Research practice report

Joe Zhang
Supervised by:
Alexandre M. J. J. Bonvin
Rens Holmer
Xiaotong Xu

Modeling protein-protein interactions with graph neural network

Abstract

Protein-protein interactions (PPIs) are essential in biological processes. Understanding PPIs from structural aspects provides more information on drug discovery and protein engineering. We have developed DeepRank-GNN-ESM, a rotation-invariant deep-learning framework combined with large protein language model embeddings. It has competitive performance on various tasks but was not tested on the prediction of binding affinity changes, which can help disease diagnoses and function annotations. We benchmarked DeepRank-GNN-ESM with other prediction models on a single-point mutation dataset. The result shows that our model does not have comparable performance to others. In an attempt to improve its performance, we added steerable equivariant features into the model, allowing it to better represent the directional information. This new architecture was trained on the binary classification of biological and crystal interfaces and showed an accuracy of 0.81 on the test set. While already competent in the scope of this project, we oversee its use in more complicated tasks.

1. Introduction

Protein-protein interactions (PPIs) have diverse functions and different types in biology.[1] PPIs are vital in most cellular processes and activities, such as signal processing, enzyme catalysis, and immune response.[2] Understanding the factors that influence PPIs can shed light on biological mechanisms and drug design.[3] The strength of PPIs is measured by binding affinity or binding free energy (ΔG).[4] Many PPIs are sensitive to mutations that can alter binding affinity and specificity.[5] Residue mutation mostly affects the binding affinity by altering the conformation and thermodynamics of PPIs. The mutation effect is measured by the difference between ΔG of mutated PPIs and ΔG of wildtype PPIs:

$$\Delta\Delta G = \Delta G_{\text{mutated}} - \Delta G_{\text{wildtype}} \quad (1)$$

Experimentally, alanine scanning as a mutagenesis technique is used to determine the contribution of a specific residue involved in the interaction.[6] However, experimental approaches for solving PPIs' structure remain time-consuming and cumbersome. As a result, only 22% of the human interactome has experimental structures.[7] Experimental PPI thermodynamic determinations are usually laborious and expensive.[8] On the one hand, *in-silico* prediction models could make the most of existing data to predict features of unlabeled data, for example, ranking docking poses and predicting interaction interface given PPI structures.[9–11] On the other hand, those models could filter experimental false positives, such as distinguishing biological PPI interfaces with crystal ones generated during X-ray crystallography.[10,11] During the development of a new computational method, it is crucial to find the proper protocol to represent PPI structure information so that the model can make more precise predictions. Deep learning models have the benefit of working on raw or slightly processed data, whereas traditional machine learning methods heavily depend on feature engineering.[12] Graph neural networks (GNNs) have been widely applied in protein structure analysis.[13–15] Most models use distance as the edge feature of the protein graph. DeepRank-

GNN is a GNN tool developed in our lab, which has highly competitive performance in multiple applications related to PPI analysis.[10,11] It generates the protein interaction interface's graph from the structure of a complex, where the edge represents the distance between two residues while the node features contain residue-level information such as physicochemical properties and evolutionary profiles.[10] Evolutionary Scale Modeling-2 (ESM-2) embeddings have been recently combined into node features as evolutionary information to improve the model performance.[11,16] For now, this framework was benchmarked on docking pose ranking and classification of interfaces. In the scope of this study, we aim to assess the ability of DeepRank-GNN to detect the contribution of a residue in PPIs and predict the point mutation effect. We compared the GNN model with other tools of different types, including a force-field based tool (HADDOCK 3.0), a linear regression model with structure properties (Prodigy), and a random forest model combining chemical energy terms and evolutionary profiles (iSEE).[17–19] The result shows that the force-field method can better capture mutation effects. Coevolutionary patterns are less relevant to the task. DeepRank-GNN framework does not have comparable performance as other tools. To improve it with more directional information, we explored using steerable features in the model.[20] It is an equivariant model that combines 3D vectors into edge features. It is named steerable equivariant graph interaction neural network (SEGIN). We benchmarked it on a simple binary classification task of classifying biological and crystal PPI interfaces.

2. Data and methods

2.1 Mutation effect analysis

2.1.1 dataset

The experimental ΔΔGs were extracted from SKEMPI v2.0 dataset.[21], where only single-point mutations on dimeric complexes were selected to reduce the computational complexity. The selected subset contains 486 point-mutation entries spread over 64 proteins. Among them, 362 entries are alanine mutations (the residue is mutated into alanine), while the remaining 124 are non-alanine mutations (the residue is mutated into other amino acids).

2.1.2 Software

DeepRank-GNN-ESM

DeepRank-GNN-ESM improves over DeepRank-GNN by replacing the position-specific scoring matrix (PSSM) with the ESM-2 embeddings as one of the node features. The PPI interface graphs were generated with an intramolecular contact distance cutoff of 3 Å (within one protein) and an intermolecular one of 8.5 Å (between two proteins).[10] The contact distance was defined by the smallest atom distance between two residues' heavy atoms.[10] Node features include residue type, polarity, charge, BSA, and embeddings from the last layer of ESM-2. Buried Surface Area (BSA) is calculated with open-source solvent-accessible surface areas python library (FreeSASA), which uses Lee & Richards' algorithm to calculate surface area geometry.[22] Distances between interacting residues are calculated and used as edge features. In mutation effect analysis, we downloaded model weights from GitHub (https://github.com/DeepRank/DeepRank-GNN-esm). The model aims to predict the fraction of conserved native contacts (FNAT), which is the number of contacts in the predicted PPI interface with respect to that from the experimental reference interface. The mutation effect was calculated as follows:

$$\Delta\text{FNAT} = \text{FNAT}_{\text{mutated}} - \text{FNAT}_{\text{wildtype}} \quad (2)$$

HADDOCK 3.0

The High ambiguity-driven docking approach (HADDOCK) is a docking software using the Crystallography and NMR system (CNS) as the core engine.[17] We used the alanine scanning module from HADDOCK 3.0, which mutates residues in PDB files (https://github.com/haddocking/haddock3). CNS can build the residue topology and coordinates after mutation.[23] In our experiments, the interface contact cutoff was set to 8.5 Å. HADDOCK would refine 10 structural models with the Water refinement module (mdref); the final score was calculated based on the mean values of 10 structural models. The output HADDOCK score is calculated as follows in Water refinement:

$$1.0\ Evdw + 0.2\ Eelec + 1.0\ Edesol \quad (3)$$

$Evdw$: Van der Waals intermolecular energy. $Eelec$: electrostatic intermolecular energy. $Edesol$: desolvation energy.

iSEE

Interface Structure, Evolution and Energy-based ΔΔG predictor (iSEE) is a machine-learning based random-forest model trained to predict point mutation ΔΔGs.[19] Input features include HADDOCK-based features and evolutionary features from PSSM.[24] The predicted ΔΔGs of the SKEMPI S487 test set were directly extracted from the iSEE repository (https://github.com/haddocking/iSee), based on the refined top1 model generated by HADDOCK (water refinement web service).

Prodigy

Protein binding energy prediction (Prodigy) is a linear regression model trained to predict binding affinity.[18] We ran Prodigy with default settings. In the development of Prodigy, different types of inter-residue contacts (ICs) and percentages of non-interacted surfaces (%NIS) were tested. The terms significantly correlated with binding affinity were selected as predictors. The final model is as follows:

$$\begin{aligned}\Delta G = {} & 0.09459\ ICs_{charged/charged} + 0.10007\ ICs_{charged/apolar} - 0.19577\ ICs_{polar/polar} \\ & + 0.22671\ ICs_{polar/apolar} - 0.18681\%NIS_{apolar} - 0.13810\%NIS_{charged} \\ & + 15.9433 \end{aligned} \quad (4)$$

The type of amino acids defines different types of ICs and %NIS, for example, $\text{ICs}_{\text{charged/charged}}$: inter-residue contacts between charged and charged residue. $\%\text{NIS}_{\text{apolar}}$: percentage of non-interacted surfaces from apolar residue.

2.2 Steerable Equivariant model development

2.2.1 dataset

The SEGIN was benchmarked on the MANY/DC datasets.[25,26] This is a binary classification task used in DeepRank-GNN.[10] The model will be trained to classify PPI interfaces between biological and crystal ones. Being crystal means the two proteins are artificially combined during X-ray crystallography experiments, whereas biological contacts exist in organisms.[27] The MANY dataset contains 5739 PPI structures, 2911 crystal interfaces, and 2828 biological interfaces, while the DC dataset has 161 structures, 81 crystal interfaces, and 80 biological interfaces. The MANY dataset was used for training, 20% of which is the validation set, while the DC dataset was used as the test set.

2.2.2 model development

In DeepRank-GNN, the edge between contact residues is distance, a scalar value. It is invariant with rotations in 3D space. However, this representation loses the directional information (vectors) that may be beneficial to better represent the interactions between residues. The goal is to represent vectors equivariantly with neural networks. Specifically, the neural network should be equivariant against the 3D rotation (Fig 1).

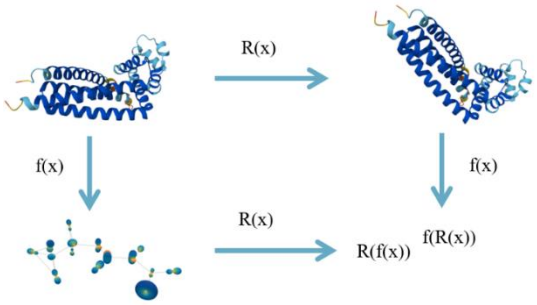


Fig 1. Equivariance property. f(x) is the neural network representation of the protein structure. R(x) is the 3D rotation. Being equivariant is that R(f(x)) should be equal to f(R(x)), which means the order of functions does not matter and f(x) is efficient to capture structural information independent of R(x).

Directly inputting vectors into a neural network is not equivariant with rotations. Spherical harmonics can solve this problem. The vector is decomposed into subspace representations. Rotations on vectors are equal to rotations on all subspaces (Fig 2).[28] More details are in Supplementary materials (Sup. (1)).

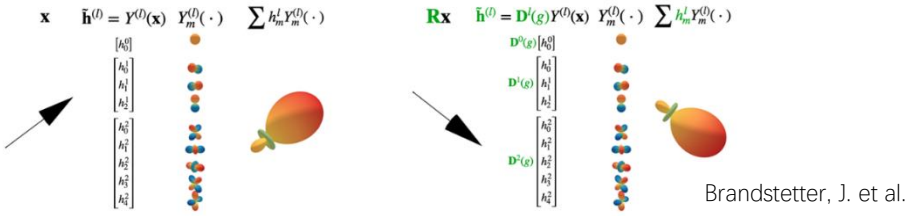


Fig 2. In this case, vector X is decomposed into $\tilde{h}^{(l)}$, 3 subspaces ($Y_m^{(l)}(\cdot)$, l =2). l is the highest order of the subspace. From l = 0 to l = 2, subspace dimensions range from 1 to 5 ($2\times l + 1$). Rotation on X is equal to rotations acting on all subspaces ($D^l(g)$). The steerable representation of vector is the sum of all subspaces' representations $\sum h_m^l Y_m^{(l)}(\cdot)$ Since dimensions of subspaces are increasing, it is called 'steerable'.[20]

Message passing among spherical harmonics requires the Clebsh-Gordan tensor product to keep equivariance, which combines two sets of subspaces with different l and returns another set. Detailed formulas are explained in Supplementary materials (Sup. (1)).

The SEGIN model was built based on *e3nn* python3 library, where spherical harmonics and Clebsh-Gordan tensor product are implemented.[29] The attention mechanism is also combined into the model (Sup. (2)). The model first passes messages through internal edges within each protein, then through PPI's contacts (Fig 3).

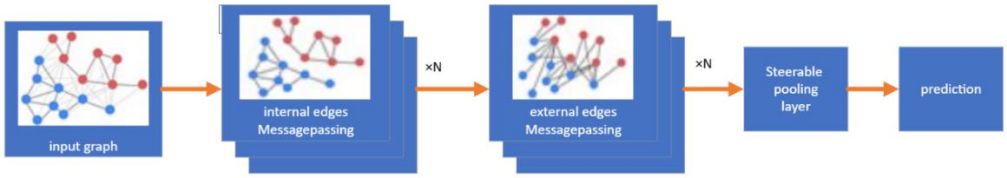


Fig 3. The input PPI interface graph is processed with internal edges and external edges message-passing blocks, both of which are repeated N times. The final prediction is based on a steerable pooling layer.

In our test, SEGIN node features are residue type, polarity, charge, and BSA. To reduce the computations, ESM-2 embeddings were not included. Intramolecular and intermolecular contacts are defined with the same cutoffs as DeepRank-GNN-ESM, an intramolecular contact distance cutoff of 3 Å and an intermolecular one of 8.5 Å. Two hyperparameters must be tuned in this model architecture: *N*, the number of repeated blocks (Fig 3), and Lmax_h, the highest order of the steerable subspace (*l*) (Fig 2). All models were trained for 150 epochs. The model weights that led to the minimum validation loss were selected for test. All models were trained on a single 11GB GTX 1080Ti GPU.

3. Results

3.1 mutation effect analysis

Model performance was evaluated by the Pearson Correlation Coefficient (PCC). Correlations between model predictions are also measured (Fig 4).

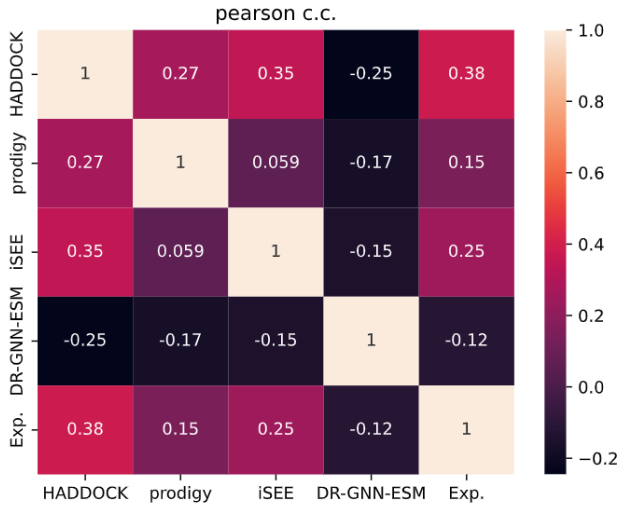


Fig 4. Pearson correlation coefficients among different models and experimental data. Among all models, HADDOCK has the highest correlation coefficients against the experimental dataset. DeepRank-GNN-ESM has a negative correlation coefficient since positive DeepRank-GNN-ESM prediction (ΔFNAT) means more contacts, which generally corresponds to negative experimental data (ΔΔG).

In Fig 4, HADDOCK 3.0 has the highest correlation coefficient against experimental data. Interestingly, iSEE, which combines HADDOCK terms and evolutionary terms, is less successful than HADDOCK. DeepRank-GNN-ESM has a small negative correlation against experimental data (Fig 4). Intuitively, positive ΔFNAT means more contacts, which generally corresponds to negative ΔΔG. But there is no strict correlation between ΔFNAT and ΔΔG.

All models perform poorly, with PCCs that never exceed 0.4 against the experiments (Fig 4). We took a detailed look at each residue's model performance on the alanine mutation dataset (362 entries). The PCC of each model prediction against experimental data for each residue is shown (Fig 5). Given a specific residue mutation to alanine, it can give us some hints about tool selection.

Clearly, DeepRank-GNN-ESM cannot achieve promising performance. After all, it is not trained to predict ΔΔG. However, we still want to improve the protein structure representation in the model to enhance the performance. We want to represent the edges between contact residues with higher resolution.

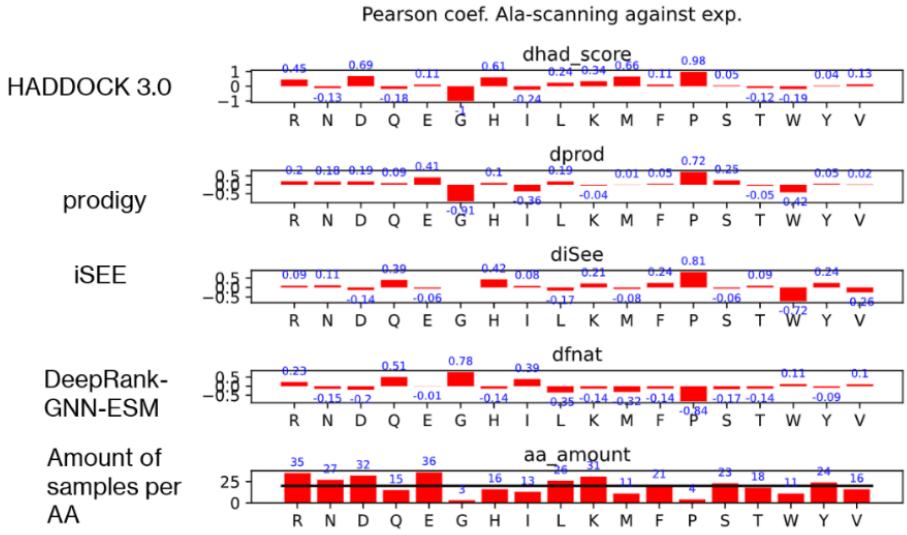


Fig 5. Per-residue PCCs against experimental data on the alanine mutation dataset. Cysteine mutation is not in the dataset. The last plot shows the number of samples of each amino acid (AA) in the dataset. The black line is the average amount of samples per residue. Glycine (G) and Proline (P) are much less than the average (AA amount <10). Other plots show the correlations (PCC) between each model prediction and experimental data. Only several AAs are significant for each model (PCC > 0.5). The significant residues of HADDOCK are Aspartic acid (D), Histidine (H), Methionine (M), Proline (P), and Glycine (G). In Prodigy, Glycine (G) and Proline (P) are significant. Tryptophan (W) and Proline (P) are significant for iSEE. Glutamine (Q), Glycine (G), and Proline (P) are significant for DeepRank-GNN-ESM.

3.2 SEGIN development

We combined steerable equivariant features into the model. Model performances with different hyperparameters were checked on a binary classification task (Crystal vs biological) (Table 1).

SEGIN	Test accuracy (%)	The epoch with the lowest validation loss	Training time per epoch (Sec)
N = 1, Lmax_h = 1	**81**	128	**26**
N = 1, Lmax_h = 2	80	138	50
N = 2, Lmax_h = 1	74	104	36
N = 2, Lmax_h = 2	71	134	80
N = 2, Lmax_h = 3	67	125	210
N = 3, Lmax_h = 2	69	104	105

Table 1. Test accuracies for models with different hyperparameters on the MANY/DC dataset. N is the number of message-passing blocks. Lmax_h is the highest order of steerable subspace, and its highest dimension is 2 * Lmax_h + 1.

As the number of blocks (N) increased, training time per epoch increased, and the test accuracy of SEGIN decreased on the MANY/DC dataset. This is the same as Lmax_h. Based on the result, N=1, Lmax_h=1 is the best-performed model, which was used for comparison with other models (Table 2).

Model	Model type	Test accuracy (%)
PISA[30]	Linear Regression	79
PRODIGY-crystal[31]	Random Forest	74
DeepRank	3D CNN + PSSM	**86**
DeepRank-GNN	GNN + PSSM	82
DeepRank-GNN-ESM	GNN + ESM-2	83
SEGIN (N = 1, Lmax_h = 1)	SE-GNN	81

Table 2. Test accuracies for different models on MANY/DC dataset. 3D CNN: Convolutional Neural Network in 3D space. GNN: Graph Neural Network. ESM: Evolutionary Scale Modeling-2.[16] SE-GNN: Steerable Equivariant Graph Neural Network.

SEGIN shows comparable performance to other models. DeepRank-related models include coevolutionary patterns or large protein language model embeddings (ESM-2). SEGIN only uses chemical information of residue and coordinates as the input feature, which requires less feature engineering.

4. Discussion

All models' performance on mutation effect analysis is not promising. The benchmarked models are not designed to perform mutation effect analysis except iSEE. The state-of-the-art mutation effect prediction model, SSIPe, can achieve 0.68 PCC on the whole SKEMPI dataset.[32,33] It is a linear regression model combining physical energy terms from EvoEF, sequence-based, and structure-based profiles. EvoEF and other top-performing models are all force-field based.[34–36] EvoEF can achieve 0.55 PCC on whole SKEMPI dataset.[21] The Force-field method is dominant in mutation effect analysis. Here we have the same conclusion since HADDOCK performs the best. Though iSEE combines HADDOCK energy terms and evolutionary profile, its performance is even worse in this case. The coevolutionary pattern from PSSM may not help. More advanced sequence or structure profiling is required for enhancement, like those used in SSIPe.

In per-residue analysis, Glycine and Proline have high correlations among all models. However, they have fewer samples (respectively 3 and 4), which indicates that the high PCCs may not be statistically representative. HADDOCK covers a wider range of well-predicted residues compared to other models. All models have weak correlations on residues with polar uncharged side-chain residue except that DeepRank-GNN-ESM has a high correlation on Glutamine. This indicates that DeepRank-GNN-ESM can learn orthogonal information compared with others. This result can give suggestions for integrating models to make mutation effect predictions since different model performs significantly on specific residues.

In steerable equivarant model development, combining steerable features into DeepRank-GNN aims to make it better represent interresidue geometry with higher resolution since residue interactions mainly depend on such geometry. Unfortunately, due to computation and time limitations, we did not manage to benchmark the model on the SKEMPI dataset. SEGIN shows comparable performance without coevolutionary patterns or huge embedding features (ESM-2 embeddings) in the binary classification (Table 2). We found training accuracy reaches 100% and training loss is nearly zero during training in several cases, which indicates SEGIN overfits the training data (Sup. (3)). This can be understood by the limited size of the MANY

dataset composed of 5739 entries. The model architecture could be used for more complex tasks such as binding affinity prediction or docking rank. In addition, node features can also be 3D or even higher-dimensional features with steerable equivariant message passing, for example, velocity or force.[20] It provides more possibilities for representing PPI structures and even some structure dynamics. Regarding computational time, since the Clebsh-Gordan tensor product requires more steps than simple matrix multiplication, the model needs more training time. Regarding future work, AlphaFold 3 can predict mutated PPI structures.[37] It has been shown that AlphaFold 3 can improve other model performances in mutation effect analysis.[33] The next step is to combine the models we used with PPI structures from AlphaFold3. In addition, the performance of SEGIN on mutation effect analysis and further tasks needs to be benchmarked.

5. Supplementary material

(1) Steerable equivariant components

Spherical harmonics:

Spherical harmonics are a set of orthonormal bases that can represent the function on a sphere. They can be considered high-dimension Fourier transform (https://www.thefouriertransform.com/). When the dimension of the highest base goes into infinity, the representation is totally precise. For more information, see Wikipedia (https://en.wikipedia.org/wiki/Spherical_harmonics). The basis table of spherical harmonics is also included (https://en.wikipedia.org/wiki/Table_of_spherical_harmonics).

Here is the equivariant formula for spherical harmonics:

$$Y_m^{(l)}\left(n_{\psi,\theta}\,\hat{r}\right) = \sum_{m'} \frac{1}{\sqrt{2l+1}} D^{(l)}_{m_0 m'}\left(R_{\psi,\theta,\phi}\right)\; Y_{m'}^{(l)}(\,\hat{r}) \tag{5}$$

$Y_m^{(l)}$ is the spherical harmonics function, l is the order of the steerable vector, m is the scalar value between -l and l. $n_{\psi,\theta}$ denote rotations with degree ψ and θ around unit basis x and y respectively. Spherical harmonics is invariant with ϕ by definition. $\hat{r}$ is a vector in 3D Euclidean space. Matrix representations of SO(3) group can be reduced to an equivalent block diagonal matrix; each block is called a Wigner-D matrix. In this case, each Wigner-D matrix acts on a steerable vector space. $D^{(l)}_{m_0 m'}\left(R_{\psi,\theta,\phi}\right)$ is the corresponding Wigner-D matrix representation of rotations with degree ψ, θ, and ϕ. $\frac{1}{\sqrt{2l+1}}$ is a normalization term. Equation (5) shows that spherical harmonics function is equivariant to SO(3) group. Rotating $\hat{r}$ is equal to rotate all subspace of spherical harmonics.

Clebsh-Gordan tensor product:

Generally, the Clebsh-Gordan tensor product maps two steerable spaces into one with the desired highest dimensions.

$$(\tilde{h}^{(l_1)} \otimes \tilde{h}^{(l_2)})_m^{(l)} = w \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C^{(l,m)}_{(l_1,m_1)(l_2,m_2)} h^{(l_1)}_{m_1} h^{(l_2)}_{m_2} \tag{6}$$

Where w is the learnable weight, $\tilde{h}$ is a steerable space, and $C^{(l,m)}_{(l_1,m_1)(l_2,m_2)}$ is the Clebsch-Gordan coefficient. l is the order of the steerable vector. m is the scalar value between $-l$ and l. This tensor product can take any steerable vectors with order l_1 and l_2,

and output combined steerable vector with order l.

(2) SEGIN Algorithm

Algorithm SEGINET

def SEGINET($pos, V_{node\_feature}, E_{internal}, E_{external}, N = 3, l_{pos\_max} = 3$) $E : dist$

Internal part

1. $d_{ij} = pos_i - pos_j \quad i, j \in A \mid i, j \in B$
2. $\tilde{a}_{in}^{(l)} = (Y_m^{(l)}(\frac{d_{ij}}{\|d_{ij}\|}))_{m=-l,\dots,l}^{T} \quad Y_m^{(l)} : spherical \quad harmonics, \quad \tilde{}: steerable$
3. $\tilde{V}_{in\_message} = MessagePassing_{(sum)}(V_{node-feature}, \tilde{a}^{(l)})$
4. $\tilde{h}_{node\_feature} = V_{node\_feature} \otimes_{cg}^{w} \tilde{V}_{message}$
5. $for\ i\ in\ range(N)$:
   $\tilde{h}_{internal} = SEMessagePassing(\tilde{h}_{node\_feature}, E_{internal}, \tilde{a}_{in}^{(l)}, \tilde{V}_{in\_message})$

external part

1. $d_{ij} = pos_i - pos_j \quad i \in A, j \in B | i \in B, j \in A$
2. $\tilde{a}_{ex}^{(l)} = (Y_m^{(l)}(\frac{d_{ij}}{\|d_{ij}\|}))_{m=-l,\dots,l}^{T} \quad Y_m^{(l)} : spherical \quad harmonics, \quad \tilde{}: steerable$
3. $\tilde{V}_{ex\_message} = MessagePassing_{(sum)}(\tilde{h}_{internal}, \tilde{a}^{(l)})$
4. $\tilde{h}_{node\_feature} = V_{node\_feature} \otimes_{cg}^{w} \tilde{V}_{ex\_message}$
5. $for\ i\ in\ range(N)$:
   $\tilde{h}_{external} = SEMessagePassing(\tilde{h}_{node\_feature}, E_{external}, \tilde{a}_{ex}^{(l)}, \tilde{V}_{ex\_message})$

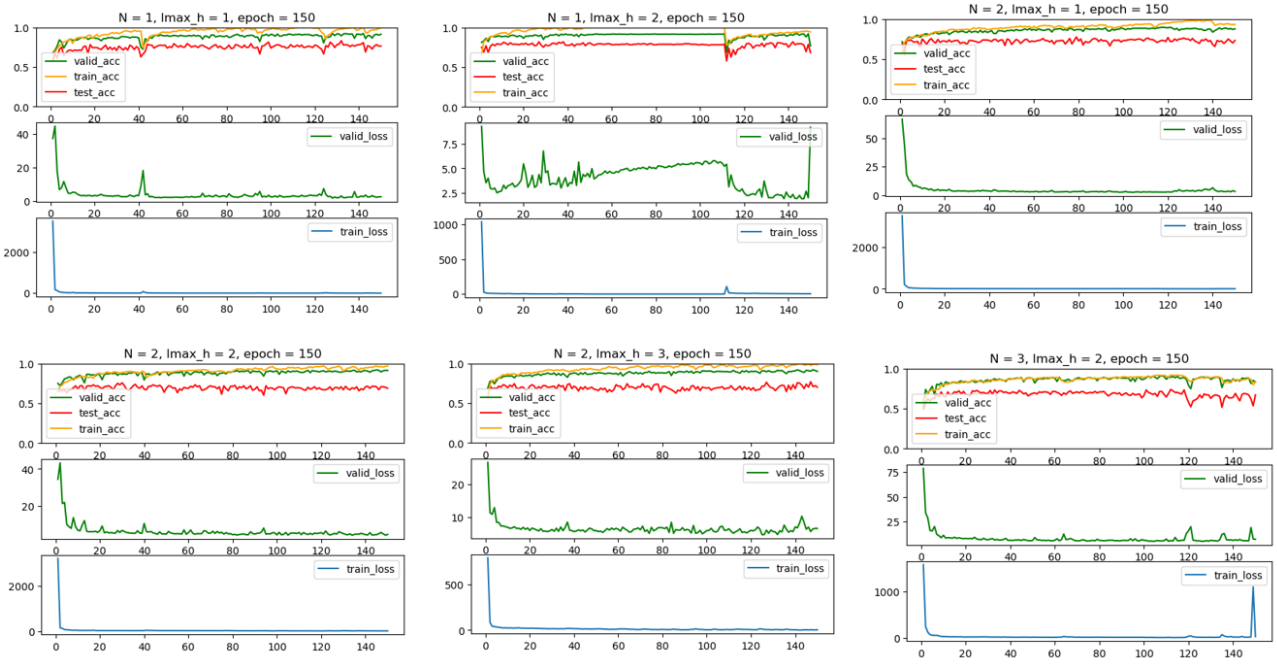
output part

1. $\hat{h}_{\text{pre}} = \text{Gate}(\hat{h}_{\text{external}} \otimes_{cg}^{w} \hat{V}_{\text{ex\_message}})$
2. $m = \text{Global\_add\_pool}(\hat{h}_{\text{pre}})$
3. $\hat{h}_{\text{post}_1} = \text{Gate}(\text{MLP}(m))$
4. $\hat{h}_{\text{post}_2} = \text{BatchNorm}(\text{MLP}(m))$
5. $P_{\text{pred}} = \text{Softmax}(\hat{h}_{\text{post}_2})$
6. $\text{return} P_{\text{pred}}$

def SEMessagePassing($\hat{h}_{\text{node\_feature}}, E, \hat{a}^{(l)}, \hat{V}_{\text{message}}$)

1. $\tilde{m}_1 = Gate((\hat{h}_i, \hat{h}_j, E) \otimes_{cg}^{w} \hat{a}^{(l)})$
2. $\tilde{m}_2 = \text{Gate}(\tilde{m}_1 \otimes_{cg}^{w} \hat{a}^{(l)})$
3. $\text{Att} = \text{Sigmoid}(\text{BatchNorm}(\tilde{m}_2 \otimes_{cg}^{w} \hat{a}^{(l)}))\backslash$
4. $\tilde{m} = \tilde{m}_2 * Att$
5. $\hat{h}_{\text{node\_feature}} = Softmax(\tilde{m}) * \hat{h}_{\text{node\_feature}}$
6. $\hat{h}_{\text{out}} = (\hat{h}_{\text{node\_feature}}, \tilde{m}) \otimes_{cg}^{w} \hat{V}_{\text{message}}$
7. return $\hat{h}_{\text{out}}$

(3) SEGIN training plots with different hyperparameters



5. Data and code availability

Residue mutation analysis is available at
https://github.com/haddocking/point-mutation-analysis
SEGIN is available at
https://github.com/haddocking/segin

6. Reference

1. Irene M.A.N. & Janet M.T. Diversity of protein-protein interactions. The EMBO Journal Vol. 22 No. 14 3486-3491 (2003).
2. Geng, C. et al. Structural bioinformatics iScore: a novel graph kernel-based function for scoring protein-protein docking models. Bioinformatics 36, 112–121 (2019).

3. Gromiha, M. M., Yugandhar, K. & Jemimah, S. Protein–protein interactions: scoring schemes and binding affinity. Current Opinion in Structural Biology vol. 44 31–38 Preprint at https://doi.org/10.1016/j.sbi.2016.10.016 (2017).
4. Stephanie L. & Ernesto F. Direct measurement of protein binding energetics by isothermal titration calorimetry. Biophysical methods 560-566 (2001).
5. Jubb, H. C. et al. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. Progress in Biophysics and Molecular Biology vol. 128 3–13 Preprint at https://doi.org/10.1016/j.pbiomolbio.2016.10.002 (2017).
6. Kim, L. M. & G, A. W. Combinatorial Alanine-Scanning Current Opinion in Chemical Biology, 5:302–307 (2001).
7. Petrey, D., Zhao, H., Trudeau, S. J., Murray, D. & Honig, B. PrePPI: A Structure Informed Proteome-wide Database of Protein–Protein Interactions. J Mol Biol 435, (2023).
8. Geng, C., Xue, L. C., Roel-Touris, J. & Bonvin, A. M. J. J. Finding the ΔΔG spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it? Wiley Interdisciplinary Reviews: Computational Molecular Science vol. 9 Preprint at https://doi.org/10.1002/wcms.1410 (2019).
9. Krapp, L. F., Abriata, L. A., Cortés Rodriguez, F. & Dal Peraro, M. PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. Nat Commun 14, (2023).
10. Réau, M., Renaud, N., Xue, L. C. & Bonvin, A. M. J. J. DeepRank-GNN: a graph neural network framework to learn patterns in protein–protein interfaces. Bioinformatics 39, (2023).
11. Xu, X. & Bonvin, A. M. J. J. DeepRank-GNN-esm: A graph neural network for scoring protein-protein models using protein language model. Bioinformatics Advances 4, (2024).
12. Goldblum, M. et al. Perspectives on the State and Future of Deep Learning - 2023. arXiv preprint arXiv:2312.09323 (2023). (2023).
13. Gligorijević, V. et al. Structure-based protein function prediction using graph convolutional networks. Nat Commun 12, (2021).
14. Zhang, Z. et al. Protein Representation Learning by Geometric Structure Pretraining. (2022).
15. Jha, K., Saha, S. & Singh, H. Prediction of protein–protein interaction using graph neural networks. Sci Rep 12, (2022).
16. Lin, Z. et al. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. Science vol. 379 https://www.science.org (2023).
17. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125, 1731–1737 (2003).
18. Vangone, A. & Mjj Bonvin, A. Contacts-based prediction of binding affinity in protein-protein complexes. doi:10.7554/eLife.07454.001.
19. Geng, C., Vangone, A., Folkers, G. E., Xue, L. C. & Bonvin, A. M. J. J. iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations. Proteins: Structure, Function and Bioinformatics 87, 110–119 (2019).
20. Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. J. & Welling, M. Geometric and Physical Quantities Improve E(3) Equivariant Message Passing. In ICLR (2021).
21. Jankauskaite, J., Jiménez-García, B., Dapkunas, J., Fernández-Recio, J. & Moal, I. H. SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. Bioinformatics 35, 462–469 (2019).
22. Mitternacht, S. FreeSASA: An open-source C library for solvent accessible surface area calculations. F1000Res 5, (2016).

23. Brunger, A. T. et al. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. Acta Cryst vol. 54 (1998).
24. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. Nucleic Acids Research vol. 25 https://academic.oup.com/nar/article/25/17/3389/1061651 (1997).
25. Duarte, J. M., Srebniak, A., Schärer, M. A. & Capitani, G. Protein Interface Classification by Evolutionary Analysis. http://www.biomedcentral.com/1471-2105/13/334 (2012).
26. Baskaran, K., Duarte, J. M., Biyani, N., Bliven, S. & Capitani, G. A PDB-Wide, Evolution-Based Assessment of Protein-Protein Interfaces. BMC Structural Biology vol. 14 http://www.biomedcentral.com/1472-6807/14/22 (2014).
27. Elez, K., Bonvin, A. M. J. J. & Vangone, A. Distinguishing crystallographic from biological interfaces in protein complexes: Role of intermolecular contacts and energetics for classification. BMC Bioinformatics 19, (2018).
28. Thomas, N. et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. In NIPS (2018).
29. Geiger, M. & Smidt, T. e3nn: Euclidean Neural Networks. arXiv preprint arXiv:2207.09453 (2022).
30. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. J Mol Biol 372, 774–797 (2007).
31. Jiménez-García, B. et al. PRODIGY-crystal: A web-tool for classification of biological interfaces in protein complexes. Bioinformatics 35, 4821–4823 (2019).
32. Huang, X., Zheng, W., Pearce, R., Zhang, Y. & Zhang, Y. SSIPe: Accurately estimating protein-protein binding affinity change upon mutations using evolutionary profiles in combination with an optimized physical energy function. Bioinformatics 36, 2429–2437 (2020).
33. Lu, W., Zhang, J., Rao, J., Zhang, Z. & Zheng, S. AlphaFold3, a secret sauce for predicting mutational effects on protein-protein interactions. doi:10.1101/2024.05.25.595871.
32. Barlow, K. A. et al. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. Journal of Physical Chemistry B 122, 5389–5399 (2018).
33. Schymkowitz, J. et al. The FoldX web server: An online force field. Nucleic Acids Res 33, (2005).
34. Huang, X., Pearce, R. & Zhang, Y. EvoEF2: Accurate and fast energy function for computational protein design. Bioinformatics 36, 1135–1142 (2020).
35. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500 (2024).