SPIN: SE(3)-Invariant Physics Informed Network for Binding Affinity Prediction

Seungyeon Choi[a,1]**, Samgmin Seo**[a] **and Sanghyun Park**[a,*]

[a]Department of Computer Science, Yonsei University, Seoul, Republic of Korea

Abstract. Accurate prediction of protein-ligand binding affinity is crucial for rapid and efficient drug development. Recently, the importance of predicting binding affinity has led to increased attention on research that models the three-dimensional structure of protein-ligand complexes using graph neural networks to predict binding affinity. However, traditional methods often fail to accurately model the complex's spatial information or rely solely on geometric features, neglecting the principles of protein-ligand binding. This can lead to overfitting, resulting in models that perform poorly on independent datasets and ultimately reducing their usefulness in real drug development. To address this issue, we propose **SPIN**, a model designed to achieve superior generalization by incorporating various inductive biases applicable to this task, beyond merely training on empirical data from datasets. For prediction, we defined two types of inductive biases: a geometric perspective that maintains consistent binding affinity predictions regardless of the complex's rotations and translations, and a physicochemical perspective that necessitates minimal binding free energy along their reaction coordinate for effective protein-ligand binding. These prior knowledge inputs enable the **SPIN** to outperform comparative models in benchmark sets such as CASF-2016 and CSAR HiQ. Furthermore, we demonstrated the practicality of our model through virtual screening experiments and validated the reliability and potential of our proposed model based on experiments assessing its interpretability.

1 Introduction

The binding affinity (BA) of a small molecule to a target protein associated with a disease indicates whether the small molecule (a ligand) can become a new drug [9, 26]. The higher the BA, the stronger the binding to the target protein and the desired effect, which can be measured using biological tests. However, these methods are expensive and time consuming. Therefore, compared with biological testing, BA prediction using machine learning (ML) and deep learning (DL) may accelerate drug development and reduce costs [37]. Powered by this need, numerous studies have been conducted to predict BA, especially with a recent increase in efforts to predict BA by analyzing the three-dimensional structure of complexes formed between proteins and ligands [22]. This trend stems from research indicating that since the BA between a protein and a ligand is based on their interaction in 3D space, predicting BA through information derived from the three-dimensional structural characteristics of the protein and ligand is a rational approach [2, 18].

---
* Corresponding Author. Email: sanghyun@yonsei.ac.kr
1 First Author. Email: tmddus1553@yonsei.ac.kr

DL models utilize various methods to represent the 3D structures of protein–ligand complexes for BA prediction and are primarily categorized into grid- and graph-based representations. The grid-based approach involves voxelizing the structure of the protein-ligand complex into a 3D grid format, after which it is commonly processed using 3D convolutional neural network (CNN) to predict the binding affinity [33, 41]. However, most grid-based representations do not
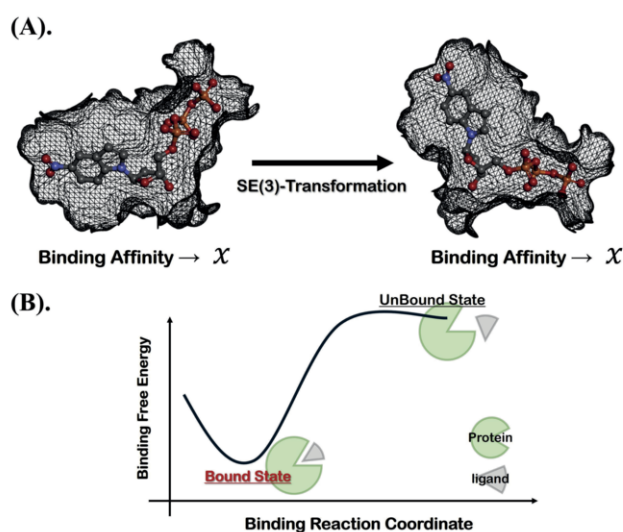

Figure 1. Two inductive biases of the protein–ligand BA prediction model: (A). Geometric inductive bias, i.e., the BA of a complex remains constant despite undergoing SE(3)-transformations, such as translations and rotations. (B). Physicochemical inductive bias, i.e., protein-ligand complex should be positioned at the point where binding free energy is minimal among their possible reaction coordinates.

completely encapsulate the structural information of the complexes. Consequently, the CNN architecture becomes relatively inefficient in depicting the geometric intricacies of protein-ligand complexes. Furthermore, employing a 3D rectangular grid representation can generate high-dimensional sparse 3D matrices, potentially resulting in significant computational costs. To address these issues, recent research has focused on representing the structure of complexes in a graphical format to predict BA. This approach defines the atoms of the protein and ligand as nodes in a graph and assigns edge information to atom pairs that are within a certain distance threshold. Consequently, the complex is represented as either a single or multiple graphs, enabling the operation of graph neural network models [14, 25, 30, 40]. How-

ever, traditional methods that consider only the connectivity information between atoms fail to adequately model the spatial information of complexes. Additionally, the need to define numerous preset rules to handle complex geometric features limits the model's flexibility when encountering unseen samples during training. Therefore, designing a model that efficiently models the complex information of the protein-ligand complex, ensuring it has excellent generalization capabilities, is essential.

Applying the concept of Physics-Informed Neural Network (PINN) [8, 13] to BA prediction can be a promising approach to enhancing model generalization performance by incorporating domain knowledge inherent in the protein-ligand complex into the prediction model. Specifically, beyond relying solely on the empirical distribution of data for model training, defining the immutable governing laws of the real world as the inductive bias of the prediction model ensures robustness against noise in the training dataset. Furthermore, the same domain knowledge can be applied to unseen samples during training to achieve superior generalization performance [3]. In this research, the governing laws, or inductive biases, are defined in two categories as illustrated in figure 1. The first inductive bias, as shown in figure 1.(A), is based on the fundamental principle that the binding affinity of a protein-ligand complex remains constant despite SE(3)-transformations, such as rotations and translations. This inherently sensible law allows for more efficient modeling of the spatial information of the complex compared to some binding affinity prediction models that only input three-dimensional coordinates. The second inductive bias, illustrated in figure 1.(B), relies on the physicochemical principle that the binding free energy of a protein-ligand complex is located at the point where the energy is minimized among all possible geometric configurations of the protein and ligand [39]. These two governing laws represent domain knowledge that can be explained within the context of protein-ligand complexes and are immutable truths applicable equally in both training and testing scenarios. In this way, we propose **SPIN**, a graph transformer model that incorporates the SE(3)-invariant principle, which is invariant to rotation and translation, and the principle of minimal binding free energy, as its geometric and physicochemical inductive biases, respectively, to achieve high generalization performance. Consequently, we have validated that the **SPIN** outperforms other prediction models on two benchmark datasets. Extensive experiments have also allowed us to assess the interpretability and practicality of our proposed model. Our main contributions can be summarized as follows:

- We propose **SPIN**, a novel binding affinity prediction model that is infused with essential prior knowledge applicable to various protein-ligand interactions, designed to achieve superior generalization performance.
- To the best of our knowledge, this is the first model for predicting binding affinity where geometric and physicochemical knowledge are defined as inductive biases. This approach integrates empirical learning from deep learning with immutable truths applicable in reality, allowing for extreme efficiency with limited data. Furthermore, The trained prediction model inherently reflects the fundamental principles of protein-ligand binding.
- **SPIN** achieves state-of-the-art (SOTA) performance in the CASF-2016 and CSAR-HiQ benchmarks, demonstrating the model's excellent interpretability and practicality through various experiments.

2 Related Work

2.1 Structure-based BA Prediction

The prediction of protein–ligand BA is indispensable in drug screening, wherein the selection of molecular structures that are viable as drug candidates from numerous molecular structures is crucial [26]. Recognizing this importance, numerous studies have proposed various methodologies for BA prediction owing to the advancements in ML and DL. Moreover, recent research, such as that involving AlphaFold [12, 35], which models the tertiary structure of proteins, has significantly enhanced the potential and possibility of studies that aim to predict binding affinities by modeling the three-dimensional structures of protein-ligand complexes [36]. This approach can be categorized into the following three types: ML-, CNN-, and GNN-based methods.

ML-based methods typically utilize predefined rules to extract descriptors from interactions within a certain distance between protein and ligand atoms, using ML models such as support vector regression and random forest, to predict BA [1]. However, these methods only consider atom pair interactions between proteins and ligands, neglecting the interactions within themselves and between multiple atom types. Thus, they fail to capture the rich spatial correlations in protein-ligand complexes. CNN-based methods transform the protein-ligand complex into a 3-D rectangular grid representation and implement a 3-D CNN model to extract features for BA prediction [33, 41]. Although this grid representation enables the mining of spatial and local correlations, it generate empty grid points where no atoms exist, causing inefficient computation and unnecessary memory usage. Furthermore, grid representations lack distance awareness and rotational invariance, which can render the prediction performance unstable. To address these issues, GNN-based methods have been developed for BA prediction [25, 40, 17, 24]. The GNN approaches define the atoms of proteins and ligands as nodes in a graph, and atom pairs within a certain distance threshold are connected with edge information, representing the complex as a single graph or multiple graphs for processing by GNN models. Recent efforts also directly input the 3D coordinates of atoms to predict binding affinity [5], and additional research aims to map various spatial features of the complex to more deeply learn its geometric characteristics [14, 15]. However, these methods can be sensitive to rotations and translations of the complex or may struggle to flexibly handle various geometric combinations not encountered during training due to too many predefined rules. To overcome the absence of generalization, our goal is to introduce various inductive biases into the predictive model. These inductive biases are intended to enable the model to account for unseen geometrical configurations and ensure robustness against variations in complex orientation and position, thus enhancing the model's predictive accuracy and generalization capabilities across different protein-ligand complexes.

2.2 PINN

Deep learning models have achieved tremendous success in various domains such as computer vision and natural language processing. However, the traditional approach of learning from large amounts of data still faces challenges in extracting interpretable information. Additionally, while existing data-driven approaches may align well with observations in training data, predictions from models can be physically inconsistent or result in impossible values, leading to reduced generalization performance. This issue is especially pronounced in domains like physics and fluid dynamics, where data collection is

relatively difficult, unlike in fields with vast amounts of data such as image processing or natural language processing.

To address these issues, some domains are actively researching physics-informed neural network (PINN), which integrate governing physical laws and domain knowledge into deep learning models [13, 23, 28, 4]. Integrating domain knowledge into DL models means defining laws applicable to a domain as inductive biases during the model's training process. This ensures that the DL model implicitly satisfies these predefined laws during learning and inference. Even if the training dataset contains outliers or noise, integrating these immutable truths applicable to the real world can enhance the model's robustness. Consequently, the trained model can make more realistic predictions by training on not just empirical information from the dataset but also incorporating prior domain knowledge. This not only improves model generalization but also uses prior knowledge to illuminate the inner mechanisms of deep learning and provide theoretical insights [31, 7]. Especially in studies predicting the interactions between proteins and ligands, research has been conducted that integrates physicochemical prior knowledge related to binding free energy to apply the PINN framework [24]. This research achieved high generalization performance in the model by defining the physical law that the positions of the atoms in the ligand within the experimentally validated protein-ligand complex structure are located at the local minimum of the potential energy as prior knowledge, i.e., inductive bias, of the prediction model. However, this model has a clear limitation in that it considers only the connection information between the protein and the ligand, failing to model the geometric information between the two structures.

3 Preliminaries

Definition 1. **Protien-Ligand Graph Construction** Given a protein-ligand comlex as shown in Figure2.(B), we define the feature of atom node of protein and ligand as $V^P \in \mathbb{R}^{N \times D_F}$ and $V_L \in \mathbb{R}^{N \times D_F}$ with position matrix $\mathbf{x} \in \mathbb{R}^{(N+M) \times 3}$. The protein atom features include atom type, amino acid types and whether the atoms are backbone atoms. The ligand atom features include atom type, hybridzation, formal charge, degree and whether the atoms are aromatic atoms. The *knn* graph is utilized to define edges $\mathbf{e}_{ij}$ between atoms based on the 3D coordinates of each atom. $\mathbf{e}_{ij}$ is represented as a 4-dim one-hot vector, denoting connections between protein atoms, connections between ligand atoms, protein-ligand connections, and ligand-protein connections.

Definition 2. **Protien-Ligand Binding Affinity** Our goal is to predict the binding affinity, i.e., how strongly a protein and ligand bind, given the structure of a protein-ligand complex. To achieve this, we construct and train a regression model $\phi : (V^P, V^L, \mathbf{x}) \rightarrow y$ that predicts the binding affinity using the features of the atoms composing the protein and ligand in the given complex, V^P, V^L, and the 3D coordinates of the atoms, $\mathbf{x}$.

4 Method

4.1 Overview

To achieve a predictive model with high generalizability, it is important to define as the model's inductive bias the prior knowledge that encompasses the real-world or the datasets being used, in addition to empirical learning based on datasets. For this purpose, we inject into the predictive model inductive biases from both geometric and physicochemical perspectives. The overall framework can be seen in Figure 2. More specifically, as shown in Figure 2.C, it is possible to

internalize a geometric inductive bias that satisfies SE(3)-invariant characteristics by utilizing the SE(3)-graph transformer, and in Figure 2.E$_2$, it can satisfy the physicochemical inductive bias by enforcing the condition that the binding free energy should be minimized at the given complex position.

The statement below outlines the proposition necessary for the overall framework to satisfy the characteristic of being SE(3)-invariant.

Proposition 1. *Denoting the SE(3)-transformation as T_g, let's $f(x)$ be an SE(3)-invariant function, i.e., $f(x) = f(T_g(x))$. If function $g(x)$ is SE(3)-invariant function, i.e., $g(x) = g(T_g(x))$, then we have composition function $g(f(x))$ is also SE(3)-invariant function, i.e., $g(f(x)) = g(f(T_g(x)))$.*

Here, $f(x)$ refers to the SE(3)-invariant graph transformer, and $g(x)$ denotes the SE(3)-invariant prediction function that ultimately predicts the BA. This proposition indicates the conditions necessary for the proposed framework to possess SE(3)-invariant characteristics overall when performed across multiple functions. The following sections will detail the operation of an SE(3)-invariant graph transformer (Section 4.2), how the function predicting binding affinity can satisfy physical laws (Section 4.3), and whether that function is SE(3)-invariant (Section 4.3).

4.2 SE(3)-Invariant Graph Transformer

We first encode the types (V) of the atoms composing the given protein and ligand individually through an embedding layer, as below.

$$\begin{aligned}
\mathbf{h}_P^0 &= \text{Linear}(V^P) \in \mathbb{R}^{N \times D_E} \\
\mathbf{h}_L^0 &= \text{Linear}(V^L) \in \mathbb{R}^{M \times D_E} \\
\mathbf{h}^0 &= \left[\mathbf{h}_P^0 \| \mathbf{h}_L^0 \right]
\end{aligned} \quad (1)$$

Here, the result of concatenating $\mathbf{h}_L^0$ and $\mathbf{h}_P^0$, denoted as $\mathbf{h}^0$, becomes the initial hidden representation that serves as the input to the SE(3)-Invariant graph transformer. Subsequently, to update the representation of each node, we define the SE(3)-invariant graph transformer layer, which is invariant to SE(3)-transformation, as follows.

$$\mathbf{h}_i^{l+1} = \mathbf{h}_i^l + \sum_{j \in \mathcal{V}, i \neq j} f_\mathbf{h} \left(\| \mathbf{x}_i - \mathbf{x}_j \|, \mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{e}_{ij}; \theta_\mathbf{h} \right) \quad (2)$$

where h_i^l is the hidden representation vector for atom i at the lth layer and $\| \mathbf{x}_i - \mathbf{x}_j \|^2$ is the euclidean distance between two atom i and j. The update function $f_\mathbf{h}$ computes messages for updating the node state after aggregating information from neighboring nodes using attention operations as follows.

$$\begin{aligned}
f_\mathbf{h} &= \mathbf{Attn}\left(\mathbf{q}_i, \mathbf{k}_j, \mathbf{v}_j \right) \cdot \text{Linear}\left(\mathbf{r}_{ij} \right) \\
\mathbf{q}_i &= \text{Linear}\left(\mathbf{h}_i^0 \right) \\
\mathbf{k}_j &= \text{Linear}\left([\mathbf{r}_{ij} \| \mathbf{e}_{ij} \| \mathbf{h}_i \| \mathbf{h}_j] \right) \\
\mathbf{v}_j &= \text{Linear}\left([\mathbf{r}_{ij} \| \mathbf{e}_{ij} \| \mathbf{h}_i \| \mathbf{h}_j] \right)
\end{aligned} \quad (3)$$

Here, $\mathbf{q}_i$, $\mathbf{k}_j$, and $\mathbf{v}j$ represent the query, key, and value matrices for the attention operation, respectively, and $\mathbf{r}_{ij}$ is defined as the distance embedding with radial basis functions located at 20 centers between 0Å and 10Å. The final atom hidden representation $\mathbf{h}^L$ becomes the node state where spatial information of neighboring atoms is aggregated, explicitly considering the relationships between protein and ligand atoms.
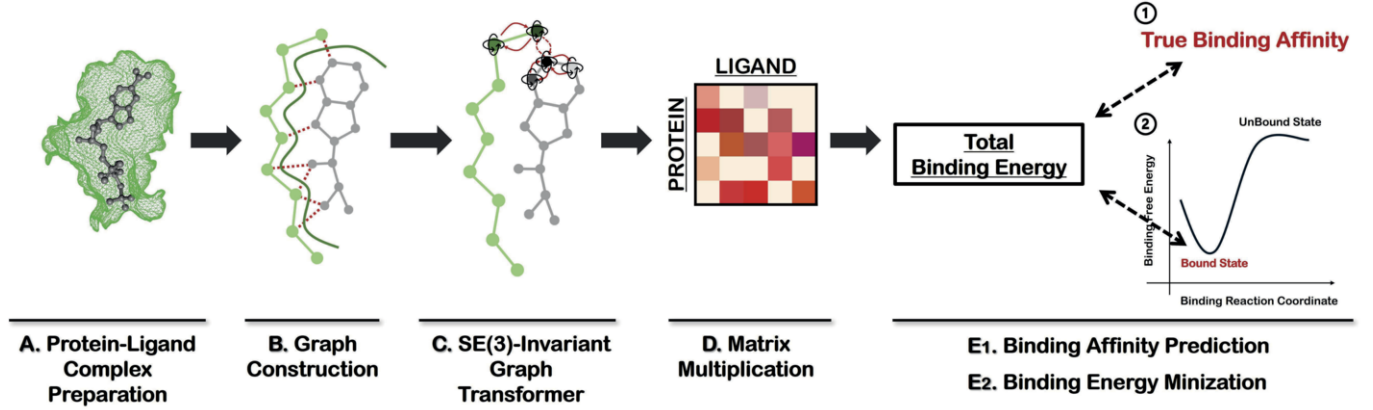
Figure 2. Overview of **SPIN**. **A.** Protein-ligand complex preparation from PDBbind Dataset **B.** Atoms constituting a complex are defined as nodes, and the connections between atoms are defined as edges, representing the entire structure as a single graph **C.** Each node feature is updated through an SE(3)-Graph transformer, in conjunction with the features of the connected edges, to model the geometric information in the three-dimensional space of the complex **D.** The protein-ligand interaction matrix is extracted through matrix multiplication of the final node representation vectors of protein and ligand atoms. The computed interaction matrix is defined in terms of pairwise energy values between protein atoms and ligand atoms. **E₁**. The binding affinity is predicted by summing the values of the extracted pairwise interaction matrix. **E₂**. The binding free energy is minimized by enforcing that the derivatives of the pairwise distances between protein and ligand atoms in the extracted pairwise interaction matrix equal zero.

4.3 SE(3)-Invariant Physics-informed Prediction

The obtained final representation vectors of protein and ligand atoms are used to model the pairwise interaction between the protein and ligand. For conciseness, we denote the final representation vector of the atoms constituting the ligand as $\mathbf{h}_M^L$, and the atoms constituting the protein as $\mathbf{h}_P^L$. The protein-ligand binding affinity can be converted as the sum of the atom pairwise van der Waals interaction energy between the protein and the ligand [6]. To achieve this, we calculate the protein-ligand interaction matrix $\mathbf{H}$ through the matrix multiplication of $\mathbf{h}_M$ and $\mathbf{h}_P$, which are outputted by the SE(3)-invariant graph transformer, as shown in Figure 4.D and as described in Equation (4).

Additionally, we model the van der Waals interaction by calculating the pairwise distance between protein atoms and ligand atoms, along with the interaction matrix $\mathbf{H}$. The van der Waals ineteraction energy can be formulated as in Equation (5) through the Lennard-Jones potential fomula.

$$\mathbf{H} = \mathbf{h}_M^L \cdot \mathbf{h}_P^{L^\top} \tag{4}$$

$$\mathrm{E}^{\mathrm{VDW}} = \sum_{i,j} c_{ij} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right)^{12} - 2 \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right)^6 \right] \tag{5}$$

Here, $\mathbf{u}_{ij}$ refers to the sum of the van der Waals radii of the i-th ligand atom and the j-th protein atom. The van der Waals radii for each atom type were obtained from the parameters of X-score [38]. Thus, $\mathrm{E}^{\mathrm{VDW}}$ can be calculated as a combination of parameters parameterized by the SE(3)-invariant transformer and the parameters of a physics formula utilizing the positional information of the protein-ligand complex.

An important point here is that for the overall framework of binding affinity prediction to be invariant to the rotation and translation of 3D objects, the sum of the protein-ligand pairwise interactions we model must also be invariant to translation and rotation. To this end, we prove below that the interaction we model is SE(3)-invariant.

Proof. Let $T_g(\mathbf{x})$ can be written explicitly as $T_g(\mathbf{x}) = \boldsymbol{R}\mathbf{x} + \mathbf{b}$, where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ is the rotation matrix and $\boldsymbol{b} \in \mathbb{R}^3$ is the translation vector.

$$\sum_{i,j} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|^2} \right)^{12} - \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|^2} \right)^6 \right]$$

$$= \sum_{i,j} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{T_g\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2\right)} \right)^{12} - \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{T_g\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2\right)} \right)^6 \right]$$

$$= \sum_{i,j} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|(\boldsymbol{R}\mathbf{x}_i + \boldsymbol{b}) - (\boldsymbol{R}\mathbf{x}_j + \boldsymbol{b})\|^2} \right)^{12} - \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|(\boldsymbol{R}\mathbf{x}_i + \boldsymbol{b}) - (\boldsymbol{R}\mathbf{x}_j + \boldsymbol{b})\|^2} \right)^6 \right]$$

$$= \sum_{i,j} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{(\mathbf{x}_i - \mathbf{x}_j)^\top \boldsymbol{R}^\top \boldsymbol{R}(\mathbf{x}_i - \mathbf{x}_j)} \right)^{12} - \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{(\mathbf{x}_i - \mathbf{x}_j)^\top \boldsymbol{R}^\top \boldsymbol{R}(\mathbf{x}_i - \mathbf{x}_j)} \right)^6 \right]$$

$$= \sum_{i,j} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{I}(\mathbf{x}_i - \mathbf{x}_j)} \right)^{12} - \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{I}(\mathbf{x}_i - \mathbf{x}_j)} \right)^6 \right]$$

$$= \sum_{i,j} \left[\left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|^2} \right)^{12} - \left(\frac{\mathbf{u}_{ij} + \mathbf{H}_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|^2} \right)^6 \right]$$

□

The optimization strategy of **SPIN** is derived in two directions from the sum of the previously calculated protein-ligand pairwise interactions (equal to the binding energy) as follows.

$$\hat{y} = \sigma \cdot \mathrm{E}^{\mathrm{VDW}}$$
$$L_{\mathrm{d}} = \sum_N (y - \hat{y})^2 \tag{6}$$

$$L_{\mathrm{p}} = \sum_N \sum_{i,j} \left[\frac{\partial\, c_{ij} \left[\left(\frac{\mathbf{r}_i + M_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right)^{12} - 2 \left(\frac{M_{ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right)^6 \right]}{\partial\, \|\mathbf{x}_i - \mathbf{x}_j\|} \right]^2 \tag{7}$$

The first direction is to minimize the mean squared error between the predicted binding energy-derived binding affinity and the experimental binding affinity as shown in Figure.E₁ (σ is a trainable parameter). The second direction involves enforcing a term that satisfies the laws

Method	CASF-2016 set				CSAR-HiQ set			
	RMSE ($\downarrow$)	MAE ($\downarrow$)	SD ($\downarrow$)	R ($\uparrow$)	RMSE ($\downarrow$)	MAE ($\downarrow$)	SD ($\downarrow$)	R ($\uparrow$)
LR [1]	1.675 (0.000)	1.358 (0.000)	1.612 (0.000)	0.671 (0.000)	2.071 (0.000)	1.622 (0.000)	1.973 (0.000)	0.652 (0.000)
SVR [1]	1.555 (0.000)	1.264 (0.000)	1.493 (0.000)	0.727 (0.000)	1.995 (0.000)	1.553 (0.000)	1.911 (0.000)	0.679 (0.000)
RF-Score [1]	1.446 (0.008)	1.161 (0.007)	1.335 (0.010)	0.789(0.003)	1.947 (0.012)	1.466 (0.009)	1.796 (0.020)	0.723 (0.007)
Pafnucy [33]	1.585 (0.013)	1.284 (0.021)	1.563 (0.022)	0.695 (0.011)	1.939 (0.103)	1.562 (0.094)	1.885 (0.071)	0.686 (0.027)
OnionNet [41]	1.407 (0.034)	1.078 (0.028)	1.391 (0.038)	0.768 (0.014)	1.927 (0.071)	1.471 (0.031)	1.877 (0.097)	0.690 (0.040)
SGCN [5]	1.583 (0.033)	1.250 (0.036)	1.582 (0.320)	0.686 (0.015)	1.902 (0.063)	1.472 (0.067)	1.891 (0.077)	0.686 (0.030)
GraphTrans [40]	1.539 (0.044)	1.182 (0.046)	1.521 (0.042)	0.714 (0.019)	1.950 (0.072)	1.508 (0.069)	1.886 (0.083)	0.687 (0.033)
NL-GCN [19]	1.516 (0.019)	1.198 (0.013)	1.511 (0.024)	0.720 (0.010)	1.840 (0.024)	1.393 (0.016)	1.817 (0.028)	0.716 (0.011)
GNN-DTI [17]	1.492 (0.025)	1.192 (0.032)	1.471 (0.051)	0.736 (0.021)	1.972 (0.061)	1.547 (0.058)	1.834 (0.090)	0.709 (0.035)
PIGNet [24]	1.428 (0.016)	1.133 (0.009)	1.425 (0.014)	0.761 (0.006)	1.532 (0.026)	1.198 (0.054)	1.512 (0.035)	0.781 (0.013)
MAT [21]	1.457 (0.037)	1.154 (0.037)	1.445 (0.033)	0.747 (0.013)	1.879 (0.065)	1.435 (0.058)	1.816 (0.083)	0.715 (0.030)
CMPNN [30]	1.408 (0.028)	1.117 (0.031)	1.399 (0.025)	0.765 (0.009)	1.839 (0.096)	1.411 (0.064)	1.767 (0.103)	0.730 (0.052)
SIGN [14]	1.316 (0.031)	1.027 (0.025)	1.312 (0.035)	0.797 (0.012)	1.735 (0.031)	1.327 (0.040)	1.709 (0.044)	0.754 (0.014)
GIANT [15]	1.269 (0.020)	0.999 (0.018)	1.265 (0.024)	0.814 (0.008)	1.666 (0.024)	1.242 (0.030)	1.633 (0.034)	0.779 (0.011)
CAPLA [10]	1.389 (0.018)	1.109 (0.021)	1.356 (0.034)	0.775 (0.009)	1.630 (0.044)	1.201 (0.025)	1.598 (0.59)	0.742 (0.21)
SPIN (Ours)	**1.258** (0.013)	**0.996** (0.021)	**1.229** (0.011)	**0.826** (0.007)	**1.288** (0.027)	**0.999** (0.034)	**1.270** (0.022)	**0.800** (0.017)

Table 1. Performance comparision with baselines on the CASF-2016 set and CSAR-HiQ set.

of physics between the protein and ligand, specifically by ensuring the derivative of the binding free energy with respect to the distance between ligand and protein atoms is zero as shown in Figure.E_2. This approach is designed to fulfill the prior knowledge that the modeled complex, as an experimentally elucidated structure, is located at the minimum point of the binding free energy at that reaction coordinate. Ultimately, the total objective function L is defined as $L_d + L_p$.

5 Experiments

5.1 Experimental Setup

Dataset For BA prediction, the PDBbind v2020 [20] dataset was used as the training dataset and it provided the 3D crystal structures of protein–ligand complexes with experimentally determined BAs. The 3D crystal structures were represented by the coordinates of the atoms constituting the protein and ligand, comprising 19,443 samples. To evaluate the performances of the proposed and comparison models, we used the CASF-2016 [34] and CSAR-HiQ [29] benchmark sets. Similar to PDBbind, these datasets provide the 3D crystal structures of protein–ligand complexes and their BAs. In accordance with previous studies, the CASF-2016 and CSAR-HiQ sets comprised 285 and 343 samples, respectively, and duplication in the training dataset was avoided by removing these samples. During training, the most recent samples were defined as the validation set to verify the generalizability, as has been done in recent studies [32].

Implement detail The key/value/query embedding in the SE(3)-invariant graph transformer, a component of SPIN, is obtained through a 2-layer MLP with layer normalization and ReLU activation. The transformer has 16 layers, and the hidden dimension and the number of heads are each 128 and 9, respectively. Additionally, the swish [27] function was used as the activation function for each layer. We have implemented an exponential decay of the learning rate, using a factor of 0.6 and setting a minimum of 1e-6 for the learning rate with Adam. If the validation loss does not improve over 20 consecutive evaluations, the learning rate is reduced. Our model training was conducted on a single NVIDIA GeForce GTX 3090 GPU.

5.2 Performance Evaluation

First, the proposed **SPIN** was compared with the existing baseline models on two benchmark datasets. As shown in Table 1, it

achieves the best performance across all four metrics for the two benchmark datasets. CNN-based methods, such as Pafnucy [33], lack generalizability on benchmark datasets because they are not invariant to rotations and translations of the complex. SGCN[5], a GNN-based method utilizes spatial structures but directly inputs coordinates for predictions, encounters similar issues as CNN-based models and is sensitive to rotations and translations, thus exhibiting comparable performance levels. The Graphtrans[40], NL-GCN[19], PIGNet[24], and CMPNN[30] models are not ideal for BA prediction because they rely solely on connectivity information without modeling the geometric characteristics of the 3D space. The SIGN[14] and GIANT[15] models incorporate various modules to better capture the spatial information of proteins and ligands by considering distance and angle information. However, since they follow a method based on predefined rules obtained from training data, they are not flexible enough to handle various complex structures that are not present in the training data. Especially, the result that the proposed model, **SPIN**, achieved a 30% improvement over the best baseline models on the CSAR set, which is comparatively difficult to generalize across all comparison models, suggests that injecting inductive biases from various perspectives into the prediction model plays a crucial role in generalizing to unseen datasets. An interesting observation is that another model, PIGNet[24], which defines physicochemical laws as the inductive bias of the prediction model, also achieved relatively high performance on the challenging CSAR set compared to other comparison models. This reaffirms the importance of defining physicochemical information as the inductive bias of the prediction model from the perspective of applications that require predicting binding affinity for various complexes. However, because this model does not clearly model the spatial information of complex structures, it still shows inferior performance in the CASF-2016, indicating it is not ideal as a model for predicting binding affinity.

PIGNet, which defines physicochemical laws as the inductive bias of the prediction model, significantly outperformed the other models on the challenging CSAR set. This validates the importance of defining physicochemical information as the inductive bias of the prediction model based on BA prediction for various complexes. PIGNet did not represent the spatial information of complex structures; however, it exhibited inferior performance on CASF-2016, indicating its unsuitability for BA prediction.
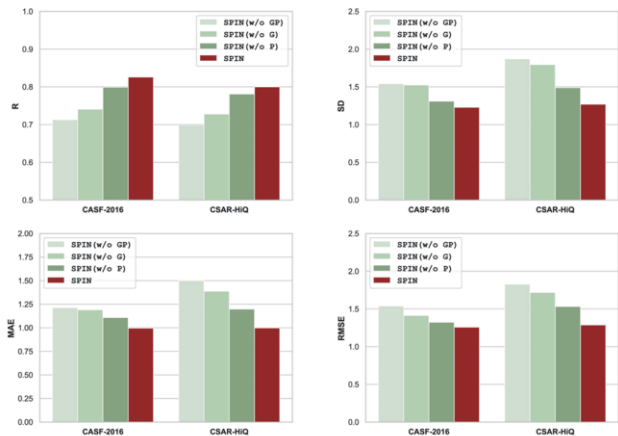
Figure 3. Ablation studies on CASF-2016, CSAR HiQ set. Performance for four evaluation metrics are presented for four cases of the SPIN model: the complete model with all inductive biases injected (`SPIN`), the case without geometric inductive bias (`SPIN[w/o G]`), the case without physicochemical inductive bias (`SPIN[w/o P]`), and the case with both biases removed (`SPIN[w/o GP]`).

5.3 Ablation Study

In this section, we conduct an ablation study on the **SPIN** model to verify the importance of each component. `SPIN(w/o G)` is a model from which the geometric inductive bias that satisfies SE(3)-invariance has been removed, `SPIN(w/o P)` is a model from which the physicochemical inductive bias concerning the Binding free energy minimum has been removed, and finally, `SPIN(w/o GP)` is a model with both types of inductive bias removed. We measured the performance of these four variant models, including the complete `SPIN` model, on two benchmark datasets. The results, as shown in Figure 3, confirm that both geometric and physicochemical inductive biases play essential roles in predicting binding affinity. Specifically, when comparing the performance of `SPIN(w/o G)` and `SPIN(w/o P)`, it is revealed that the geometric inductive bias, namely the condition that the binding affinity remains invariant to the rotation and translation of the protein-ligand complex, is the most critical component for prediction performance. It is an interesting observation that the physicochemical inductive bias shows better synergy when conditions for reasonably modeling the complex's geometric information are met. This outcome, considering that the binding free energy calculations are based on the complex's position in three-dimensional space, provides significant insights into the interplay between the two types of inductive biases.

5.4 Virtual Screening

Along with predicting the BA of a protein–ligand complex, correctly ranking multiple ligands based on their binding strength to the target protein is crucial in drug development. This facilitates the proper listing of the most promising drug candidates, thereby making the drug development process more efficient. For this purpose, we adopt the methodology of previous studies [16, 34] about CASF-2016 benchmark set to validate the practicality of **SPIN**. Specifically, we measure the ranking power, which is the ability of a scoring function to accurately rank known ligands of a specific target protein based on binding affinity when the precise binding pose is given. This approach allows us to rigorously assess the efficacy of **SPIN** in predicting binding affinities. The CASF2016 set comprises 57 protein

clusters, each cluster containing five complexes that are bound to the same protein but exhibit markedly different binding affinities. Each cluster has pre-defined rankings based on the binding affinities for different ligands. In each cluster, rankings are established based on the binding affinities for different ligands, and we measure how accurately the predictive model inferences these rankings using the Spearman's rank correlation coefficient. The average of these coefficients across all clusters is defined as the ranking power for this experiment. We selected several scoring functions that are utilized in actual docking protocols as comparative models. Additionally, the Random Forest model, already trained on the CASF-2016 dataset, was excluded from this comparison [34]. As demonstrated in Figure 4, **SPIN** exhibits superior ranking power compared to the scoring functions of other docking programs. This indicates that our proposed model can effectively prioritize candidate compounds in real-world drug development scenarios.
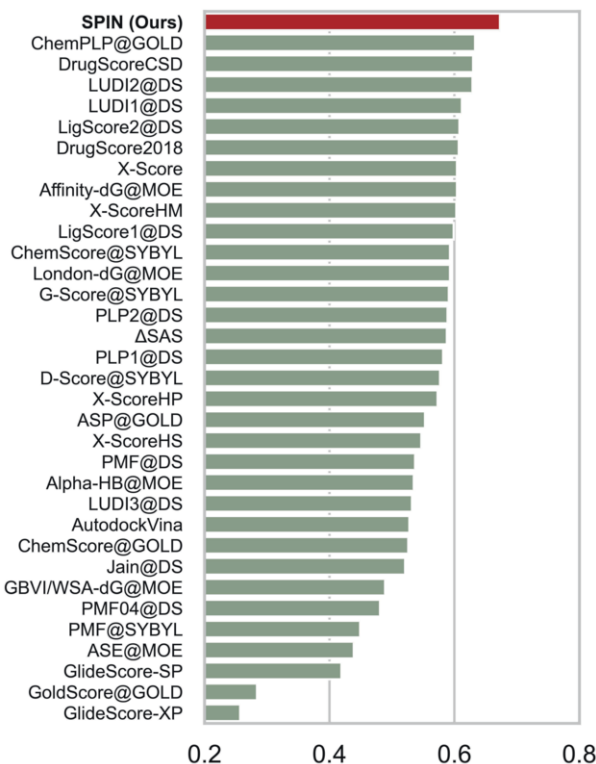

Figure 4. Average Spearman correlation coefficient obtain on 57 target proteins by each scoring function included proposed **SPIN** in the ranking power test.

5.5 Interpretability

To determine if the predictions output by a predictive model can be actively utilized in the drug development process with sufficient reliability, it is crucial to analyze the interpretability of the predictive model. Independent of the predictive performance in actual drug development, result validation can be challenging if the basis of the predicted BA is ambiguous, hindering the efficient development of subsequent processes, such as lead optimization [11]. To evaluate this, we utilize the pairwise interaction matrix **H** within the **SPIN** framework, which involves interactions between protein and ligand

atoms. The protein-ligand interaction matrix $\mathbf{H}_{ij}$ represents the binding energy between protein atom i and ligand atom j, where a lower value indicates stronger interaction between the two substructures. We extract the amino acids of the protein corresponding to the lowest 10% of the energy values in the interaction matrix output by the trained predictive model. To substantiate the interpretability of **SPIN**, we focus our analysis on the 3bu1(PDB ID) protein-ligand complex, as illustrated in Figure 5.(a), by visualizing amino acids that engage in strong interactions. To confirm that these extracted amino acids are indeed engaged in actual intermolecular interactions, we compare our results with the findings from Discovery Studio's intermolecular interaction profiler. The amino acids corresponding to the lowest 10% of the energy values in **SPIN**'s interaction matrix were specifically *21.A TYR, 37.A VAL, 51.A TYR, 94.A ASP, 96.A VAL, and 105.A TRP* (*[residue index]*.A *[amino acid]*). These residues are hypothesized to be pivotal in the protein-ligand interactions during the predictive modeling performed by **SPIN**. Subsequent analysis con-
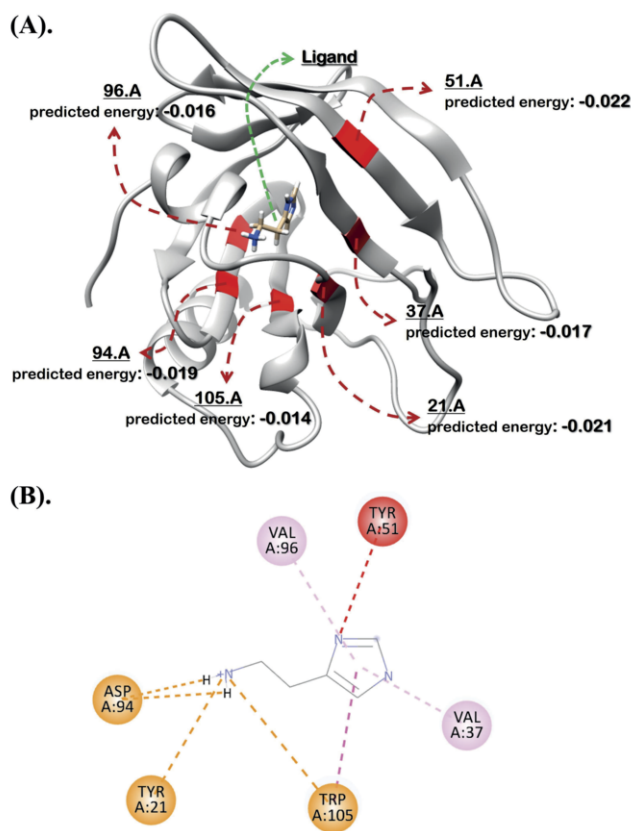

Figure 5. (A). Visualization of the 3bu1 PDB sample: Protein amino acids corresponding to the lowest 10% of the predicted protein-ligand interaction energy are highlighted in red. Additionally, the specific energy values are annotated alongside these amino acids. (B). Results from the Discovery Studio interaction profiler for the 3bu1 PDB sample show that a total of six amino acids are involved in critical interactions.

firmed that these amino acids matched precisely with those identified in Discovery Studio's profiling results, as depicted in Figure 5.(b). This concordance strongly indicates that **SPIN** not only accurately predicts binding affinity but also reliably identifies the biologically relevant interactions that underpin these predictions. This capability to discern and rationalize complex interactions in biological systems underscores the robust interpretability of **SPIN**, affirming its utility

in predictive modeling within the biochemical research field. This validation lends significant credence to **SPIN**'s application in computational drug discovery, demonstrating its potential to contribute effectively in the field by providing insights that are both predictive and interpretable.

6 Conclusion

In this work, we propose SPIN, a binding affinity prediction model infused with various inductive biases, designed to achieve excellent generalization performance from limited data. The model incorporates a geometric inductive bias that assumes binding affinity remains constant regardless of rotations and translations in three-dimensional space, and a physicochemical inductive bias that posits binding occurs at minimal binding free energy. Through rigorous validation across multiple benchmark sets, the superiority of our proposed model is confirmed. Additionally, the model's practicality in virtual screening during actual drug development processes is demonstrated. Finally, by visualizing the interpretability of the prediction model, we ensure the reliability of the values it predicts. The proposed model shows potential in the screening process for selecting molecules that strongly bind to a target protein among various molecular structures. Furthermore, using this prediction model to generate molecules with desired properties in the field of generative models presents an interesting research direction.

Acknowledgements

References

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

[2] M. Batool, B. Ahmad, and S. Choi. A structure-based drug discovery paradigm. *International journal of molecular sciences*, 20(11):2783, 2019.

[3] Z. Cui, T. Gao, K. Talamadupula, and Q. Ji. Knowledge-augmented deep learning and its applications: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[4] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

[5] T. Danel, P. Spurek, J. Tabor, M. Śmieja, Ł. Struski, A. Słowik, and Ł. Maziarka. Spatial graph convolutional networks. In *International Conference on Neural Information Processing*, pages 668–675. Springer, 2020.

[6] X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, and S.-Q. Liu. Insights into protein–ligand interactions: mechanisms, models, and methods. *International journal of molecular sciences*, 17(2):144, 2016.

[7] M. Geiger, A. Jacot, S. Spigler, F. Gabriel, L. Sagun, S. d'Ascoli, G. Biroli, C. Hongler, and M. Wyart. Scaling description of generalization with number of parameters in deep learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(2):023401, 2020.

[8] Z. Hao, S. Liu, Y. Zhang, C. Ying, Y. Feng, H. Su, and J. Zhu. Physics-informed machine learning: A survey on problems, methods and applications. *arXiv preprint arXiv:2211.08064*, 2022.

[9] A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees, and C. H. Reynolds. The role of ligand efficiency metrics in drug discovery. *Nature reviews Drug discovery*, 13(2):105–121, 2014.

[10] Z. Jin, T. Wu, T. Chen, D. Pan, X. Wang, J. Xie, L. Quan, and Q. Lyu. Capla: improved prediction of protein–ligand binding affinity by a deep learning approach based on a cross-attention mechanism. *Bioinformatics*, 39(2):btad049, 2023.

[11] W. L. Jorgensen. Efficient drug lead discovery and optimization. *Accounts of chemical research*, 42(6):724–733, 2009.

[12] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873): 583–589, 2021.

[13] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[14] S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou, and H. 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*, pages 975–985, 2021.

[15] S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou, and H. Xiong. Giant: Protein-ligand binding affinity prediction via geometry-aware interactive graph neural network. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[16] Y. Li, L. Han, Z. Liu, and R. Wang. Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results. *Journal of chemical information and modeling*, 54(6):1717–1736, 2014.

[17] J. Lim, S. Ryu, K. Park, Y. J. Choe, J. Ham, and W. Y. 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.

[18] E. Lionta, G. Spyrou, D. K Vassilatis, and Z. Cournia. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry*, 14(16):1923–1938, 2014.

[19] M. Liu, Z. Wang, and S. Ji. Non-local graph neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 44(12): 10270–10276, 2021.

[20] Z. Liu, M. Su, L. Han, J. Liu, Q. Yang, Y. Li, and R. Wang. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of chemical research*, 50(2):302–309, 2017.

[21] Ł. Maziarka, T. Danel, S. Mucha, K. Rataj, J. Tabor, and S. Jastrzębski. Molecule attention transformer. *arXiv preprint arXiv:2002.08264*, 2020.

[22] R. Meli, G. M. Morris, and P. C. Biggin. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: A review. *Frontiers in bioinformatics*, 2:885983, 2022.

[23] C. Meng, S. Seo, D. Cao, S. Griesemer, and Y. Liu. When physics meets machine learning: A survey of physics-informed machine learning. *arXiv preprint arXiv:2203.16797*, 2022.

[24] S. Moon, W. Zhung, S. Yang, J. Lim, and W. Y. Kim. Pignet: a physics-informed deep learning model toward generalized drug–target interaction predictions. *Chemical Science*, 13(13):3661–3673, 2022.

[25] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh. Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021.

[26] M. D. Parenti and G. Rastelli. Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnology advances*, 30(1):244–250, 2012.

[27] P. Ramachandran, B. Zoph, and Q. V. Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.

[28] P. Sharma, W. T. Chung, B. Akoush, and M. Ihme. A review of physics-informed machine learning in fluid mechanics. *Energies*, 16(5):2343, 2023.

[29] R. D. Smith, J. B. Dunbar Jr, P. M.-U. Ung, E. X. Esposito, C.-Y. Yang, S. Wang, and H. A. Carlson. Csar benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *Journal of chemical information and modeling*, 51(9):2115–2131, 2011.

[30] Y. Song, S. Zheng, Z. Niu, Z.-H. Fu, Y. Lu, and Y. Yang. Communicative representation learning on attributed molecular graphs. In *IJCAI*, volume 2020, pages 2831–2838, 2020.

[31] S. Spigler, M. Geiger, S. d'Ascoli, L. Sagun, G. Biroli, and M. Wyart. A jamming transition from under-to over-parametrization affects generalization in deep learning. *Journal of Physics A: Mathematical and Theoretical*, 52(47):474001, 2019.

[32] H. Stärk, O. Ganea, L. Pattanaik, R. Barzilay, and T. Jaakkola. Equibind: Geometric deep learning for drug binding structure prediction. In *International conference on machine learning*, pages 20503–20521. PMLR, 2022.

[33] M. M. Stepniewska-Dziubinska, P. Zielenkiewicz, and P. Siedlecki. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics*, 34(21):3666–3674, 2018.

[34] M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, Y. Li, and R. Wang. Compar-

[35] M. Varadi and S. Velankar. The impact of alphafold protein structure database on the fields of life sciences. *Proteomics*, 23(17):2200128, 2023.

[36] D. D. Wang, W. Wu, and R. Wang. Structure-based, deep-learning models for protein-ligand binding affinity prediction. *Journal of Cheminformatics*, 16(1):2, 2024.

[37] H. Wang. Prediction of protein–ligand binding affinity via deep learning models. *Briefings in Bioinformatics*, 25(2):bbae081, 2024.

[38] R. Wang, L. Lai, and S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of computer-aided molecular design*, 16:11–26, 2002.

[39] H.-J. Woo and B. Roux. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences*, 102(19):6825–6830, 2005.

[40] Z. Wu, P. Jain, M. Wright, A. Mirhoseini, J. E. Gonzalez, and I. Stoica. Representing long-range context for graph neural networks with global attention. *Advances in Neural Information Processing Systems*, 34:13266–13279, 2021.

[41] L. Zheng, J. Fan, and Y. Mu. Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS omega*, 4(14):15956–15965, 2019.

ative assessment of scoring functions: the casf-2016 update. *Journal of chemical information and modeling*, 59(2):895–913, 2018.