

Geometric Deep Learning for Protein-Protein Binding Affinity Prediction

Arne Schneuing, Ilia Igashov
École polytechnique fédérale de Lausanne

February 9, 2022

Introduction

Proteins are involved in several biological reactions by means of interactions with other proteins or with other molecules such as nucleic acids, carbohydrates, and ligands. Among these interaction types, protein-protein interactions (PPIs) are considered to be one of the key factors as they are involved in most of the cellular processes [1].

The binding of two proteins can be viewed as a reversible and rapid process in an equilibrium that is governed by the law of mass action. Binding affinity is the strength of the interaction between two (or more than two) molecules that bind reversibly (interact). It is translated into physico-chemical terms in the dissociation constant K_d , the latter being the concentration of free protein at which half of all binding sites of the second protein type are occupied [2].

Predicting the affinity of protein-protein complexes has been a topic of active research for more than two decades. The availability of experimental data on binding affinity prompted researchers to explore the principles and develop methods for prediction [1]. However, the amount of experimentally observed three-dimensional protein-protein complexes with known binding constants still remains extremely limited [3, 4], which complicates the application of modern deep learning methods in this task. The most recent computational research on PPI binding affinity prediction is mainly built around the idea of utilizing standard statistical [5, 6] and machine learning methods [4, 7, 8, 9] trained on various handcrafted descriptors such as QSAR features [8], inter-residue contacts and buried surface area [5], surface tension area and hydrophobicity [4], and sequence-based descriptors [6, 9].

Recent advances in adjacent computational problems such as protein structure prediction [10] or protein-protein interaction prediction [11] demonstrated how powerful deep learning methods can be as long as enough training data is provided and correct inductive biases are set up.

In this work, we will apply geometric deep learning methods for predicting protein-protein binding affinity. We believe that geometry is a clue for understanding protein-protein interactions, and aim to notably move forward the state of the art in binding affinity prediction with the aid of graph neural networks. To the best of our knowledge, geometric deep learning methods have never been applied to the protein-protein binding affinity prediction problem so far.

Objectives

Three main objectives of this work can be formulated as follows:

- Refine PDBbind [12] data and a standard binding affinity dataset [3], and compile a novel benchmark of PPIs with known binding affinity values
- Employ graph-learning toolset to predict binding affinities of PPIs from the new dataset

- Benchmark the resulting method against existing state-of-the-art approaches

Data & Metrics

In this work, we will operate on experimentally-observed three-dimensional structures of protein-protein complexes annotated with the binding affinity values. Two main sources of data are the following:

- PDBbind dataset [12] that includes around 2k PPIs
- Standard dataset introduced in [3] that includes 144 PPIs

As main regression metrics, we suggest to consider Mean Squared Error (MSE), Mean Absolute Error (MAE) and Pearson correlation.

Requirements

- Python and PyTorch (medium level and higher), Git, Bash
- Fluent English
- Background in computational biology is a plus

References

- [1] K Yugandhar and M Michael Gromiha. Computational approaches for predicting binding partners, interface residues, and binding affinity of protein-protein complexes. In *Prediction of Protein Secondary Structure*, pages 237–253. Springer, 2017.
- [2] Panagiotis L Kastritis and Alexandre MJJ Bonvin. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *Journal of The Royal Society Interface*, 10(79):20120835, 2013.
- [3] Panagiotis L Kastritis, Iain H Moal, Howook Hwang, Zhiping Weng, Paul A Bates, Alexandre MJJ Bonvin, and Joël Janin. A structure-based benchmark for protein-protein binding affinity. *Protein Science*, 20(3):482–491, 2011.
- [4] Raquel Dias and Bryan Kolaczowski. Improving the accuracy of high-throughput protein-protein affinity prediction may require better training data. *BMC bioinformatics*, 18(5):7–18, 2017.
- [5] Anna Vangone and Alexandre MJJ Bonvin. Contacts-based prediction of binding affinity in protein-protein complexes. *elife*, 4:e07454, 2015.
- [6] K Yugandhar and M Michael Gromiha. Protein-protein binding affinity prediction from amino acid sequence. *Bioinformatics*, 30(24):3583–3589, 2014.
- [7] Iain H Moal, Rudi Agius, and Paul A Bates. Protein-protein binding affinity prediction on a diverse set of structures. *Bioinformatics*, 27(21):3002–3009, 2011.
- [8] Feifei Tian, Yonggang Lv, and Li Yang. Structure-based prediction of protein-protein binding affinity with consideration of allosteric effect. *Amino Acids*, 43(2):531–543, 2012.
- [9] Wajid Arshad Abbasi, Adiba Yaseen, Fahad Ul Hassan, Saiqa Andleeb, and Fayyaz Ul Amir Af-sar Minhas. Island: in-silico proteins binding affinity prediction using sequence information. *BioData Mining*, 13(1):1–13, 2020.
- [10] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

- [11] Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, D Boscaini, MM Bronstein, and BE Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020.
- [12] Renxiao Wang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. The pdbind database: methodologies and updates. *Journal of medicinal chemistry*, 48(12):4111–4119, 2005.