fang, X., Lu, Y., Wang, S. (2004). The pdbbind database: Collection of binding affinities for protein- ligand complexes with known three-dimensional structures. *Journal of medicinal chemistry*, *47*(12), 2977–2980.

Wu, F., & Bai, X. (2021). Insertgnn: Can graph neural networks outperform humans in toefl sentence insertion problem? *arXiv preprint arXiv:2103.15066*.

Wu, F., Zhang, Q., Jin, X., Jiang, Y., Li, S.Z. (2022). A score-based geometric model for molecular dynamics simulations. *arXiv preprint arXiv:2204.08672*.

Zhang, Q., Wang, Z., Han, Y., Yu, H., Jin, X., Chen, H. (2022). Prompt-guided injection of conformation to pre-trained protein model. *arXiv preprint arXiv:2202.02944*.

Zhang, Z., Liu, Q., Wang, H., Lu, C., Lee, C.-K. (2021). Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, *34*.

Zhang, Z., Xu, M., Jamasb, A., Chenthamarakshan, V., Lozano, A., Das, P., Tang, J. (2022). Protein representation learning by geometric structure pretraining. *arXiv preprint arXiv:2203.06125*.