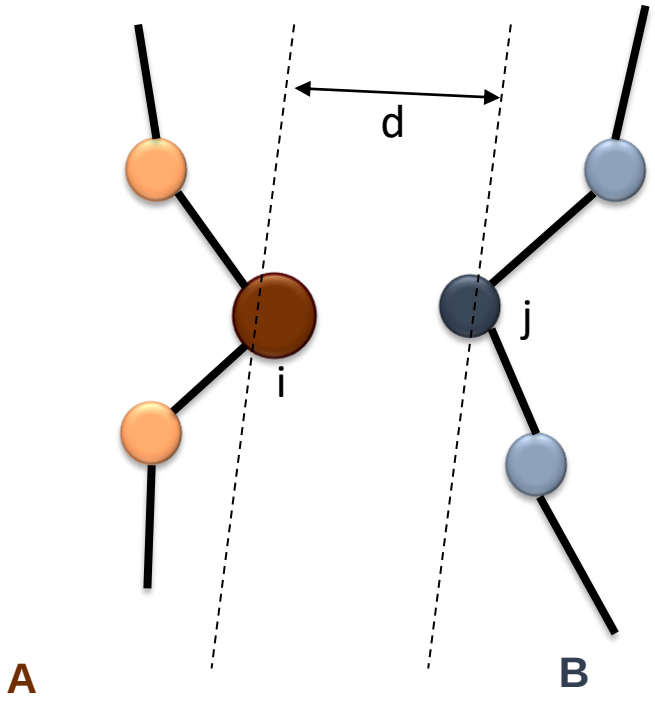
# Project Presentation for AI-powered Molecular Modeling (CS 6825, Spring 2022)

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

Presented by: Rahmatullah Roche

COMPUTER SCIENCE
VIRGINIA TECH

# Interface



# 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

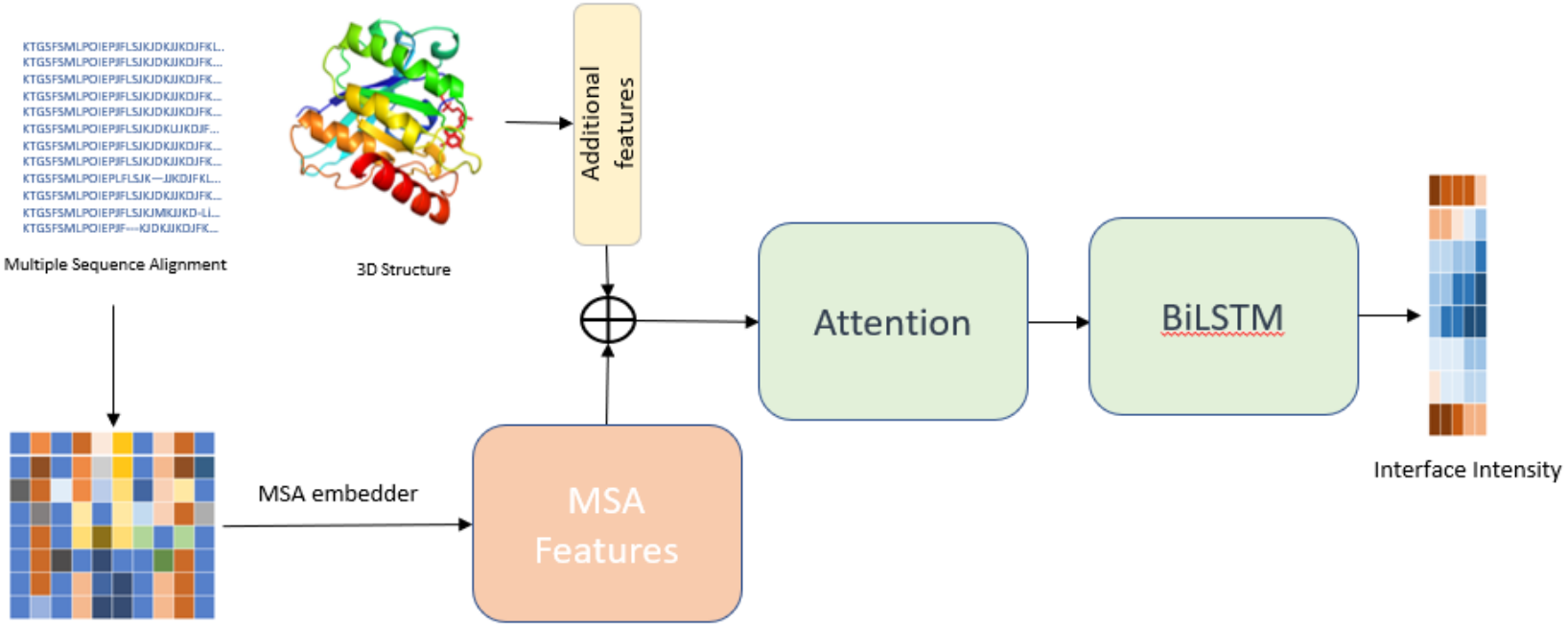
# Primary Approach

- Dataset
  - 3D complex(http://shmoo.weizmann.ac.il/elevy/3dcomplexV6/Home.cgi)
    - 760 complexes
    - QS30 train (1200 monomers)
    - QS30 validation (296 monomers) set
  - DB5(https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/H93ZKK)
    - 230 complexes

# Proposed Method

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

# Method Overview



# Features

- Sequence-based features: FASTA sequence, PSSM, MSA
- Structure-based features: 3D-coordinates, SS, SA, contact count, angles

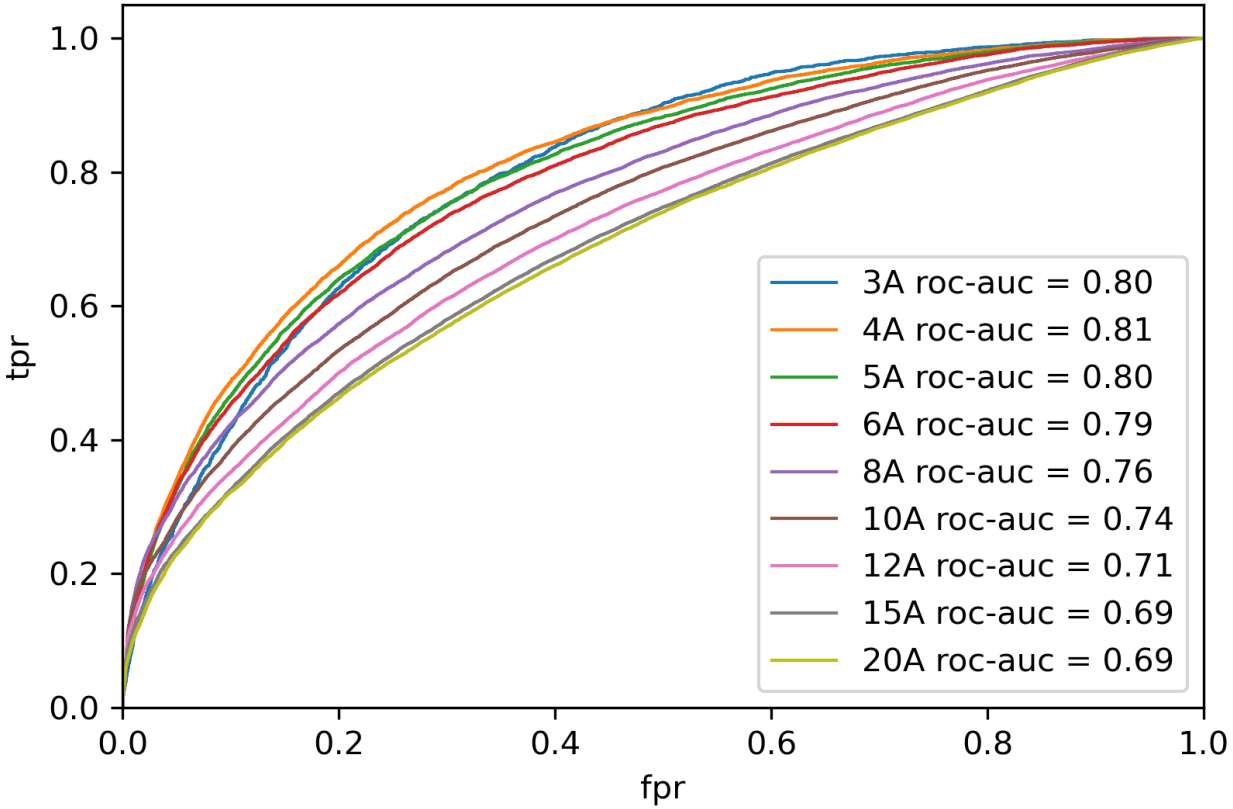
# Output

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

# Methods to Compare

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

# Validation on QS30 validation



# Validation on QS30 validation

| Threshold | Prec | Recall | F1 |
|---|---|---|---|
| 3 | 0.42 | .02 | .04 |
| 4 | .57 | .18 | .27 |
| 5 | .59 | .2 | .3 |
| 6 | .63 | .24 | .35 |
| 8 | .68 | .3 | .42 |
| 10 | .67 | .34 | .46 |
| 12 | .65 | .43 | .51 |
| 15 | .61 | .72 | .66 |
| 20 | .68 | .99 | .8 |

# Comparison with other methods (ZDOCK)

| Method | Precision | ROC-AUC |
|---|---|---|
| Baseline | .55 | .65 |
| PInet | .492 | .75 |
| BIPSPI | .39 | .82 |

## Working on more benchmarking

# Ablation Study

On features

| Feature | ROC-AUC | PR-AUC | Precision |
|---|---|---|---|
| Baseline | 79 | 46 | 63 |
| w/o MSA-embed | 78 | 45 | 59 |
| w/o sequences + pssm | 76 | 41 | 58 |
| w/o structural | 57 | 20 | 42 |

On Architectures

| Feature | ROC-AUC | PR-AUC | Precision |
|---|---|---|---|
| Baseline | 79 | 46 | 63 |
| Transformer | 69 | 33 | 61 |
| BiLSTM | 76 | 42 | 53 |
| BiLSTM+Attention | 55 | 23 | 16 |

# 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?