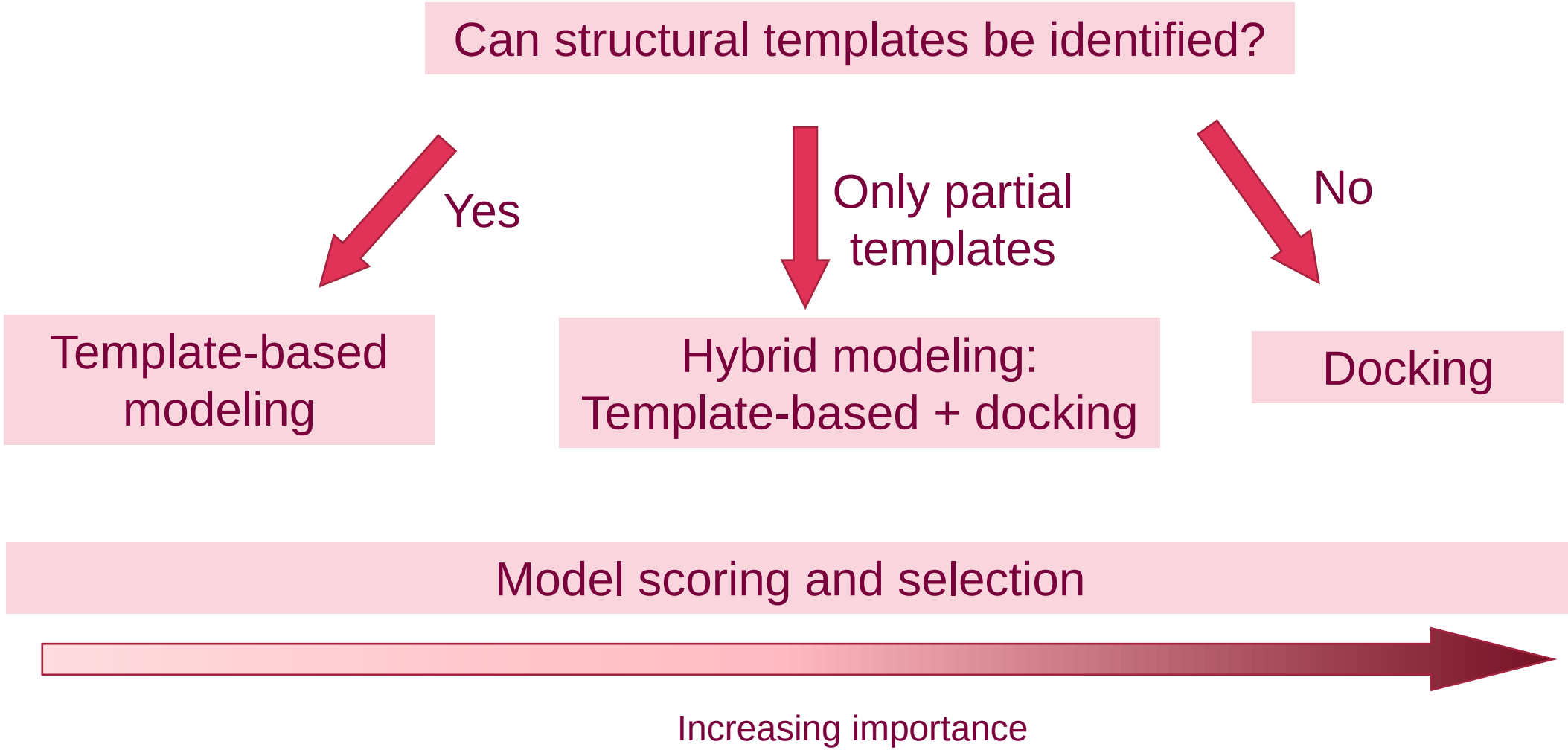
Vilnius University

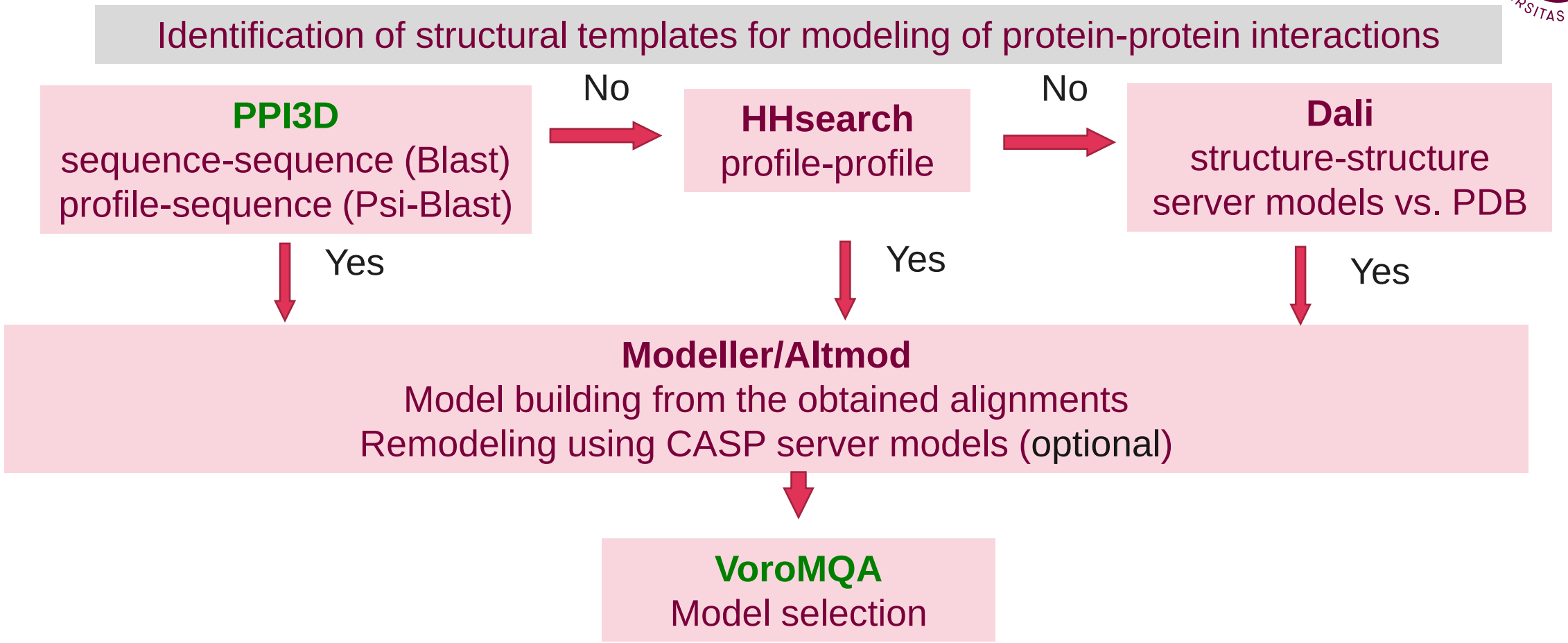
# Prediction of diverse protein assemblies in CASP14-CAPRI: from template-based modeling to free docking

Česlovas Venclovas
Institute of Biotechnology, Life Sciences Center, Vilnius University, Lithuania
http://bioinformatics.lt

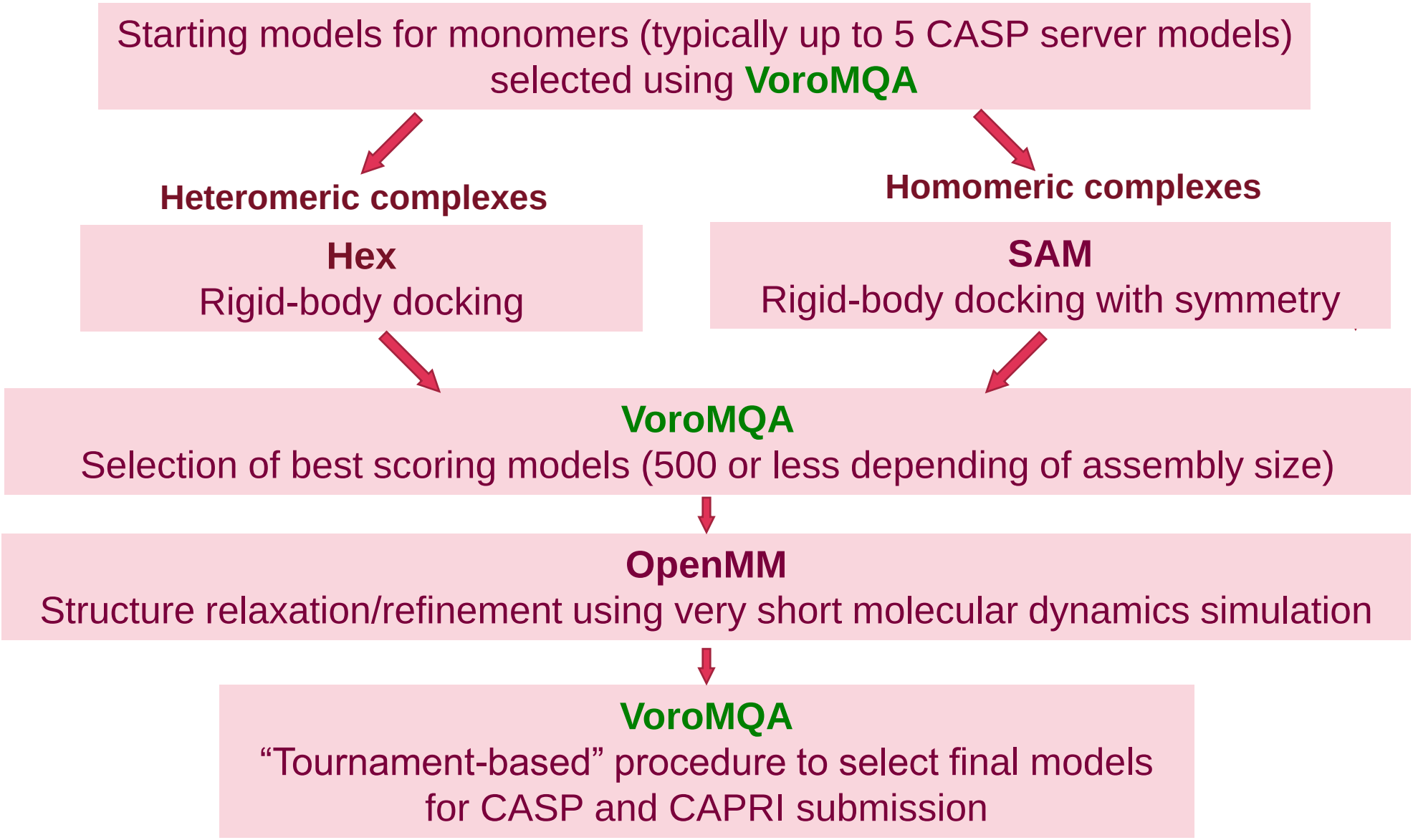
# Choice of modeling aproaches



# Template-based modeling workflow



# Docking workflow



# Hybrid modeling

Targets with partial coverage of templates, large complexes, coiled-coils

↓

**Obtaining models for subcomplexes**

- Template-based pipeline
- Docking pipeline
- Custom-built procedures

↓

**Combining subcomplexes**

- Structure superposition (TM-align)
- Rigid-body docking
- Custom-built procedures

↓

**Model relaxation and selection**

# Key methods

## Protein-protein interaction templates

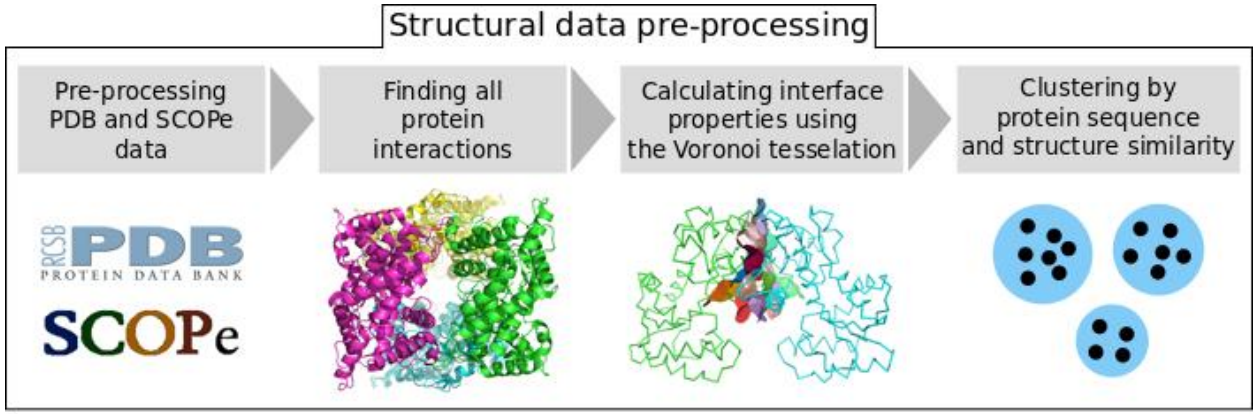
- **PPI3D** (**P**rotein-**P**rotein **I**nteractions in **3D**): clustered protein-protein interfaces derived from PDB biological assemblies
  - Sequence-based searches
  - Comprehensive analyses of interfaces
  - Template-based modeling

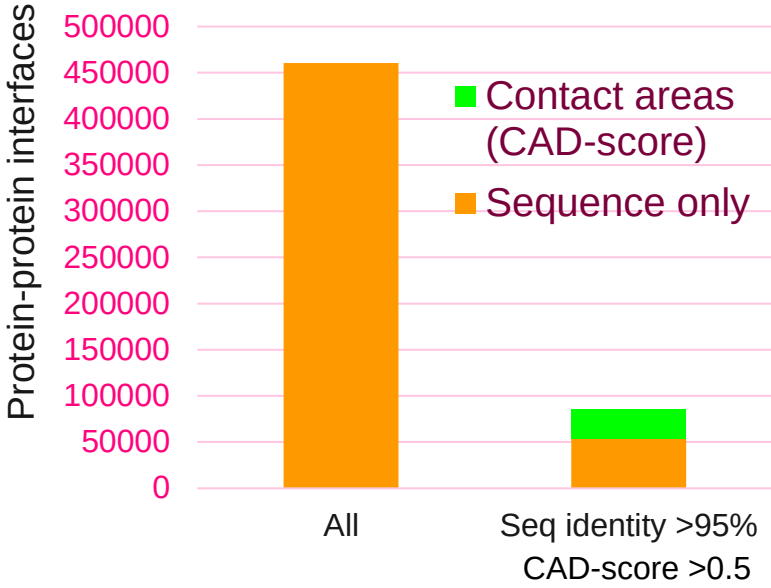
## Model selection

- **VoroMQA** (**Voro**noi tessellation-based **M**odel **Q**uality **A**ssessment)
  - Combines interatomic contact areas and statistical potentials
  - Can assess both monomeric and multimeric structures
  - Can assess interfaces

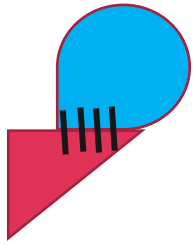
# PPI3D: a web server for searching, analyzing and modeling pairwise interactions

http://bioinformatics.ibt.lt/ppi3d/

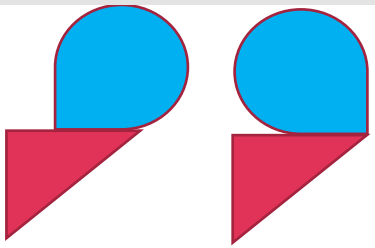


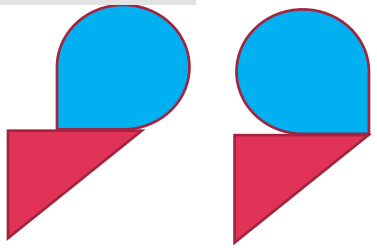

## Interface clustering using CAD-score variants



➢**PPI3D: CAD-score (contact area similarity)**
➢CASP: ICS/F1



➢**PPI3D: CAD-score (interface site similarity)**
➢CASP: IPS/Jaccard



➢**PPI3D: CAD-score (interface area similarity)**
➢CASP: No correspondence

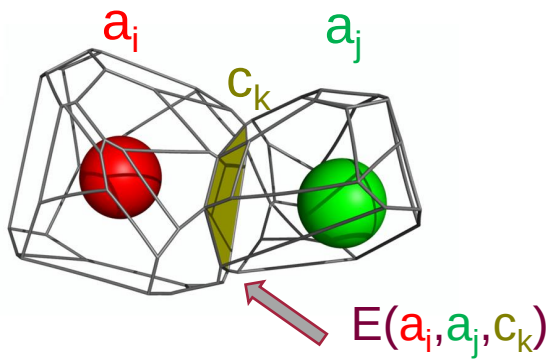
PPI3D provides a set of interaction interfaces:
- ➢ non-redundant
- ➢ comprehensive

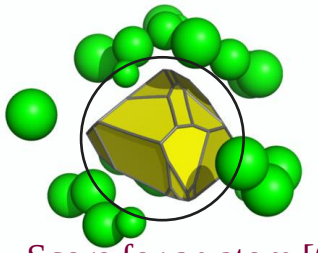
Dapkūnas et al. (2017) The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinformatics* 33:935

# VoroMQA: a method for assessing structures of proteins and protein complexes

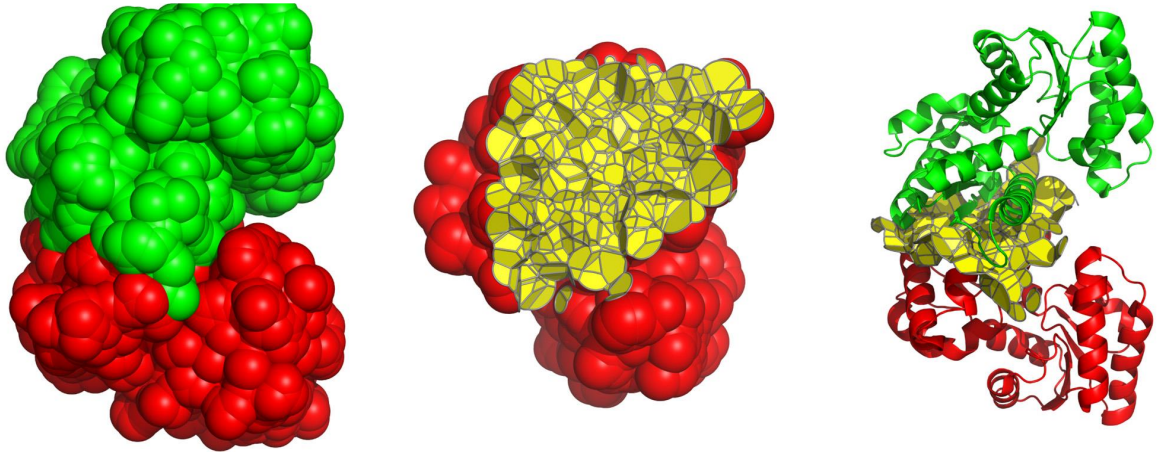
- VoroMQA estimates energy at the level of individual interatomic contacts
- VoroMQA design enables getting scores at different structural levels
  - Scores for individual atoms
  - Scores for individual residues
  - Global score for protein structure/complex
- VoroMQA also provides direct scoring of the protein-protein interaction interface
  - Interface score (VoroMQA score for the interface atoms)
  - Interface pseudoenergy (total VoroMQA pseudo-energy for the inter-subunit contacts)

Each interatomic contact (Voronoi face) can be assigned a pseudo-energy value



Score for an atom [0,1]: transformed normalized sum of pseudoenergies of contact areas



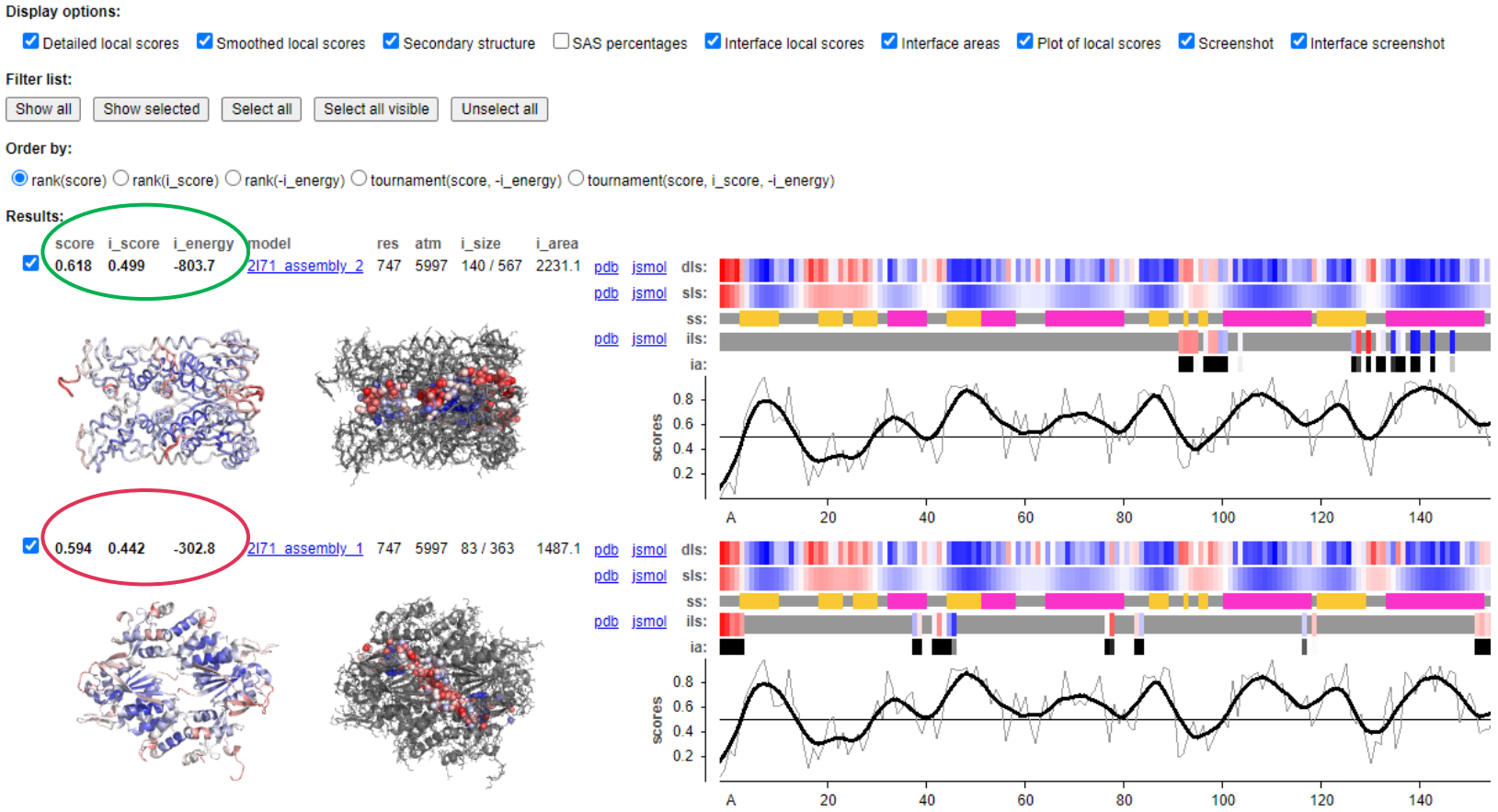
$$E(a_i, a_j, c_k) = \log \frac{P_{exp}(a_i, a_j, c_k)}{P_{obs}(a_i, a_j, c_k)} = \log \frac{F_{exp}(area(a_i), area(a_j), area(c_k))}{F_{obs}(area(a_i, a_j, c_k))}$$


Olechnovič & Venclovas (2017) VoroMQA: Assessment of protein structure quality using interatomic contact areas. *Proteins*. 85(6):1131-1145.

Olechnovič & Venclovas (2019) VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic Acids Res. 47(W1):W437-W442*

# VoroMQA can be used to directly assess protein-protein interfaces

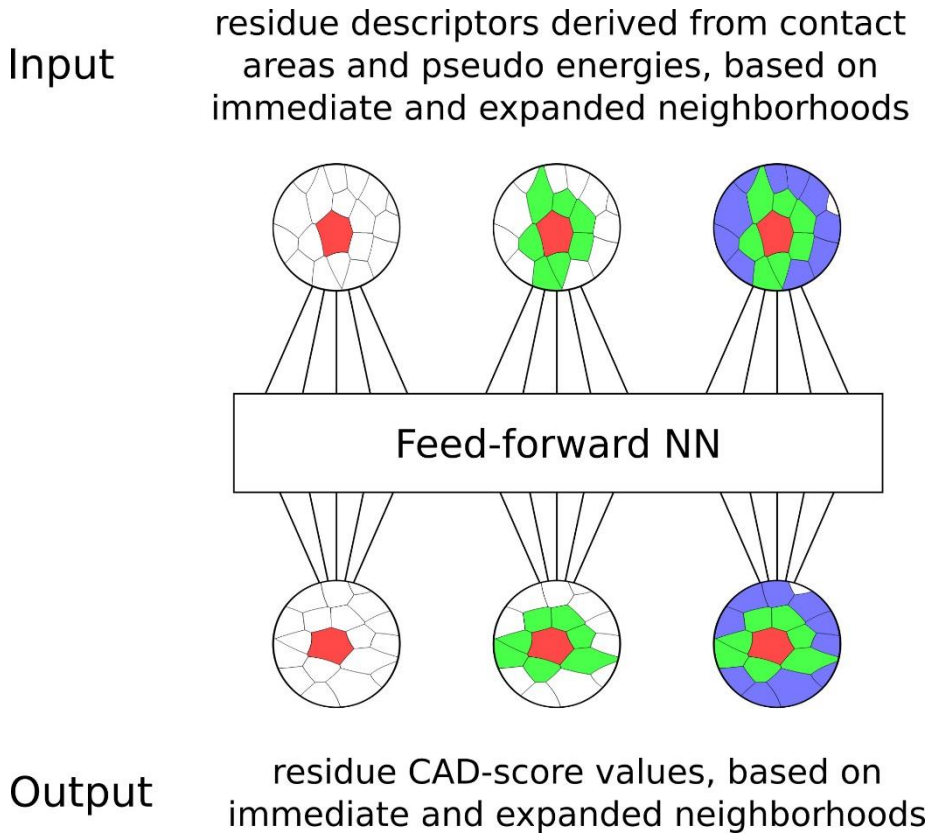
http://bioinformatics.ibt.lt/wtsam/voromqa




- Assessment of models for protein complexes
- Assessment of biological relevance of interfaces in crystal structures of protein complexes

Olechnovič & Venclovas (2019) VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic Acids Res*. 47(W1):W437-W442

# VoroMQA-dark: a new version of VoroMQA used in CASP14-CAPRI

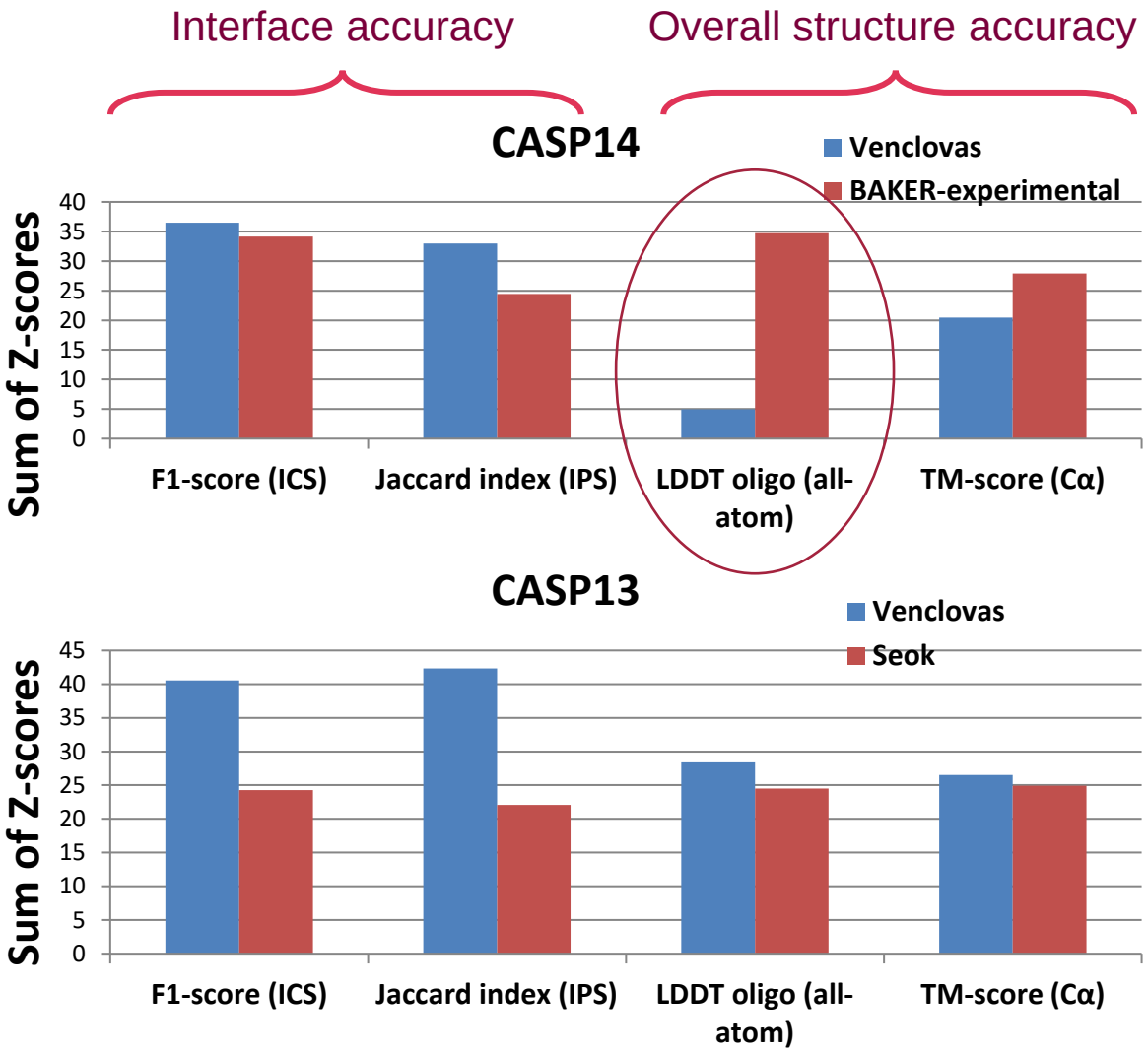



- VoroMQA-dark uses a neural network (NN) trained to predict local (per-residue) CAD-score values using three expanding shells of residue neighborhood
- Input data are the same as in the original VoroMQA: Voronoi tessellation-based contact areas and the corresponding contact potential values

# Results

## What went right, what went wrong and why

# Overall results: Interface and overall structure prediction




**What went right?**

- Both interface patches (Jaccard) and interface contacts (F1-score) were predicted relatively well (our main focus)

**Why?**

- Effective template identification
- Improved model selection procedure
  - New improved version of VoroMQA
  - VoroMQA interface energy score having more weight than the global score
  - Short MD simulations with OpenMM, improving stereochemistry

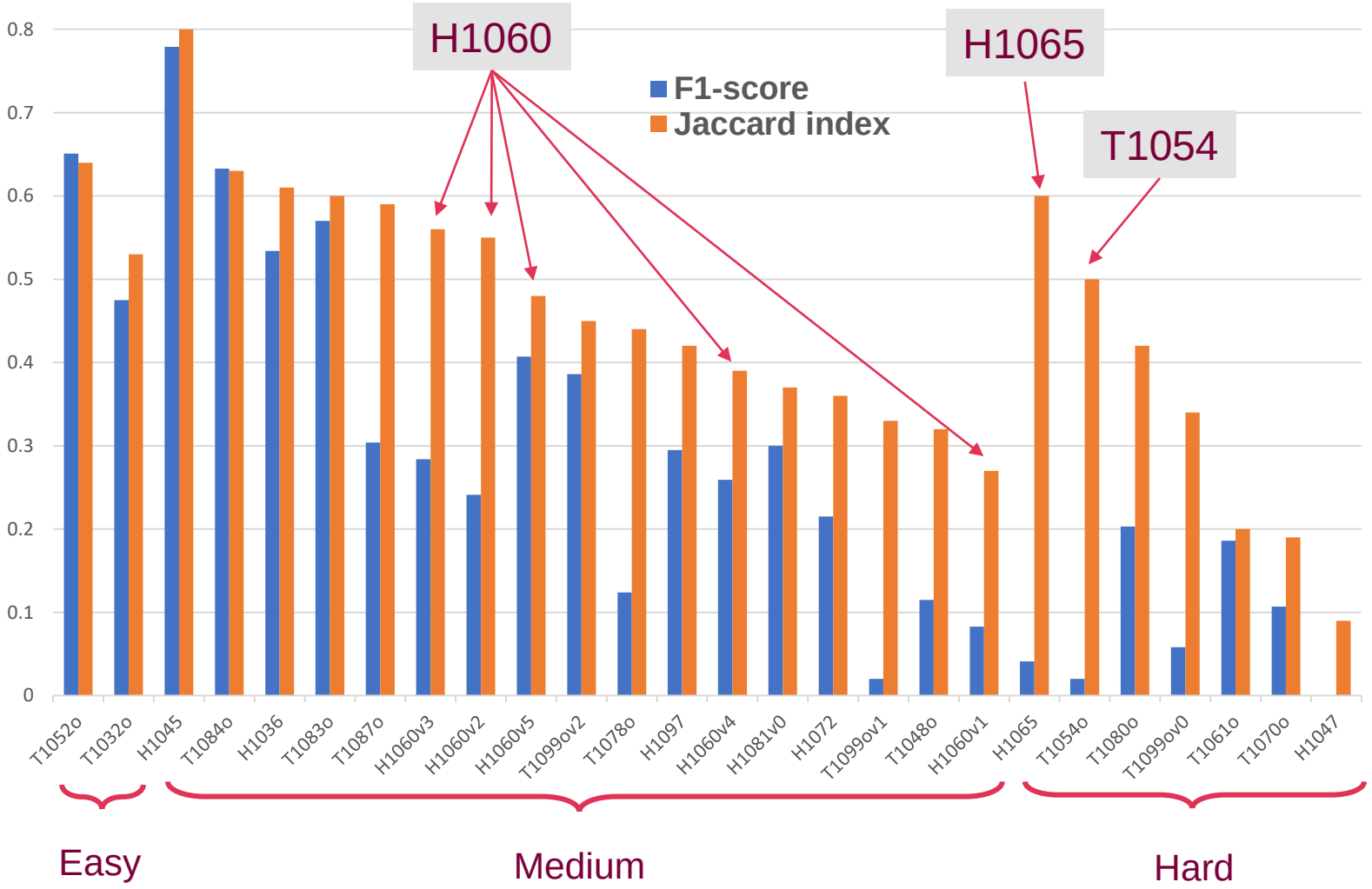
**What went wrong?**

- Overall structure accuracy is not that great
- All-atom structure accuracy is even worse

**Why?**

- CASP14: More frequently used Modeller/AltMod for building models
- CASP13: Most often used monomers selected from CASP server models

# Interface accuracy of the best models (F1-score/ICS & Jaccard index/IPS)




**Interesting cases**

Accurate interface patch but not contacts

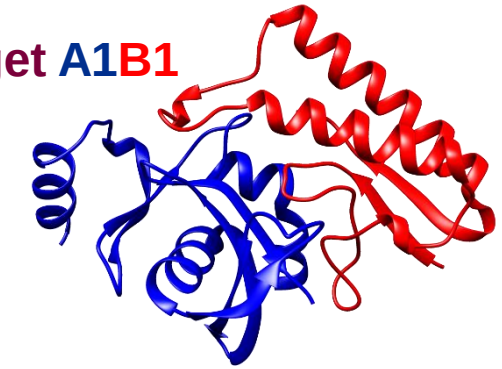
- H1065 (docking)
- T1054 (docking)

Variable success in assembling a large complex

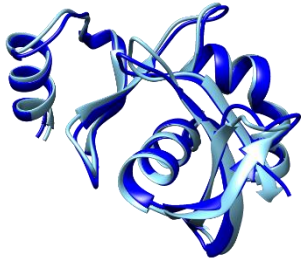
- H1060 (hybrid modeling)

# H1065 heterodimer: N4-Cytosine Methyltransferase Docking

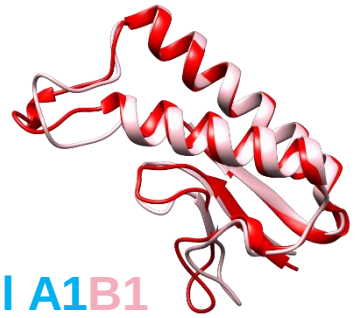
Target A1B1



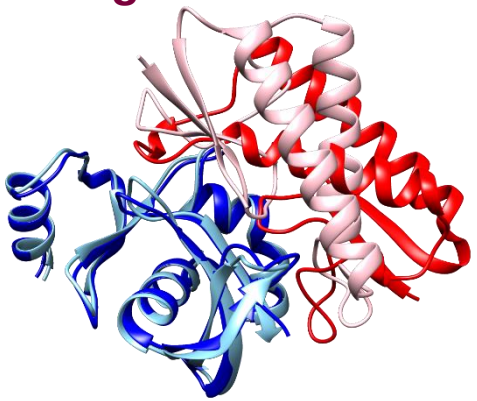
Target A1/Model A1



Target B1/Model B1



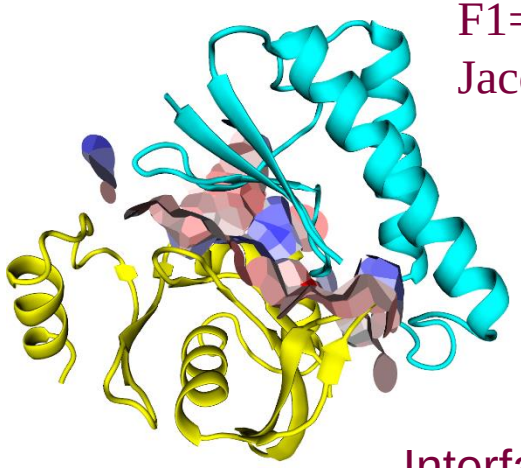
Target A1B1/Model A1B1



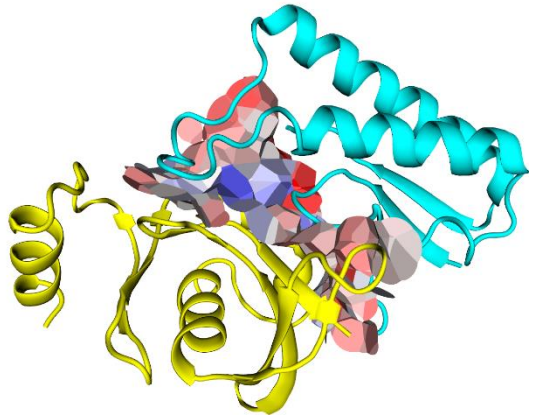
B1 rotated ~180º

Model

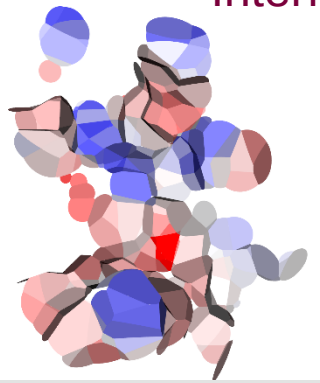
H1065TS029_1
F1=4%
Jaccard=0.57

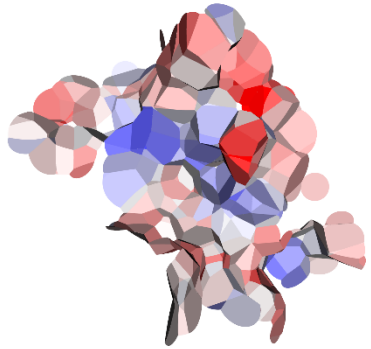


Target



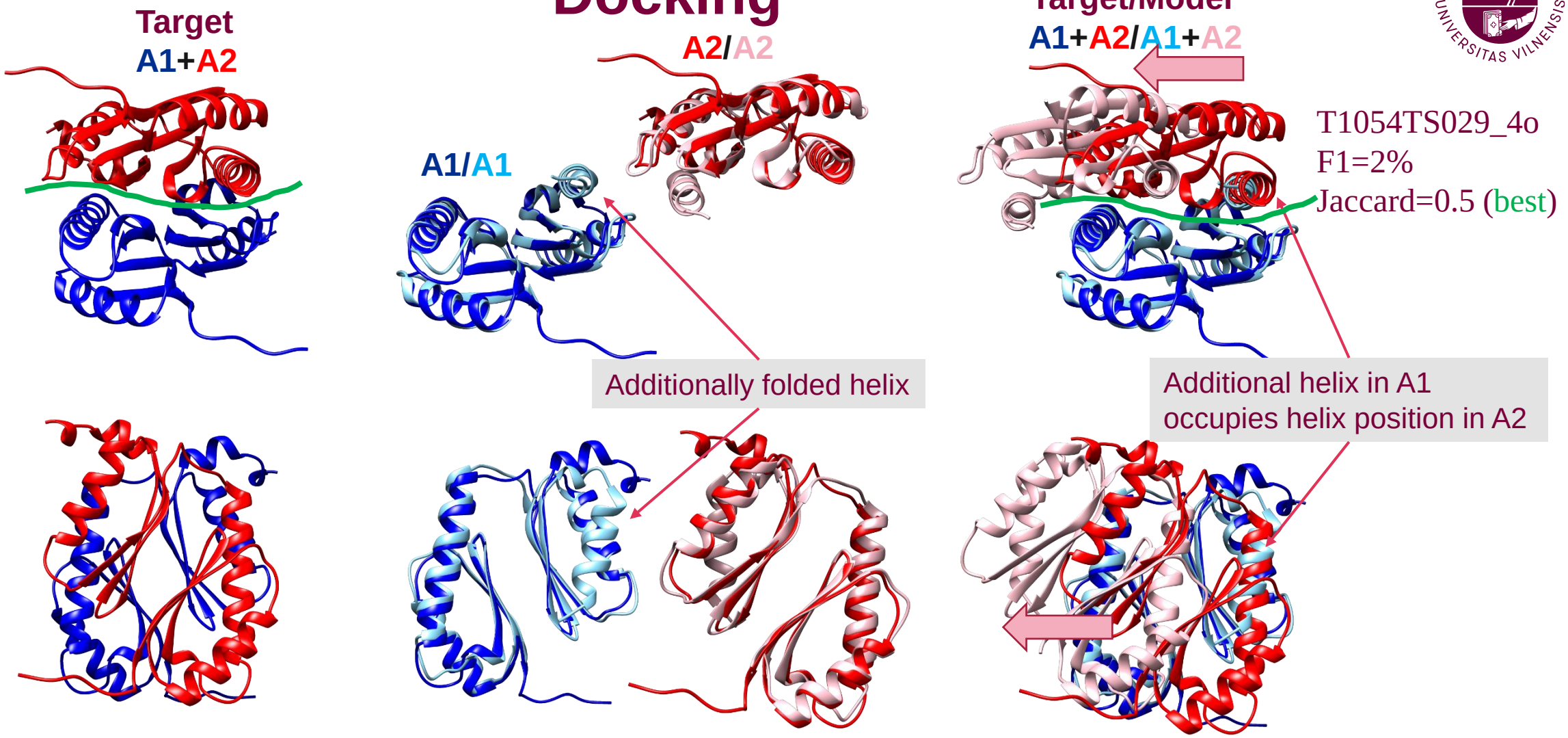
Interfaces similar
Interface energy good





Selection identifies correct interface patch, but not the correct orientation:
Scoring problem? Need additional constraints – contacts?

# T1054 homodimer: BON domain containing protein Docking



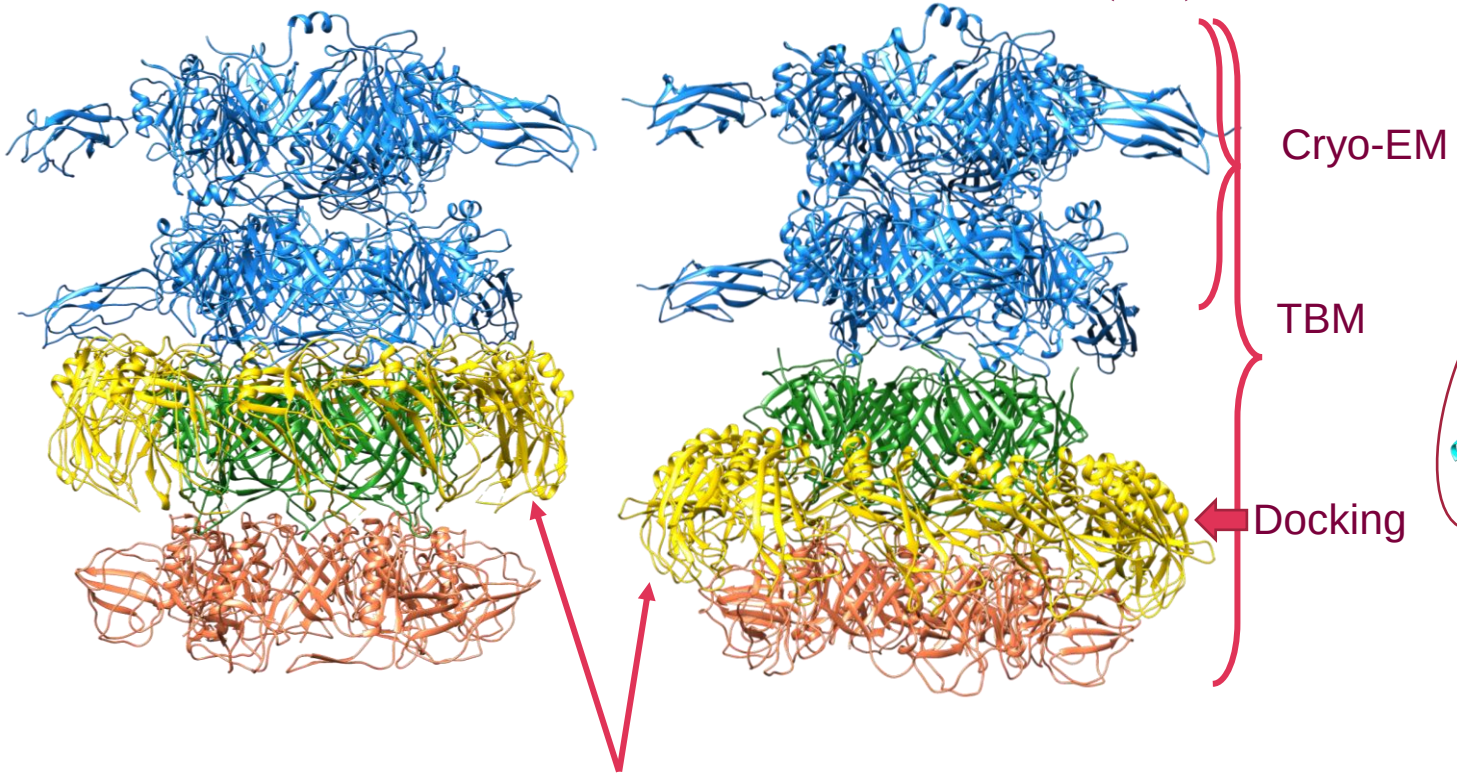

Monomeric structure “too well-folded” to be compatible with the native dimer: a shift is necessary

# H1060:T5 phage tail subcomplex A6B3C12D6
# Hybrid modeling

Target

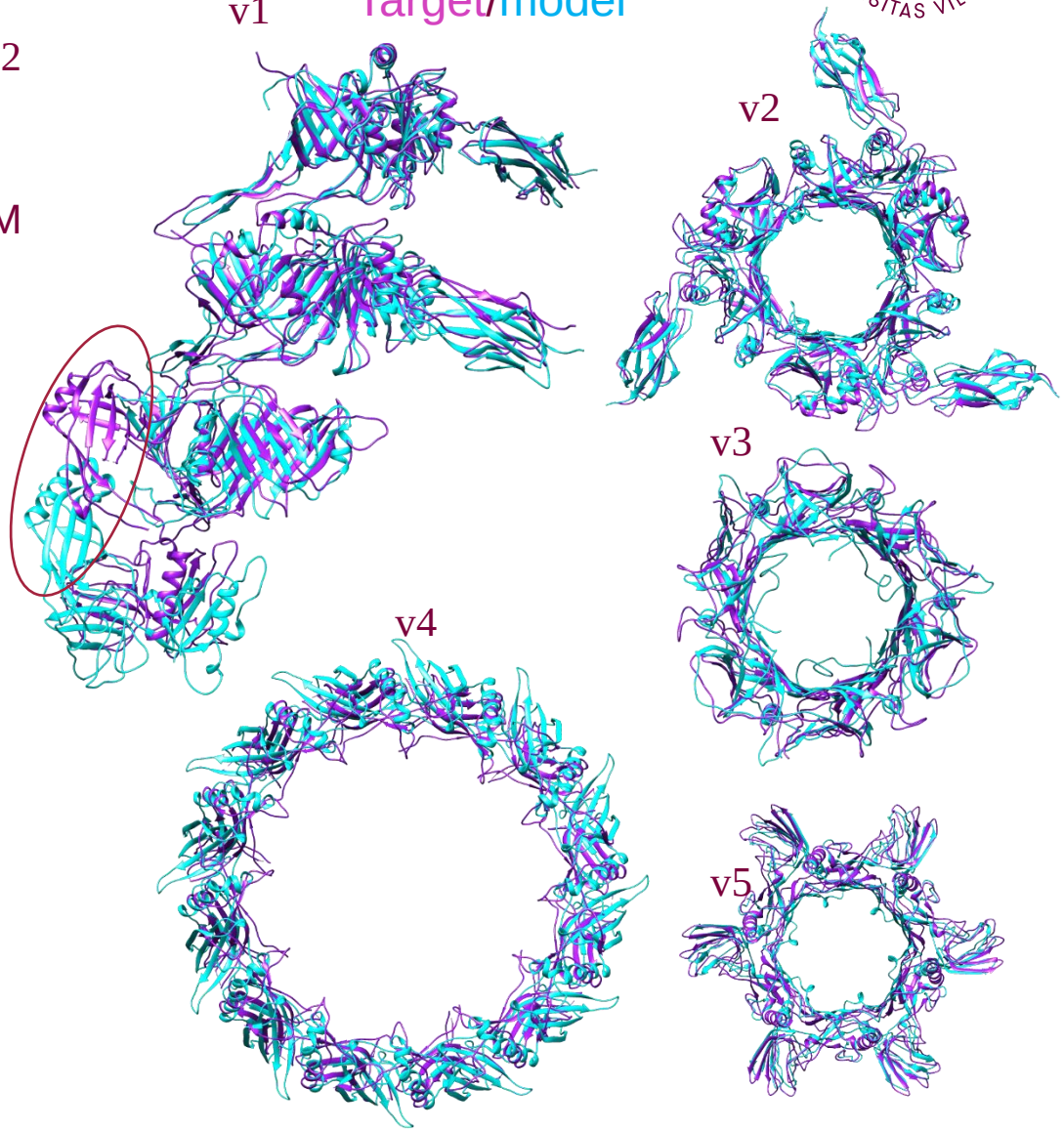
Model

H1060v0TS029_1 F1=40.8% (best); Jaccard=0.72




Relative position of the largest ring missed

Target/model

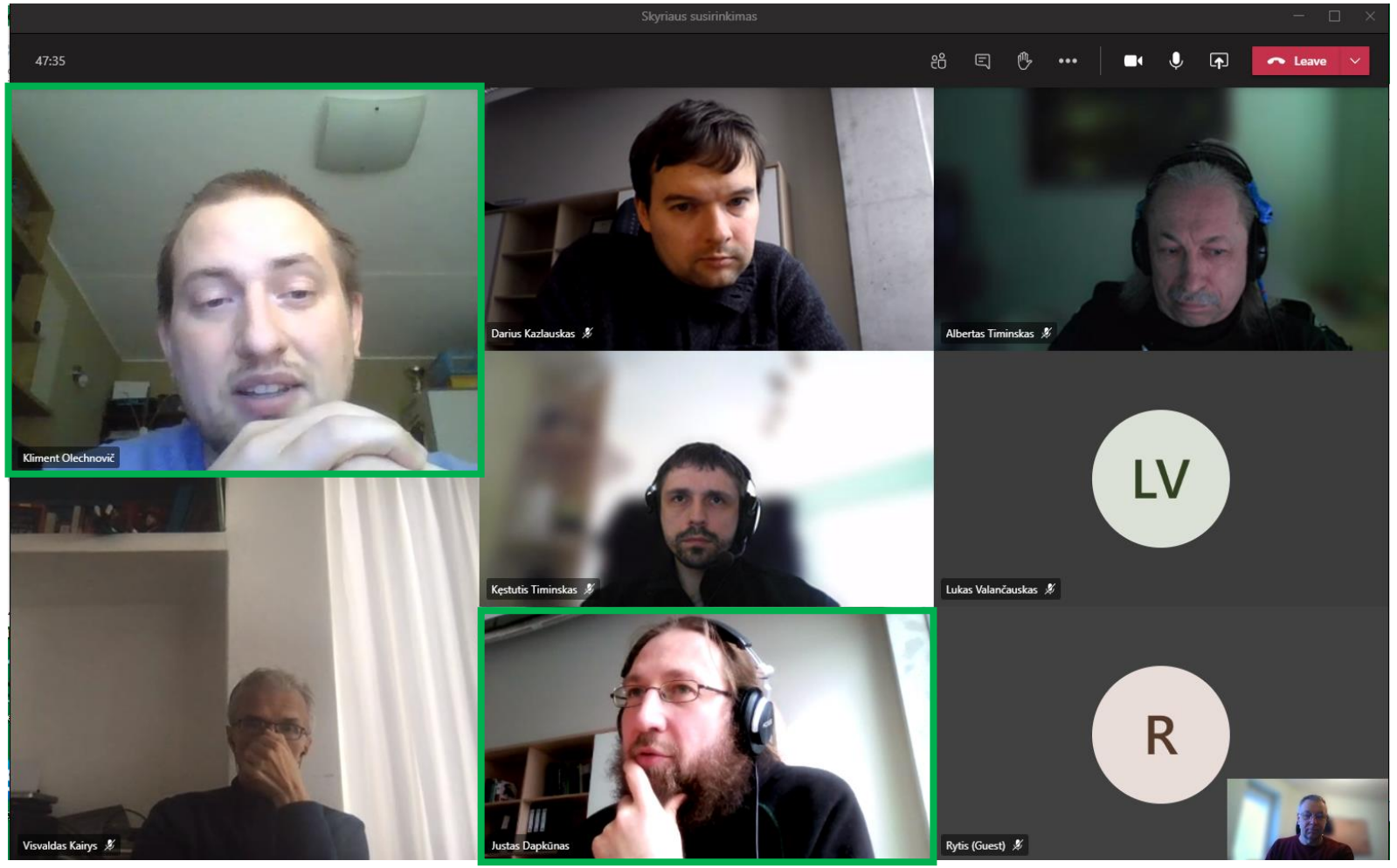



- Individual rings modeled fairly accurately
- Some problems with mutual arrangement

# Summary

- Interface prediction was relatively more successful than the overall structure modeling
- Reasons for relatively accurate interface prediction
  - Effective template identification
  - Robust selection of docking models using VoroMQA
- Reasons for relatively poor overall structure accuracy
  - Application of Modeller/Altmod to generate models
- Docking is much better in predicting interface patches than the orientation of subunits

# Acknowledgements




**Kliment Olechnovič**

**Justas Dapkūnas**


Laboratory of Bioinformatics, Institute of Biotechnology, LSC Vilnius University


CASP & CAPRI organizers

Experimentalists

## Funding


Lietuvos mokslo taryba