Project Proposal for AI-powered Molecular Modeling (CS 6825, Spring 2022)

Project name: Sequence-driven protein-protein interface intensity prediction

Presented by: Rahmatullah Roche

Motivation

- Protein-protein interaction highly important for drug discovery and design
- Experimentally: cost, time …
- Recent Development in protein structure prediction
- Computationally: Predict protein-protein interface [1-4]
- Sequence-driven model can improve interface intensity

Primary Approach

- Dataset
  - QS30 train (1400 monomers)
  - QS30 validation (120 monomers) set
  - DB5 (ZDOCK 230 complexes)

Proposed Method

- Sequential property of protein (sequence of amino acids …)
- AI method that leverage sequential information
- LSTM/GRU

Features

- Sequence-based features: FASTA sequence, PSSM
- Structure-based features: 3D-coordinates, motifs: SS, SA, …

Output

- Interface intensity
- Will discretize the distances to 10 bins
- <3, <4, <5, <6, <7, <8, <9, <10, <12, <15, <20, and >20 Angstroms

Methods to Compare

- two recent state-of-the-art interface prediction methods:
  - BIPSPI[1]
  - PINet[2]

Evaluation

- Precision
- Recall
- F1-score
- ROC-AUC
- PR-AUC
- Pearson correlations

Proposed Tasks and Timeline

Milestone Task	Date (From Month Date – To Month Date)
Feature extraction: Generating sequence- and structure-based features for all dataset	March 4 – March 7
Model creation: Exploring different RNN (or sequence-based) models	March 8 – March 14
Model Training	March 15 – March 27
Testing and validation	March 28 – April 3
Comparison with other methods	March 4 – April 9
Ablation studies	April 10 – April 12
Writing report	April 13 – May 4
Presenting result	April 13

References

- [1] R. Sanchez-Garcia, C. O. S. Sorzano, J. M. Carazo, and J. Segura, "BIPSPI: a method for the prediction of partner-specific protein–protein interfaces," *Bioinformatics*, vol. 35, no. 3, pp. 470–477, Feb. 2019, doi: 10.1093/bioinformatics/bty647.
- [2] B. Dai and C. Bailey-Kellogg, "Protein interaction interface region prediction by geometric deep learning," *Bioinformatics*, vol. 37, no. 17, pp. 2580–2588, Sep. 2021, doi: 10.1093/bioinformatics/btab154.
- [3] F. ul A. A. Minhas, B. J. Geiss, and A. Ben-Hur, "PAIRpred: partner-specific prediction of interacting residues from sequence and structure," *Proteins*, vol. 82, no. 7, pp. 1142–1155, Jul. 2014, doi: 10.1002/prot.24479.
- [4] R. Townshend, R. Bedi, P. Suriana, and R. Dror, "End-to-End Learning on 3D Protein Structure for Interface Prediction," in *Advances in Neural Information Processing Systems*, 2019, vol. 32. Accessed: Mar. 04, 2022.

Questions?