

Predicting structural heterogeneity of protein contacts from singular static conformational models

Dr. Kliment Olechnovič

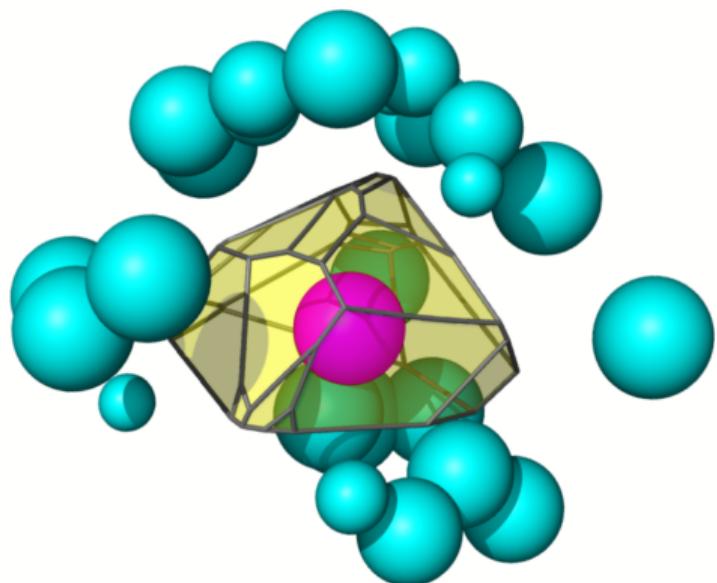
CNRS Laboratoire Jean Kuntzmann, Grenoble, France

2025-06-27

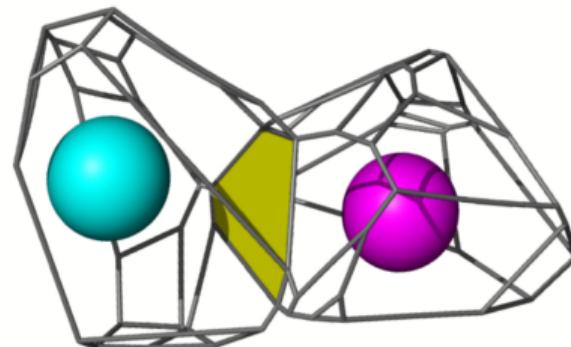


Voronoi tessellation-based analysis of structures

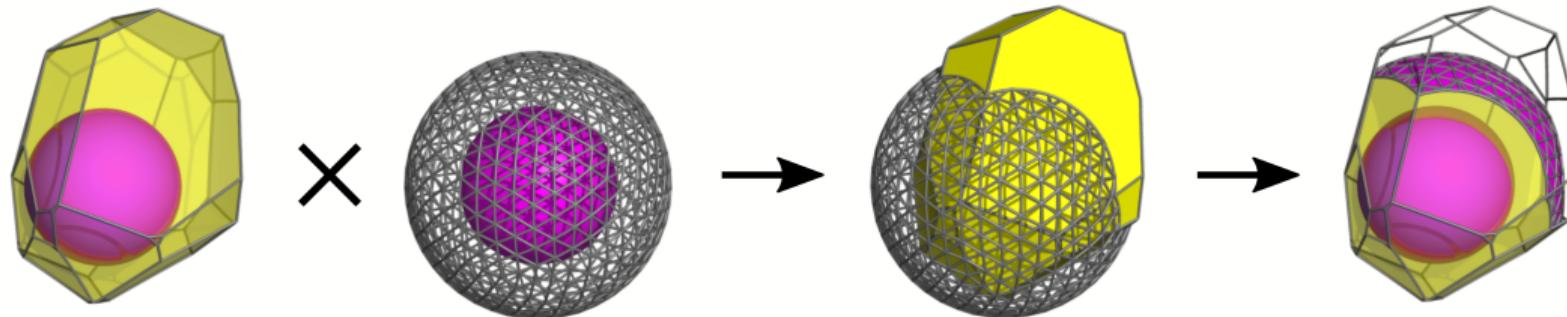
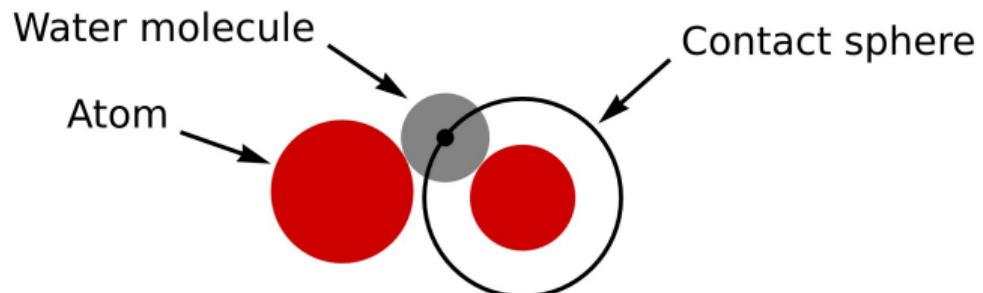
Voronoi cell of an atom surrounded by its neighbors



Atom-atom contact surface defined as the face shared by two adjacent Voronoi cells.

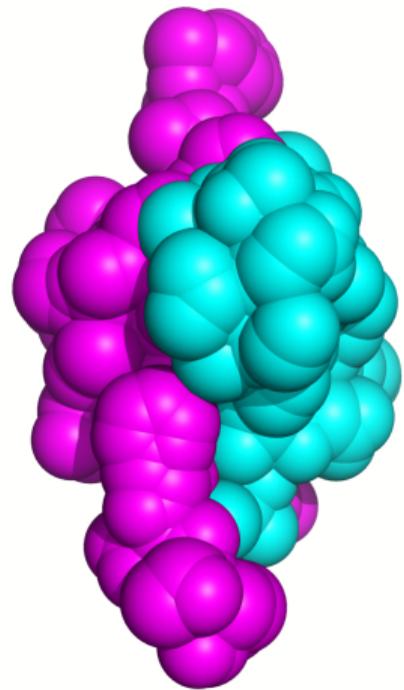


Constrained contacts

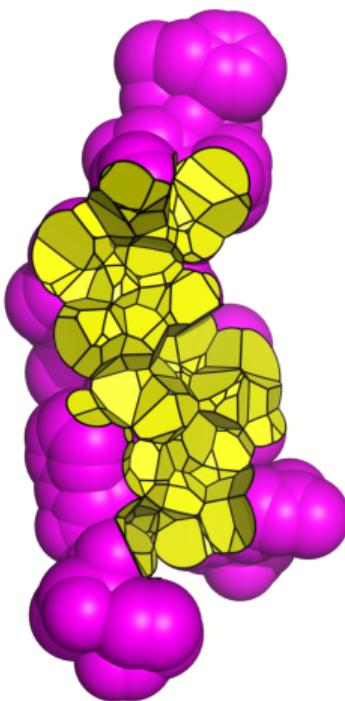


Inter-chain contacts

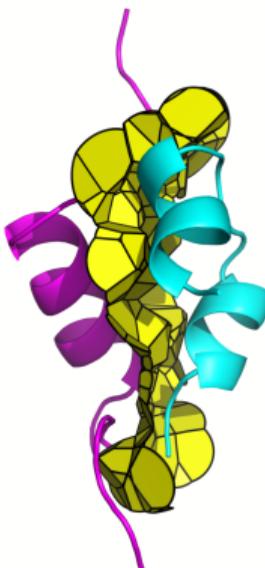
Solvent-accessible surface
of an insulin heterodimer
PDB:4UNG colored by subunit



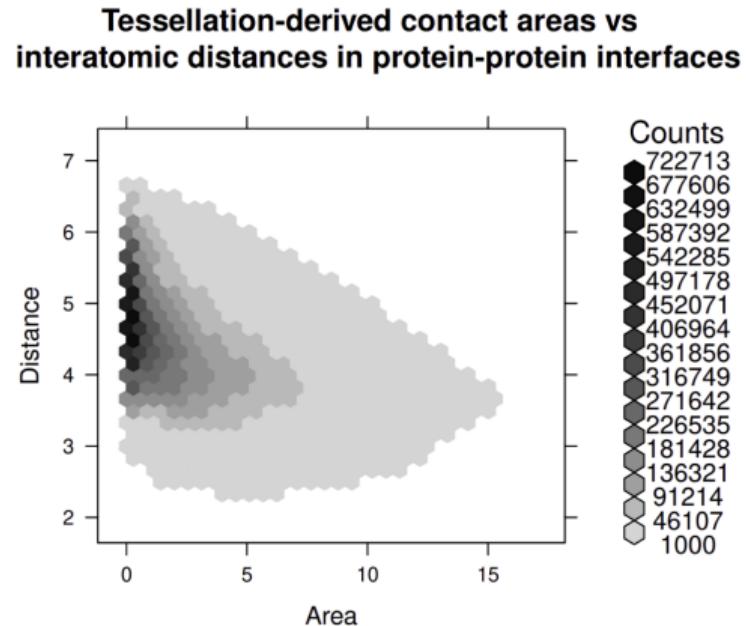
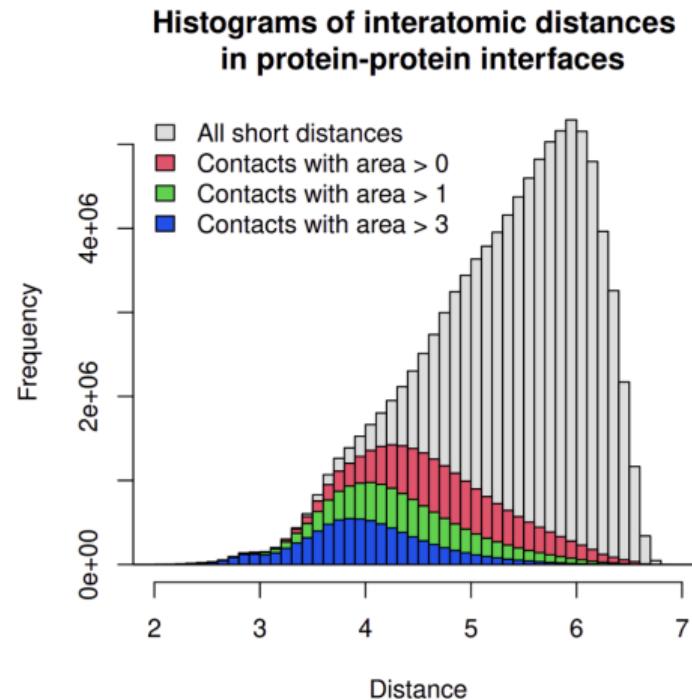
The intersubunit interface
shown together with the
SAS of one subunit



The intersubunit interface
shown together with
both subunits represented
as cartoons



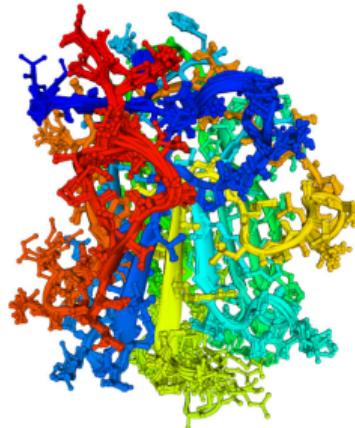
Inter-chain contact areas vs distances, PDB-based statistics



$$\text{corr}(\text{area}, \text{distance}) \approx -0.43$$

Contact area persistence values for PDB ensembles

Ensemble of conformations from PDB, based on a 90% sequence identity cluster

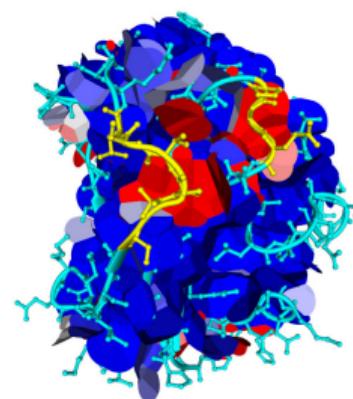


Residue-residue contact areas of every conformation, minimum seq. sep. = 6, colored by area

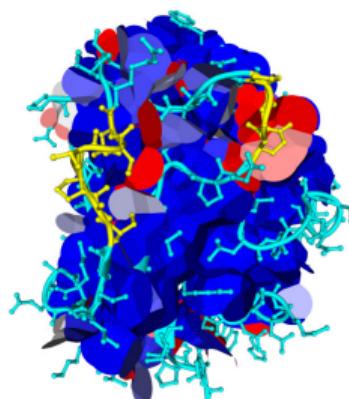


Residue-residue contact areas colored by ensemble-wide persistence

PDB ID = 1D2S



PDB ID = 6PYB

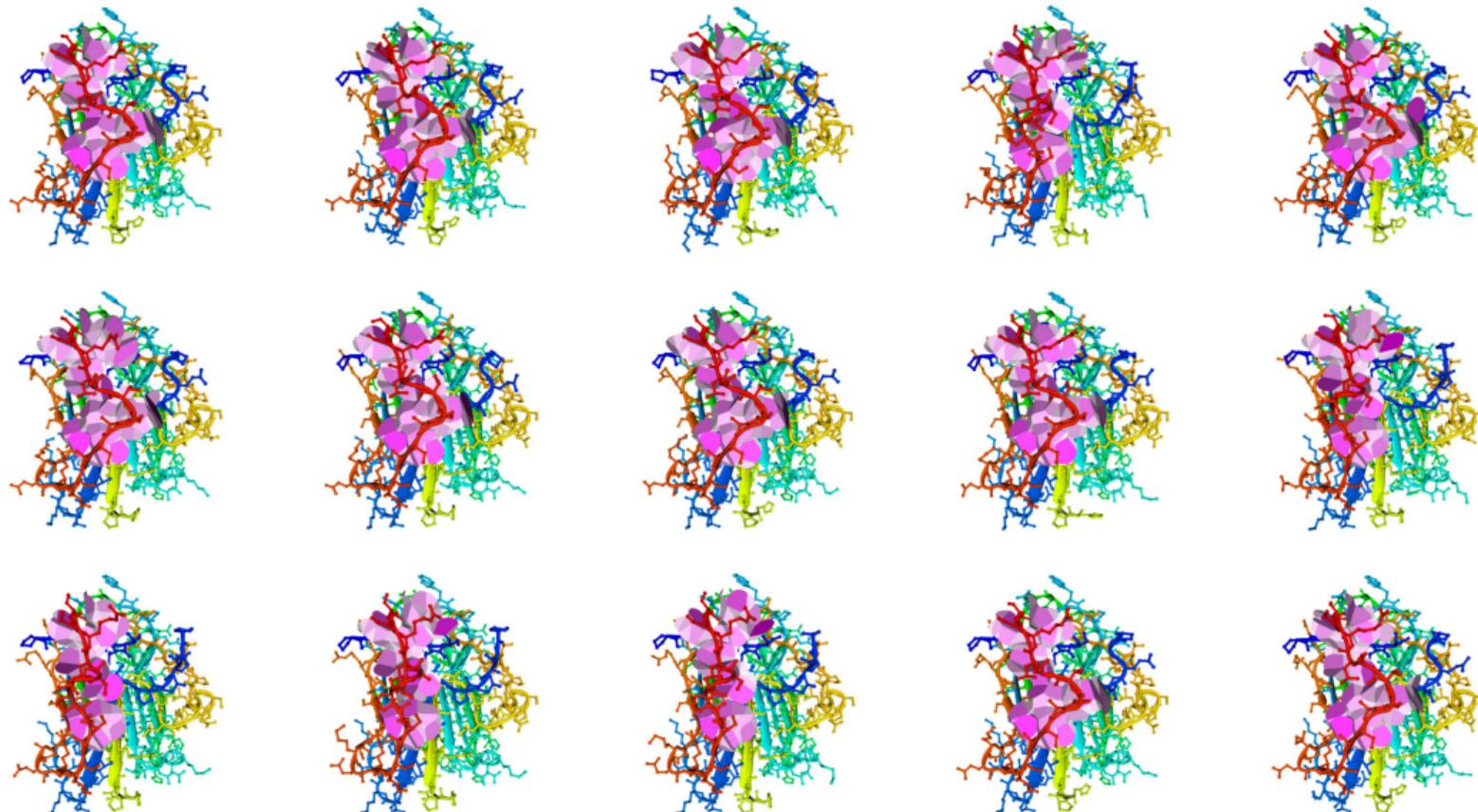


If a contact with ID $\kappa = (\text{residue}_a, \text{residue}_b)$ is present in at least one conformation, and both residue_a and residue_b are present in at least two conformations, then $\text{area}_{\min}(\kappa) \geq 0$ and $\text{area}_{\max}(\kappa) > 0$ are available.

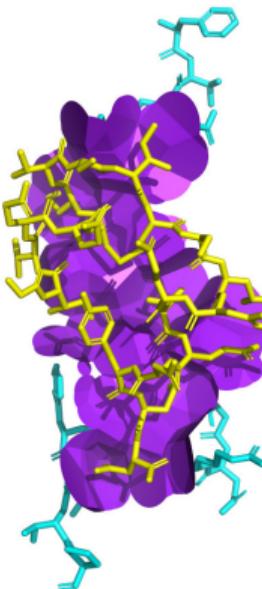
Then the ensemble-wide contact area persistence is defined as $\text{persistence}(\kappa) = \frac{2 \cdot \text{area}_{\min}(\kappa)}{\text{area}_{\min}(\kappa) + \text{area}_{\max}(\kappa)}$

Ensemble contacts for a sub-interface — risky animation

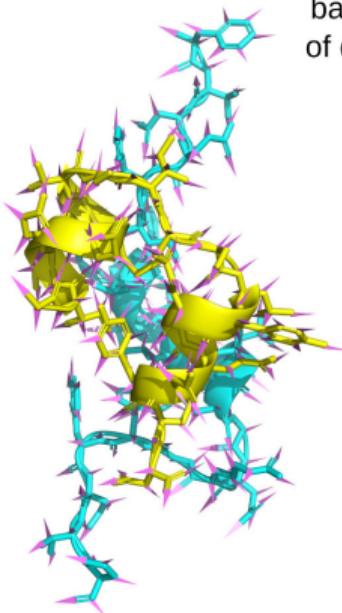
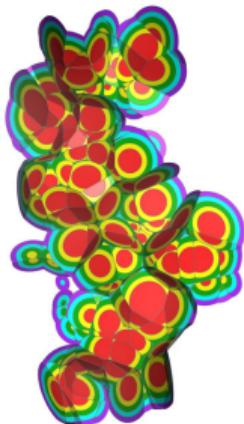
Ensemble contacts for a sub-interface — reliable frames



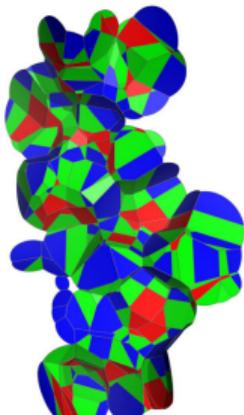
Extended geometry of contacts



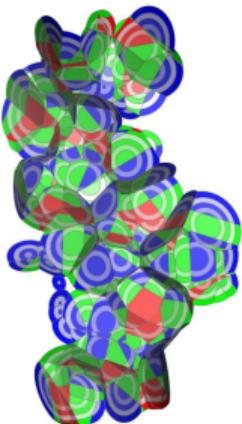
"Layers" of contacts
based on distances
from outer boundaries



"Sectors" of contacts
based on halfspaces
of directions of atoms

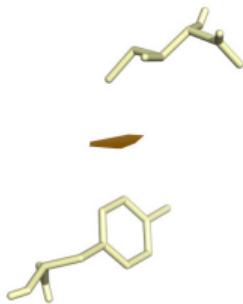


"Layers" and "Sectors"
combined to define multiple
geometric contact categories

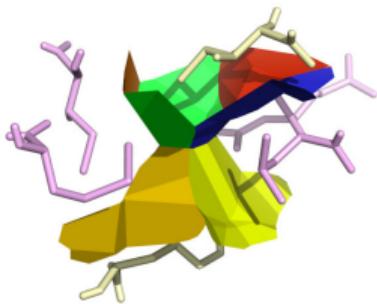


Voronoi contacts block (VCBlock)

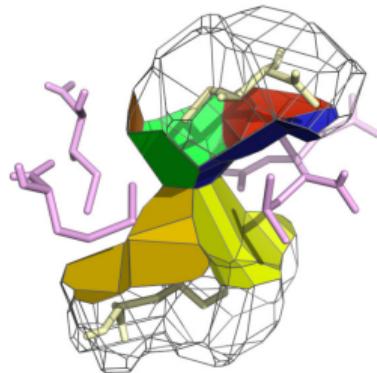
Tessellation-derived contact surface between two main residues



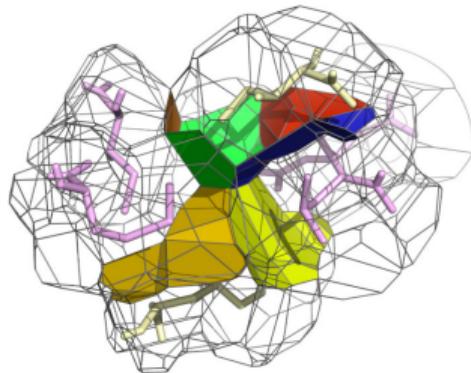
+ Adjacent contacts with shared residues (side residues)



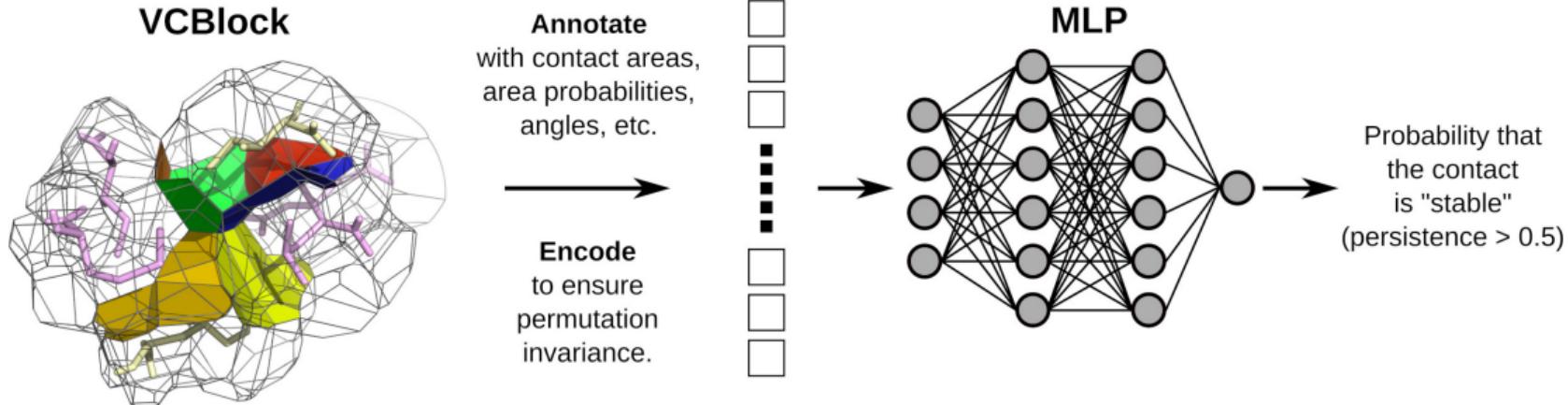
+ Interface of the main residues with the environment



+ Interface of the side residues with the environment

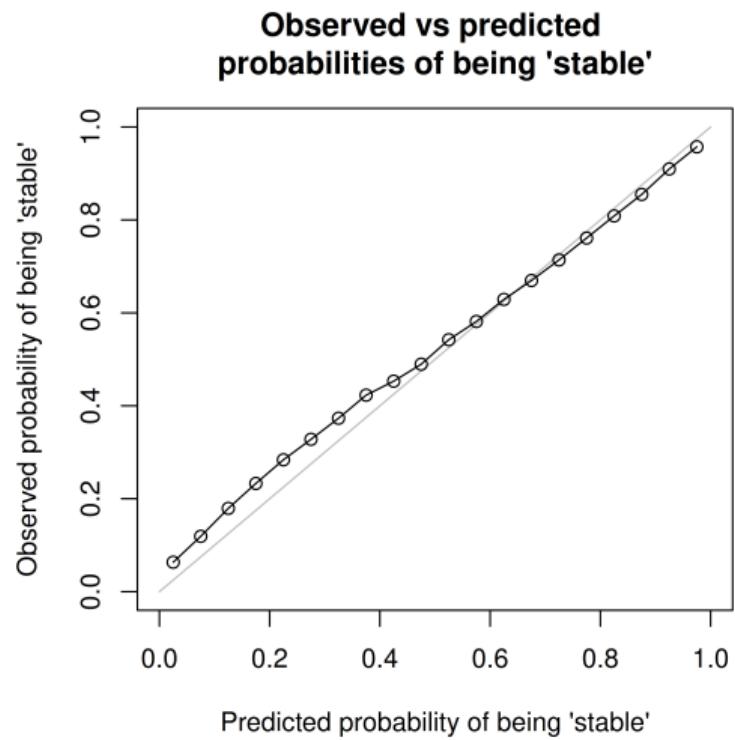
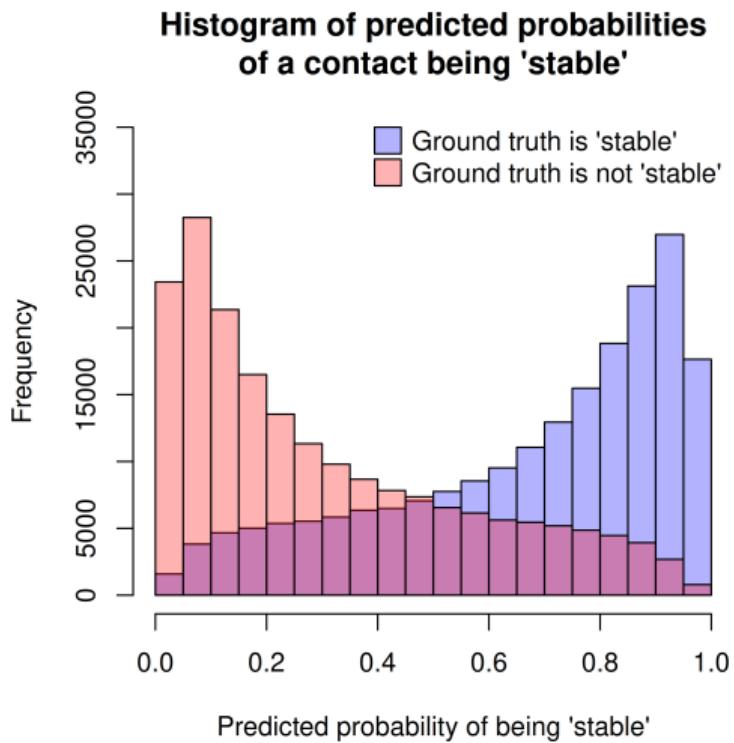


VCBlock-based contacts stability predictor



Current MLP training/validation/testing was done using a set of $\sim 2 \times 10^6$ VCBlocks sampled from $\sim 2 \times 10^4$ protein ensembles collected from PDB. The data split was done respecting the 30% sequence identity-based clustering.

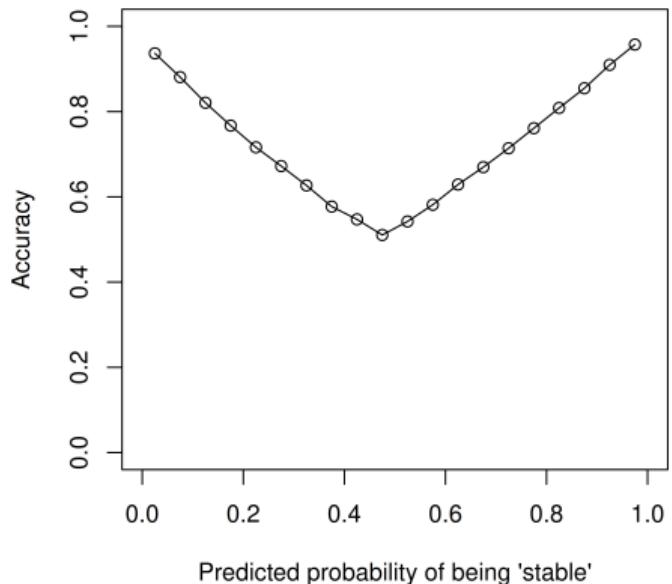
Classifier performance on the testing set



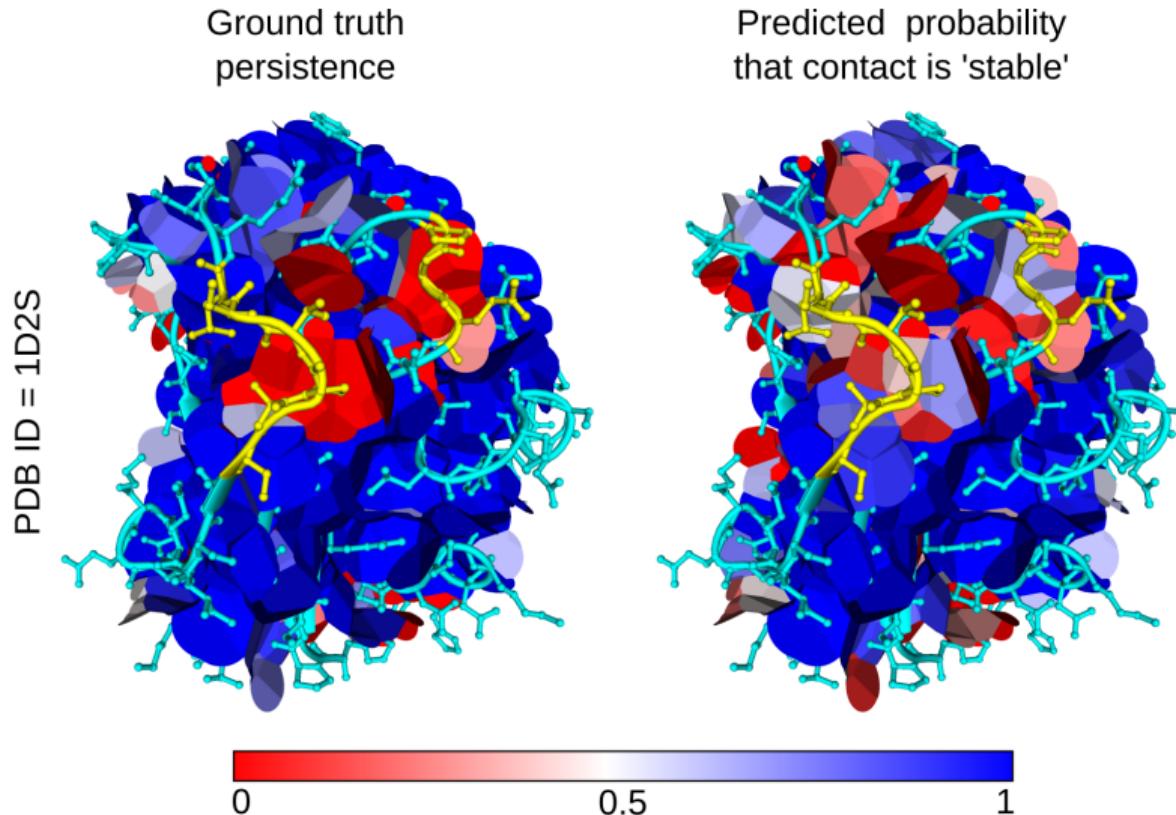
Classifier performance on the testing set

Accuracy ~ 0.78 overall, but extreme values are predicted more accurately

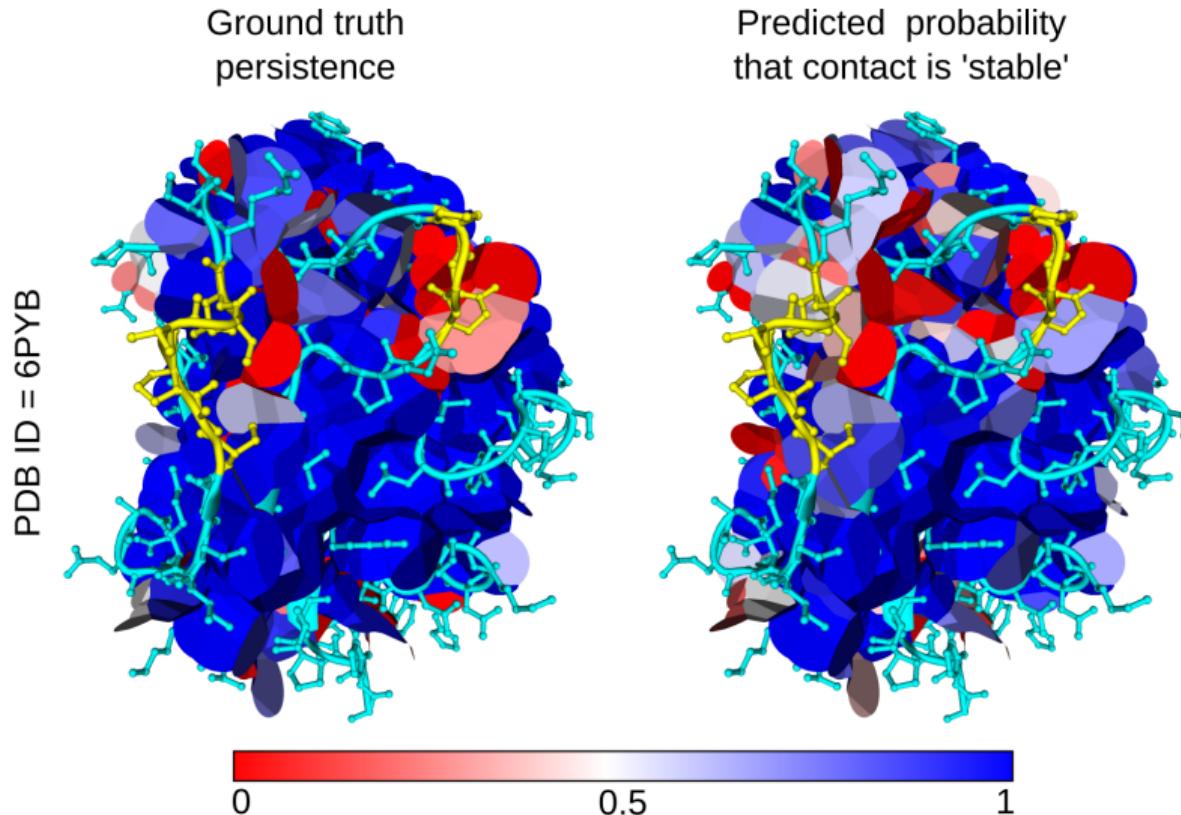
Accuracy vs predicted probability of being 'stable'



Example of MLP predictions vs ground truth

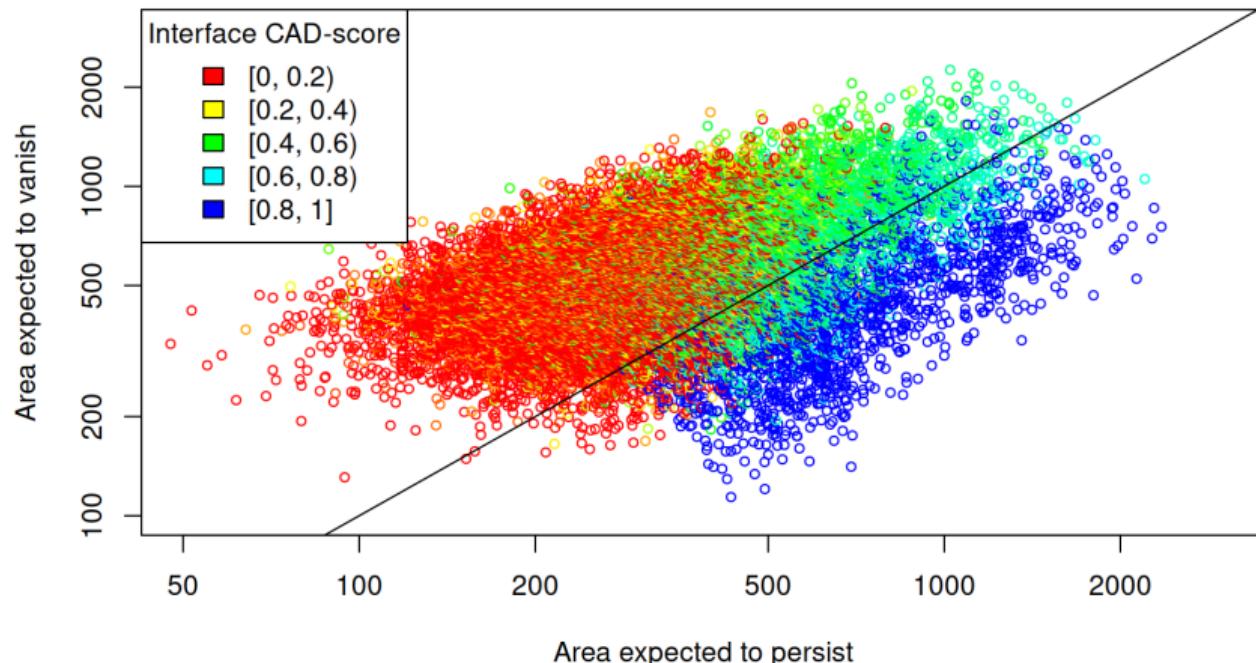


Example of MLP predictions vs ground truth



Interface area persistence predicted for native and redocked heterodimers

VoroMarmotte global scores for native and predicted inter-chain interfaces,
colored by similarity to native interface (CAD-score)



Thanks

Thank you!

CNRS Laboratoire Jean Kuntzmann:

- ▶ Sergei Grudinin
- ▶ The GruLab Team
(<https://grulab.imag.fr>)

Useful links:

- ▶ <https://www.voronota.com>
- ▶ <https://www.kliment.lt>
- ▶ <https://www.bioinformatics.lt>



Funded by
the European Union