Modeling of Protein Complexes in CASP15

Justas Dapkūnas, Kliment Olechnovič

Venclovas' group, Institute of Biotechnology, Vilnius University, Lithuania

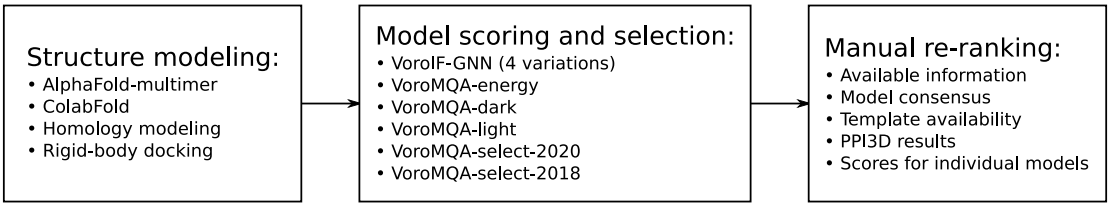
2022

Modeling of protein complexes in CASP15

- Protein assemblies
- Protein-RNA interactions
- Large multidomain proteins

Methods

Generate an ensemble of diverse models and select the best of them



Sampling using AlphaFold

- AlphaFold with `full_dbs` and `reduced_dbs` presets
- ColabFold using MSA with additional sequence databases:
  - MMseqs2 on ColabFoldDB
  - IMG/VR v3
  - Custom Mgnify coding sequences DB
  - UniProt TREMBL
  - NCBI NR
- More ColabFold:
  - Pairing of multiple sequence alignments
  - Number of recycles (3-100)
  - Multimer modeling using AlphaFold-ptm

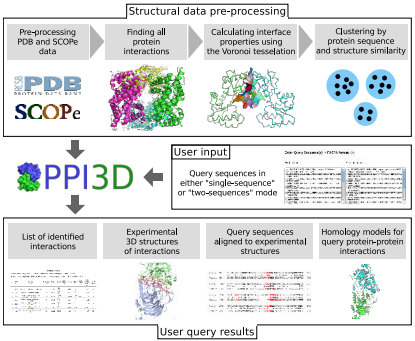
Poster 23: Challenges in modeling protein assemblies with AlphaFold: a case of anti-CRISPR complexes

Analysis of the available data for manual model re-ranking

- Search and model protein complexes using PPI3D
- Sequence-based homology search by HHpred and COMER
- Structure-based homology search with DALI
- RCSB Advanced Search
- UniProt
- Scientific literature
- Disorder prediction

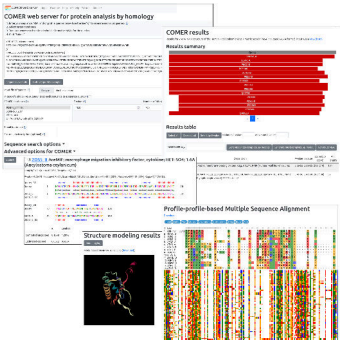
Methods from our lab: PPI3D and COMER

`http://bioinformatics.lt/ppi3d/`



Dapkūnas et al., *Bioinformatics*, **2017**, 33:935

`http://bioinformatics.lt/comer/`



Dapkūnas & Margelevičius, **2022**, submitted

Poster 6: PPI3D

Docking

We used rigid body docking methods:

- FTDock and Hex for hetero-oligomers
- SAM for homo-oligomers

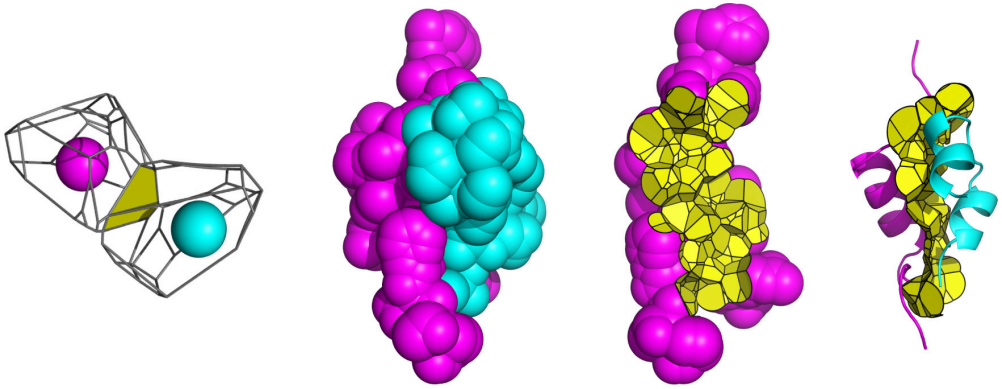
Our docking strategy:

- Select a diverse set of monomers from AF2 and template-based models.
- Dock as exhaustively as possible.
- Score and rank all the generated complexes with our interface-focused methods.
- Relax top 300 docking models with OpenMM.
- Pool the relaxed docking models together with AF2 and template-based models, then score and rank everything.

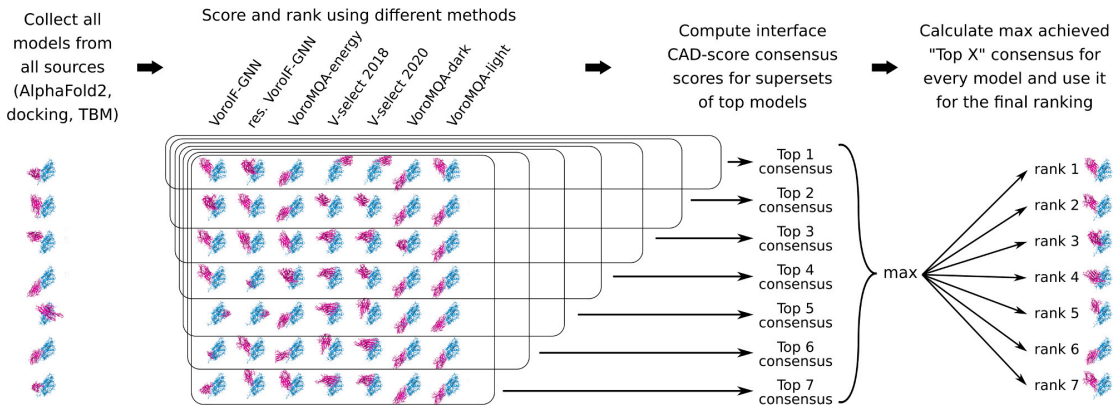
Interface-focused scoring

We used our methods that analyze Voronoi-tessellation derived interface contacts:

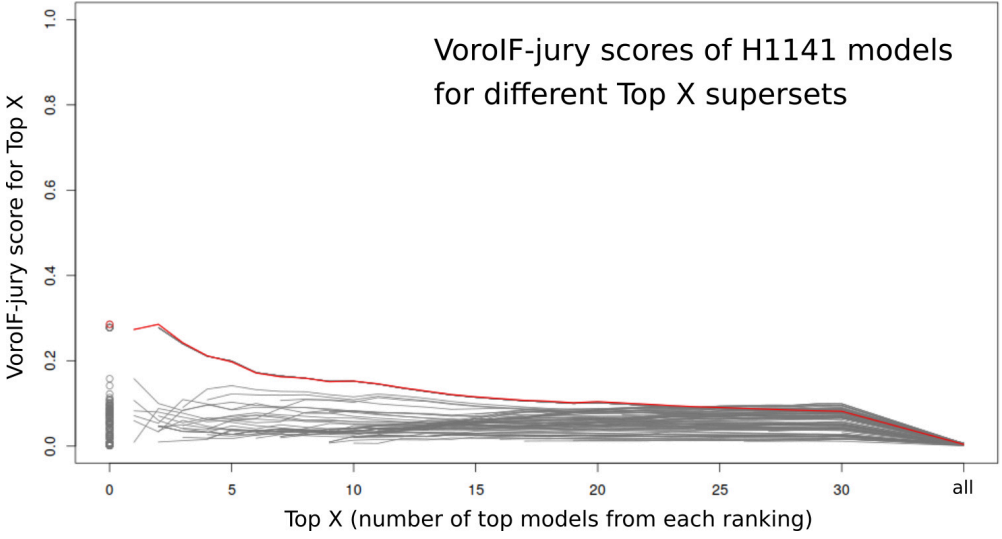
- fast interface VoroMQA-energy and (a bit slower) VoroIF-GNN for quality assessment
- fast interface CAD-score (contact area difference score) for comparison and clustering



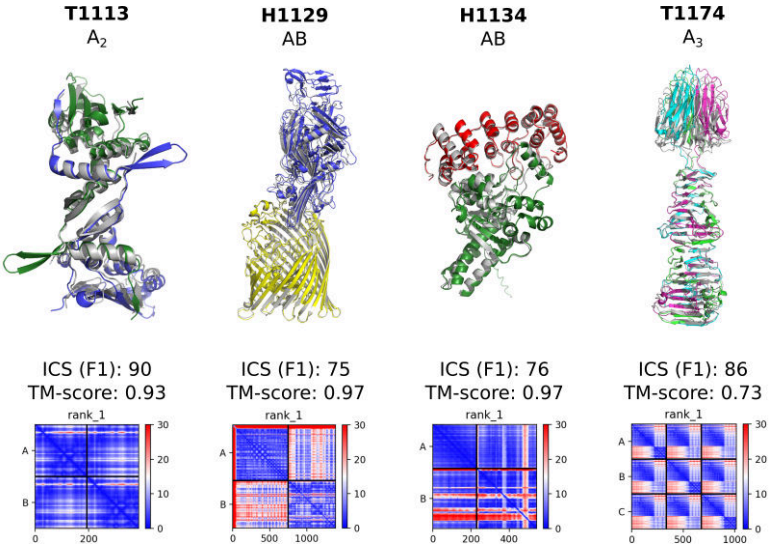
Ranking with VoroIF-jury



VoroIF-jury scores plot



AlphaFold produces high-quality models for multimers

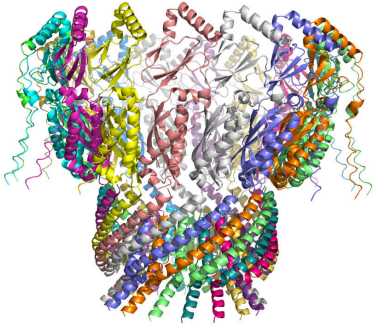


AlphaFold does not model very large complexes

T1115 (model T1115TS494_1o)

A_{16}

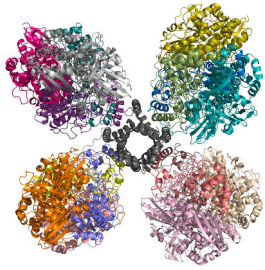
Symmetry docking



AlphaFold

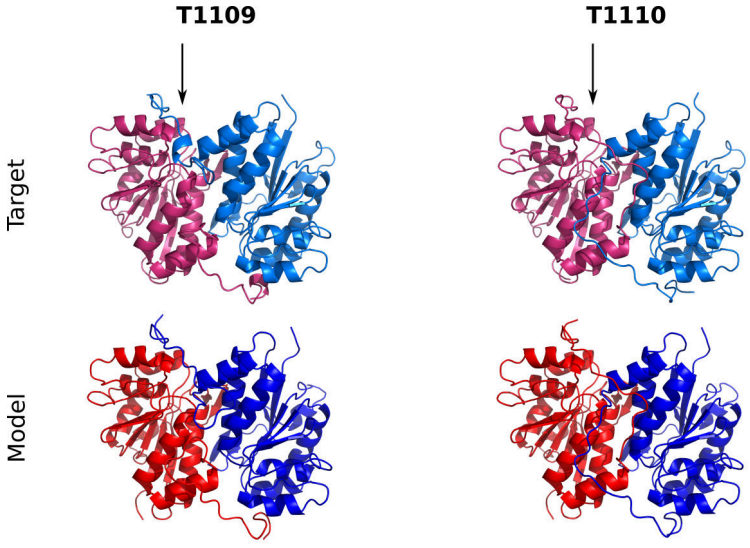
H1114 (model H1114TS494_1)

$A_4B_8C_8$

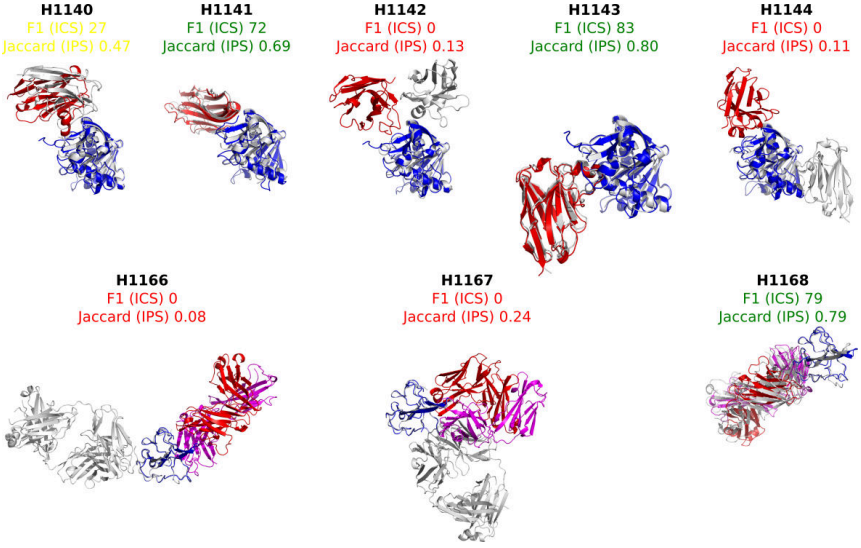


- A_4: AlphaFold
- A_4B_2: AlphaFold
- B_2C_2: AlphaFold
- Align and refine

Manual selection of models is sometimes useful



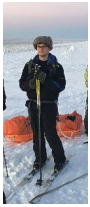
Both AlphaFold and docking work or fail for antibody-antigen interactions



Conclusions

- AlphaFold predicts structures of protein complexes well
- AlphaFold fails to model large protein assemblies
- Selection of models using different scores and other information is important
- Antibody-antigen interactions are hard to predict

Acknowledgements



CASP and CAPRI team: Kliment Olechnovič, Lukas Valančauskas, Česlovas Venclovas

COMER method for sequence profile-profile comparison: Mindaugas Margelevičius

Funding: Research Council of Lithuania