

Bs'd

# Context-dependent design of induced-fit enzymes using deep learning generates well-expressed, thermally stable and active enzymes

Research conducted  
during my tenure at



Chen Brestel | 5785 | 2025

Dr. Chen Brestel



Independent  
AI Scientist



Paper: Zimmerman et al. PNAS 2024

# How to design a novel enzyme?

What are enzymes good for?

Industrial applications

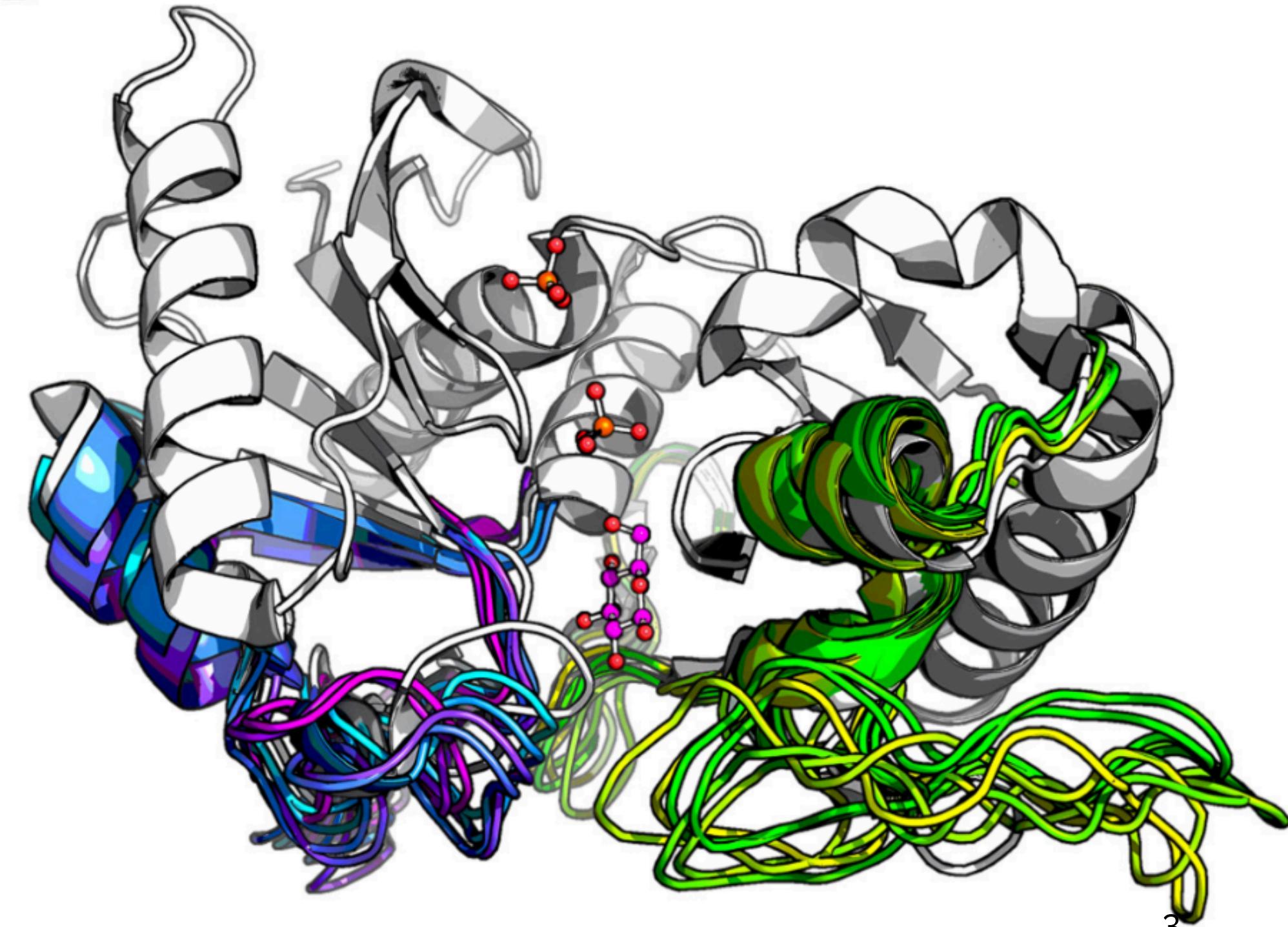
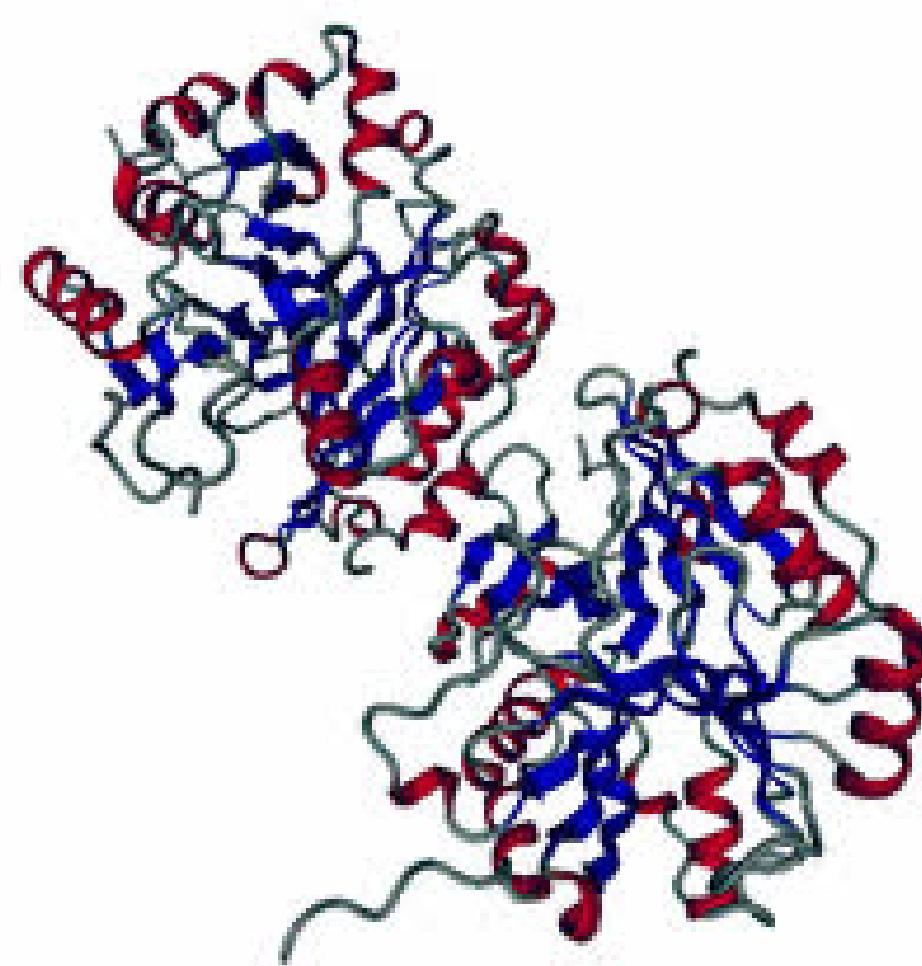
- Healthcare
- Foodtech
- Energy
- Agtech
- Environment
- Military/HLS
- etc.

Why we need novel enzymes?

- Different substrate
- Different temperature
- Better activity
- etc.

# How does enzymes look?

B



AndrewGNF

# Current Approaches, Fields & Tools

- Fast evolution
- Protein folding
  - Rosetta
  - David Baker, Washington, Lab tools
  - AlphaFold 2 & 3 - Deepmind
- Protein language models
  - ESM2, ESM3 - Facebook
  - ProteinGPT
  - ProGen - Salesforce & Stanford

The Nobel Prize in Chemistry  
2024



Ill. Niklas Elmehed © Nobel Prize Outreach  
**David Baker**

Prize share: 1/2



Ill. Niklas Elmehed © Nobel Prize Outreach  
**Demis Hassabis**

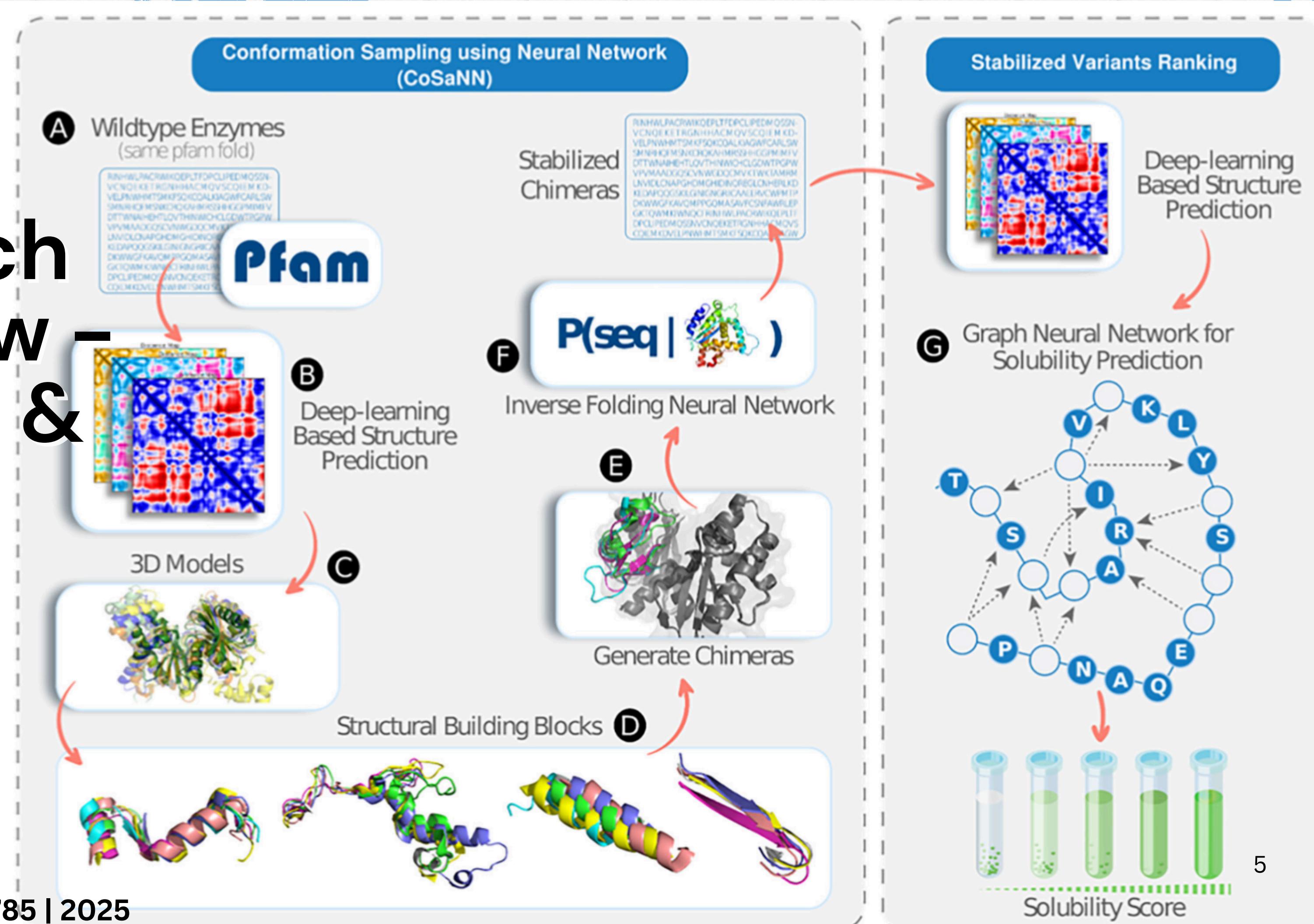
Prize share: 1/4



Ill. Niklas Elmehed © Nobel Prize Outreach  
**John Jumper**

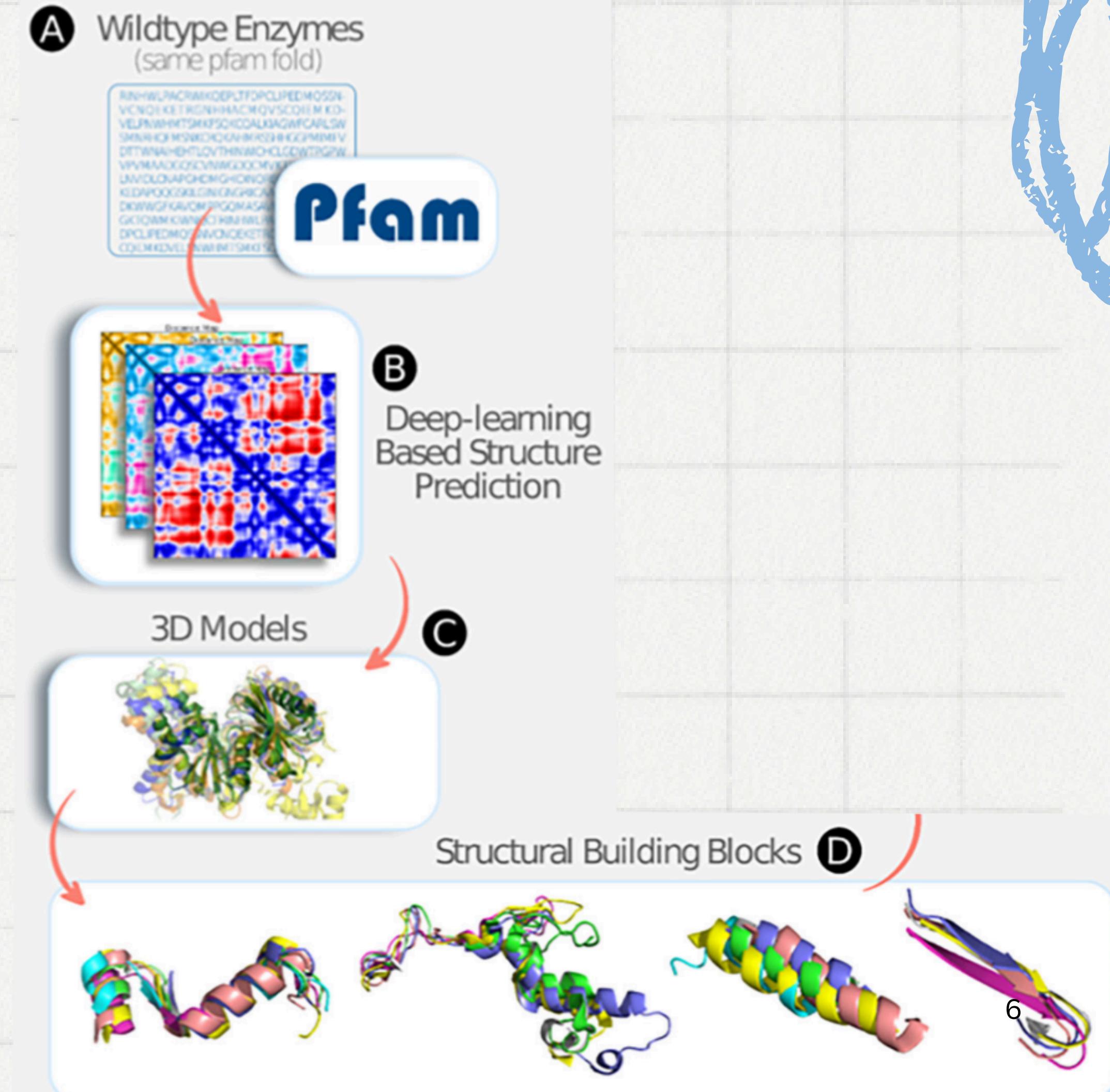
Prize share: 1/4

# Approach Overview – CoSaNN & SolvIT



# How to find structural building blocks?

- Why blocks?



# How to generate stabilized chimeras?

- What are chimeras?
- Why chimeras?
- What are stabilized chimeras?
- Why stabilized chimeras?

Stabilized  
Chimeras

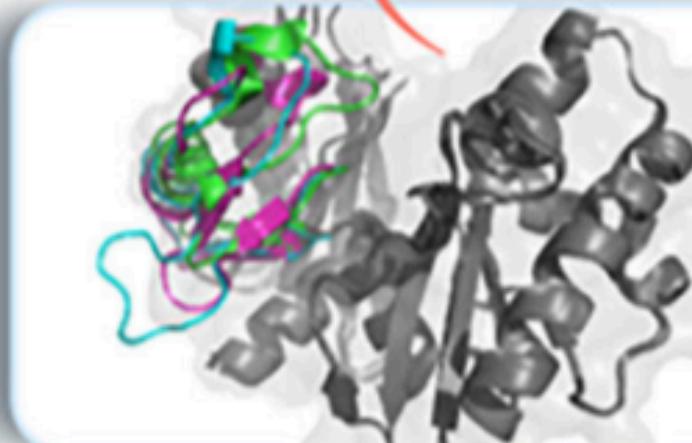
RHHWLPACRAIKQELTIDPCLPEDMQSN-  
VCNQEKEFRGNHHAQM QVSCDEH KDV-  
VELPNWHNTSMKFSOKODAKAGAVICARLSW  
YMNHHGIMSNKDHOKAHMRSSHHGGPM MIV  
DTIWNAHEHTLQVTHINWYHOLGOWTPPW  
VIPVMAADGGSKVNWYDQOMVKEWKIAMHM  
UNVIDLONAPGHOMQHIDNOREGLQHRLKD  
KEDAFQDGSKLEGNIENGRKAALEHVCHWMTR  
DKWNGFKAVQMPGOMASAVFCNFIAWLF  
GKTQWMIKWNQDQKHHLWLPNQWKAQPLF  
DPCLUPEMDQSENVNQNEKETRGNHAKMVS  
COEMKQEVLPNWIITSMGTSOKODAEGIN

F



Inverse Folding Neural Network

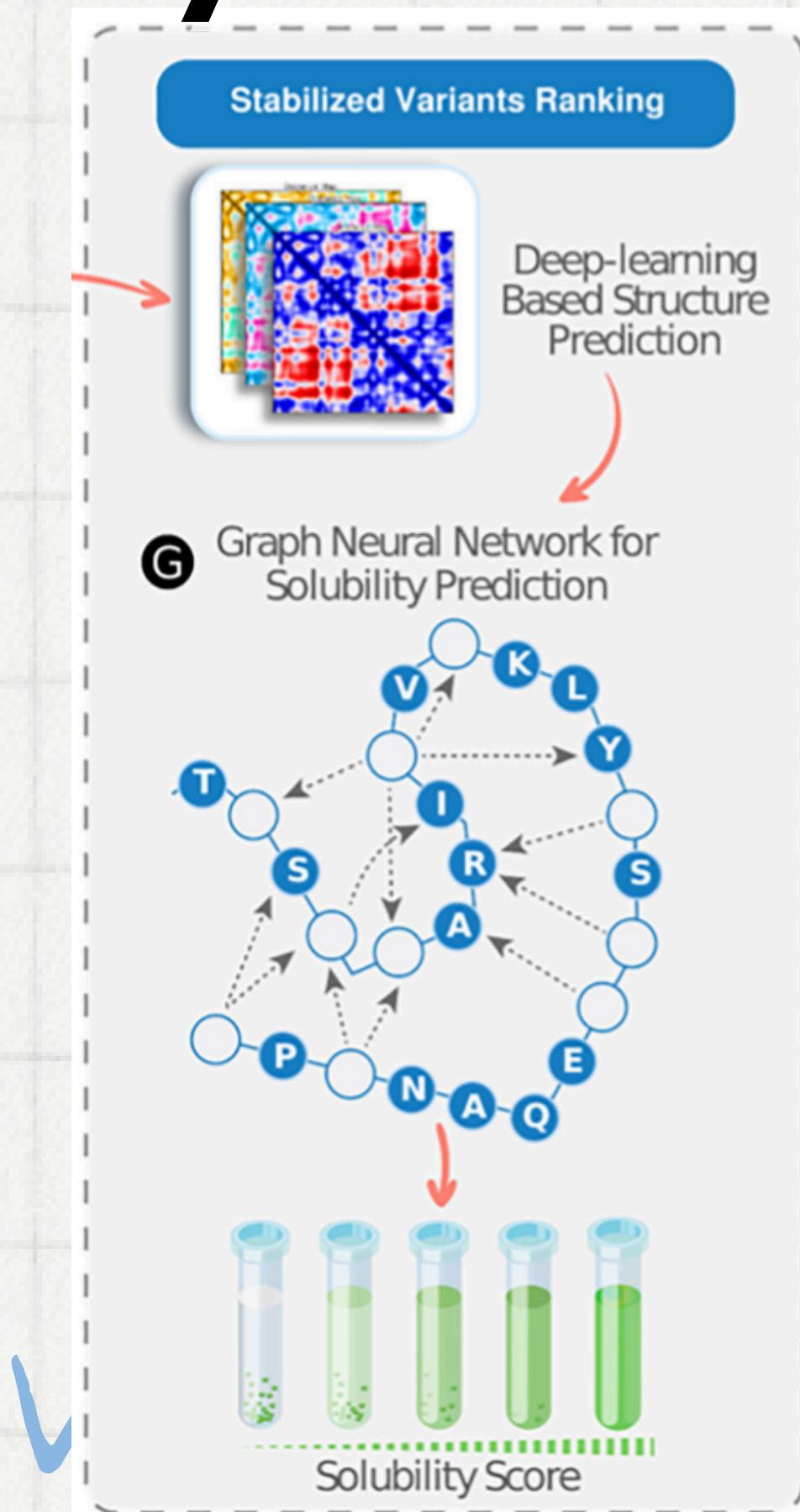
E



Generate Chimeras

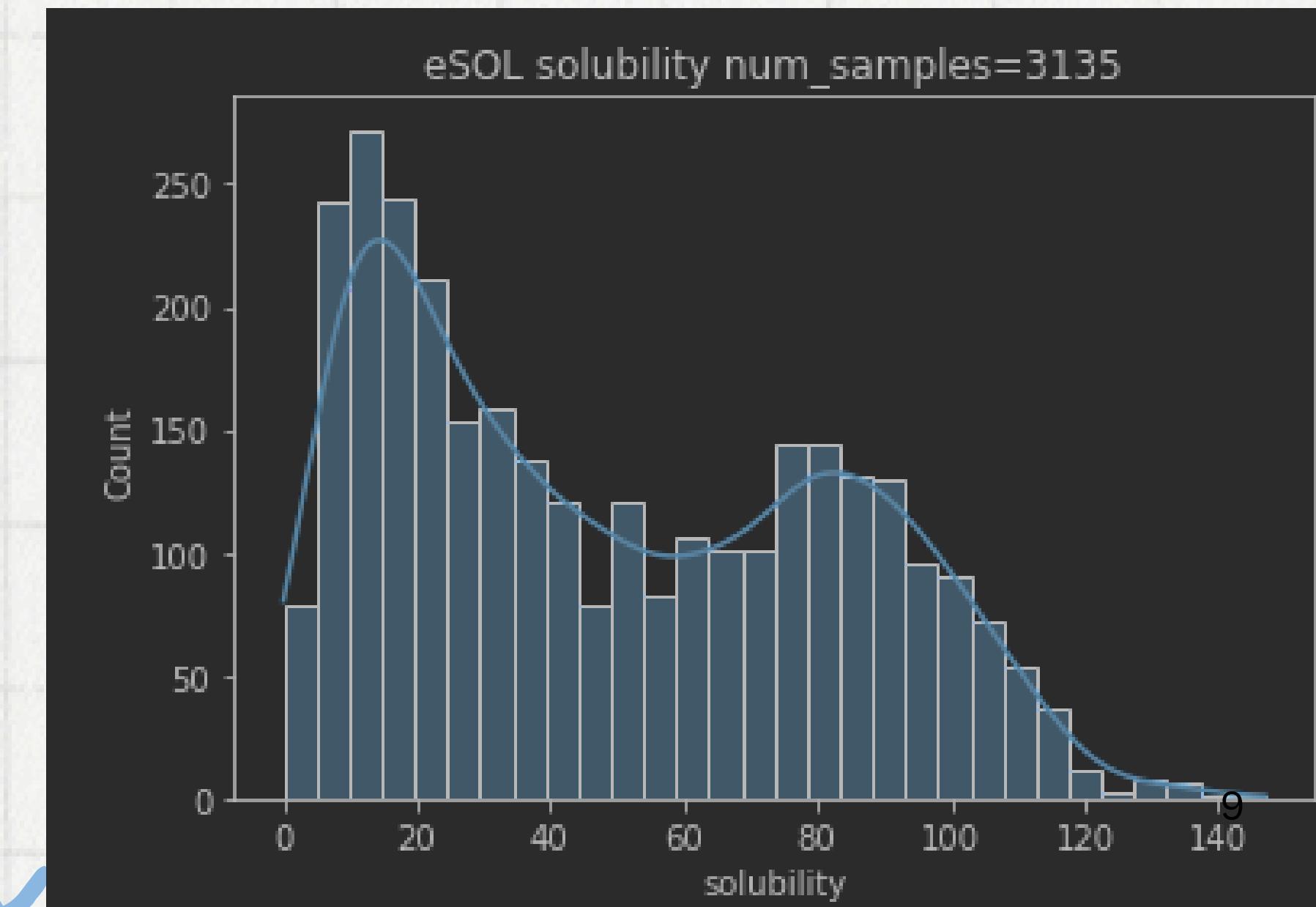
# How to get solubility-based ranking

- Why ranking?
- What is solubility-based ranking?



# What is the dataset used for training SolvIT?

- What is the target of solubility predictor?
- What is the training purpose?
- eSOL Dataset Niwa2009
- How to clean the data?
- Enzymit Dataset.
- How to ensure the prediction confidence?
- How to ensure the prediction generality?



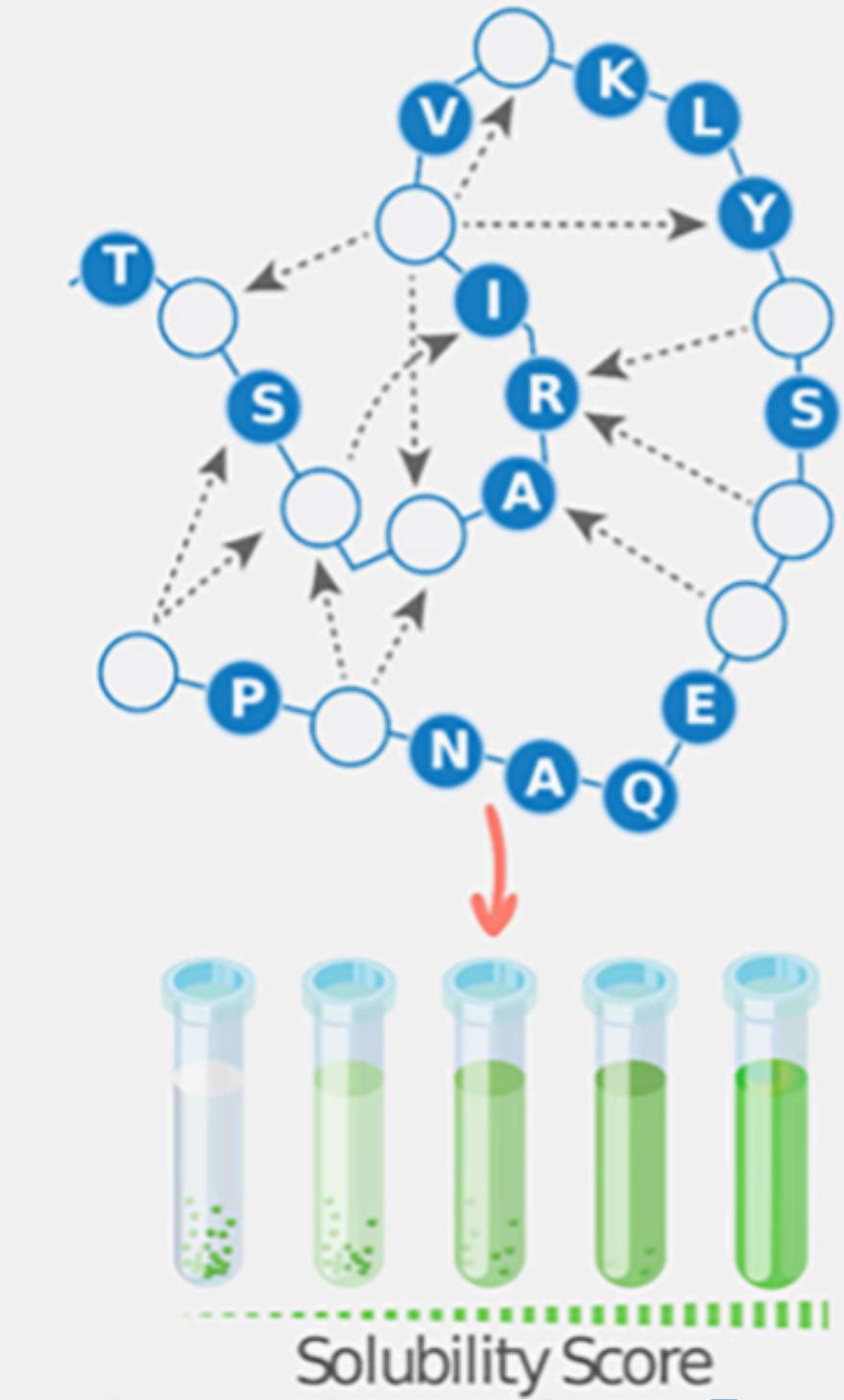
# Training experiments

- Why multiple training experiments?
- What parameters were tuned?
  - Heterogeneous/homogeneous GNN
  - Type of convolution layer
  - Type of global pooling layer
  - Number of heads
  - Number of convolution layers
  - Batchnorm
  - Neighbour distance threshold for edges
  - Binary/quantized/float rosetta features
  - ESM features

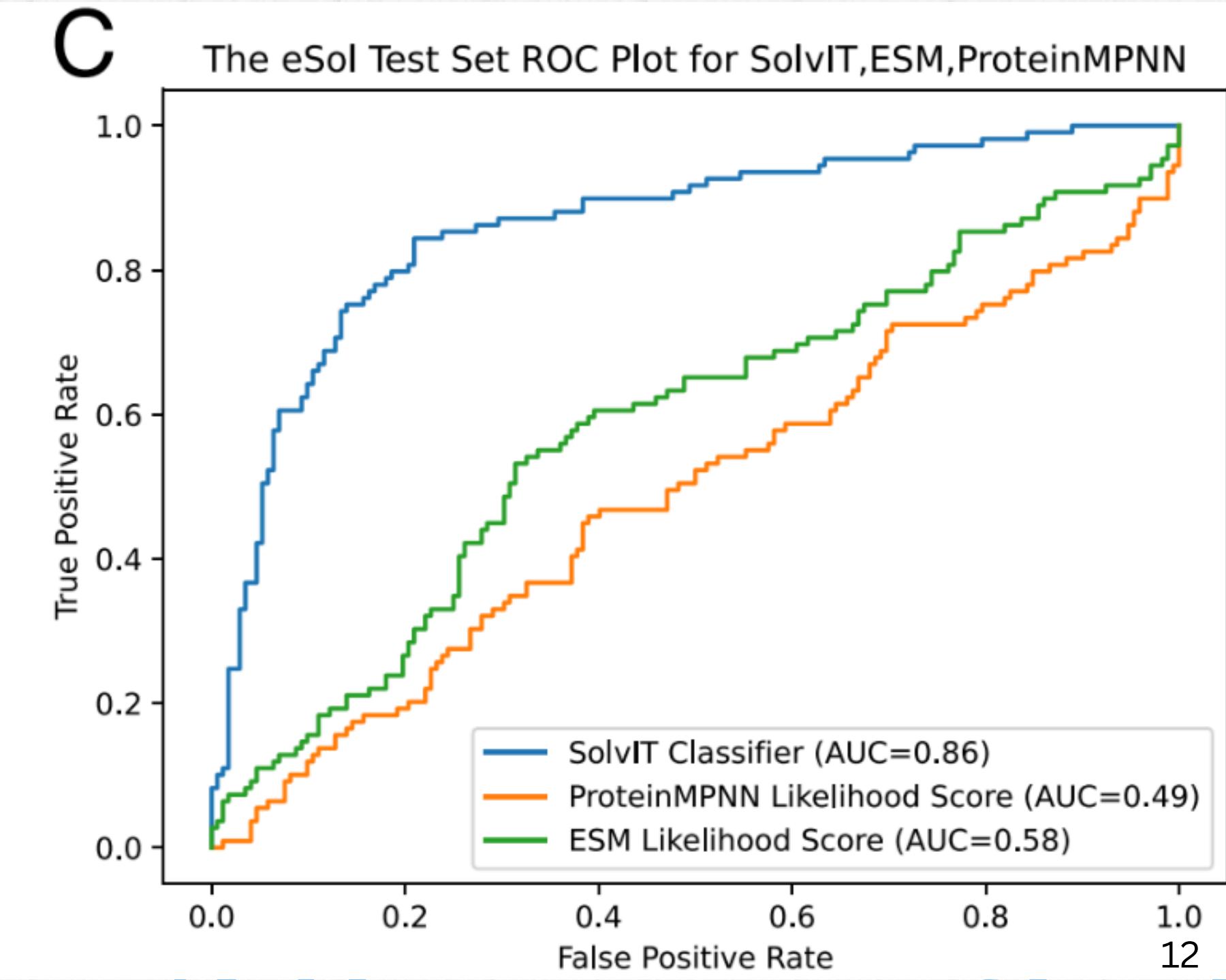
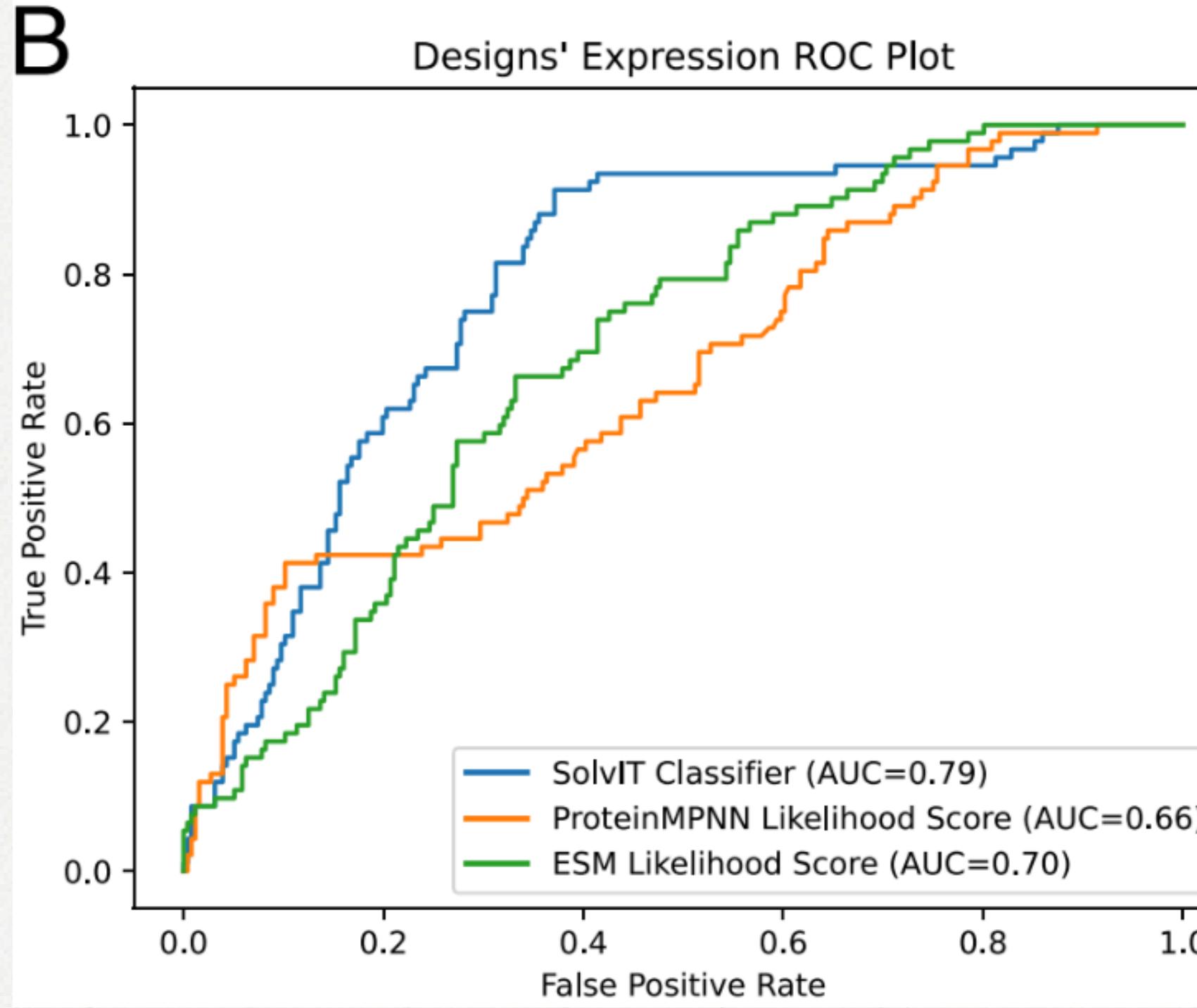
# Model architecture

- Graph neural network (GNN)
- Convolution layer: GATv2Conv [Brody2021]
- Global pooling layer: Global Graph Multiset Transformer pooling operator [Baek2021]
- Number of heads: 8
- Number of convolution layers: 2
- Batchnorm: Off
- Neighbour distance threshold for edges: 5 [Å]
- Amino-acid encoding: one-hot
- ESM embedding reduced size: Off
- Model target: binary

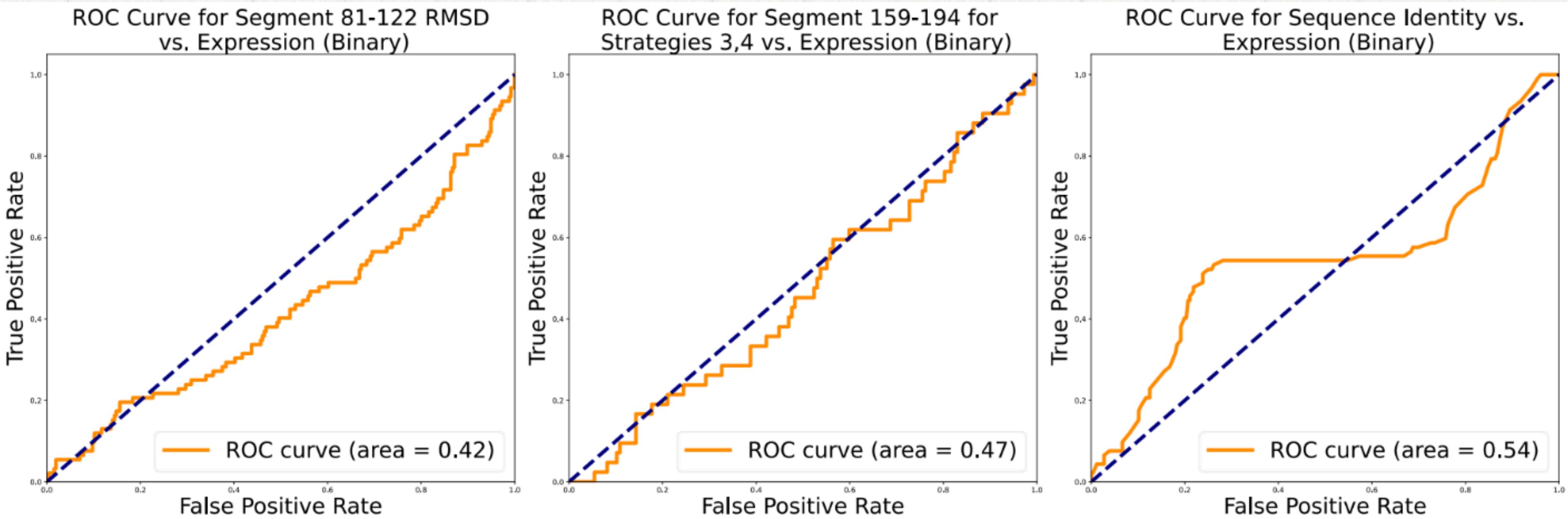
G Graph Neural Network for Solubility Prediction



# What is the performance of SolvIT?

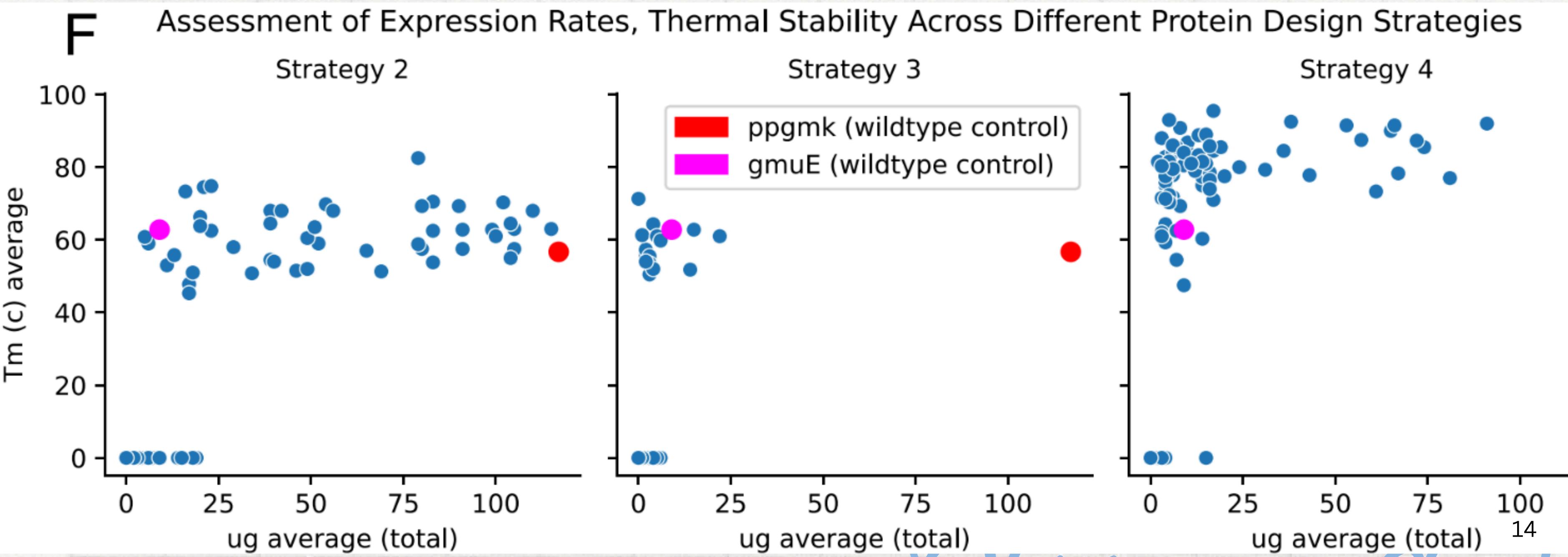


# Could we achieve same performance with simpler features?



# Does the melting temperature change?

- Higher melting temperature - 83% of novel designs
- Strategy 4 - 78% of high melting-temp. novel designs

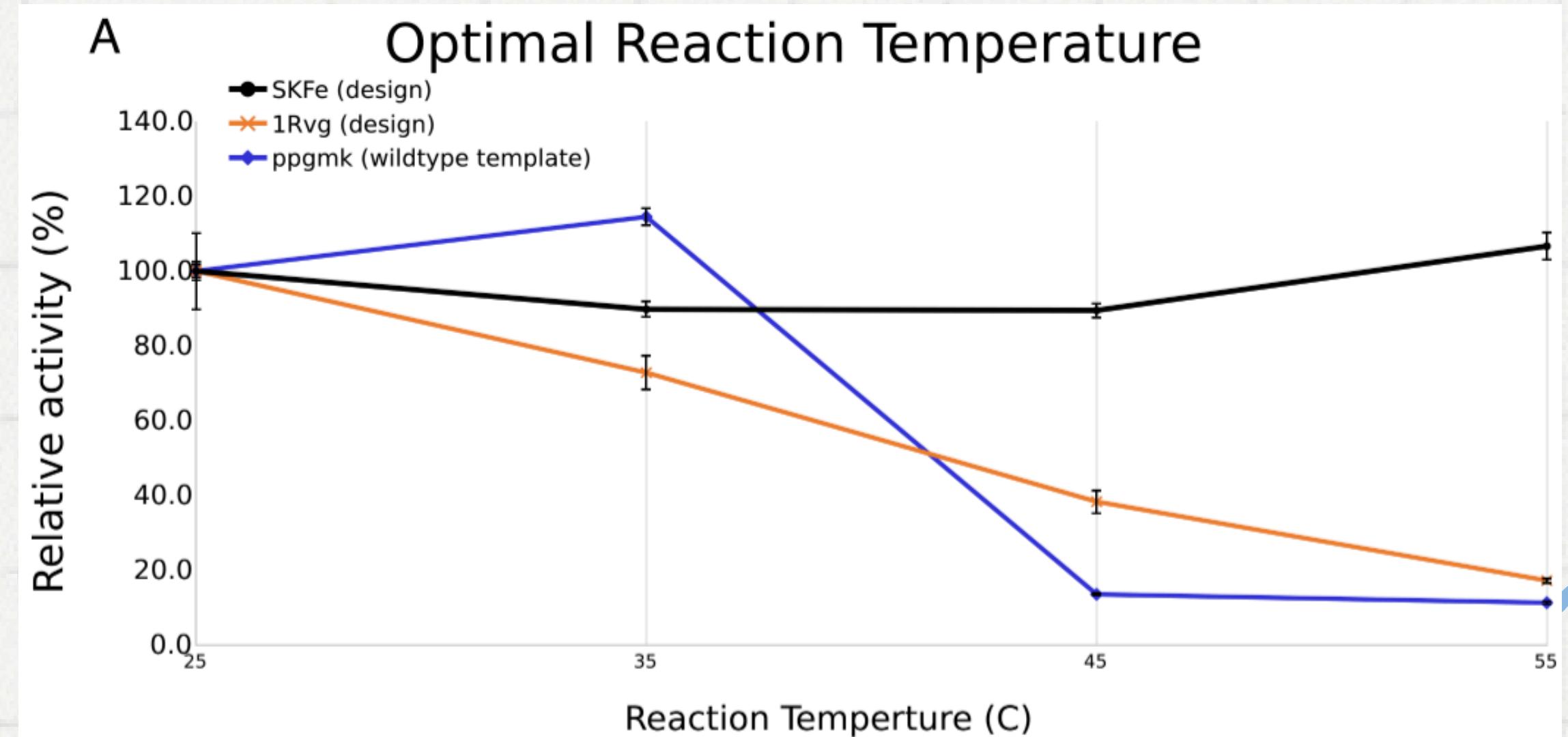


# Are the enzymes active?

- 60 designs active with either hexametaphosphate or ATP
- 8 (13%) designs showed obligatory polyphosphate-dependent activity
- Template enzyme can utilize both ATP and inorganic polyphosphate
- Obligatory polyphosphate activity is not common in natural enzymes. Has been documented in a handful of cases
- Underscores the potential of our method to introduce non-natural enzymatic traits

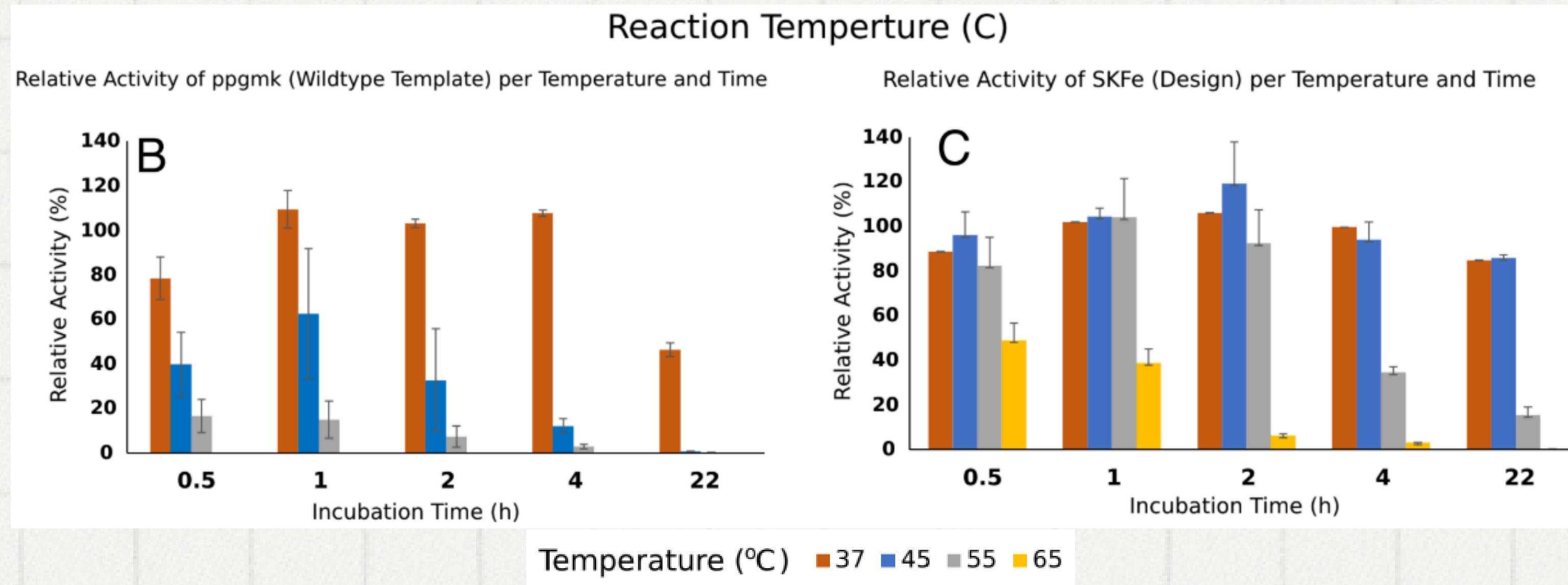
# What is the activity of novel enzymes as temp. increase?

- Above 35C novel designs activity is higher
- SKFe retains 40% activity at 95C
- Wildtype has no activity at 95C



# What is the activity after prolonged temperature stress?

- 55C: WT non-active after 4h | SKFe 80% activity after 22h
- 65C: WT non-active after 30min | SKFe 40% activity after 30min



# Conclusion

- CoSaNN – Enable rapid and robust novel enzyme designs
- SolvelT – Enable high performance solubility prediction
- Tested in wet lab
- Achieve improved novel designs with
  - Higher melting temperature
  - Higher activity
  - Higher robustness to prolonged temperature stress
  - Possible non-natural enzymatic traits

# Thank You



Research conducted  
during my tenure at



**Dr. Chen Brestel**



**Independant**

**AI Scientist**

**[https://chenbrestel.github.io/science/protein\\_engineering](https://chenbrestel.github.io/science/protein_engineering)**

**chen.brestel@gmail.com**



Chen Brestel | 5785 | 2025