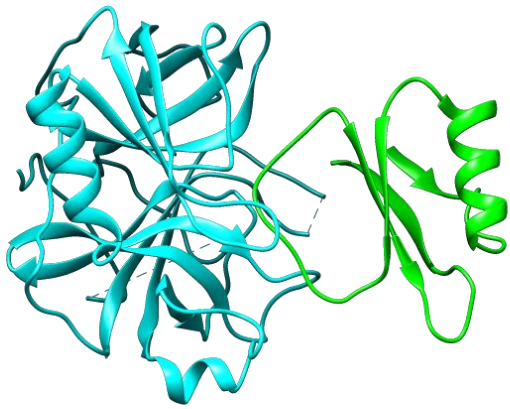


# **Estimation of interfacial quality of protein complex models**

Md Hossain Shuvo  
Virginia Tech

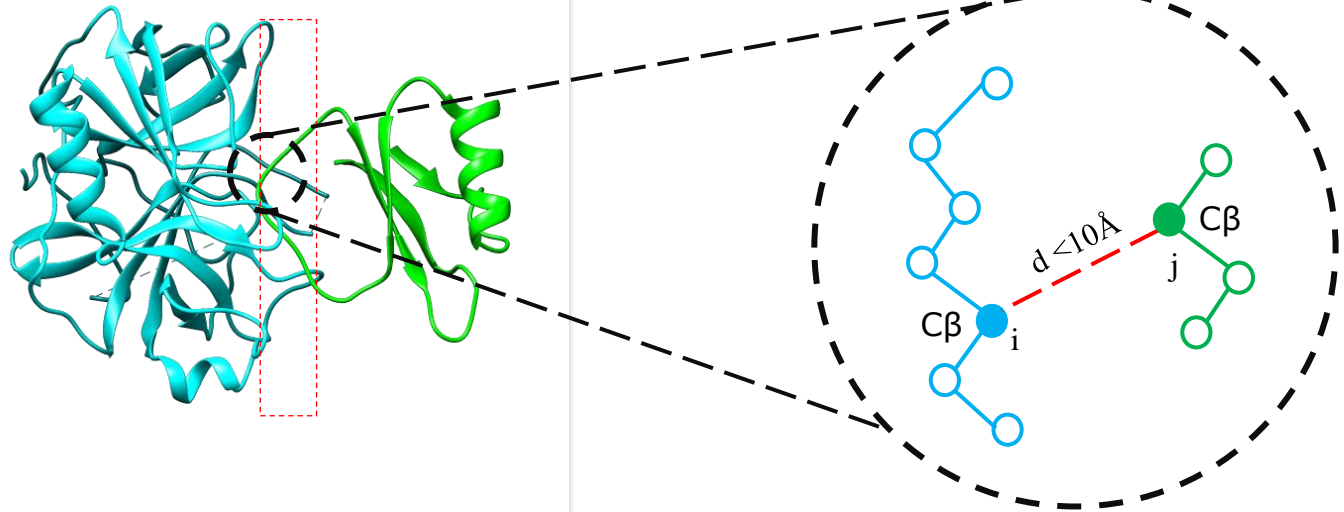
# Background

Protein Complex

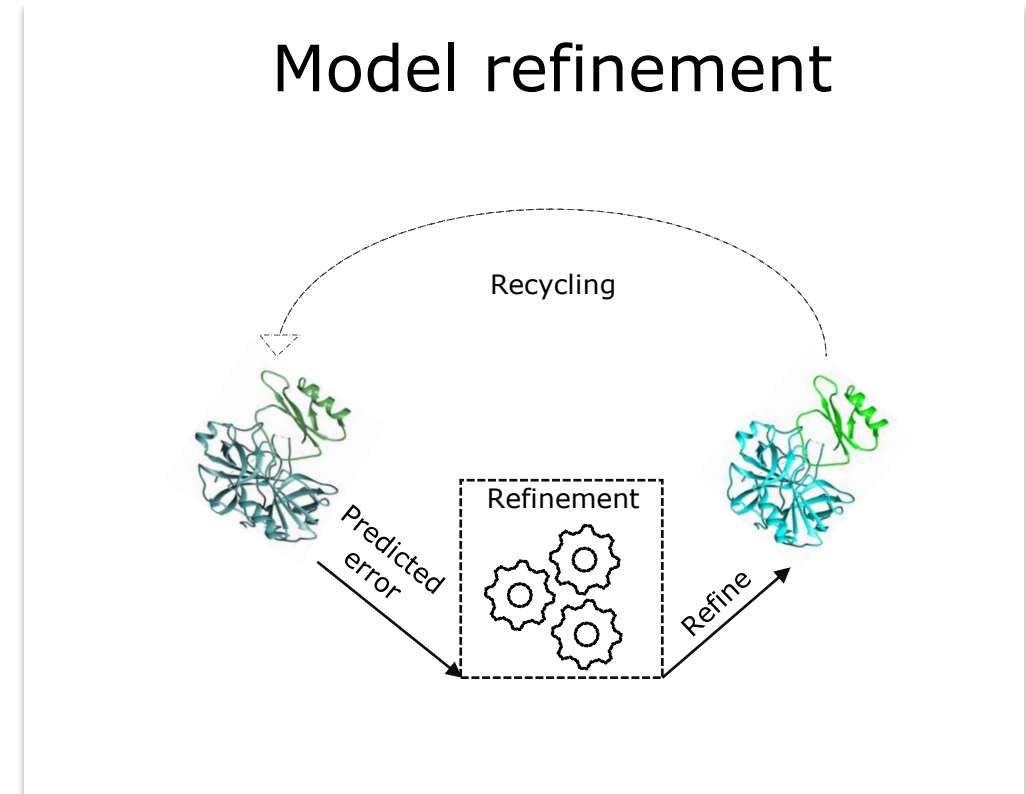
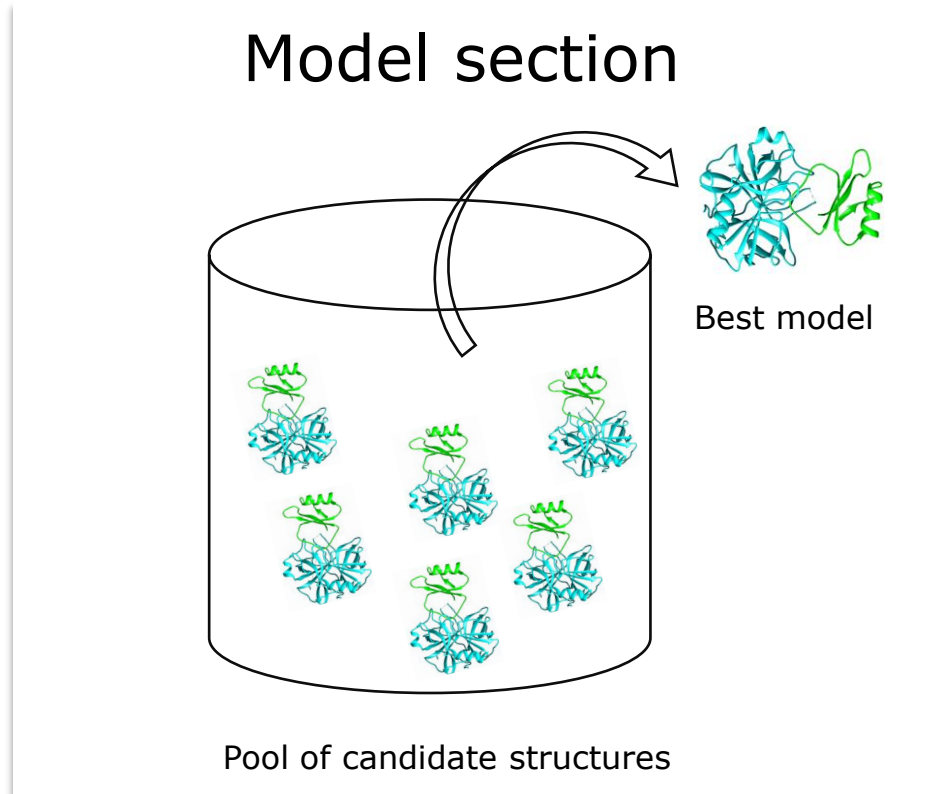


Crystal structure of 1ACB

Interfacial region



# Motivation



Helps in accurately guiding the process of protein complex prediction

# Approach

- Dataset curation
- Feature extraction
- Model training
- Quality estimation

# Dataset

## Training

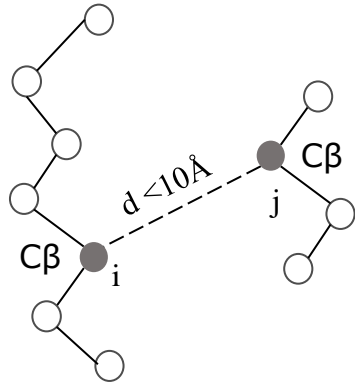
- DOCKGROUND docking decoy set 2
- 180 complex targets
- ~18000 docking decoys

## Testing

- DOCKGROUND docking decoy set 1
- 23 complex targets
- ~2600 docking decoys

# Feature extraction

Graph  
representation  
 $G(\text{nodes, edges})$



Node features

- Amino acids encoding
- Secondary structure and solvent accessibility encoding
- Relative residue positioning
- MSA based features (NEFF)

Edge features

- Dihedral angles
  - Phi
  - Psi
- Orientation between the connecting nodes
  - theta
  - Omega
  - phi

# Quality estimation

## Initial deep learning model

- Graph neural network
- Ideal for learning for graph representation
- Regression problem

## Target label

- For each edge (local quality)
- $d_i = 10$

$$s\_score = \frac{1}{1 + \left(\frac{d}{d_i}\right)^2}$$

- Global quality

$$global_{quality} = \frac{\sum_1^e s\_score_e}{n}$$

# Evaluation metrics

- Ground truth:
  - Observed s-score
  - iRMSD (if time permits)
- Pearson correlation between  $\text{global}_{\text{quality}}$  and the s-score
- Spearman correlation between  $\text{global}_{\text{quality}}$  and the s-score
- Kendall's Tau correlation between  $\text{global}_{\text{quality}}$  and the s-score



# Competing methods (if time permits)

- DOVE
- GNN-DOVE
- DeepRank

# Milestones

Table1: Tentative timeline for the individual task for the proposed project

<b>Task</b>	<b>Tentative completion date</b>
Data curation	3/11/2022
Feature generation	3/25/2022
Initial network construction	4/1/2022
Network training	4/8/2022
Evaluation	4/13/2022

# Acknowledgement



**Debswapna Bhattacharya, Ph.D.**  
**Associate Professor**  
**Virginia Tech**

# References

1. Rao VS, Srinivas K, Sujini GN, Kumar GNS. Protein-Protein Interaction Detection: Methods and Analysis. *Int J Proteomics* 2014;2014:147648.
2. Kundrotas PJ, Anishchenko I, Dauzhenka T, Kotthoff I, Mnevets D, Copeland MM, Vakser IA. Dockground: A comprehensive data resource for modeling of protein complexes. *Protein Sci* 2018;27(1):172–181.
3. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. *IEEE Transactions on Neural Networks* 2009;20(1):61–80.
4. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins* 2018;86 Suppl 1:257–273.
5. Wang X, Terashi G, Christoffer CW, Zhu M, Kihara D. Protein docking model evaluation by 3D deep convolutional neural networks. *Bioinformatics* 2020;36(7):2113–2118.
6. Wang X, Flannery ST, Kihara D. Protein Docking Model Evaluation by Graph Neural Networks. *Frontiers in Molecular Biosciences* 2021;8.
7. Renaud N, Geng C, Georgievskaja S, Ambrosetti F, Ridder L, Marzella DF, Réau MF, Bonvin AMJJ, Xue LC. DeepRank: a deep learning framework for data mining 3D protein-protein interfaces. *Nat Commun* 2021;12(1):7068.