EGGNet, a generalizable geometric deep learning framework for protein complex pose scoring

Zichen Wang[1], Ryan Brand[1], Jared Adolf-Bryfogle[2], Jasleen Grewal[1], Yanjun Qi[1], Steven A. Combs[2], Nataliya Golovach[2], Rebecca Alford[2], Huzefa Rangwala[1*] and Peter M. Clark[2*]

[1]Machine Learning Solutions Lab, Amazon, New York, 10001, NY, USA.
[2]Janssen Biotherapeutics, The Janssen Pharmaceutical Companies of Johnson & Johnson, Spring House, 19002, PA, USA.

*Corresponding author(s). E-mail(s): rhuzefa@amazon.com; pclark3@its.jnj.com;
Contributing authors: zichewan@amazon.com; brandry@amazon.com; jadolfbr@its.jnj.com; gjaslee@amazon.com; yanjunqi@amazon.com; scombs1@its.jnj.com; ngolovac@its.jnj.com; ralford@its.jnj.com;

Abstract

Computational prediction of molecule-protein interactions has been key for developing new molecules to interact with a target protein for therapeutics development. Literature includes two independent streams of approaches: (1) predicting protein-protein interactions between naturally occurring proteins and (2) predicting the binding affinities between proteins and small molecule ligands (aka drug target interaction, or DTI). Studying the two problems in isolation has limited computational models' ability to generalize across tasks, both of which ultimately involve non-covalent interactions with a protein target. In this work, we developed Equivariant Graph of Graphs neural Network (EGGNet), a geometric deep learning framework for molecule-protein binding predictions that can handle three types of molecules for interacting with a target protein: (1) small molecules, (2) synthetic peptides and (3) natural proteins.

EGGNet leverages a graph of graphs (GoGs) representation constructed from the molecule structures at atomic-resolution and utilizes a multi-resolution equivariant graph neural network (GNN) to learn from such representations. In addition, EGGNet gets inspired by biophysics and makes use of both atom- and residue-level interactions, which greatly improve EGGNet's ability to rank candidate poses from blind docking. EGGNet achieves competitive performance on both a public protein-small molecule binding affinity prediction task (80.2% top-1 success rate on CASF-2016) and an synthetic protein interface prediction task (88.4% AUPR). We envision that the proposed geometric deep learning framework can generalize to many other protein interaction prediction problems, such as binding site prediction and molecular docking, helping to accelerate protein engineering and structure-based drug development.

Keywords: Drug discovery, Molecular interactions, Geometric deep learning, Graph of graphs

1 Introduction

Protein-protein interactions (PPIs) are key to many fundamental biological processes. Proteins mostly perform their functions via non-covalent interactions with three kinds of molecules including proteins, nucleotides and small molecules. The mechanisms of actions for most drugs involve interacting with protein targets to modulate their biological functions and activities. Being able to design drugs, either small molecules or biologics, to selectively bind a protein target with a desirable affinity is critically important for structure-based drug design.

The ability to modulate PPIs with small molecules or synthetic peptides is a component of therapeutics development. There exist four classes of problems in structure-based drug design for which molecular interactions can be tackled by machine learning (ML) [1], including (i) protein complex property prediction, (ii) binding site/interface identification, (iii) docking (binding pose generation), and (iv) *de novo* design. Property prediction for protein complexes is a task that can also serve as an integral component for other tasks, such as evaluating binding poses generated by docking algorithms or predicting the affinity for computationally designed novel drug candidates. Many ML methods have been developed for two popular groups of property prediction tasks, PPIs and drug-target interactions (DTIs), with the structure of the protein complex provided as input. However, to the best of our knowledge, none of the existing methods can be used across both problems. In this work, we developed a unifying geometric deep learning (GDL) framework for protein complex pose scoring, which encompasses both DTI affinity prediction and PPI interface prediction.

The proposed approach provides a generalizable and scalable representation of macromolecular complexes that can efficiently represent both protein-small

molecule complexes and protein-protein complexes. We also desire the representation to be capable of handling synthetic peptides composed of both natural and non-canonical amino acids. 3D graphs are commonly used to represent the 3D structures of individual proteins and protein-small molecule complexes [1]. To increase its generalizbility and scalability, we extend the 3D graph to 3D graph of graphs (GoG) representation.

GoG, also known as Network of Networks (NoN), is a special graph where the nodes in the top-level graph are also graphs [2, 3]. GoGs can be constructed naturally by connecting independent lower-level graphs by pre-defined associations [4], or similarity measured by a graph kernel [5]. GoG can also emerge from partitioning a large graph into different subgraphs based on topology or pre-defined rules. Message-passing GNNs have also been extended to operate on both levels of graphs in GoG [4–6]. The GoG data structure has been used to represent drug-drug interaction networks and metabolite networks [4, 6]. However, it has not been applied to model the 3D structures of macromolecules.

In this study, our contributions are two-fold: (i) we developed a featurization procedure to use the GoG data structure to efficiently unify the representations of all types of molecules, including small molecules, intermediate molecules such as peptides, and macromolecules. The 3D GoG representation retains both atomic- and residue-level information of molecular complexes. (ii) We developed EGGNet, an end-to-end geometric deep learning architecture based on an equivariant GNN to learn from the GoG representations of the 3D structures of protein complexes, optionally integrating physics-informed inductive bias to learn atomic-level interactions [7]. Our architecture can be used to predict both DTIs and PPIs. Notably, we achieved the state-of-the-art protein-small molecule binding affinity predictive performance. We further analyzed the effects of different choices of lower-level molecule graph models and evaluated the potential of transfer learning between the DTI and PPI prediction tasks. We also showed that our model improves the outputs of blind docking models.

2 Related work

Prior approaches for structure-based prediction of DTIs and PPIs with ML represent 3D structures of protein complexes using three common representations of protein structures based on 3D grids, 3D surfaces, and 3D graphs [1]. Among them, the 3D graph representation is the only one that preserves all of the information from the input protein structure.

For DTI prediction, all three types of representations and some of their combinations have been explored. For instance, KDEEP [8] estimates binding affinities by representing the protein-ligand complex as a 3D grid and learns from this representation using a 3D convolutional neural network (CNN). PotentialNet [9] builds 3D graphs for the protein-ligand complex using their atoms as nodes and chemical bonds and non-covalent interactions as edges, and then leverages message passing graph neural networks (GNN) [10] to learn

from the 3D heterogenous graph. PIGNet [7] extended PotentialNet by adding physics-informed inductive bias to a gate-augmented graph attention network (GAT) [11]. The physics-informed inductive bias is encoded by parameterized energy equations calculating a few non-covalent forces based upon the corresponding inter-atomic distances. HoloProt [12] combines the graph and surface representations of proteins. These studies decompose amino acid residues into atoms and bonds, making it difficult to incorporate potentially useful residue-level features for the ML model to learn from, such as embeddings from protein language models [13–16].

3D graph representations that capture the structural information at atomic resolution are more costly for PPIs than DTIs; therefore, 3D grids are the most prevalent representation of 3D protein complexes in protein-protein interface prediction. For instance, DeepRank [17] is a 3D CNN-based deep learning method that first maps the amino acid residues at the protein-protein interface to a 3D grid centered on the interface. More recently, Deep Local Analysis (DLA) [18] extends such 3D grid representations to an ensemble of 3D grids.

Table 1 Table of notations.

Symbol	Definition
$G = (V, E)$	an atomic graph of a molecule, with atoms as nodes $atom \in V$ and covalent bonds as edges $bond \in E$.
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	a GoG where the node set is composed of lower-level graphs $\mathcal{V} = \{G_1, G_2, \ldots, G_n\}$
$f_\theta : G \to \mathbb{R}^d$	a residue featurizer parameterized by θ, mapping an atomic graph to a vector
$\mathbf{s} \in \mathbb{R}^m$	scalar features on a node or an edge of a graph
$\mathbf{x} \in \mathbb{R}^{\nu \times 3}$	vector features on a node or an edge of a graph
$\mathbf{h} = (\mathbf{s}, \mathbf{x})$	features on a node or an edge of a graph
$\mathbf{H}^{\mathcal{V}} \in \mathbb{R}^{n \times (m+d)}, \mathbb{R}^{n \times \nu \times 3}$	all the node features on a GoG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
$\mathbf{H}^{\mathcal{E}} \in \mathbb{R}^{\lvert\mathcal{E}\rvert \times m'}, \mathbb{R}^{\lvert\mathcal{E}\rvert \times \mu \times 3}$	all the edge features on a GoG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
$\mathbf{Z} \in \mathbb{R}^{n \times d'}$	node embeddings on the higher-level graph of a GoG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
$f_\phi : \mathcal{G} \to \mathbb{R}^{n \times d'}$	a GVP-GNN parameterized by ϕ
$E^{(\cdot)} \in \mathbb{R}$	energy

3 Methods

3.1 Overview of the EGGNet approach

Overall, EGGNet takes the a 3D structure of a protein complex (also known as pose) as the input and makes predictions about the global properties of the pose, such as the binding affinity between two molecules in the protein complex. We refer to this problem as protein complex pose scoring. The input protein complex is comprised of a protein molecule and an interaction partner that takes one of the three types: (1) small molecules, (2) synthetic

peptides and (3) natural proteins. EGGNet first converts the input protein complex pose into a GoG to unify the representation of different types of molecules (Section 3.2). Next, EGGNet's architecture can learn from the GoG representation by coupling equivariant message-passing GNN with different lower-level molecule-to-vector methods (Section 3.3). Further, EGGNet incorporates physics-informed energy inductive biases (Section 3.4).

3.2 GoG representation of molecule complexes

Formally, a GoG is a higher-level graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set is composed of lower-level graphs $\mathcal{V} = \{G_1, G_2, \ldots, G_n\}$. The lower-level graph is used to represent the molecule's building blocks at atomic resolution, hence, $G_i = (V_i, E_i)$ is a graph with atoms as nodes $atom \in V$ and covalent bonds as edges $bond \in E$. For convenience we represent all notations used in this paper in Table 1.

Given a molecular structure, one can directly construct the atomic resolution graph for the entire molecule by connecting atoms with covalent bonds. We denote this graph as $G = (V, E)$. To construct a GoG from G, we first perform edge-cut graph partitioning to partition V into disjoint subsets $V_1 \cup \cdots \cup V_n = V$ of all the atoms $v_u \in V$ in graph G, resulting in n subgraphs, each of which corresponds to a lower-level graph $G_i = (V_i, E_i)$. We perform the graph partition by cutting the molecular graph at peptide bonds (Fig 1). Next, we construct the higher-level graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ using k-nearest neighbor (kNN) algorithm. We measure the Euclidean distance between the lower-level graphs using the C-alpha or the geometric centroids of heavy atoms, if the subgraph is not an amino acid residue.

With this procedure, we can unify the representations of small molecules and macromolecules: small molecules is a special case of GoG where the higher-level graph is composed of a single node (i.e. $\mathcal{G} = (\mathcal{V} = \{G_1\}, \mathcal{E})$). The GoG representation has the advantage over one-hot encoding with fixed vocabulary commonly used in protein language models [13–16], because it can generalize to non-canonical amino acids such as penicillamine used in synthetic peptides.

We also use GoG to represent a protein in complex with other molecules, including small molecules, peptides, or proteins. To do that, we simply apply the edge-cutting graph partition procedure to every polymer chain to derive the lower-level graphs, then construct the higher-level graph by using kNN to connect spatially proximal lower-level graphs (Fig 2).

3.3 Model architectures for protein complex pose scoring

This section describes the novel deep learning architectures for protein complex pose scoring. The overall architectures for EGGNet are shown in Fig.3

We formulate the protein complex pose scoring problem as a supervised graph-level prediction task. The goal of this task is to learn a function $\hat{y} = f(\mathcal{G})$ mapping a GoG $\mathcal{G}$ representing the protein complex's structure (pose) to a

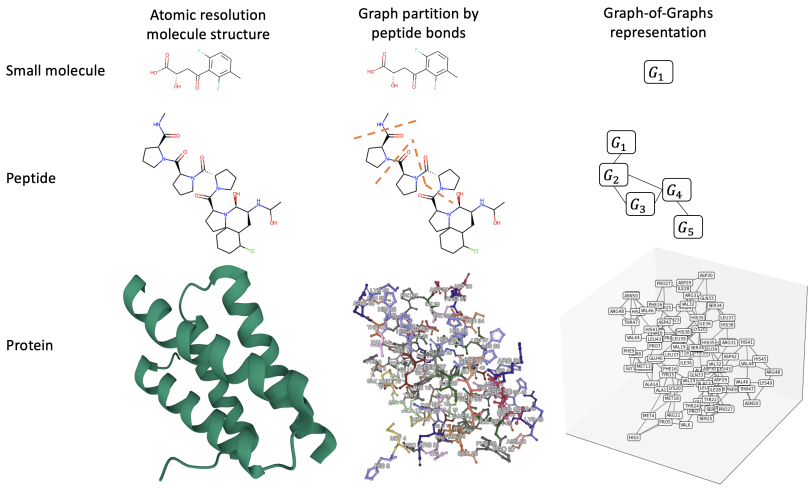

Fig. 1 Graph-of-Graphs (GoGs) representation of molecular structures. To construct GoG from molecular structure (left column), we first perform graph partition by cutting peptide bonds (middle column), then connect subgraphs using kNN, leading to a higher-level graphs (right column). The figure shows how GoG representation can generalize to small molecules, peptides with non-canonical amino acid residues, and proteins

scalar value y, representing the global property of the pose such as the binding affinity between the molecules.

Let $\mathbf{s}_i^{residue} \in \mathbb{R}^d$ be the feature vector extracted from the lower-level graph/residue. There are many possible choices of the residue featurizer $f_\theta : G \rightarrow \mathbb{R}^d$ including 1) chemical fingerprints such as MACCS and ECFP/Morgan [19], 2) GNN models trained on small molecule graphs, and 3) language models trained on the SMILES strings of small molecules such as MolT5 [20].

We denote the node features on higher-level graph as $\mathbf{h}_i^{\mathcal{G}} = (\mathbf{s}_i^{\mathcal{G}}, \mathbf{x}_i^{\mathcal{G}}) \in \mathbb{R}^{m+d}, \mathbb{R}^{\mu\times 3}$, for simplicity, we denote the node and edge features as $\mathbf{H}^{\mathcal{V}} \in \mathbb{R}^{n\times(m+d)}, \mathbb{R}^{n\times\nu\times 3}$ and $\mathbf{H}^{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|\times m'}, \mathbb{R}^{|\mathcal{E}|\times\mu\times 3}$, respectively.

GVP-GNN [21] is a $SE(3)$-equivariant GNN in which all node and edge embeddings are tuples $(\mathbf{s}, \mathbf{x})$ of scalar features $\mathbf{s} \in \mathbb{R}^m$ and geometric vector features $\mathbf{x} \in \mathbb{R}^{\nu\times 3}$. GVP-GNN operates on the node and edge features from the higher-level graph:

$$\mathbf{Z} = f_\phi(\mathbf{H}^{\mathcal{V}}, \mathbf{H}^{\mathcal{E}}) \tag{1}$$

Next, we use a readout network/prediction head to take the learned equivariant node representations $\mathbf{Z} \in \mathbb{R}^{n\times d'}$ to make graph-level prediction:

$$\hat{y} = READOUT(\mathbf{Z}) \tag{2}$$

We also design a multi-stage variant of our model architecture to have two GVP-GNNs to learn from two GoGs representing the two interacting molecules $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$, which are two subgraphs of the complex GoG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$ and $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{int}$. $\mathcal{E}_{int}$ denotes the edges representing

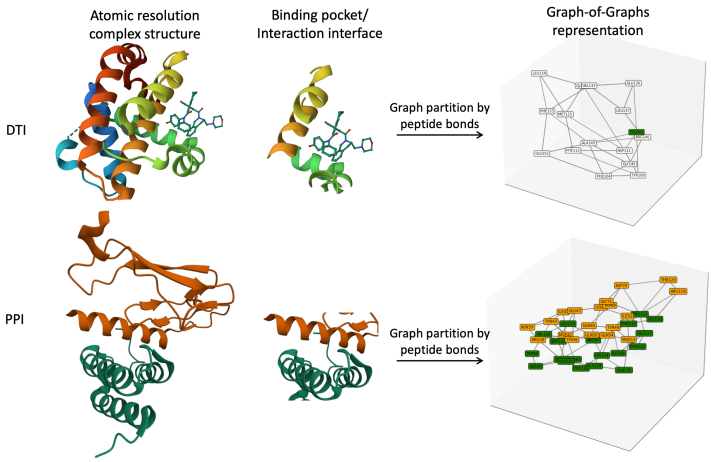

Fig. 2 Graph-of-Graphs (GoGs) representation of protein complexes. To construct GoG from the structures of protein complexes (left column), we first identify the binding pocket or interaction interface between molecules (middle column), then perform graph partition by cutting peptide bonds followed by connecting subgraphs using kNN, leading to a higher-level graphs (right column). The nodes in the higher-level graphs of GoGs are colored by the origin of the molecules. The figure shows how GoG representation can generalize to drug-target interaction (DTI) and protein-protein interaction (PPI).

intermolecular interactions.

$$\mathbf{Z}_1 = f_{\phi_1}(\mathbf{H}^{\mathcal{V}_1}, \mathbf{H}^{\mathcal{E}_1}) \tag{3}$$

$$\mathbf{Z}_2 = f_{\phi_2}(\mathbf{H}^{\mathcal{V}_2}, \mathbf{H}^{\mathcal{E}_2}) \tag{4}$$

The second stage GVP-GNN then computes node representations:

$$\mathbf{Z} = f_{\phi_3}(concat(\mathbf{Z}_1, \mathbf{Z}_2), \mathbf{H}^{\mathcal{E}_{int}}) \tag{5}$$

3.4 Objective functions and model training

We use mean squared error (MSE) and binary cross-entropy for binding affinity regression and binary interaction prediction tasks, respectively. Additionally, we adopted the physics-informed energy inductive biases from PIGNet [7]. Briefly, we use an energy decoder as the readout function to approximate 4 types of non-covalent interaction energies (van der Waals interactions, hydrogen bonds, metal-ligand interactions, and hydrophobic interactions) from a protein complex using parameterized equations.

$$\hat{E} = EnergyDecoder(\mathbf{Z}) \tag{6}$$

$$\mathcal{L} = MSE(\hat{E}, y) \tag{7}$$

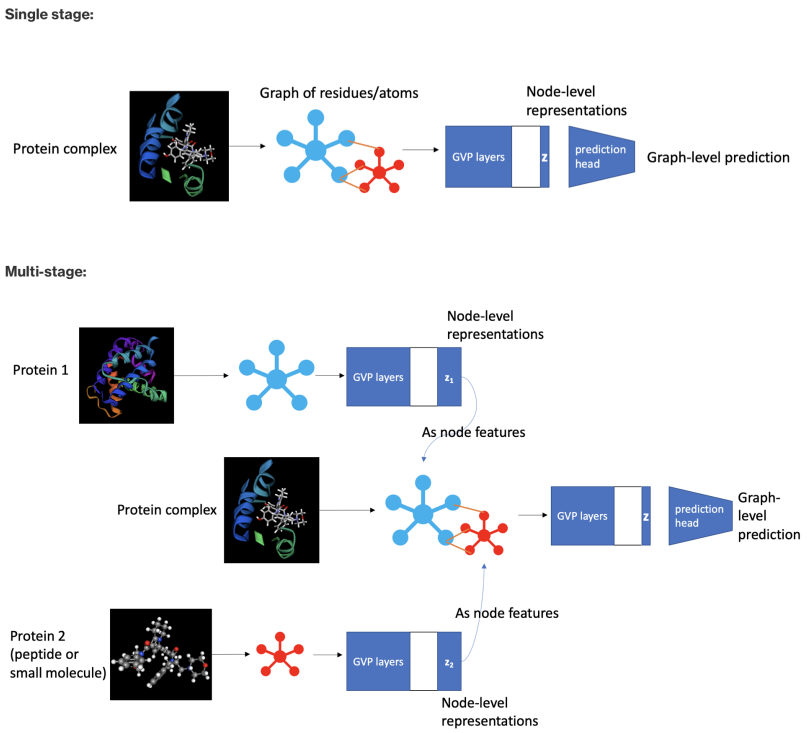

Fig. 3 Model architectures. Top: single-stage EGGNet. Bottom: multi-stage EGGNet.

All the parameterized equations takes the atom-level representation $\mathbf{z}_i, \mathbf{z}_j \in \mathbb{R}^{d'}$ as inputs. For instance, van der Waals interaction takes the following form:

$$E^{vdW} = \sum_{i,j} c_{ij}[(\frac{d'_{ij}}{d_{ij}})^{12} - 2(\frac{d'_{ij}}{d_{ij}})^{6}] \tag{8}$$

where $d_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\|$ and $d'_{ij} = r_i + r_j + MLP([\mathbf{z}_i, \mathbf{z}_j])$. r_i denotes the radii or the i-th node.

We also decompose the energy from the whole protein complex to intramolecular (monomer) and intermolecular energies $E = E^{(1)} + E^{(2)} + E^{int}$ as novel physics inductive bias.

We use stochastic gradient descent (SGD) to learn the parameters ϕ, and optionally the residue featurizer parameter θ, to optimize the objective function. When using SGD to learn both ϕ and θ, we essentially allow the end-to-end training of the lower-level residue featurizer and the higher-level GVP-GNN. Similar end-to-end joint training of models operating at different data modalities has also been used for protein function prediction [22].

3.5 Datasets and experiments

We use PDBbind [23] and ProtCID [24] datasets for training and validation of our models. The two datasets corresponds two otherwise distinctive tasks, DTI and PPI prediction, respectively, that can be unified by our approach.

For the DTI binding affinity regression task, we used the same setup of the PDBbind and CASF-2016 [25] datasets as the PIGNet study [7]. Briefly, the PDBbind 2019 refined set provides quantified binding affinity data (in pKd, where Kd is the experimentally measured dissociation constant) and corresponding structure of protein–ligand complexes deposited in the protein data bank (PDB) [26]. We used 4514 samples for the training set. For hold-out evaluation, we used the CASF-2016 benchmark dataset, which originated from the PDBbind 2016 core set. We used 283 samples for the scoring and ranking, 22,340 samples for the docking benchmark. All protein-ligand complexes were processed, to focus on the pocket-ligand structure, which removes amino acid residues whose minimum distance between the ligand is greater than 5Å to remove amino acid residues distant to the protein's binding pocket. We adopt the evaluation metrics recommended in the CASF-2016 dataset [25]. Briefly, we calculate the Spearman's correlation coefficient *rho* as the evaluate metric for the ranking setting in CASF-2016. For the docking setting, we used the predicted binding affinities to rank candidate poses from each protein-ligand pair, then define a success docking if the top-1 ranked pose has the lowest root-mean-square deviation (RMSD) among candidate poses (decoys). We then compute the top-1 success rate across all protein-ligand pairs as the fraction of successful docking from all docking experiments in the CASF-2106 benchmark.

ProtCID dataset is used for the binary binding classification task. Here, the aim is to distinguish physiological binders with non-physiological ones. Physiological binders were defined as those homo- and hetero-dimers which had more than 5 crystal forms of a Uniport ID in the PDB. If an interface is only seen in one crystal form of this UniProt ID in addition to not being in any common cluster, then it is likely to be non-physiological. For data preprocessing, we remove amino acid residues whose C-beta distance is greater than 6Å to any amino acids from the protein chain of the binding partner. We also removed large interfaces with more than 10^5 atoms, leading to 15,736 and 3,889 protein interfaces for training set and hold-out test set, respectively. The train/test split was performed by sequence clusters resulting from MMseq2 [27] with identity cutoff of 60%.

We evaluate the model performance on the binary classification task using the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR).

For EGGNet models trained on both datasets, we set the GVP-GNN to have 3 GVP convolutional layers with node and edge hidden dimensions to be (200, 32) and (64, 2), respectively. We used the following hyperparameters for model training: learning rate of 1e-4, per-GPU batch size of 16. To avoid overfitting, we employed an early stopping criterion with patience=50 epochs and trained for a maximum of 1000 epochs. ADAM [28] optimizer with $\beta_1 =$

0.9, $\beta_2 = 0.999$ is used for optimizing the learnable parameters. All models are implemented using the Pytorch deep learning library [29] and training are performed using Pytorch-Lightning library with 16-bit mixed precision training using 4 NVIDIA V100 GPUs with 16GB of memory each on Amazon SageMaker.

4 Experimental Results

4.1 GoG representation of molecules allows flexible transfer learning of molecule-to-vector (Mol2Vec) methods

In this work, we leverage the GoG data structure to represent the 3D structures of small- and macro- molecules. Macro-molecules, such as proteins and nucleic acids, are composed of small molecule building blocks (amino acids and nucleotides) connected by covalent bonds. Inspired by this observation, we view a macro-molecule as a graph of small molecule residues, which can be represented by graph of atoms connected by chemical bonds. The GoG representation of molecules provides the flexibility in modeling macro-molecules with out-of-vocabulary residues such as post-translational modifications (PTMs) and non-canonical amino acids. It also unifies the chemical space of small- and macro-molecules by applying Mol2Vec methods to the lower-level graphs.

We first demonstrate that EGGNet is able to take advantage of any Mol2Vec methods on the GoG representation of protein in complex with small molecule ligands. We experimented with three classes of Mol2Vec methods: (i) chemical fingerprint approaches including MACCS keys and Morgan fingerprints [19], (ii) GNNs pretrained on a small molecule chemical library using structures and properties, and (iii) language models pretrained on SMILES strings of small molecules. We evaluated the performance of single-stage EGGNet with different Mol2Vec methods on the PDBbind/CASF-2016 binding affinity regression task. Our results show that Mol2Vec methods based on deep learning such as Graph Isomorphism Network (GIN) [30] and MolT5 [20] achieved better performance than chemical fingerprints (Table 2). We also observe that EGGNet with the MolT5 model achieves an improved docking success rate than with GIN, suggesting the more powerful pretrained Mol2Vec model with can further boost the overall model performance. However, MolT5 models (MolT5-small: ~77M parameters, MolT5-base: ~250M parameters) are significantly larger than GIN (~1.8M parameters), we use GIN for the later experiment for its training efficiency.

Table 2 Residue-level representation models improve binding affinity prediction. The following experiments use GVP backbone with energy inductive biases and identical training hyperparameters and setups, except for the residue-level model.

Mol2Vec methods	Docking top1 success rate	Ranking ρ
Baseline (zero vector)	0.802	0.675
MACCS	0.820	**0.727**
Morgan/ECFP4	0.806	0.670
GIN	0.816	0.715
MolT5-small	0.760	0.702
MolT5-base	0.813	0.715
MolT5-small with joint training	**0.827**	0.722

4.2 Performance of EGGNet on protein-molecule binding prediction benchmarks

Next, we evaluated EGGNet on two distinctive protein complex pose scoring tasks, including protein-small molecule binding affinity prediction and protein-protein interface classification.

4.2.1 Protein small molecule binding affinity regression

On the binding affinity prediction task, we train our model on PDBbind and evaluated on CASF-2016 ranking and docking benchmarks. Our results show that EGGNet with joint training and energy inductive biases outperforms PIGNet [7] in the same setting on both docking and ranking benchmarks (Table 3). We found our best performing model on the docking setting achieved 80.2% top 1 success rate over 77.4% reported in PIGNet.

Interestingly, we found the joint training of the GNNs at both the lower-level and higher-level graphs of the GoG improves the models' performance on the ranking task, while slightly decreasing performance on the docking task. The ranking task examines a model's ability to generalize to different combinations of different protein-ligand pairs, which corresponds to different composition and geometry of the GoGs. Meanwhile, the docking task probes the models' ability to rank candidate poses of the same protein-ligand pair, which corresponds to GoGs with the same composition of lower-level graphs but different geometries. Therefore, we reason that the model with the ability to adjust the embeddings of the lower-level residue representations are only beneficial to tasks that prioritize in distinguishing the composition, rather than the geometry of the GoGs. This also explains the observation that the multi-stage variant of EGGNet is more performant on the docking setting (Table 3). After all, the multi-stage variant has more learnable parameters focusing on the inter-molecular geometries.

We also confirmed the findings from Moon *et al* [7] that using the non-covalent interaction energy as inductive bias for the deep learning model can significantly improve binding affinity prediction. Without the energy inductive

Table 3 Model performances on the CASF-2016 binding affinity regression task.

Model	Lower-level GNN	Model backbone	Energy inductive bias	Docking top1 success rate	Ranking ρ
Ours	GIN	GVP	None	0.230±0.005	0.653±0.036
	GIN	MS-GVP	None	0.148±0.019	0.566±0.031
	GIN joint training	GVP	None	0.205±0.009	0.715±0.019
	GIN joint training	MS-GVP	None	0.200±0.023	0.690±0.009
	GIN	GVP	E	0.759±0.020	0.719±0.025
	GIN	MS-GVP	E	**0.802±0.020**	0.658±0.032
	GIN joint training	GVP	E	0.779±0.021	**0.765±0.023**
	GIN joint training	MS-GVP	E	0.788±0.007	0.760±0.015
3D GNN	None	GNN	E	0.299	0.604
PIGNet	None	GNN	E	0.774	0.672

bias, the model can only achieve 22.9% top-1 success rate, compared to 75.8% with the energy inductive bias (Table 3). As the energy E was calculated for the entire protein-ligand complex, we hypothesized that the inductive bias is more informative to the model if we decompose it to $E = E^{(1)} + E^{(2)} + E^{int}$. That is, for 2-molecule complex, the non-covalent interaction energy can be decomposed to the sum of intra-molecule energy and inter-molecule energy. However, our results on the PDBbind/CASF-2016 experiments do not support this hypothesis (Table A1). This is probably due to the inaccurate estimation of the protein's energy $E^{(1)}$ because we had to use the binding pocket rather than the entire protein structure to construct the GoG to make the training memory efficient.

4.2.2 Protein-protein interface classification

Next, we evaluated EGGNet on another common type of protein complex pose prediction task: distinguishing whether a protein-protein interaction interface is physiological or an artifact of the crystallization process. This task can be useful to deduce solution-state quaternary states (dimers, trimers, etc.) or to identify novel functions of protein families that were originally deemed crystal artifacts [31]. In addition, models trained for this task could be fine-tuned for detection of binders and non/weak – binders for protein interface design. We train and validated our models on the ProtCID [24] dataset. We found different variants of our models are able to significantly outperform CAMP [32], which only takes the sequence information of the protein pairs (Table 4). We also found that the joint training of GNNs at different levels of the GoGs improve the predictive performance over freezing the lower-level GIN by at least 5.8% in AUROC.

The evaluation of the task is similar to the ranking setting of the PDBbind/CASF-2016, where the model is tasked to classify different protein-protein interaction pairs rather than the same protein-protein pair with

Table 4 Binary protein-protein interaction prediction on ProtCID.

Model	Lower-level GNN	Model backbone	AUROC	AUPR
Ours	GIN	GVP	0.681	0.875
	GIN	MS-GVP	0.672	0.869
	GIN joint training	GVP	0.739	0.878
	GIN joint training	MS-GVP	**0.746**	**0.884**
CAMP	None	CNN	0.500	0.800

Table 5 Binding affinity predicted by our model improves ranking of DiffDock pose candidates.

Method	Top-1 RMSD % < 2	Top-5 RMSD % < 2
DiffDock(40) confidence model	32.2	40.6
EGGNet predicted pKd	20.4	35.0
Rank product of model outputs	**32.8**	**40.9**

different poses. Therefore, we observed that the energy inductive bias is not helpful in improving the binary classification performance (Table A2).

Thanks to the uniform GoG representation of the protein complexes, our model architectures used for the protein-protein interface prediction and the protein-small molecule binding affinity prediction are identical. We next ask if model trained on the binding affinity prediction task can improve protein-protein interface prediction via transfer learning. However, our results show that transfer learning from the model trained to predict small molecule binding affinity doesn't help with this task compared to a model trained from scratch (Table A3). We speculate that this is due to the distributions of the input GoGs and the labels between these datasets are disparate.

4.3 EGGNet improves ranking of docking poses from DiffDock

Although our model is developed to predict the property of a protein complex using the 3D structure as input, we assess whether our model is helpful to tasks beyond property prediction, such as blind docking, another important task in structural-based drug discovery. Binding affinity prediction is the next step for blind docking, which generates poses for how a drug candidate would bind to a protein target.

We applied our model to predict the binding affinity on candidate poses generated by the state-of-the-art blind docking algorithm DiffDock [33]. DiffDock is a composed of a generative model to generate candidate poses for a small molecule with respect to a protein target, and a confidence model to rank the quality of each pose. We compared the rankings of candidate poses by DiffDock's confidence model, the predicted binding affinity from our model, as well as the combination of both using rank product (Table 5).

We found that although the predicted binding affinity alone from our model is not as good as prioritizing the best poses, combining the predicted binding affinity with DiffDock's confidence measures slightly improves the prioritization of poses, suggesting predicted binding affinity are helpful information in identifying the experimentally validated poses of protein-small molecule complexes. We also envision using our model architecture to jointly learn the RMSD and binding affinity of each candidate pose to further improve generative ML models of blind docking.

5 Discussion

In this work, we first developed a biologically inspired data structure GoG to represent the structures of molecules and molecular complexes. We construct GoGs by applying edge cutting graph partition on the atomic graphs of molecules, which is inspired by the fact that macromolecules are composed of small molecule residues connected by common covalent bonds such as peptide bonds for proteins and phosphodiester bonds for nucleic acids. Next, we developed EGGNet, an equivariant graph neural network to learn from the GoG representations of molecular complexes to predict their properties. The unifying representation allows EGGNet to perform both DTI and PPI prediction tasks, achieving competitive performances. We also demonstrated the value of EGGNet in improving results from blind docking.

Acknowledgments. We thank Sandeep Somani from Janssen and Tong Wang, Nkechinyere Agu from AWS for insightful discussions.

Data availability. The PDBBind/CASF-2016 data can be downloaded from https://zenodo.org/record/6047984.

Code availability. The source code for this study are available for research and non-commercial use at: https://github.com/aws-samples/eggnet-equivariant-graph-of-graph-neural-network

Appendix A Appendix

Table A1 Decomposing energy terms doesn't improve the model performances on the CASF-2016 binding affinity regression task. $E = [E^{vdW}, E^{hbond}, E^{metal}, E^{hydrophobic}]$

Energy inductive bias	Docking top1 success rate	Ranking ρ
E	**0.827**	**0.722**
$\mathbf{W}([E^{(1)}, E^{(2)}, E^{int}])$	0.583	0.638
$\mathbf{W}(BatchNorm([E^{(1)}, E^{(2)}, E^{int}]))$	0.442	0.677
$\mathbf{W}(concat([E^{(1)}, E^{(2)}, E^{int}]))$	0.474	0.694
$\mathbf{W}(BatchNorm(concat([E^{(1)}, E^{(2)}, E^{int}])))$	0.378	0.670

Table A2 Energy inductive bias doesn't help with the ProtCID task.

Energy inductive bias	AUROC	AUPR
None	**0.681**	**0.875**
E	0.405	0.719

Table A3 Transferring model weights learned from PDBBind to ProtCID.

Lower-level GNN	Model training	Model init	AUROC	AUPR
GIN	trained with lr=1e-4	Random	0.681	0.875
GIN	no training	PDBBind	0.333	0.691
GIN	trained with lr=1e-4	PDBBind	0.405	0.719
GIN joint training	trained with lr=1e-4	Random	**0.739**	**0.878**
GIN joint training	no training	PDBBind	0.345	0.695
GIN joint training	trained with lr=1e-4	PDBBind	0.713	0.859

References

[1] Isert, C., Atz, K. & Schneider, G. Structure-based drug design with geometric deep learning. *arXiv preprint arXiv:2210.11250* (2022) .

[2] D'Agostino, G. & Scala, A. *Networks of networks: the last frontier of complexity* Vol. 340 (Springer, 2014).

[3] Ni, J., Tong, H., Fan, W. & Zhang, X. *Inside the atoms: ranking on a network of networks*, 1356–1365 (2014).

[4] Harada, S. *et al.* Dual graph convolutional neural network for predicting chemical networks. *BMC bioinformatics* **21** (3), 1–13 (2020) .

[5] Wang, Y., Zhao, Y., Shah, N. & Derr, T. *Imbalanced graph classification via graph-of-graph neural networks*, 2067–2076 (2022).

[6] Wang, H., Lian, D., Zhang, Y., Qin, L. & Lin, X. Gognn: Graph of graphs neural network for predicting structured entity interactions. *arXiv preprint arXiv:2005.05537* (2020) .

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

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

[9] Feinberg, E. N. *et al.* Potentialnet for molecular property prediction. *ACS central science* **4** (11), 1520–1530 (2018) .

[10] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. *Neural message passing for quantum chemistry*, 1263–1272 (PMLR, 2017).

[11] Veličković, P. *et al.* Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017) .

[12] Somnath, V. R., Bunne, C. & Krause, A. Multi-scale representation learning on proteins. *Advances in Neural Information Processing Systems* **34**, 25244–25255 (2021) .

[13] Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M. & Church, G. M. Unified rational protein engineering with sequence-based deep representation learning. *Nature methods* **16** (12), 1315–1322 (2019) .

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

[15] Rives, A. *et al.* Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences* **118** (15), e2016239118 (2021) .

[16] Elnaggar, A. *et al.* Prottrans: towards cracking the language of life's code through self-supervised deep learning and high performance computing. *arXiv preprint arXiv:2007.06225* (2020) .

[17] Renaud, N. *et al.* Deeprank: a deep learning framework for data mining 3d protein-protein interfaces. *Nature communications* **12** (1), 1–8 (2021) .

[18] Mohseni Behbahani, Y., Crouzet, S., Laine, E. & Carbone, A. Deep local analysis evaluates protein docking conformations with locally oriented cubes. *Bioinformatics* **38** (19), 4505–4512 (2022) .

[19] Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling* **50** (5), 742–754 (2010) .

[20] Edwards, C., Lai, T., Ros, K., Honke, G. & Ji, H. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817* (2022) .

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

[22] Wang, Z. *et al.* Lm-gvp: an extensible sequence and structure informed deep learning framework for protein property prediction. *Scientific reports* **12** (1), 1–12 (2022) .

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

[24] Xu, Q. & Dunbrack, R. L. Protcid: a data resource for structural information on protein interactions. *Nature communications* **11** (1), 1–16 (2020) .

[25] Su, M. *et al.* Comparative assessment of scoring functions: the casf-2016 update. *Journal of chemical information and modeling* **59** (2), 895–913 (2018) .

[26] Berman, H. M. *et al.* The protein data bank. *Nucleic acids research* **28** (1), 235–242 (2000) .

[27] Steinegger, M. & Söding, J. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* **35** (11), 1026–1028 (2017) .

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

[29] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32** (2019) .

[30] Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018) .

[31] Xu, Q. *et al.* Identifying three-dimensional structures of autophosphorylation complexes in crystals of protein kinases. *Science signaling* **8** (405), rs13–rs13 (2015) .

[32] Lei, Y. *et al.* A deep-learning framework for multi-level peptide–protein interaction prediction. *Nature communications* **12** (1), 1–10 (2021) .

[33] Corso, G., Stärk, H., Jing, B., Barzilay, R. & Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776* (2022) .