

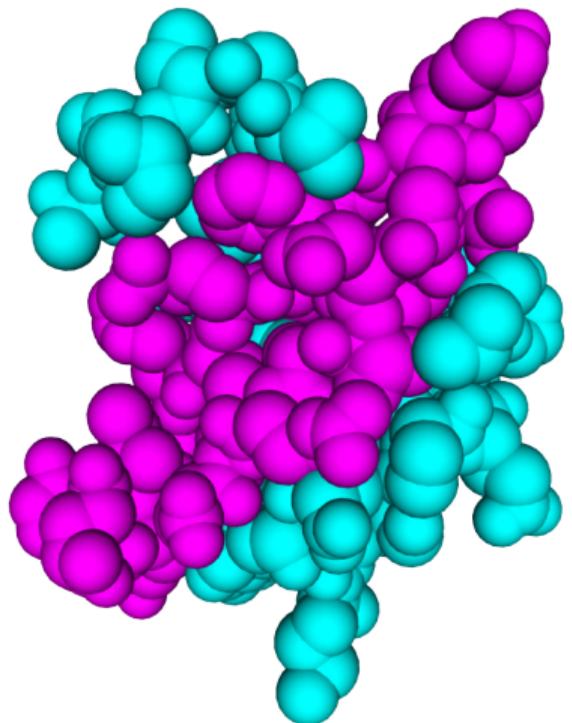
Inferring contact area heterogeneity from static protein conformations

Dr. Kliment Olechnovič

CNRS Laboratoire Jean Kuntzmann, Grenoble, France

2025-08-23





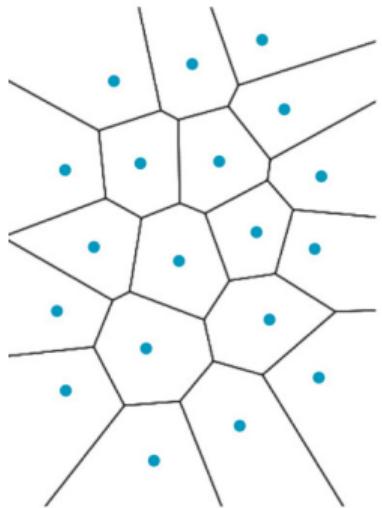
Common problems:

- ▶ analyzing how different parts in a molecule interact
- ▶ selecting the best model of a multimeric complex

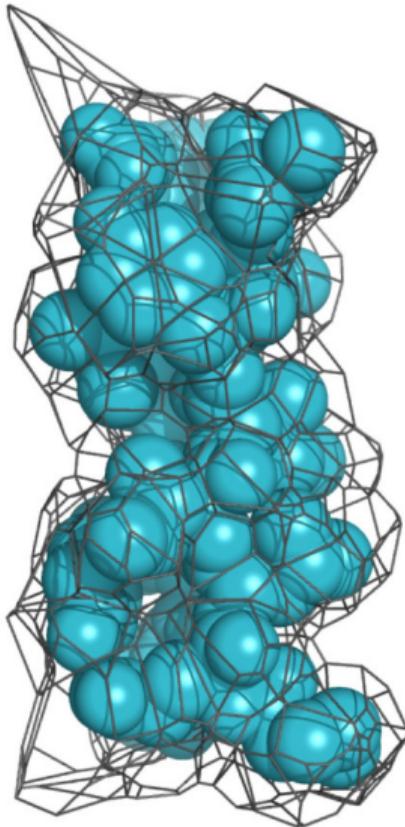
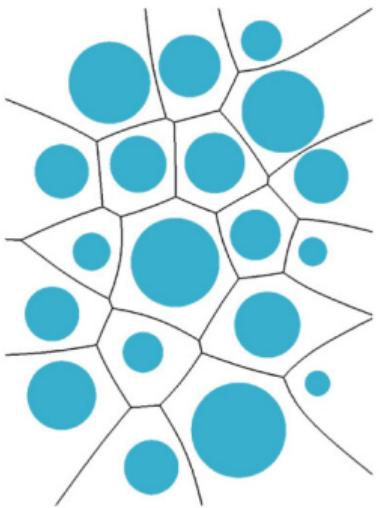
Describing interactions in molecular conformations using the
Voronoi tessellation

Voronoi diagram of points and balls

"Classic" Voronoi diagram
of points

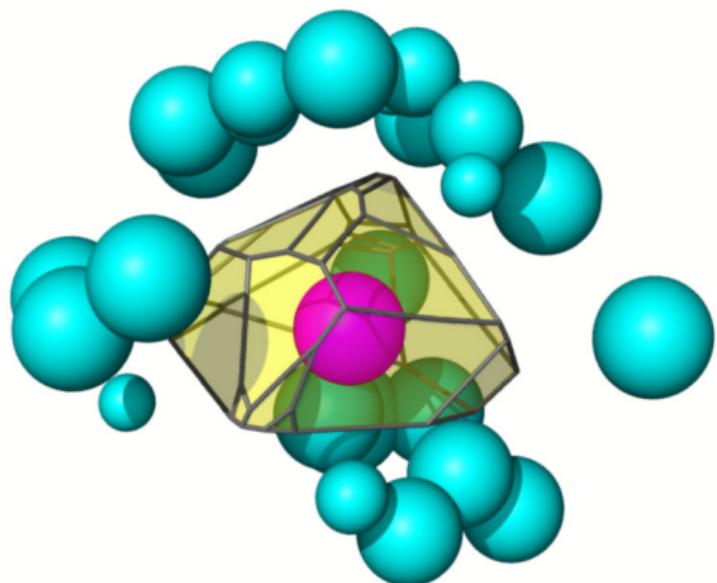


Voronoi diagram
of balls

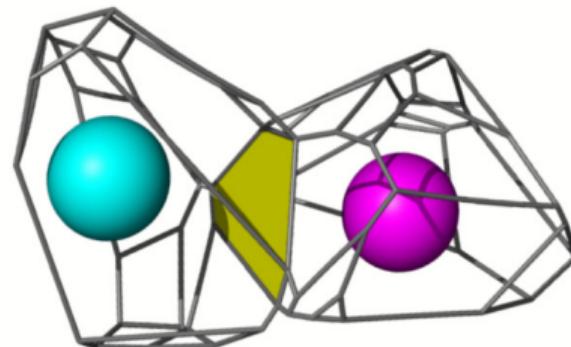


Voronoi tessellation-based analysis of structures

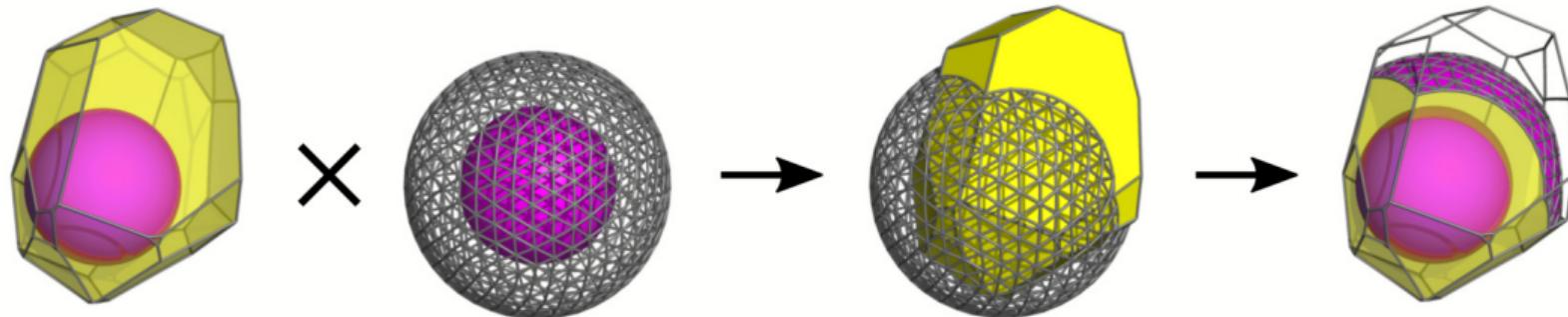
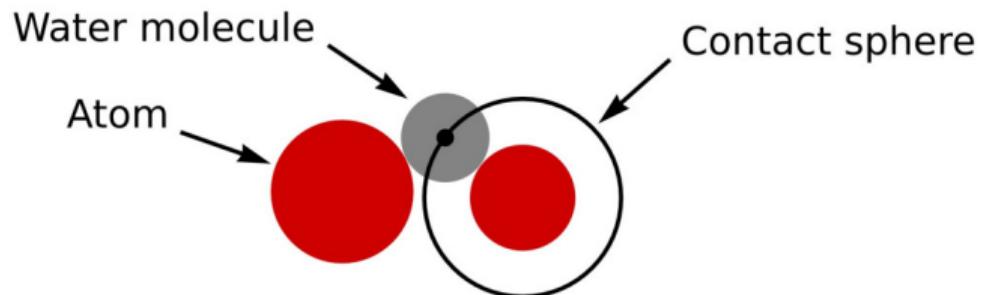
Voronoi cell of an atom surrounded by its neighbors



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

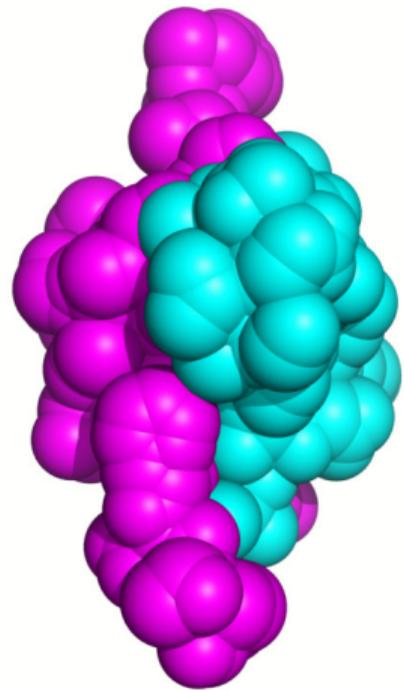


Constrained contacts

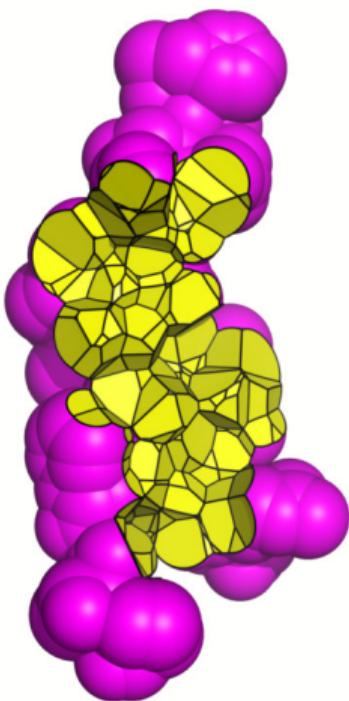


Inter-chain contacts

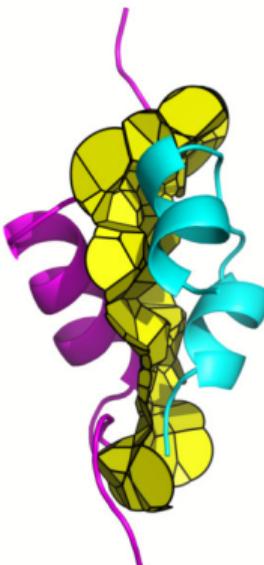
Solvent-accessible surface
of an insulin heterodimer
PDB:4UNG colored by subunit



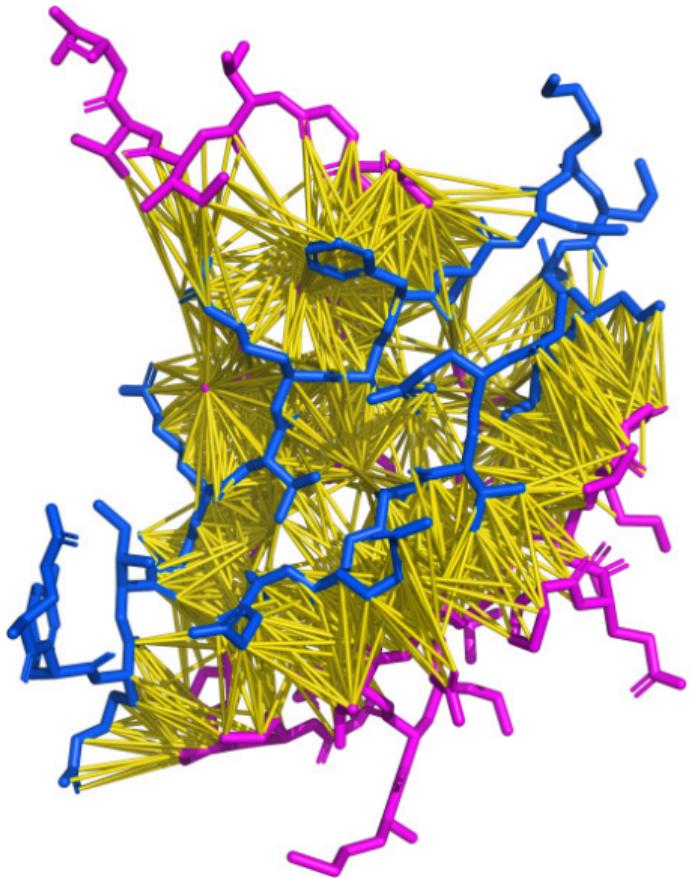
The intersubunit interface
shown together with the
SAS of one subunit



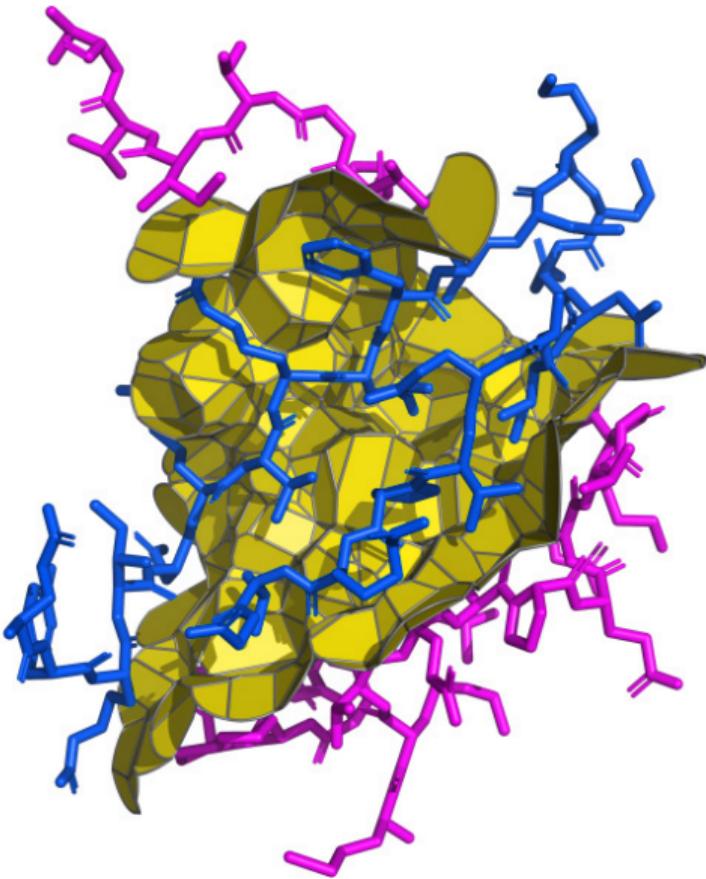
The intersubunit interface
shown together with
both subunits represented
as cartoons



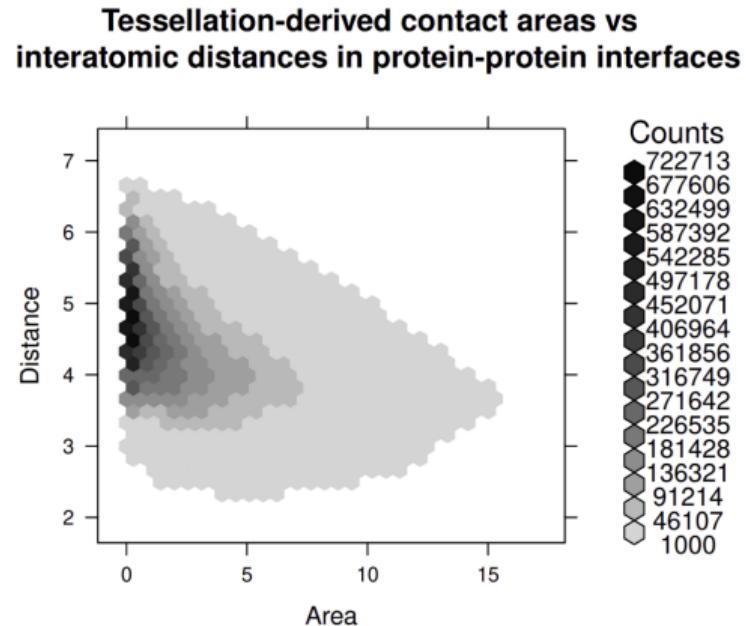
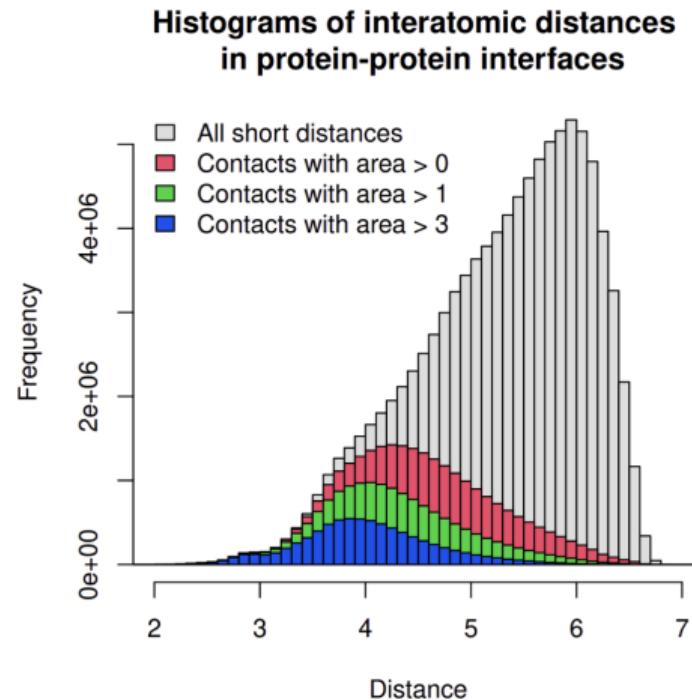
Inter-chain contact areas vs distances



VS



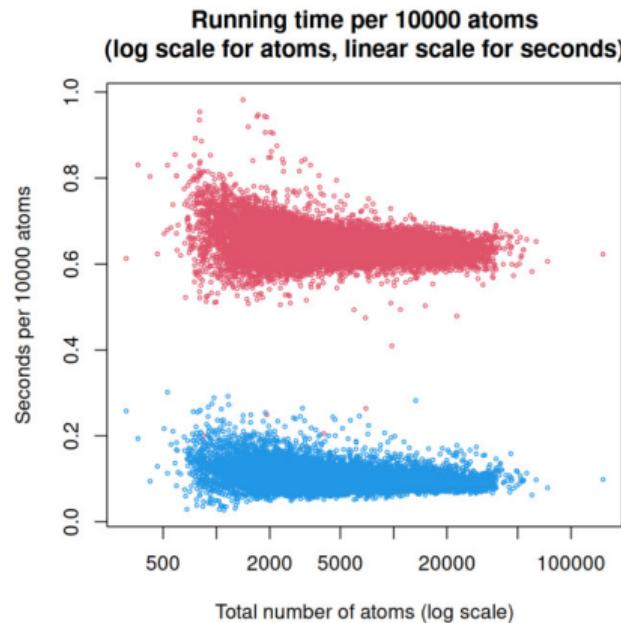
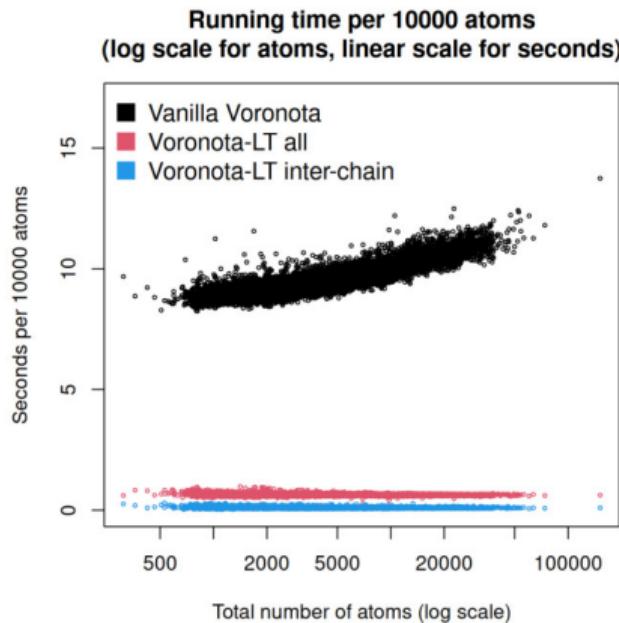
Inter-chain contact areas vs distances, PDB-based statistics



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

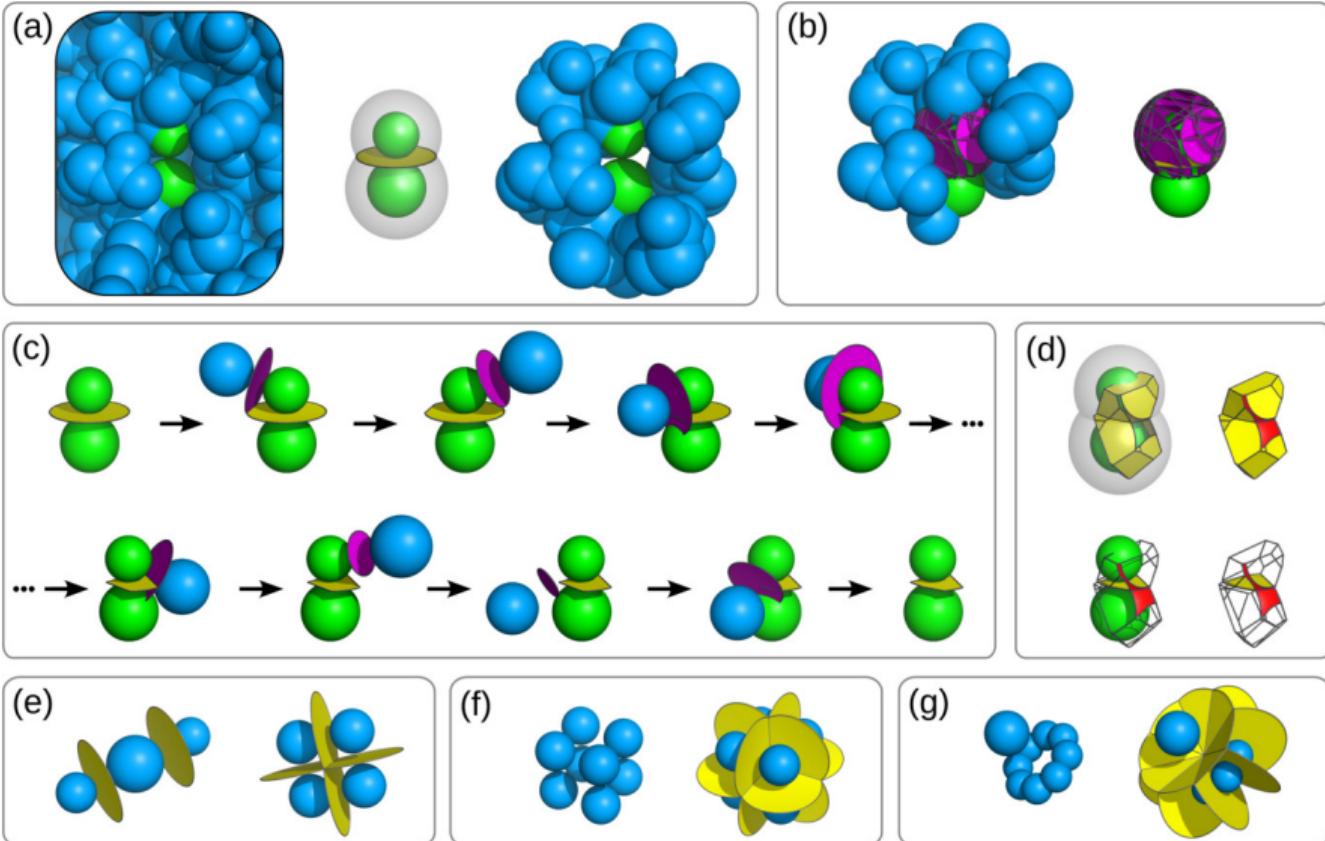
Voronota-LT

Voronota-LT is a new fast software for constructing tessellation-derived atomic contact areas and volumes. It is significantly faster than its predecessor, Voronota:



Olechnovic and Grudinin. *Voronota-LT: efficient, flexible and solvent-aware tessellation-based analysis of atomic interactions*. JCC (2025)

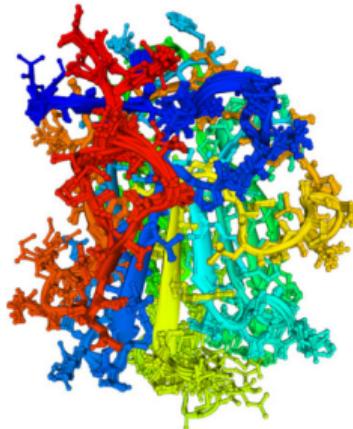
Voronota-LT



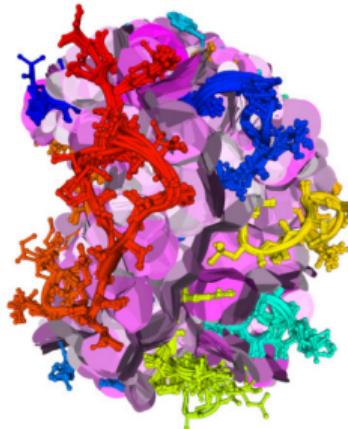
Predicting structural heterogeneity of protein contacts from
singular static conformational models

Contact area persistence values for PDB ensembles

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

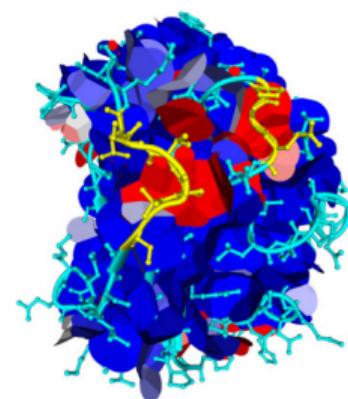


Residue-residue contact areas of every conformation, minimum seq. sep. = 6, colored by area

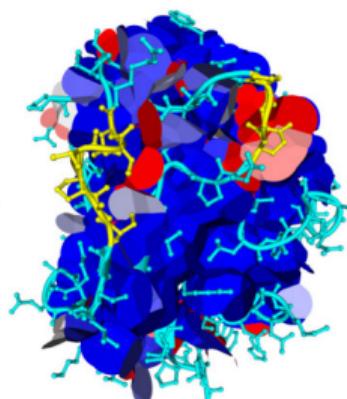


Residue-residue contact areas colored by ensemble-wide persistence

PDB ID = 1D2S

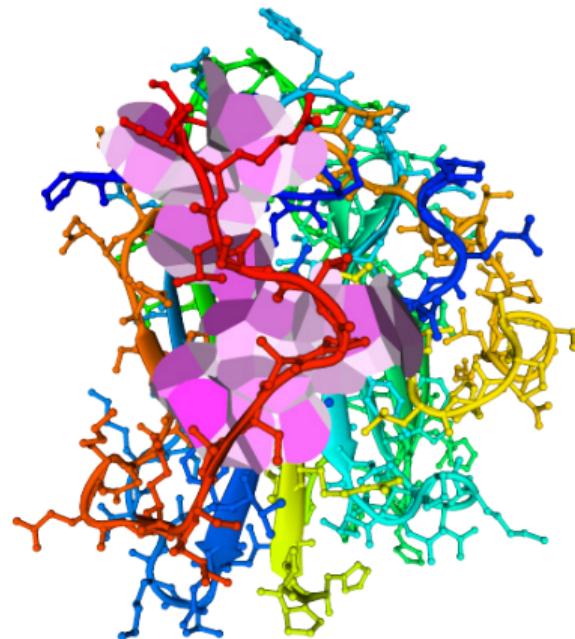


PDB ID = 6PYB

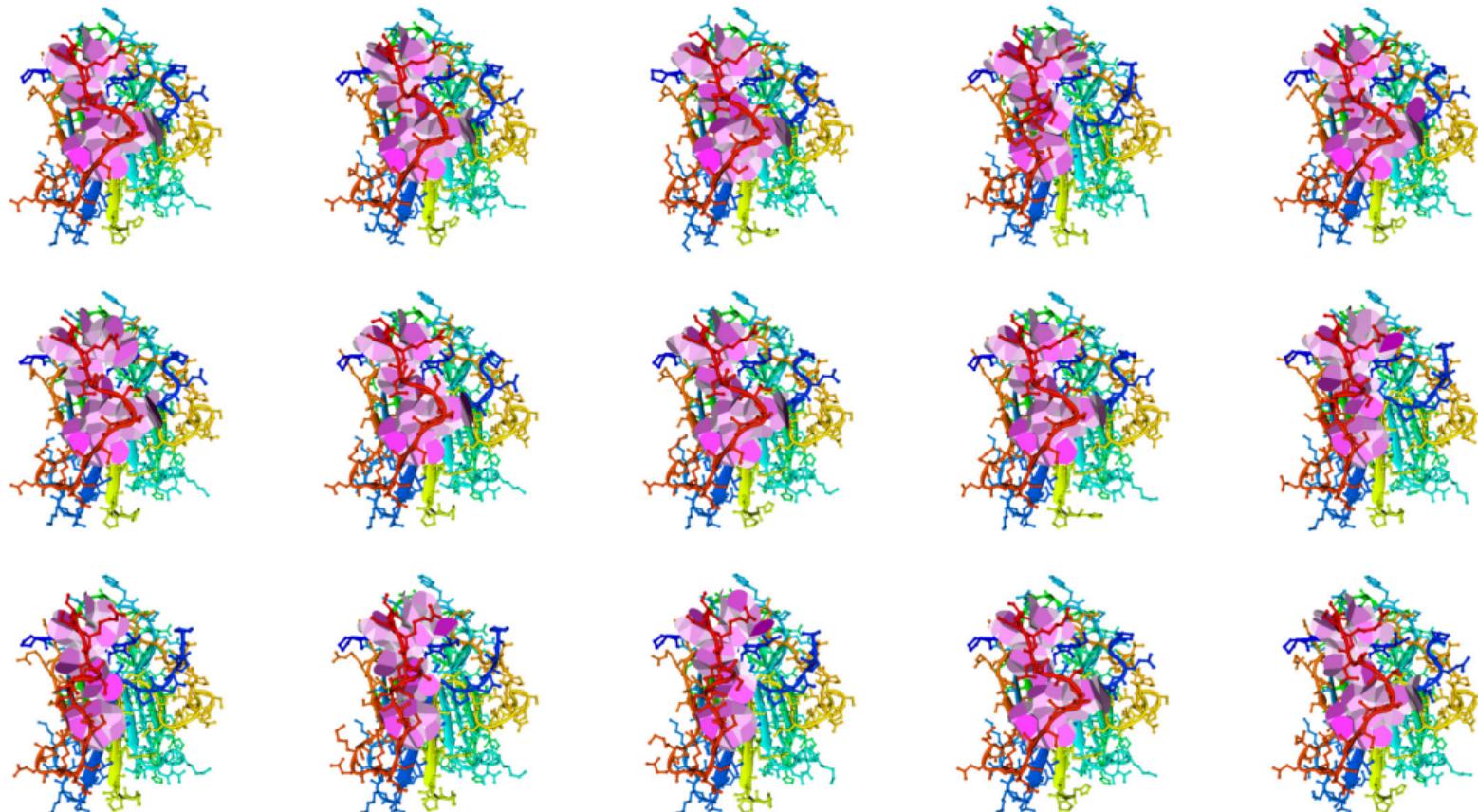


If a contact with $\kappa = (\text{residue}_a, \text{residue}_b)$ is present in at least one conformation, and both residue_a and residue_b are present in at least two conformations, then $\text{area}_{\min}(\kappa) \geq 0$ and $\text{area}_{\max}(\kappa) > 0$ are available. Then the ensemble-wide contact area persistence is defined as $\text{persistence}(\kappa) = \frac{2 \cdot \text{area}_{\min}(\kappa)}{\text{area}_{\min}(\kappa) + \text{area}_{\max}(\kappa)}$

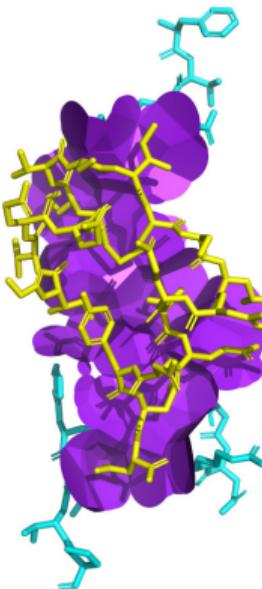
Ensemble contacts for a sub-interface — risky animation



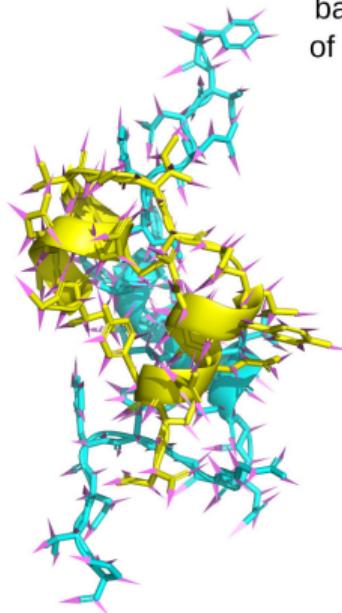
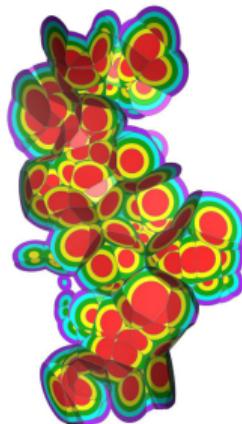
Ensemble contacts for a sub-interface — reliable frames



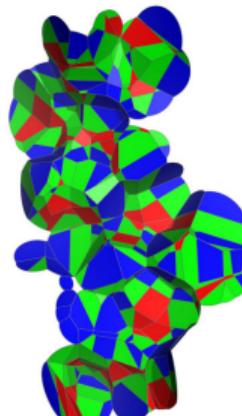
Extended geometry of contacts



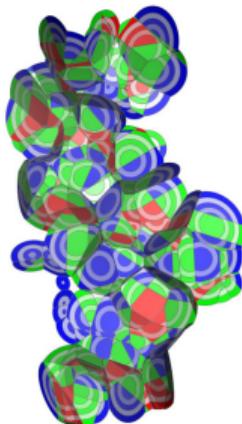
"Layers" of contacts
based on distances
from outer boundaries



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

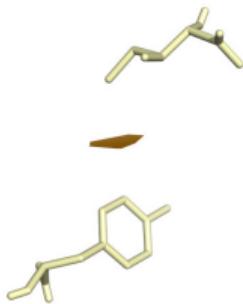


"Layers" and "Sectors"
combined to define multiple
geometric contact categories

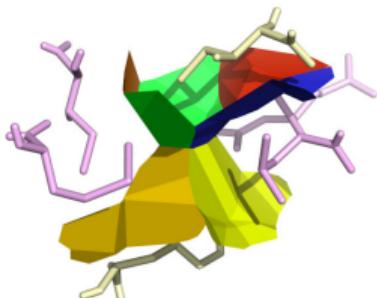


Voronoi contacts block (VCBlock)

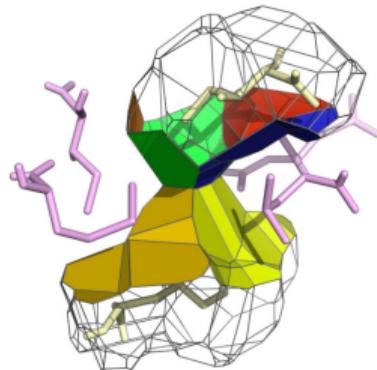
Tessellation-derived contact surface between two main residues



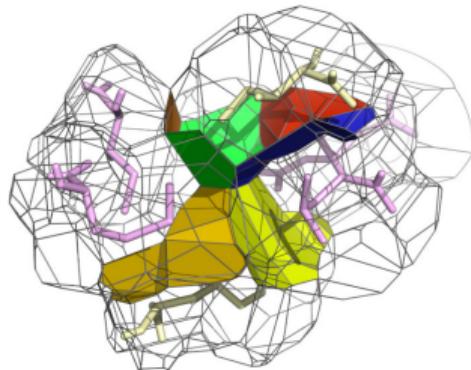
+ Adjacent contacts with shared residues (side residues)



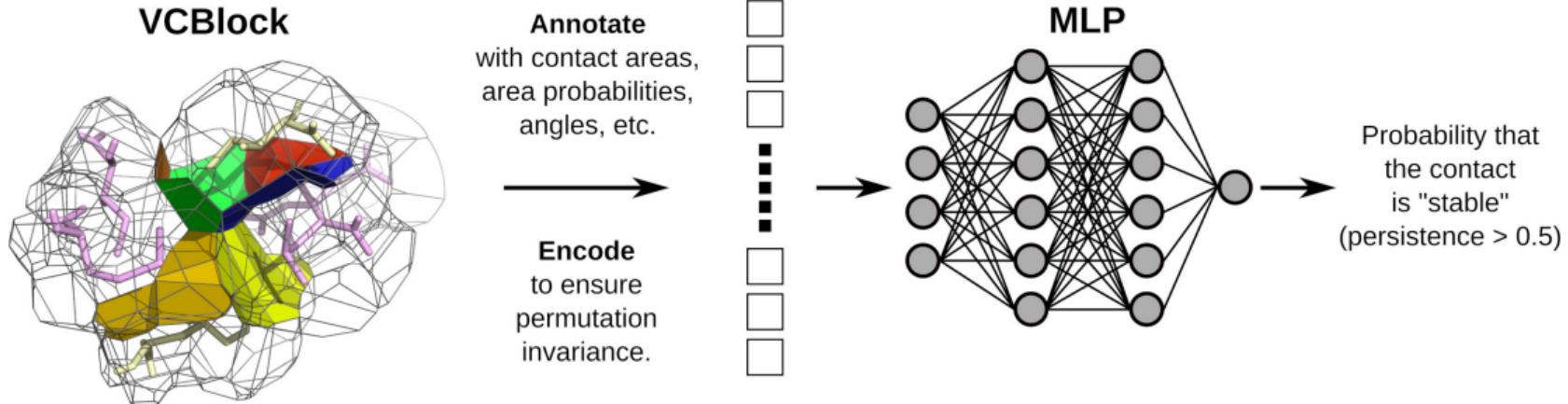
+ Interface of the main residues with the environment



+ Interface of the side residues with the environment

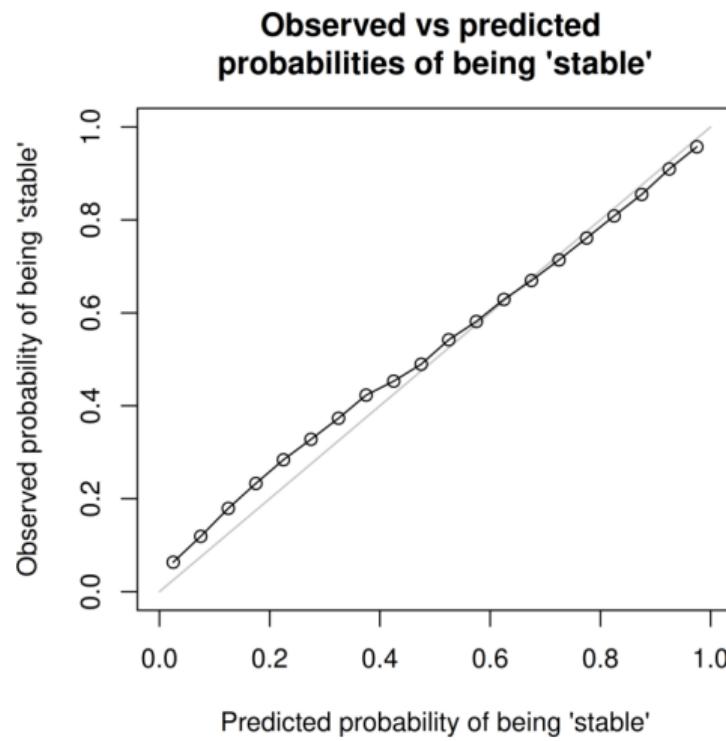
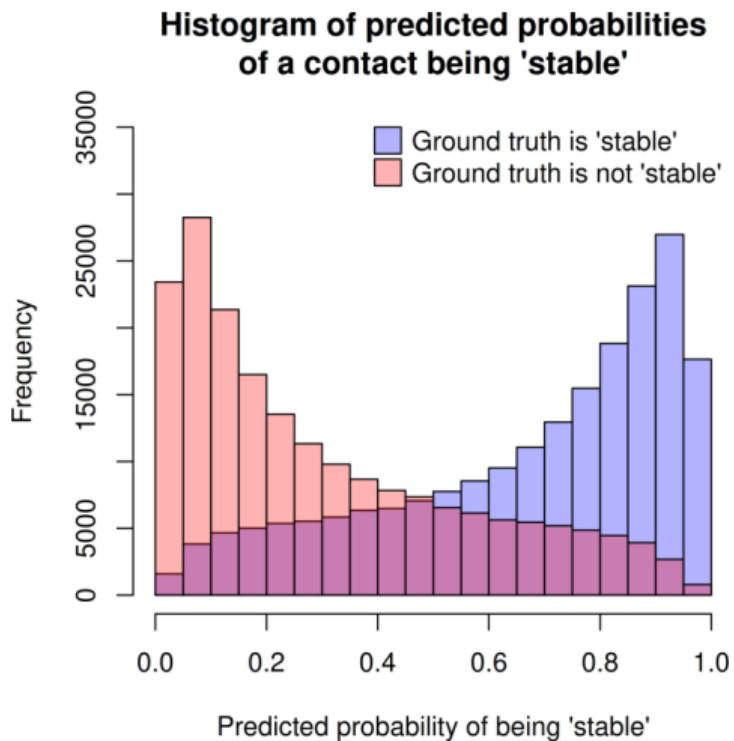


VCBlock-based contacts stability predictor



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.

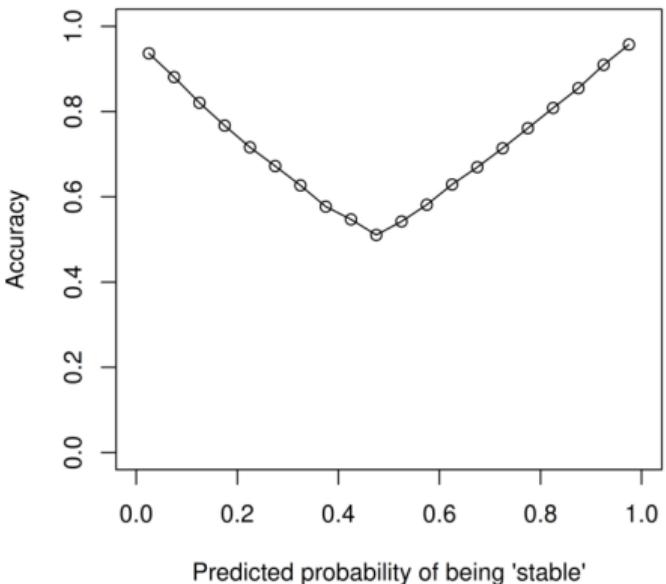
Classifier performance on the testing set



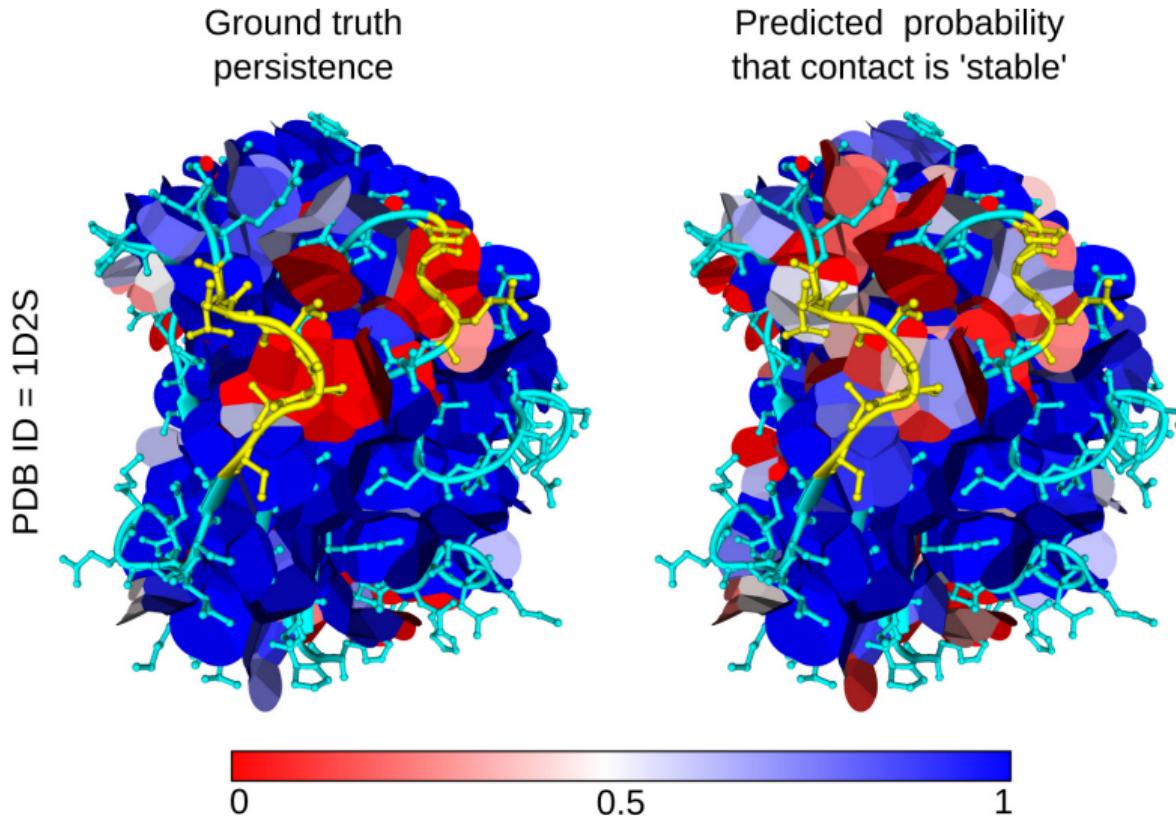
Classifier performance on the testing set

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

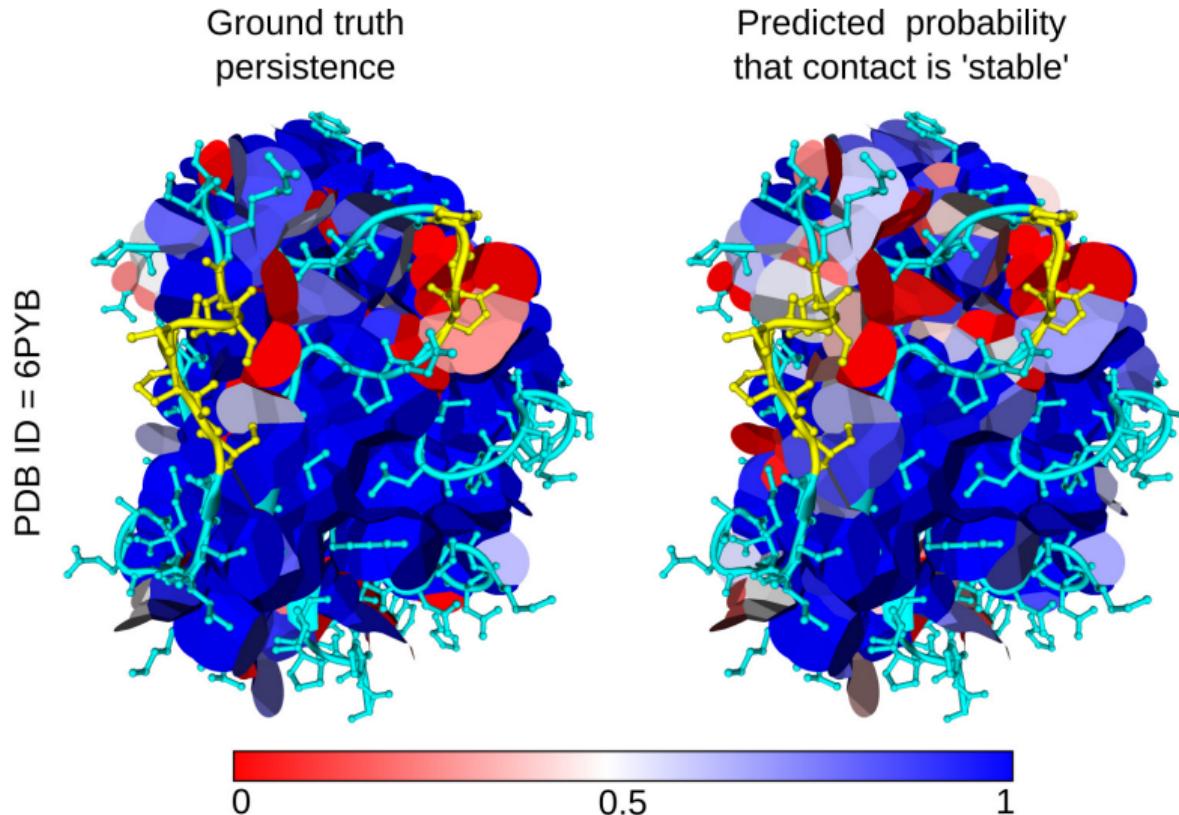
Accuracy vs predicted probability of being 'stable'



Example of MLP predictions vs ground truth



Example of MLP predictions vs ground truth



VoroMarmotte pseudoenergy

The VoroMarmotte pseudoenergy for a set of contacts K is calculated as follows:

$$E(K) = \sum_{k \in K} \text{area}(k) \cdot \log \frac{P_{\text{predicted}}(k \text{ persists}|k \text{ occurs})}{1 - P_{\text{predicted}}(k \text{ persists}|k \text{ occurs})} \quad (1)$$

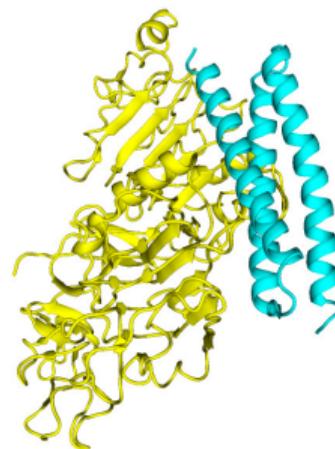
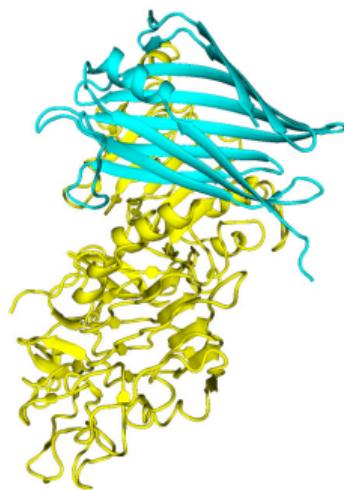
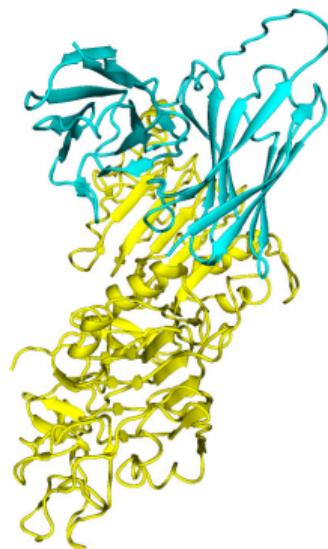
We work with probabilities, but there are some relations to physics (constants omitted):

$$\Delta G \sim E(K) \quad (2)$$

Looking at some examples from EGFR Protein Design Competition

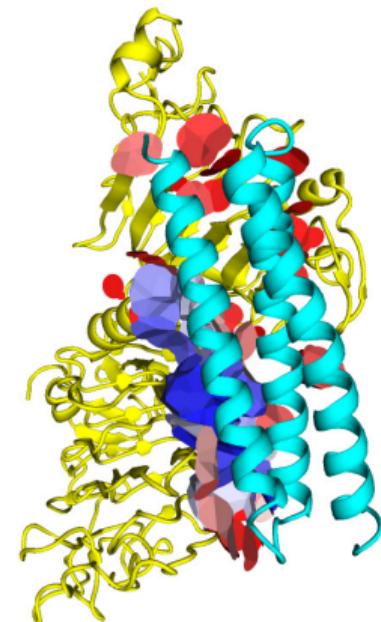
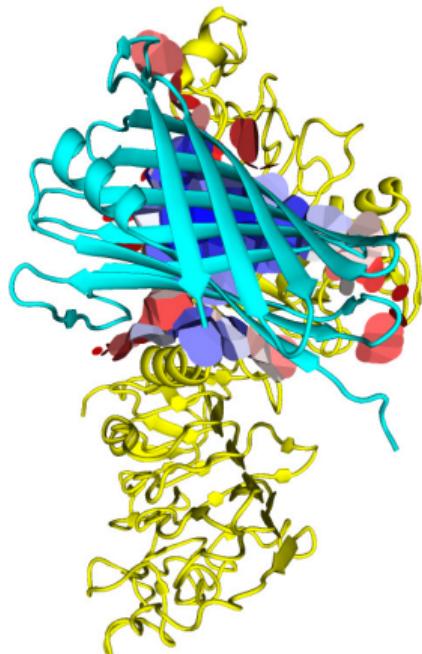
<https://foundry.adaptyvbio.com/competition>

Strong binder (design 1), binder (design 2), non-binder (design 3):



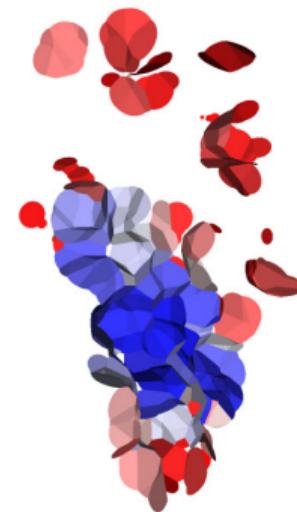
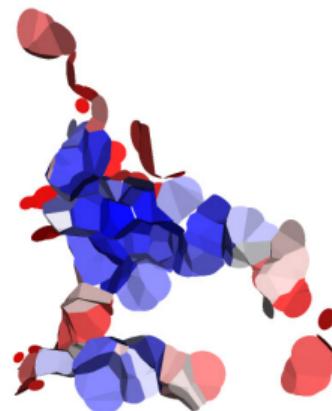
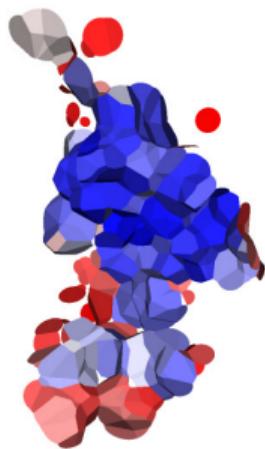
ID	modified	pseudoenergy	area	best_core_pseudoenergy	best_core_area
design1.pdb	no	-1088.83224665428	1311.29595	-1312.60234731125	1165.38884
design2.pdb	no	-483.386123040848	1399.2419	-1047.28512420783	878.30827
design3.pdb	no	-79.5982378233236	1377.05465	-923.84868490561	798.82894

Looking at some examples from EGFR Protein Design Competition



<https://github.com/kliment-olechnovic/voromarmotte-app>

Looking at some examples from EGFR Protein Design Competition



<https://github.com/kliment-olechnovic/voromarmotte-app>

Final remarks (Oasis-like)

Definitely:

- ▶ Tessellation-derived contact area descriptors can be used to collect information about contact stability from ensembles of conformations in PDB, this information can improve scoring of protein-protein interfaces.
- ▶ The Voronoi Contacts Block can be used to predict stability of contacts in an ensemble.

Maybe:

- ▶ The predicted stability of contacts can be used to select better binders.
- ▶ The tessellation-based methodology for scoring interfaces can be used for protein design-related tasks.

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