HeMeNet: Heterogeneous Multichannel Equivariant Network for Protein Multi-task Learning

Rong Han[1], Wenbing Huang[2*], Lingxiao Luo[1], Xinyan Han[1], Jiaming Shen[1], Zhiqiang Zhang[3], Jun Zhou[3], Ting Chen[1*]

[1] BNRist, Department of Computer Science and Technology, Tsinghua University
[2] Gaoling School of Artificial Intelligence, Renmin University of China
[3] Ant Group CO., Ltd.
hanr21@mails.tsinghua.edu.cn, hwenbing@126.com, tingchen@tsinghua.edu.cn

Abstract

Understanding and leveraging the 3D structures of proteins is central to various biological and drug discovery tasks. While deep learning has been applied successfully for structure-based protein function prediction tasks, current methods usually employ distinct training for each task. However, each of the tasks is of small size, and such a single-task strategy hinders the models' performance and generalization ability. As some labeled 3D protein datasets are biologically related, combining multi-source datasets for larger-scale multi-task learning is one way to overcome this problem. In this paper, we propose a model to address multiple tasks jointly upon the input of 3D protein structures. In particular, we first construct a standard structure-based multi-task benchmark called Protein-MT, consisting of 6 biologically relevant tasks, including affinity prediction and property prediction, integrated from 4 public datasets. Then, we develop a novel graph neural network for multi-task learning, dubbed **He**terogeneous **M**ultichannel **E**quivariant **N**etwork (HeMeNet), which is E(3) equivariant and able to capture heterogeneous relationships between different atoms. Besides, HeMeNet can achieve task-specific learning via the task-aware readout mechanism. Extensive evaluations on our benchmark verify the effectiveness of multi-task learning, and our model surpasses state-of-the-art models.

Code — https://github.com/hanrthu/HeMeNet
Extended version — https://arxiv.org/abs/2404.01693

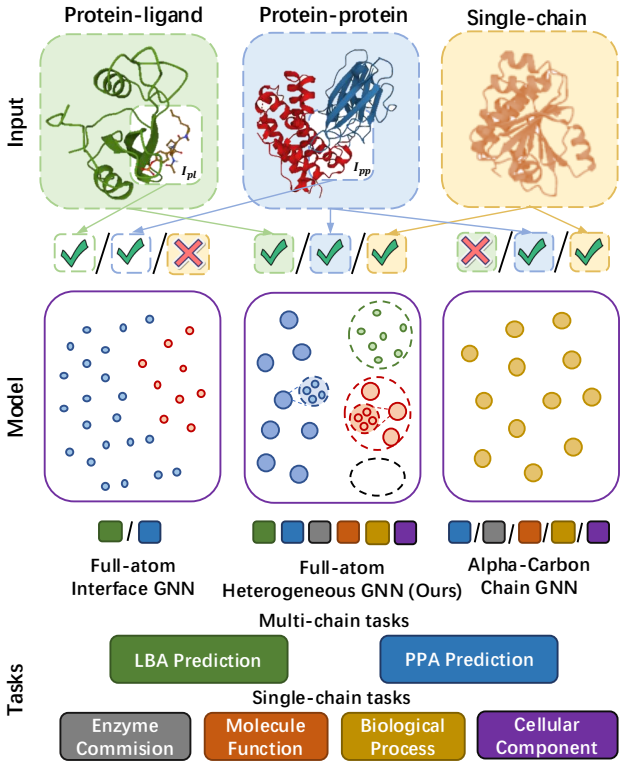

Figure 1: Comparison of different models with tasks.

Introduction

Proteins consist of one or more chains of amino acids, and they are vital in many biological systems. The 3D structure of a protein sets the foundation of its interaction with other molecules, which finally determines its functions. In recent years, learning-based methods have been applied widely to leverage the 3D structures of proteins for various tasks such as property prediction (Wang et al. 2023a), affinity prediction (Li et al. 2021), rigid docking (Ganea et al. 2022), and antibody generation (Kong, Huang, and Liu 2023a), owing to their superior efficiency and lower cost compared to those wet-lab approaches. A major part of learning-based methods resort to Graph Neural Networks (GNNs) (Xu et al. 2019), which naturally encode the 3D structures of proteins by modeling atoms or residues as nodes and the connections in between as edges. In addition, certain GNNs are geometry-aware and designed to capture the symmetry of E(3) transformations for better predictions (Satorras, Hoogeboom, and Welling 2021; Huang et al. 2022).

Despite significant progress in geometric-aware GNNs for protein tasks, existing methods usually employ one model for one task. A clear drawback of such a single-task training strategy is that the model should be re-trained for each new task. However, structural datasets with annotations are often limited in size due to the expensive cost of

*Corresponding authors.

acquiring protein 3D structures and labels via wet lab experiments, especially for affinity prediction. For example, in PDBbind (Wang et al. 2004), only 2852 complexes for the Protein-Protein Affinity (PPA) are experimentally annotated. Due to the sparsity of labeled structures, conducting model training on a single-task dataset of small size usually leads to defective performance and an inability to generalize.

To deal with the sparsity of labeled data issues, some previous works leverage multi-task learning, which designs a model and trains it with multiple related tasks. Collecting samples with related tasks can bring more information to improve the performance (Zhang et al. 2023a; Wang et al. 2022). However, most of these works are sequence-based (Xu et al. 2022; Capel et al. 2022), and the formulations are separated annotations for different samples. The work by(Capel, Feenstra, and Abeln 2022) is a structure-based multi-task method, but it mainly focuses on residue-level interface prediction.

Recent research shows that some protein properties potentially imply the protein's binding activity. For example, gene ontology would contain knowledge for protein-protein interaction (Wang et al. 2022); Enzyme commission and gene ontology can provide molecular context for protein-ligand binding affinity (LBA) (Zhang et al. 2023a). These indicate that some single-chain functions may benefit the prediction of complex-level affinity and vice versa. Motivated by this fact, we propose combining affinity and property prediction datasets in the framework of joint training.

A key problem hindering structural data integration is the lack of appropriate models for various inputs and tasks. As shown in Figure 1, many affinity prediction models (Li et al. 2021; Kong, Huang, and Liu 2023b) utilize full-atom information at the binding interface, which loses the information of the whole chain. While many function prediction models (Zhang et al. 2023b; Jing et al. 2021) utilize alpha-carbon to predict chain-level functions, which loses the detailed atom interaction information for affinity prediction.

In this paper, we propose a structure-based multi-task learning paradigm. We use a heterogeneous full-atom model for multiple tasks upon various 3D protein inputs. Nevertheless, accomplishing structural multi-task training is challenging. The first challenge is that there is no available benchmark. The ideal benchmark should cover a sufficient range of data and biologically related tasks, with a fully labeled test set to compare how a model performs on the same input for different task outputs. The second challenge is that it is nontrivial to design a generalist model that is capable of processing the complicated 3D structures of input proteins of various types, including single-chain, protein-protein, and protein-ligand, and it should perform well across different tasks, including protein affinity and property predictions. By achieving structure-based full-atom protein multi-task learning, we make the following contributions:

- To the best of our knowledge, we are the first to propose the concept of structure-based protein multi-task learning. We carefully integrate the structures and labels from 4 public datasets with our proposed standard process and construct a new benchmark named **Protein M**ultiple **T**asks (**Protein-MT**), which consists of 6 representative tasks upon 3 different types of inputs.
- We propose a novel model for protein structure learning, dubbed **He**terogeneous **M**ultichannel **E**quivariant Network (**HeMeNet**), which is E(3) equivariant and able to capture various relationships between different atoms owing to the heterogeneous multichannel graph construction of proteins. Additionally, we develop a task-aware readout mechanism by associating the output head of each task with a learnable task prompt for different tasks.
- For the experiments on Protein-MT, HeMeNet surpasses other state-of-the-art methods in most tasks under both the single-task and multi-task settings. Particularly on the LBA and PPA tasks, we find that the multi-task HeMeNet is significantly better than its single-task counterpart.

Related Works

Protein Interaction and Property Prediction Predicting the binding affinity and properties for proteins with computational methods is of growing interest (Wang et al. 2022; Zhao et al. 2020). Previous research learns protein representations by information different forms, most of which take amino acid sequence (Alley et al. 2019; Rao et al. 2019), multiple sequence alignment (Rao et al. 2021) or 3D structure (Hermosilla et al. 2021; Zhang et al. 2023b) as input. Many works encode the information of a protein's 3D structure by GNNs (Gligorijević et al. 2021a; Zhang et al. 2023b). (Li et al. 2021) take full-atom geometry at the interaction interface, and (Zhang et al. 2023b) take residue-level geometry of the protein for property prediction. Our method utilizes full-atom geometry on the whole protein to address affinity and property prediction tasks together.

Equivariant GNNs Many equivariant GNNs have emerged recently with the inductive bias of 3D symmetry, modeling various tasks including docking, molecular function prediction and sequence design (Gasteiger, Groß, and Günnemann 2020; Satorras, Hoogeboom, and Welling 2021; Zhang et al. 2024; Cen et al. 2024; Han et al. 2024). To empower the model with the ability to handle the complicated full-atom geometry, some models design multi-channel equivariant message passing for atom sets, such as GMN (Huang et al. 2022) and dyMEAN (Kong, Huang, and Liu 2023a). We propose a powerful heterogeneous equivariant GNN capable of handling various incoming message types.

Protein Multi-Task Learning Multi-task learning takes advantage of knowledge transfer across tasks, achieving a better generalization performance (Wu, Zhang, and Ré 2020). In the field of protein, several works leverage multi-task learning on the task of interaction prediction and property prediction, most of which are sequence-based. (Shi et al. 2023) design three Enzyme Commission number-related hierarchical tasks to train the model. (Wang et al. 2023b) introduce a multi-task protein pre-training method with prompts. (Xu et al. 2022; Capel et al. 2022) are sequence-based multi-task benchmarks for protein function prediction. (Capel, Feenstra, and Abeln 2022) improves the

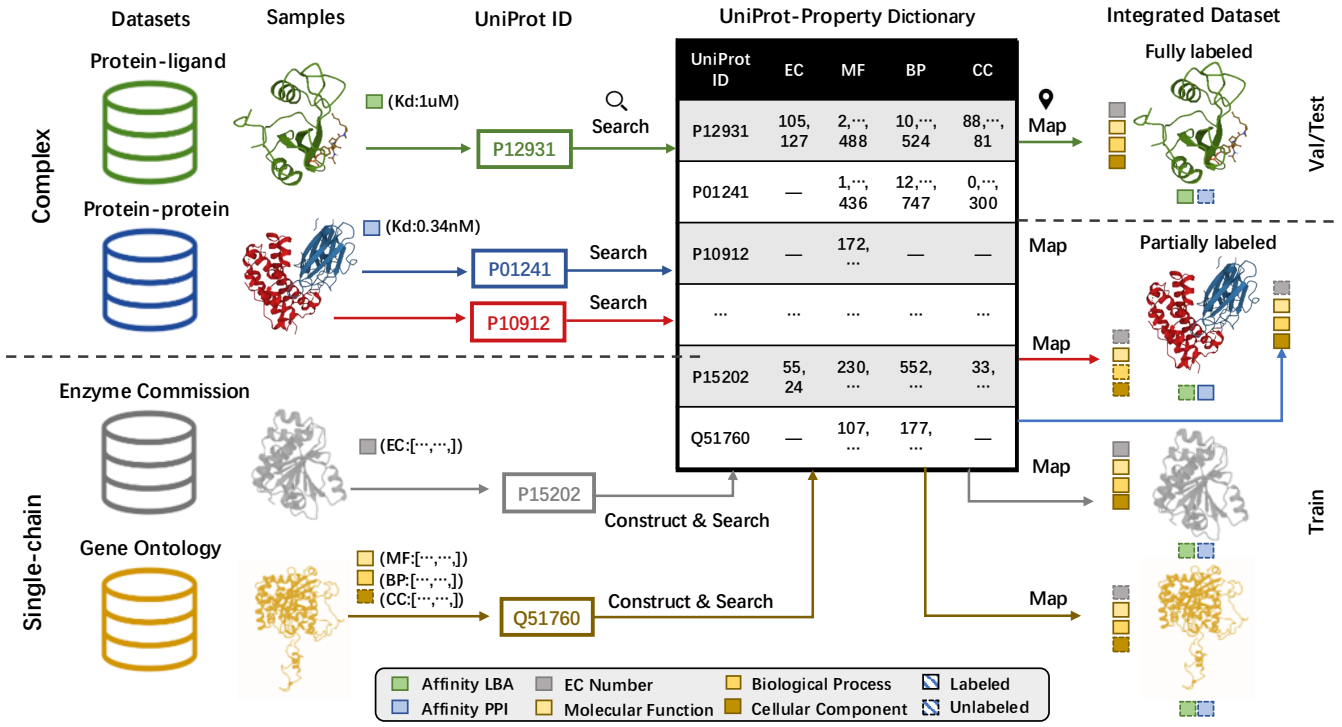

Figure 2: **Construction of Protein-MT**. We first extract the UniProt ID for each chain and construct a UniProt-Property dictionary to map each protein chain with functional labels. With this dictionary, we can extract each chain's UniProt ID and map it with its labels. The complex with one affinity label and all property labels for each chain is defined as fully labeled.

protein interaction interface prediction by structural multi-task auxiliary learning. To our knowledge, we are the first to combine structure-based interaction prediction and property prediction in a multi-task setting.

New Dataset: Protein-MT

Based on the observation that protein property and binding affinity tasks may benefit each other, we construct a new dataset called Protein Multiple Tasks (Protein-MT) for protein multi-task learning. Protein-MT is composed of different types of tasks on 3D protein structures: the prediction of Ligand Binding Affinity (**LBA**) and Protein-Protein Affinity (**PPA**) based on two-instance complexes and the prediction of Enzyme Commission (**EC**) number and Gene Ontology (**GO**) terms based on single-chain structures. Particularly, the LBA and PPA tasks originated from the PDBbind database (Wang et al. 2004) aim at regressing the affinity value of a protein-ligand complex and protein-protein complex, respectively. The EC task is constructed by (Gligorijević et al. 2021b) to describe the catalysis of biochemical reactions consisting of samples, each with 538 binary-class labels. The GO task aims to predict the hierarchically related functional properties of gene products (Gligorijević et al. 2021b): Molecular Function (MF), Biological Process (BP), and Cellular Component (CC). We treat the prediction of MF, BP, and CC as three individual tasks, resulting in six different prediction tasks in total.

One key difficulty in integrating these tasks from their sourced datasets is that samples from one task may lack the labels for others. It is crucial to obtain samples with a complete set of labels across tasks for the training and evaluation of multi-task learning methods. As shown in Figure 2, we propose a standard matching pipeline that enables transferring the labels between EC and GO, and assigning EC and GO labels for the chains of complexes in LBA and PPA as well (it is impossible to conduct the inverse direction since it is meaningless to assign LBA or PPA for those single chains in EC and GO). Specifically, we utilize the UniProt ID to uniquely identify a protein chain[1]. We first obtain the UniProt IDs of all protein chains in Protein-MT from Protein Data Bank (Berman et al. 2000). For each UniProt ID, we determine the EC and GO properties based on the labels of the corresponding chains in the corresponding datasets, resulting in a UniProt-Property dictionary. With this dictionary, for a chain missing EC or GO labels (e.g., a chain of a complex in LBA and PPA), we can supplement the labels by searching the dictionary by its UniProt ID to retrieve any known EC and GO labels. We define a complex (from either LBA or PPA) as fully labeled if the complex has one affinity label (LBA or PPA) and four function labels for each of its chains. After our above matching process, we formulate the train/validation/test split in terms of the chain-level sequence identity through the alignment methods commonly

[1]The UniProt dataset is the world's leading protein sequence and function dataset and it identifies proteins by their UniProt IDs.

used in single-chain property prediction tasks (Gligorijević et al. 2021b).

Methodology

We first introduce our heterogeneous graph representation and the multi-task formulation. Then, we design the architecture of the proposed HeMeNet, which consists of two key components: heterogeneous multi-channel equivariant message passing and task-aware readout.

Heterogeneous Graph and Task Formulation

The input of our model is of various types. It could be either a two-instance complex (protein-ligand for LBA and protein-protein for PPA) or a single chain (for EC and GO). Here, for consistency, we unify these two different kinds of input as a graph $\mathcal{G}$ composed of two sets of nodes $\mathcal{V}_r$ and $\mathcal{V}_l$. For the LBA complex input, $\mathcal{V}_r$ and $\mathcal{V}_l$ denote the receptor and the ligand, respectively, while for the PPA complex and single-chain input, $\mathcal{V}_r$ refers to the receptor protein chain and $\mathcal{V}_l$ becomes the corresponding binding protein chain. And for function prediction tasks, $\mathcal{V}_r$ refers to the protein chain and $\mathcal{V}_l$ becomes an empty set, as shown in the middle of Figure 1. We associate each node v_i with the representation $(\boldsymbol{h}_i, \vec{\boldsymbol{X}}_i)$, where $\boldsymbol{h}_i \in \mathbb{R}^d$ denotes the node feature and it is initialized as a learnable residue embedding, $\vec{\boldsymbol{X}}_i \in \mathbb{R}^{3\times c_i}$ indicates the 3D coordinates of all c_i atoms within the node. As for edge construction, we include various types of edges. In detail, for residue nodes, we allow R heterogeneous types of edge connections including sequential edges of different distances ($d = \{-2, -1, 1, 2\}$), self-loop edges, and spatial edges; for single-atom nodes from small molecules, only spatial edges are created. We present a simplified example from the LBA task in Figure 3, where we only draw a few nodes and omit the self-loop edges except for the central node for simplicity. Overall, we obtain a full-atom heterogeneous graph representation $\mathcal{G}$ for each input.

Task Formulation Given a full-atom heterogeneous graph $\mathcal{G}$, our goal is to design a model $\boldsymbol{p} = f(\mathcal{G})$ with multiple-dimensional output $\boldsymbol{p}$ that can predict the complex-level affinity and chain-level functional properties simultaneously. By making use of our proposed dataset Protein-MT, we train the model with a partially labeled training set and test it on the fully-labeled test set. Notably, the prediction should be invariant with regard to E(3) transformation (rotation/reflection/translation) of the input coordinates. To do so, we will formulate an equivariant encoder plus an invariant output layer in our model, detailed in the next subsection.

HeMeNet: Heterogeneous Multi-channel Equivariant Network

To better cope with the 3D structures of different types for different tasks, we propose a heterogeneous multi-channel E(3) equivariant graph neural network with the ability to aggregate different relational messages. After several layers of the message passing, the node representations are transformed into task-specific representations by a task-aware readout module, generating appropriate complex-level and chain-level predictions via different task heads.

Heterogeneous Multi-channel Equivariant Message Passing Inspired by dyMEAN (Kong, Huang, and Liu 2023a), we leverage a multi-channel coordinate matrix with dynamic size to record the geometric information of a node in an input graph. Moreover, we extend the setting to heterogeneous message passing along multiple types of edges to capture rich relationships between nodes. We denote the node feature and coordinates as $(\boldsymbol{h}_i^{(l)}, \vec{\boldsymbol{X}}_i^{(l)})$ in the l-th layer. The message passing is calculated as:

$$\boldsymbol{m}_{ijr} = \phi_m(\boldsymbol{h}_i^{(l)}, \boldsymbol{h}_j^{(l)}, \frac{T_R(\vec{\boldsymbol{X}}_i^{(l)}, \vec{\boldsymbol{X}}_j^{(l)})}{||T_R(\vec{\boldsymbol{X}}_i^{(l)}, \vec{\boldsymbol{X}}_j^{(l)})||_F + \epsilon}, \boldsymbol{e}_r), \quad (1)$$

$$\vec{\boldsymbol{M}}_{ijr} = T_S(\vec{\boldsymbol{X}}_i^{(l)} - \frac{1}{c_j}\sum_{k=1}^{c_j} \vec{\boldsymbol{X}}_j^{(l)}(:,k), \phi_x(\boldsymbol{m}_{ijr})), \quad (2)$$

where, $\boldsymbol{m}_{ijr}$ and $\vec{\boldsymbol{M}}_{ijr}$ are separately the invariant and equivariant messages from node j to i along the r-th edge; $\boldsymbol{e}_r$ is the edge embedding feature; ϕ_m, ϕ_x are Multi-Layer Perceptrons (MLPs) (Gardner and Dorling 1998) with one hidden layer; $||\cdot||_F$ denotes the Frobenius norm; T_R and T_S are the adaptive multichannel geometric relation extractor and geometric message scaler, in order to deal with the issue incurred by the varying shape of $\vec{\boldsymbol{X}}_i^{(l)}$ and $\vec{\boldsymbol{X}}_j^{(l)}$ since the number of atoms could be different for different nodes. With the messages, the node representation is updated by:

$$\boldsymbol{h}_i^{(l+1)} = \boldsymbol{h}_i^{(l)} + \sigma(\mathrm{BN}(\phi_h(\sum_{r\in R} \boldsymbol{W}_r \sum_{j\in\mathcal{N}_r(i)} \boldsymbol{m}_{ijr}))), \quad (3)$$

$$\vec{\boldsymbol{X}}_i^{(l+1)} = \vec{\boldsymbol{X}}_i^{(l)} + \frac{1}{\sum_{r\in R}|\mathcal{N}_r(i)|}\sum_{r\in R}\sum_{j\in\mathcal{N}_r(i)} w_r \vec{\boldsymbol{M}}_{ijr}, \quad (4)$$

where, $\boldsymbol{W}_r, w_r$ are a learnable matrix and a learnable scalar to project invariant an equivariant messages, respectively, for the r-th kind of edge; $\mathcal{N}_r(i)$ denotes the neighbor nodes of i regarding the r-th kind of edges; ϕ_h is an MLP, BN is the batch normalization operation, and σ is an activation function. During the message-passing process, our model gathers information from different relations for $\boldsymbol{h}_i$ and $\vec{\boldsymbol{X}}_i$, ensuring the E(3) equivariance.

Task-Aware Readout After L layers of the above relational message passing, we attain a set of E(3) invariant node features $\boldsymbol{H}^{(L)} \in \mathbb{R}^{n\times d_L}$, where n is the number of nodes and d_L is the feature dimension. To correlate the task-specific information with each node feature, we propose a task-aware readout function. We compute the attention weights between each node and each task-specific query, and then readout all weighted node features into dual-level representations: graph level or chain level for each task. As shown in Figure 3, the task-aware readout module is formulated as:

$$\boldsymbol{\alpha}_t = \mathrm{Softmax}(\frac{\boldsymbol{K}\boldsymbol{q}_t}{\sqrt{d_L}}), \quad (5)$$

$$\boldsymbol{f}_t = \mathrm{FFN}(\boldsymbol{\alpha}_t \boldsymbol{V} + \mathrm{Linear}(\boldsymbol{q}_t)), \quad (6)$$

where, $\boldsymbol{\alpha}_t \in [0,1]^n$ defines the attention values for task t; $\boldsymbol{q}_t \in \mathbb{R}^{d_L}$ is the learnable query for task t; $\boldsymbol{K} =$

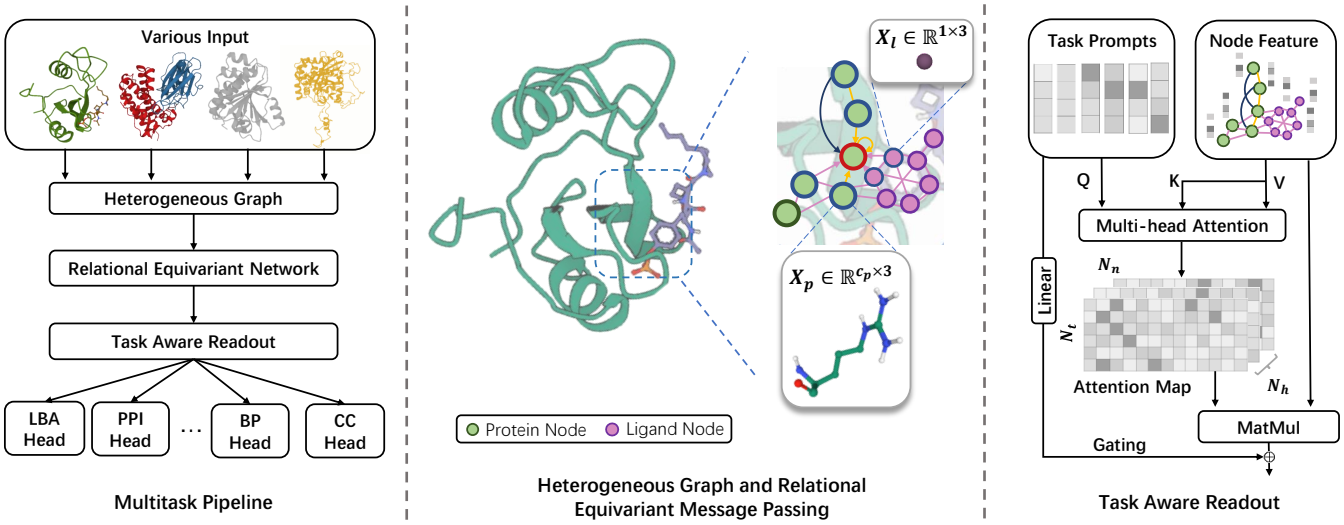

Figure 3: **Overview of our pipeline**. Left: HeMeNet takes two-instance complexes or a single chain as input and predicts complex-level affinity and chain-level properties simultaneously. Middle: An example of the heterogeneous graph and the relational equivariant message passing. Right: Task-aware readout module. We take a task prompt as the query for each task to get a multi-level readout for different downstream tasks.

$\boldsymbol{H}\boldsymbol{W}_K \in \mathbb{R}^{n\times d_L}$ and $\boldsymbol{V} = \boldsymbol{H}\boldsymbol{W}_V \in \mathbb{R}^{n\times d_L}$ are the key and value matrices, respectively; FFN is the feed-forward network containing layer normalization and linear layers; $\text{Linear}(\boldsymbol{q}_t) = \boldsymbol{W}_Q\boldsymbol{q}_t + \boldsymbol{b}$ is used before the shortcut addition. In our implementation, we apply the multi-head attention strategy by defining multiple queries for each task. We compute the attention by using the invariant features $\boldsymbol{H}^{(L)}$ as it has involved the geometric information from the 3D coordinates during the previous L-layer message passing.

Multiple Task Heads We feed the above task-specific feature $\boldsymbol{f}_i$ into different task heads implemented by MLPs, resulting in a prediction list $(\boldsymbol{p}_{\text{lba}}, \boldsymbol{p}_{\text{ppa}}, \boldsymbol{p}_{\text{ec}}, \boldsymbol{p}_{\text{mf}}, \boldsymbol{p}_{\text{bp}}, \boldsymbol{p}_{\text{cc}})$. For regression tasks $\mathcal{T}_{\text{reg}} = \{\text{LBA}, \text{PPA}\}$, we use the Mean Square Error (MSE) loss $\mathcal{L}_{\text{MSE}}$. For classification tasks $\mathcal{T}_{\text{cls}} = \{\text{EC}, \text{GO-MF}, \text{GO-BP}, \text{GO-CC}\}$, we use the Binary Cross Entropy (BCE) loss $\mathcal{L}_{\text{BCE}}$. The training loss is formulated as:

$$\mathcal{L} = \sum_{T\in\mathcal{T}_{\text{reg}}} \lambda_T \mathbf{1}_T \mathcal{L}_{\text{MSE}}(\boldsymbol{p}_T) + \sum_{T\in\mathcal{T}_{\text{cls}}} \lambda_T \mathbf{1}_T \mathcal{L}_{\text{BCE}}(\boldsymbol{p}_T), \tag{7}$$

where $\{\lambda_T\}$ are hyperparameters that balance losses for different tasks; $\mathbf{1}_T$ is the indicator function evaluating to 1 when the label for task T exists and to 0 otherwise, allowing for training on partially labeled samples. To accelerate the training convergence, we adopt a balanced batch sampling strategy that ensures each sampled mini-batch contains at least one labeled sample from each task.

Experiments

In this section, we will first introduce the experimental setup. Next, we evaluate our model on the proposed dataset Protein-MT for affinity and property prediction in both single-task and multi-task settings and compare it with other baseline models. Then, we experiment with different readout strategies and compare their performance. At last, we perform ablation experiments on different modules.

Experimental Setup

Task settings We compare HeMeNet with other models under single-task and multi-task settings using the same validation and test sets. For single-task training, models are trained on samples with labels of the corresponding task. We also remove the task-aware readout of our model for a fair comparison. For multi-task training, the models are trained on all partially labeled samples. We include samples with up to 15,000 atoms for training and evaluation.

Baselines We compared our model with nine representative baselines. **GCN** (Kipf and Welling 2017) aggregates information weighted by the degree of nodes. **GAT** (Veličković et al. 2018) utilizes an attention mechanism for message passing. **Schnet** (Schütt et al. 2017) is an invariant network with continuous filter convolution on the 3D molecular graph. **GearNet** (Zhang et al. 2023b) designs a relational message-passing network to capture information on protein function tasks. Its variant **Gearnet-fullatom** (Zhang et al. 2023c) utilizes the model to the full-atom setting. **CDConv** (Fan et al. 2022) models the geometric sequence with a continuous-discrete convolution. Besides the previous invariant models, we also compare our method with equivariant models. **EGNN** (Satorras, Hoogeboom, and Welling 2021) is a lightweight but effective E(n) equivariant graph neural network. **GVP** (Jing et al. 2021) designs an equivariant geometric vector perceptron for protein representation. **dyMEAN** (Kong, Huang, and Liu 2023a) is an equivariant model for antibody design; it takes a dynamic multichannel equivariant function for full-atom coordinates.

	Method	LBA		PPA		EC↑	GO		
		RMSE↓	MAE↓	RMSE↓	MAE↓		MF↑	BP↑	CC↑
Single-task	GCN (Kipf and Welling 2017)	2.193	1.721	7.840	7.738	0.022	0.207	0.254	0.367
	GAT (Veličković et al. 2018)	2.301	1.838	7.820	7.720	0.018	0.223	0.249	0.354
	SchNet (Schütt et al. 2017)	2.162	1.692	7.839	7.729	0.097	0.311	0.281	0.431
	GearNet* (Zhang et al. 2023b)	1.957	1.542	2.004	1.279	0.716	0.677	0.252	0.438
	GearNet-fullatom (Zhang et al. 2023b)	2.178	1.716	2.753	2.709	0.046	0.212	0.229	0.471
	EGNN (Satorras, Hoogeboom, and Welling 2021)	2.282	1.849	4.854	4.756	0.039	0.206	0.253	0.357
	GVP (Jing et al. 2021)	2.281	1.789	5.280	5.267	0.020	0.204	0.244	0.454
	dyMEAN (Kong, Huang, and Liu 2023a)	2.410	1.987	7.309	7.182	0.115	0.436	0.292	0.477
	HemeNet (Ours)	1.912	1.490	6.031	5.891	**0.863**	**0.778**	**0.404**	**0.544**
Multi-task	SchNet (Schütt et al. 2017)	1.763	1.447	1.216	1.120	0.093	0.192	0.264	0.402
	GearNet* (Zhang et al. 2023b)	2.193	1.863	1.275	1.035	0.187	0.203	0.261	0.379
	GearNet-fullatom (Zhang et al. 2023b)	1.839	1.350	1.821	1.491	0.047	0.155	0.258	0.443
	CDConv (Fan et al. 2022)	**1.579**	1.352	2.386	1.822	0.324	0.246	0.241	0.424
	EGNN (Satorras, Hoogeboom, and Welling 2021)	1.777	1.441	0.999	0.821	0.048	0.169	0.244	0.352
	GVP (Jing et al. 2021)	1.870	1.572	0.906	0.758	0.018	0.168	0.246	0.360
	dyMEAN (Kong, Huang, and Liu 2023a)	1.777	1.446	1.725	1.523	0.038	0.164	0.263	0.449
	HeMeNet* (Ours)	1.799	1.420	**0.861**	**0.719**	0.630	0.595	0.279	0.426
	HeMeNet (Ours)	1.730	**1.335**	1.087	0.912	0.810	0.727	0.379	0.436
	GPT4-turbo-1106 (ST) (Achiam et al. 2023)	2.347	1.780	**1.654**	1.343	-	-	-	-
	ESM2 (MT) (Lin et al. 2023)	2.009	1.334	1.692	**1.333**	0.917	0.764	0.389	0.533
	ESM2-HeMeNet (MT)	**1.867**	**1.661**	1.846	1.418	**0.921**	**0.796**	**0.455**	**0.567**

* represents trained under the alpha-carbon atom only setting. ST, MT is the abbreviation of single-task and multi-task, respectively.

Table 1: The mean result for three runs on the full-label test set. The upper half reports the results for the single-task setting, and the lower half reports the results for the multi-task setting. The best results are marked in bold and the second best results are underlined. In the multi-task setting, we train the models with the same size compared to their single-task models.

ESM2 (Lin et al. 2023) is a pretrained protein language model. We also conducted three-shot prompting to test GPT-4 (Achiam et al. 2023) on the two affinity prediction tasks.

Evaluation For LBA and PPA tasks, we employ the commonly used Root Mean Square Error (RMSE) and Mean Average Error (MAE) as the evaluation metrics (Townshend et al. 2021). For EC and GO tasks, we use maximum F-score (Fmax) following (Zhang et al. 2023b). Each experiment is independently run three times with different random seeds.

Results on Protein-MT

We conduct experiments under both single-task and multi-task settings. The mean results of three runs are reported in Table 1. According to the results, we draw conclusions summarized in the subsequent paragraphs.

Our model outperforms the baselines on most of the tasks under both settings. Under the single-task setting, our model surpasses other models in five of the six tasks. Under the multi-task setting, our model surpasses other models in four of the six tasks, with the remaining two tasks reaching second and third place, respectively. We also compared our model with GPT4 and ESM2 (with multi-task head fine-tuning). Our method outperforms GPT4 in both LBA and PPA. ESM2 performs well on four property prediction tasks, and combining ESM2 and HeMeNet can further improve ESM2's performance with geometric information. Notably, under the single-task setting, only the models with a heterogeneous message passing (GearNet and ours) can perform well on all of the four property prediction tasks. Under the multi-task setting, our full-atom model, benefiting from joint learning, shows superior results on different tasks, and there are two main interesting observations discussed next.

Our model benefits from the multi-task setting, especially on LBA and PPA. We observe that almost all models improve their performance on LBA and PPA tasks under the multi-task setting. In particular, our model significantly improves the PPA RMSE from 6.031 to 1.087 by utilizing a training set that is more than ten times larger (2587 for PPA single-task and 30904 for our multi-task training set). We also train our model with alpha C atom (HeMeNet*) as input, resulting in a best PPA RMSE of 0.861. As shown in Table 2, our model performs better as the training samples of PPA increase. And it performs much better than its single-task counterpart, even with a small amount of training samples, showcasing the internal transfer of information within the tasks. These results demonstrate that the model can handle challenging tasks (complex-level affinity prediction) better when more structural information is available (e.g., single-chain structures and their labels).

Our model performs harmonious multi-task training on property prediction tasks. We observe that when switching from the single-task to multi-task setting, baseline models experience performance degradation to some extent across the four property prediction tasks. This is prob-

ably because of task interference among diverse tasks, and combining different tasks for training without careful adaption can harm performance. With the guidance of our task-aware readout module, our model is able to learn from multiple tasks in a task-harmonious way, while achieving performance on the property prediction tasks comparable to their single-task counterparts with the same parameter size.

Ratio	0.00%	8.33%	16.67%	33.33%	100%
RMSE	7.764	1.651	1.161	1.126	1.087

Table 2: PPA performance with different training ratios.

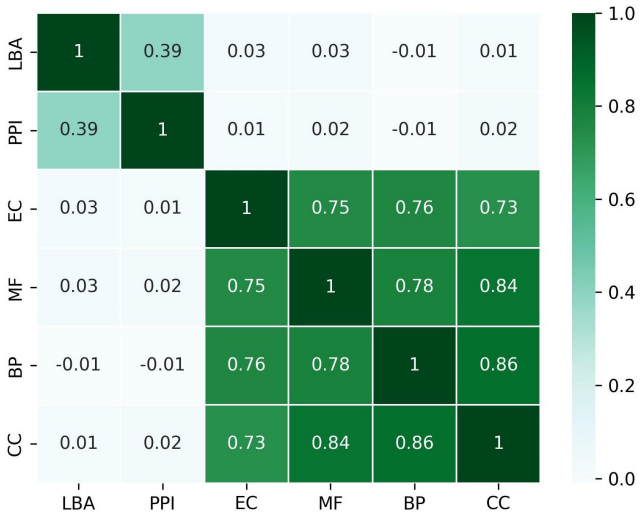

Figure 4: Task prompt correlation heatmap

Comparison of Different Readout Methods

To verify the effectiveness of our TAR module, we take HeMeNet and Gearnet as the backbone and compare the task-aware readout method with sum readout and task-prompt weighted feature (Liu et al. 2023).

Method	EC↑	GO-MF↑	GO-BP↑	GO-CC↑
Gearnet$_s$	0.187	0.203	0.261	0.379
Gearnet$_w$	0.066	0.164	0.271	0.414
Gearnet$_t$	0.421	0.310	0.287	0.403
HeMeNet$_s$	0.722	0.558	0.302	0.413
HeMeNet$_w$	0.325	0.312	0.276	**0.440**
HeMeNet$_t$	**0.810**	**0.727**	**0.379**	0.436

Table 3: Comparison of readout functions. s, w, and t represent sum, weighted feature and task-aware readout.

The results are presented in Table 3. We can conclude with the following observations: 1) Our task-aware readout injects task-related information using an attention mechanism, leading to overall improvements for various tasks, especially on the Enzyme Commission task. 2) Element-wise multiplication of the task prompt feature with nodes fails to provide sufficient guidance to learning across tasks.

To further investigate the relationship between tasks, we calculate Pearson's correlation between prompts. As shown in Figure 4, the correlations between tasks within the same category (e.g. EC and MF) are high, while the correlations between tasks from different categories (e.g. LBA and BP) are low. A high correlation between prompts indicates similar attention queries, leading to similar readout functions. Therefore, with task-aware guidance, the model employs similar readout strategies for tasks from the same category and divergent strategies from tasks from different categories.

Method	LBA↓	PPA↓	EC↑	GO		
				MF↑	BP↑	CC↑
Ours	**1.730**	1.087	**0.810**	**0.727**	**0.379**	**0.436**
- TAR	1.905	1.970	0.722	0.558	0.302	0.413
- Rs	1.790	1.446	0.547	0.663	0.359	0.391
- Atom	1.799	**0.861**	0.630	0.595	0.279	0.426

Table 4: Ablation study for components in HeMeNet.

Ablation Study

We perform ablation experiments to evaluate the necessity of different components. Specifically, the ablation of TAR replaces the TAR module with a sum readout. For e_r, W_r, and w_r, we remove different types of edges and the relational message passing weights. For full-atom ablation, we represent the coordinates of residues by their alpha C atoms.

We present the results for ablation studies in Table 4, the observations are as follows: 1) Without TAR module, significant performance degradation is observed, indicating that the tasks can hinder each other without appropriate guidance. 2) Without the heterogeneous graph, our model's performance drops on property prediction tasks, especially the Enzyme Commission number prediction. 3) Removing the full-atom geometry decreases the performance in multiple tasks. However, it improves our model's performance in PPA. We suppose that the large number of atoms in the full-atom protein-protein complex introduces excessive noise compared with input with alpha-carbon atoms only.

Conclusion

In this paper, we alleviate the problem of sparse data in structured protein datasets by a multi-task setting. First, We construct a standard multi-task benchmark Protein-MT, consisting of 6 representative tasks integrated from 4 public datasets for joint learning. To address multiple tasks in protein 3D learning, we propose a novel network called HeMeNet. Comprehensive experiments demonstrate our model's performance on the affinity and property prediction tasks. Our work brings insights for utilizing different structural datasets to train a more powerful generalist model.

Acknowledgments

This work is jointly supported by the following projects: the National Key R&D Program of China (2021YFF1201300, 2021YFF1201303, 2022YFC2703105); the National Natural Science Foundation of China (No. 62376276); Beijing Nova Program (No. 20230484278); Ant Group through CCF-Ant Research Fund (CCF-AFSG RF20220204); Guoqiang Institute of Tsinghua University; Beijing National Research Center for Information Science and Technology (BNRist); the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (23XNKJ19). The funders had no roles in study design, data collection and analysis, the decision to publish, or manuscript preparation.

References

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Alley, E. C.; Khimulya, G.; Biswas, S.; AlQuraishi, M.; Church, G. M.; and Church, G. M. 2019. Unified rational protein engineering with sequence-based deep representation learning. *Nature methods*, 16(12): 1315—1322.

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

Capel, H.; Feenstra, K. A.; and Abeln, S. 2022. Multi-task learning to leverage partially annotated data for PPI interface prediction. *Scientific Reports*, 12(1): 10487.

Capel, H.; Weiler, R.; Dijkstra, M.; Vleugels, R.; Bloem, P.; and Feenstra, K. A. 2022. ProteinGLUE multi-task benchmark suite for self-supervised protein modeling. *Scientific Reports*, 12(1): 16047.

Cen, J.; Li, A.; Lin, N.; Ren, Y.; Wang, Z.; and Huang, W. 2024. Are High-Degree Representations Really Unnecessary in Equivariant Graph Neural Networks? *arXiv preprint arXiv:2410.11443*.

Fan, H.; Wang, Z.; Yang, Y.; and Kankanhalli, M. 2022. Continuous-discrete convolution for geometry-sequence modeling in proteins. In *The Eleventh International Conference on Learning Representations*.

Ganea, O.-E.; Huang, X.; Bunne, C.; Bian, Y.; Barzilay, R.; Jaakkola, T. S.; and Krause, A. 2022. Independent SE(3)-Equivariant Models for End-to-End Rigid Protein Docking. In *International Conference on Learning Representations*.

Gardner, M. W.; and Dorling, S. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15): 2627–2636.

Gasteiger, J.; Groß, J.; and Günnemann, S. 2020. Directional Message Passing for Molecular Graphs. In *International Conference on Learning Representations*.

Gligorijević, V.; Renfrew, P. D.; Kosciólek, T.; Leman, J. K.; Berenberg, D.; Vatanen, T.; Chandler, C.; Taylor, B. C.; Fisk, I.; Vlamakis, H.; Xavier, R. J.; Knight, R.; Cho, K.; and Bonneau, R. 2021a. Structure-based protein function prediction using graph convolutional networks. *Nature Communications*, 12.

Gligorijević, V.; Renfrew, P. D.; Kosciolek, T.; Leman, J. K.; Berenberg, D.; Vatanen, T.; Chandler, C.; Taylor, B. C.; Fisk, I. M.; Vlamakis, H.; et al. 2021b. Structure-based protein function prediction using graph convolutional networks. *Nature communications*, 12(1): 3168.

Han, J.; Cen, J.; Wu, L.; Li, Z.; Kong, X.; Jiao, R.; Yu, Z.; Xu, T.; Wu, F.; Wang, Z.; et al. 2024. A Survey of Geometric Graph Neural Networks: Data Structures, Models and Applications. *arXiv preprint arXiv:2403.00485*.

Hermosilla, P.; Schäfer, M.; Lang, M.; Fackelmann, G.; Vázquez, P.-P.; Kozlikova, B.; Krone, M.; Ritschel, T.; and Ropinski, T. 2021. Intrinsic-Extrinsic Convolution and Pooling for Learning on 3D Protein Structures. In *International Conference on Learning Representations*.

Huang, W.; Han, J.; Rong, Y.; Xu, T.; Sun, F.; and Huang, J. 2022. Equivariant Graph Mechanics Networks with Constraints. In *International Conference on Learning Representations*.

Jing, B.; Eismann, S.; Suriana, P.; Townshend, R. J. L.; and Dror, R. 2021. Learning from Protein Structure with Geometric Vector Perceptrons. In *International Conference on Learning Representations*.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

Kong, X.; Huang, W.; and Liu, Y. 2023a. End-to-End Full-Atom Antibody Design. *arXiv preprint arXiv:2302.00203*.

Kong, X.; Huang, W.; and Liu, Y. 2023b. Generalist Equivariant Transformer Towards 3D Molecular Interaction Learning. *arXiv preprint arXiv:2306.01474*.

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

Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130.

Liu, Z.; Yu, X.; Fang, Y.; and Zhang, X. 2023. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the ACM Web Conference 2023*, 417–428.

Rao, R.; Bhattacharya, N.; Thomas, N.; Duan, Y.; Chen, P.; Canny, J.; Abbeel, P.; and Song, Y. 2019. Evaluating Protein Transfer Learning with TAPE. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Rao, R. M.; Liu, J.; Verkuil, R.; Meier, J.; Canny, J.; Abbeel, P.; Sercu, T.; and Rives, A. 2021. MSA Transformer. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 8844–8856. PMLR.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E(n) Equivariant Graph Neural Networks. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 9323–9332. PMLR.

Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; and Müller, K.-R. 2017. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Shi, Z.; Deng, R.; Yuan, Q.; Mao, Z.; Wang, R.; Li, H.; Liao, X.; and Ma, H. 2023. Enzyme Commission Number Prediction and Benchmarking with Hierarchical Dual-core Multitask Learning Framework. *Research*, 6: 0153.

Townshend, R.; Vögele, M.; Suriana, P.; Derry, A.; Powers, A.; Laloudakis, Y.; Balachandar, S.; Jing, B.; Anderson, B.; Eismann, S.; Kondor, R.; Altman, R.; and Dror, R. 2021. ATOM3D: Tasks on Molecules in Three Dimensions. In Vanschoren, J.; and Yeung, S., eds., *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1. Curran.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.

Wang, G.; Liu, X.; Wang, K.; Gao, Y.; Li, G.; Baptista-Hon, D. T.; Yang, X. H.; Xue, K.; Tai, W. H.; Jiang, Z.; et al. 2023a. Deep-learning-enabled protein-protein interaction analysis for prediction of SARS-CoV-2 infectivity and variant evolution. *Nature Medicine*, 1–12.

Wang, H.; Liu, H.; Ning, S.; Zeng, C.; and Zhao, Y. 2022. DLSSAffinity: protein-ligand binding affinity prediction via a deep learning model. *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)*, 24(17): 10124–10133.

Wang, R.; Fang, X.; Lu, Y.; and 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. PMID: 15163179.

Wang, X.; Zhang, Y.; Zhou, P.; and Liu, X. 2022. A supervised protein complex prediction method with network representation learning and gene ontology knowledge. *BMC bioinformatics*, 23(1): 300.

Wang, Z.; Zhang, Q.; HU, S.-W.; Yu, H.; Jin, X.; Gong, Z.; and Chen, H. 2023b. Multi-level Protein Structure Pretraining via Prompt Learning. In *The Eleventh International Conference on Learning Representations*.

Wu, S.; Zhang, H. R.; and Ré, C. 2020. Understanding and improving information transfer in multi-task learning. *arXiv preprint arXiv:2005.00944*.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *International Conference on Learning Representations*.

Xu, M.; Zhang, Z.; Lu, J.; Zhu, Z.; Zhang, Y.; Chang, M.; Liu, R.; and Tang, J. 2022. Peer: a comprehensive and multi-task benchmark for protein sequence understanding. *Advances in Neural Information Processing Systems*, 35: 35156–35173.

Zhang, C.; Zhang, X.; Freddolino, P. L.; and Zhang, Y. 2023a. BioLiP2: an updated structure database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, gkad630.

Zhang, Y.; Cen, J.; Han, J.; Zhang, Z.; Zhou, J.; and Huang, W. 2024. Improving Equivariant Graph Neural Networks on Large Geometric Graphs via Virtual Nodes Learning. In *Forty-first International Conference on Machine Learning*.

Zhang, Z.; Xu, M.; Jamasb, A. R.; Chenthamarakshan, V.; Lozano, A.; Das, P.; and Tang, J. 2023b. Protein Representation Learning by Geometric Structure Pretraining. In *The Eleventh International Conference on Learning Representations*.

Zhang, Z.; Xu, M.; Lozano, A.; Chenthamarakshan, V.; Das, P.; and Tang, J. 2023c. Physics-Inspired Protein Encoder Pre-Training via Siamese Sequence-Structure Diffusion Trajectory Prediction. *arXiv preprint arXiv:2301.12068*.

Zhao, Y.; Wang, J.; Chen, J.; Zhang, X.; Guo, M.; and Yu, G. 2020. A literature review of gene function prediction by modeling gene ontology. *Frontiers in genetics*, 11: 400.