Protein Oligomer Prediction in CASP16

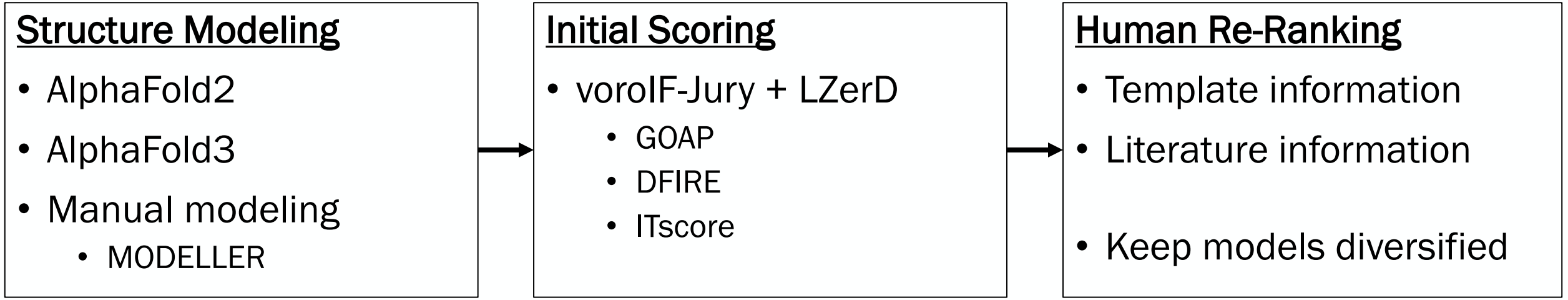
KiharaLab (294)

https://kiharalab.org

PURDUE UNIVERSITY

Overall Pipeline

- Generated Models using AlphaFold2 (AF2) with different parameters and MSAs & Alphafold3
- Manual Modeling (if necessary), e.g.
  - Large complexes: predict a chunk of complex, and merge them with MODELLER or Pymol



PURDUE UNIVERSITY

Enhanced MSA: Logan[1]

- Logan is a preprocessed dataset of all entries in Sequence Read Archive (SRA).

1. Download “Logan contigs” and filter of
   - Metagenome (Taxonomy ID: 2787823, unclassified entries)
   - Virus (Taxonomy ID: 10239, Viruses)
2. Apply Prodigal[2] to predict genes & translate
3. Remove duplicated sequences (MMseqs, 99 %)
4. Finally, this process yielded
   - Logan Meta: ~370 B seqs.
   - Logan Virus: ~1.7 B seqs.

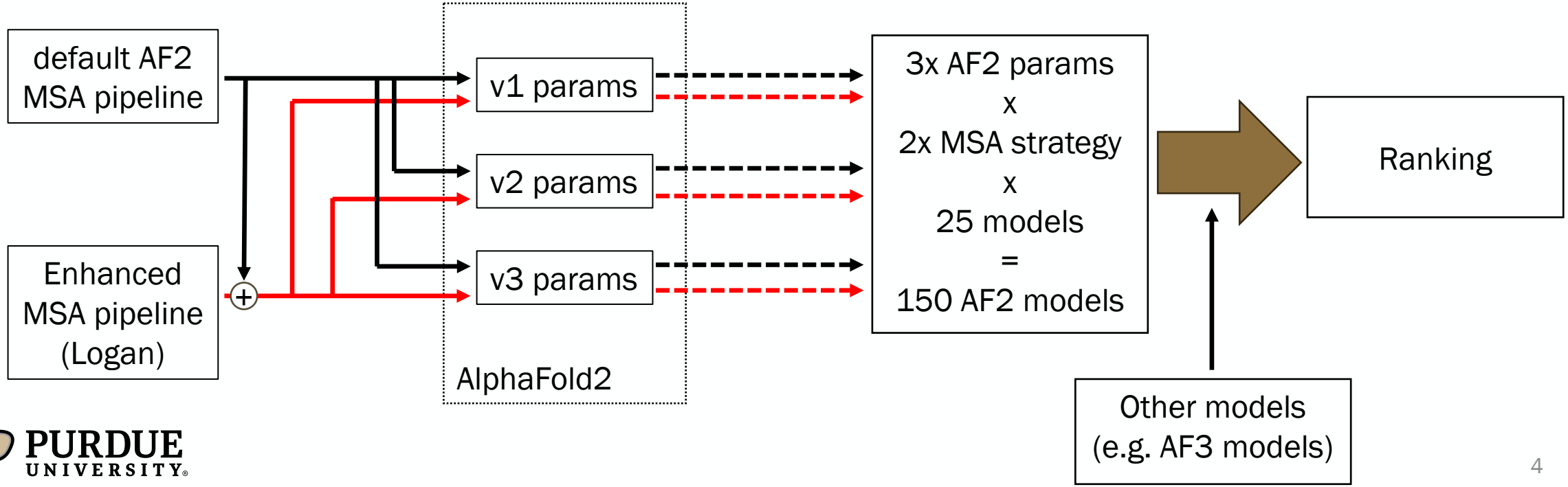
- Search with JackHMMER (1 iteration)

PURDUE UNIVERSITY

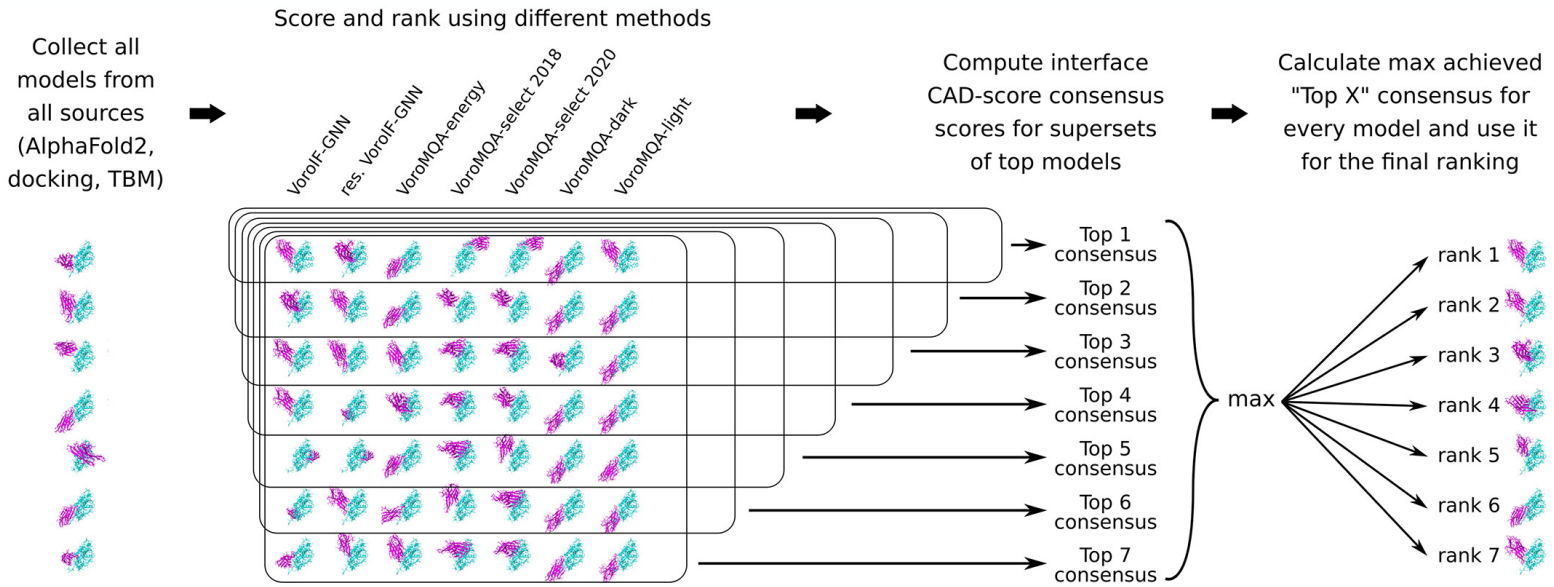
[1] Chikhi et al. bioRxiv, (2024)
[2] Hyatt et al. BMC Bioinfo. (2010)

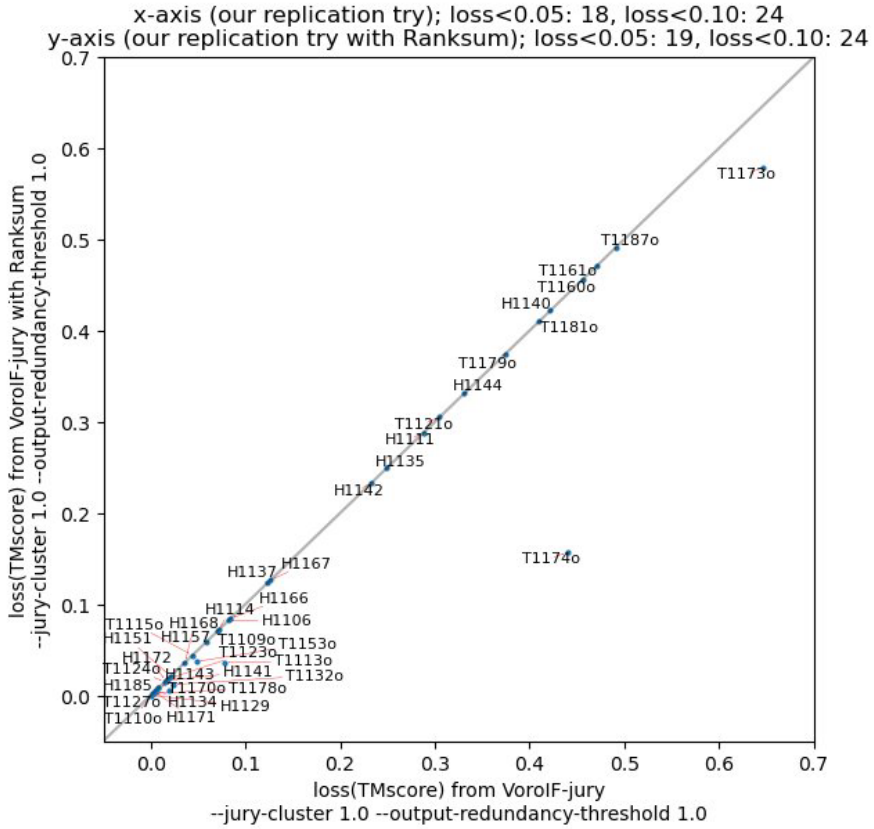
AlphaFold2 Pipeline with Enhanced MSAs

- Two sets of MSAs generated
- Each MSA was used to run AF2 with 3 different parameters
  - generated 2 x 3 x 25 = 150 models
  - If we have another MSA, 75 models were generated for that



Scoring: VoroIF-jury + LZerD Ranksum





Based on VoroIF-jury [1]

- CASP15-CAPRI winning method by the Venclovas group
- A consensus ranking method using 7 scoring functions

Added three scoring functions (components of LZerD RankSum score)

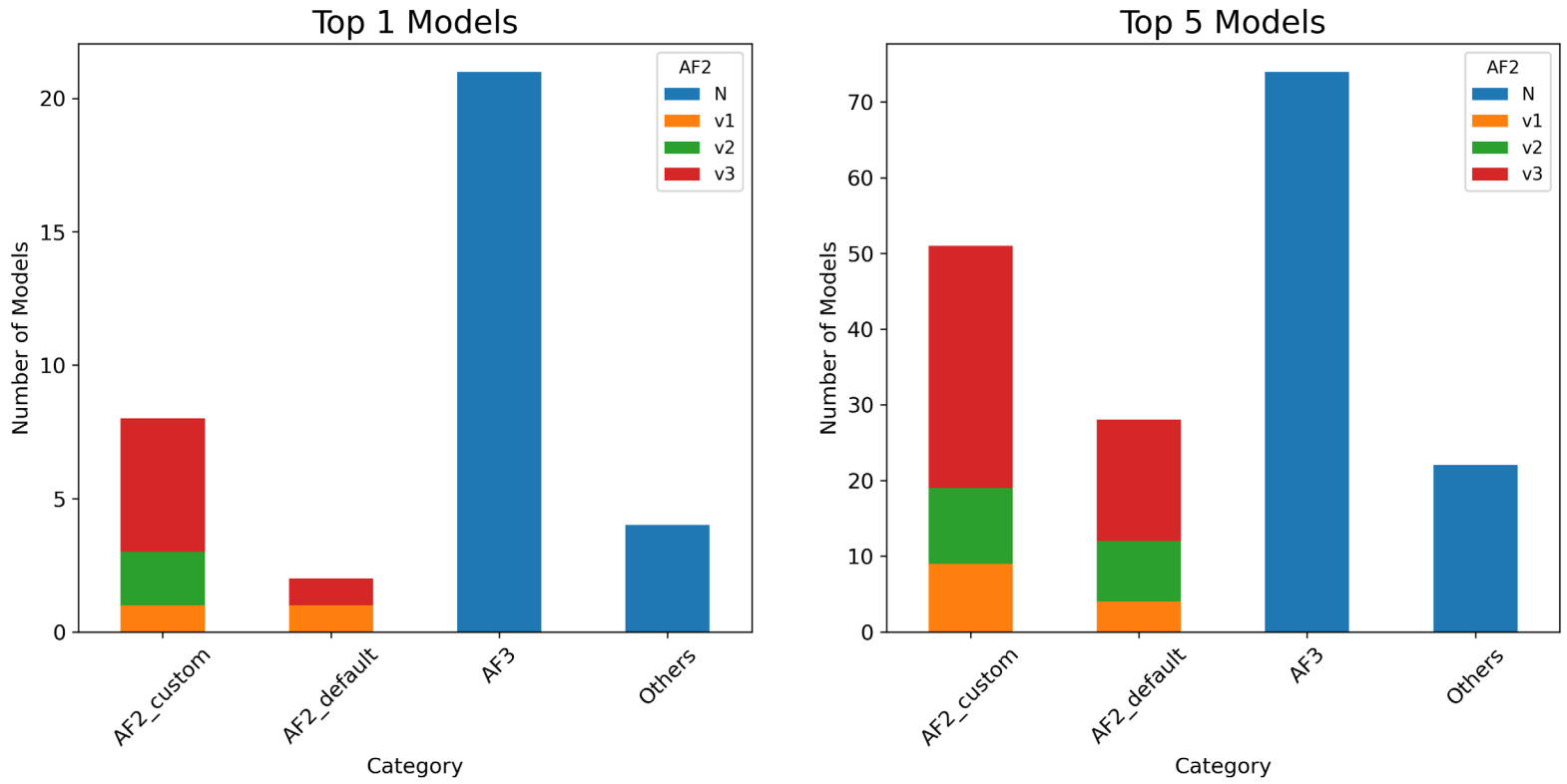
- GOAP[2], DFIRE[3], and ITScorePro[4]

[1] Olechnovič et al., Proteins, 2023
[2] Zhou et al. Biophys. J., (2011)
[3] Zhou et al. Protein Sci., (2002)
[4] Huang et al. Proteins (2002)

PURDUE UNIVERSITY

Sources of Selected Models

Final submission contained at least 1 AF2_enhancedMSA model and 1 AF3 model

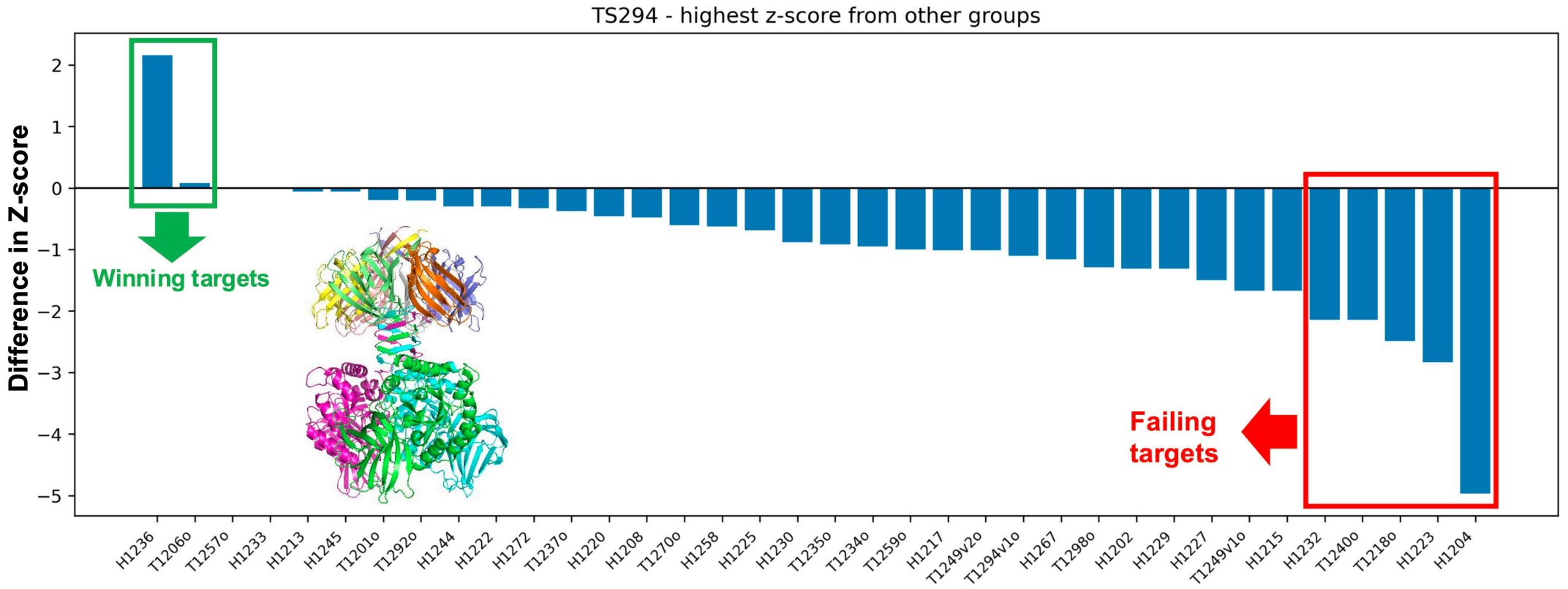


Others:

- Manual Modeling with MODELLER, PyMOL, starting from AF2/AF3 models
- Selected models from Phase 0 models, MassiveFold models

PURDUE UNIVERSITY

Performance of Each Targets



Comparing Z-score with the best team (other than ourselves)

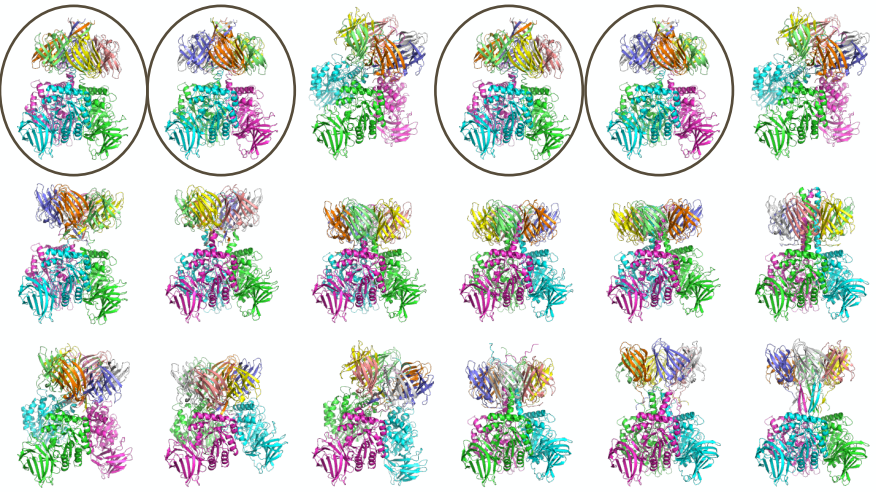
Winning on 2 targets, tie in 2 targets

PURDUE UNIVERSITY

H1236: Haloferax tailed virus 1

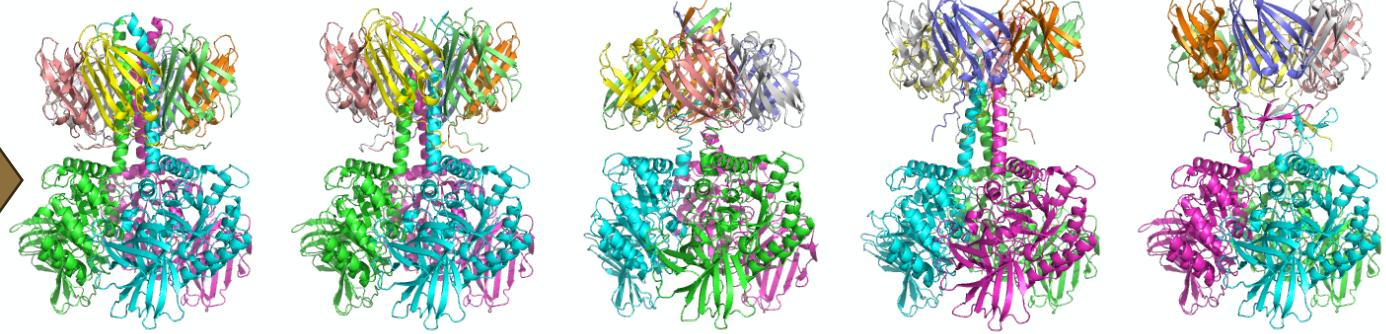
Two Subunits A3B6, No templates

A: Prokaryotic polysaccharide deacetylase
B: Unknown



AlphaFold2_v2 with Logan
AlphaFold2_v3 with Logan
AlphaFold3
Total 155 models

Ranking by
VoroIF-jury+
& Manual Selection

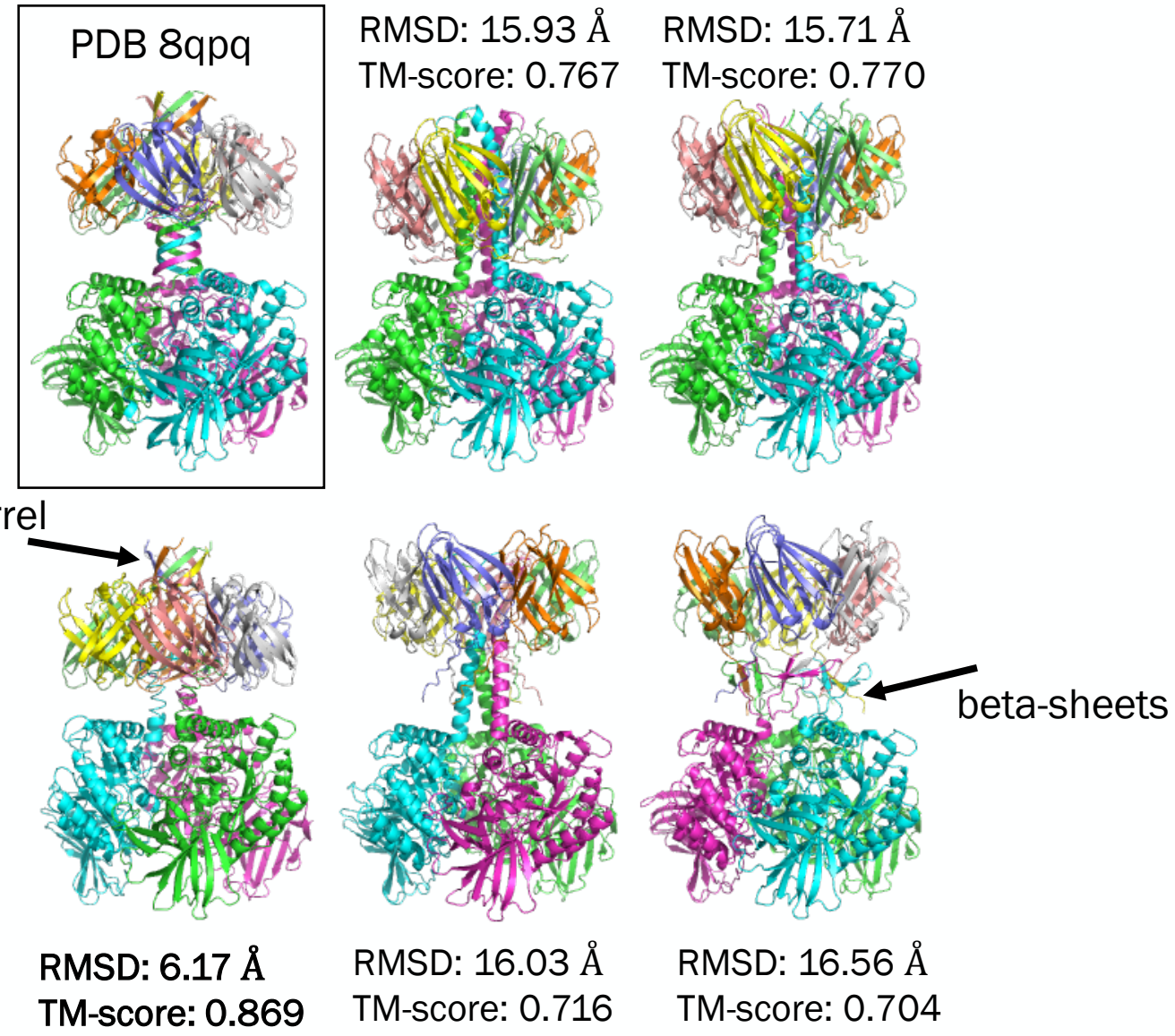


AF3 | AF3 | AF2_v3 Logan | AF2_v3 Logan | AF3

Diverse models were selected.
A3 subcomplex was almost the same.
B6 subunits were placed in different positions.
Interactions between A3 and B6 subunits are varied.

PURDUE UNIVERSITY

H1236: Haloferax tailed virus 1



What went right?

- Exploring multiple interaction patterns between A3 and B6 subunits diversified our predictions.
- Ranking by consensus helped us to identify reliable models.

H1265: TLR4 complex (A9B18)

Abstracts of the XXVI IUCr Congress
Melbourne, Australia, 22-29 August 2023

Acta Cryst. (2023). A79, C958

MS **Microsymposium**

TLR4 TIR domain higher-order assemblies reveal the structural basis of adaptor recruitment in Toll-like receptor signaling pathways

J. D. Nanson[1], T. W. Muusse[1], Y. Li[1], M. Pan[1], M. J. Landsberg[1,2], P. R. Vajjhala[1], T. Ve[3], K. J. Stacey[1], B. Kobe[1,2]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia.

[2]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia.

[3]Institute for Glycomics, Griffith University, Southport, Queensland, Australia.

j.nanson@uq.edu.au

Keywords: Innate immunity, Toll-like receptor, filament, cryo-EM

- We found one key information (1 page abstract) about this filament complex
  - No structure, no cryo-EM maps, Only one figure
- Key Findings:
  - 2 states: There are 6- and 9-stranded complex
  - B chain (MAL TIR) forms proto-filament (parallel), template: 5UZB
- Question: interface between TLR4 and MAL (not clear from the figure)

PURDUE UNIVERSITY

H1265: TLR4 complex (A9B18)

Answer:
H1265

- Approach: starting from template (PDB: 5UZB)

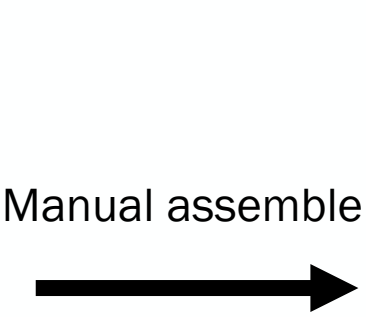
x3

MAL TIR filament
(B: 6-mer)

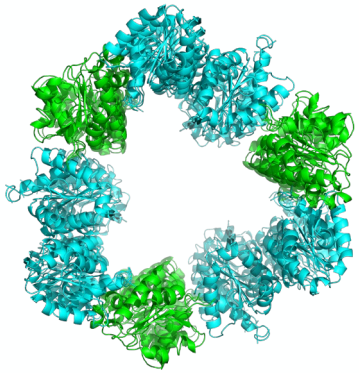
x3

TLR4 TIR filament
(A: 3-mer)
Manually modeled

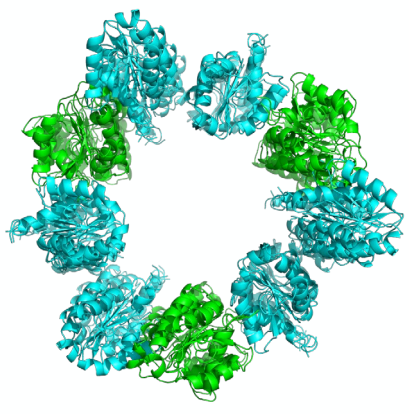
Manual assemble



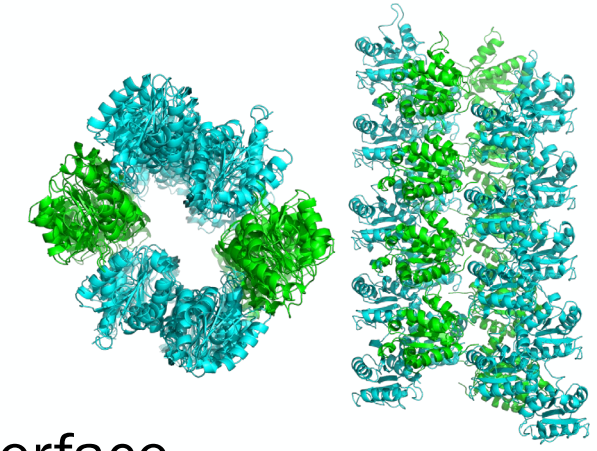
RMSD: 27.8 Å



RMSD: 22.3 Å



RMSD: 22.7 Å

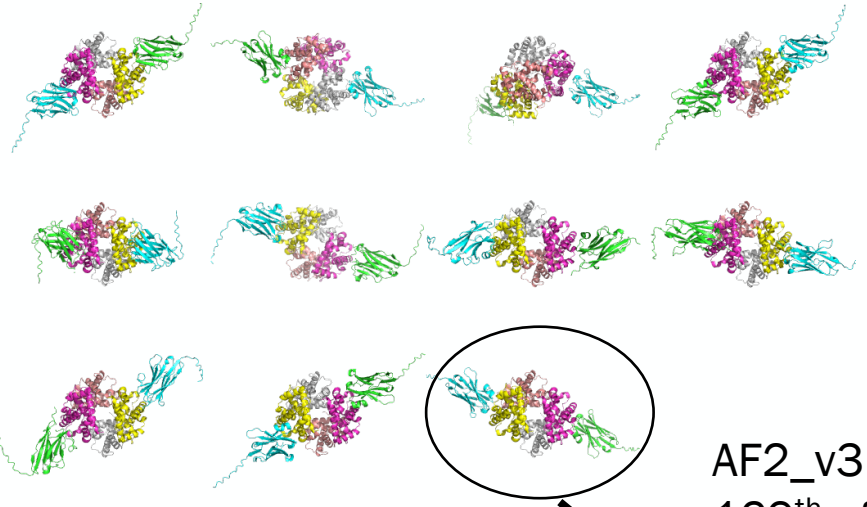


6-strands models (MODEL 4, 5)

- correct shape, wrong interface

Failed target H1204 (A2B2C2)

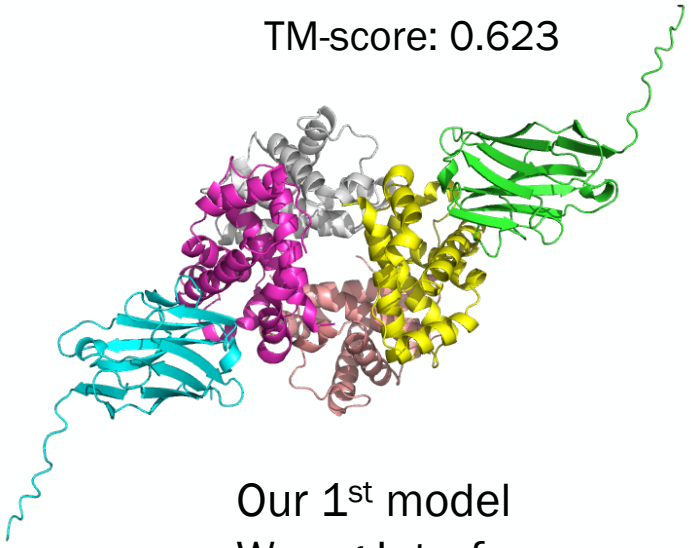
Human hemoglobin in complex with nanobody



170 models

Filtering by VoroIF-jury

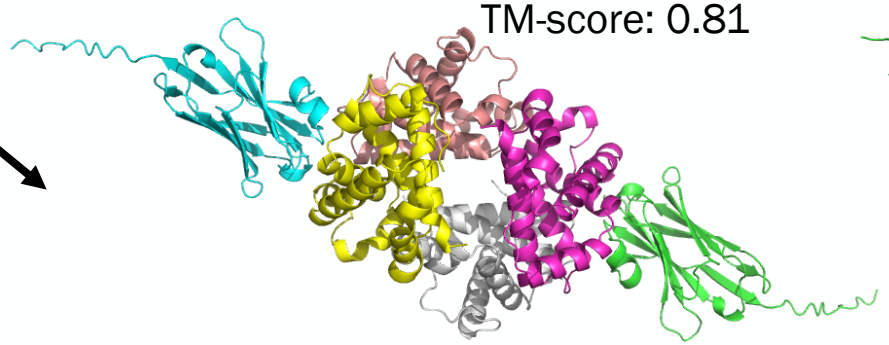
Manual Selection



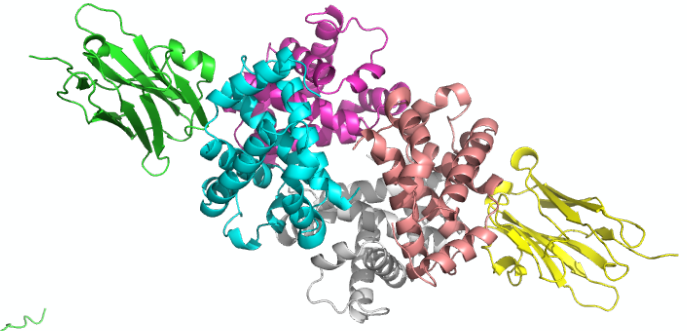
RMSD: 14.47 Å
TM-score: 0.623

Our 1st model
Wrong Interfaces

AF2_v3 default MSA
169th of VoroIF-jury ranking



RMSD: 5.66 Å
TM-score: 0.81



PDB 8VYL

What went wrong:

Our model selection process relies on the VoroIF-jury score. When high-quality models are in the minority within the model pool, our approach fails to select them.

Since the VoroIF-Jury score is significantly low, we failed to select the best model in the model pool.

PURDUE UNIVERSITY

What went well?

- Enhanced MSAs with Logan
- Scoring
  - VoroIF-jury + LZerD score worked well in both Phase 1 and Phase 2
- Group discussion
  - Literature
  - With AF2 and AF3, inexperienced students can contribute meaningfully in the team

What went wrong?

- Antibody docking. the score did not work
- We do not have an established method (and experience) for predicting stoichiometry (Phase 0)

PURDUE UNIVERSITY

Team Members

Dr. Tsukasa Nakamura

Pranav Punuru

Emilia Tugolukova

One day in a submission selection meeting