Computational modeling of protein-protein interactions

Prof. Corey S. O’Hern
Department of Mechanical Engineering, Department of Physics
Graduate Program in Computational Biology & Bioinformatics
Yale University

Jake Sumner,
CBB Ph.D. student,
Yale

Zhuoyi Liu,
ME Ph.D. student,
Yale

Naomi Brandt,
Physics Ph.D.
student, Yale

Devon Finlay,
Physics Ph.D.
student, Yale

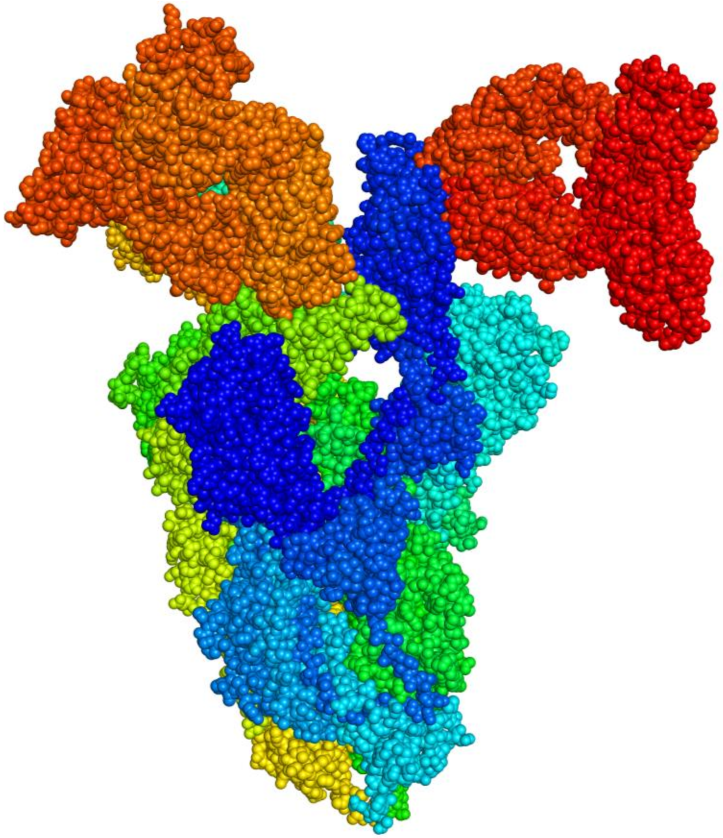
Z. Liu, A. T. Grigas, J. Sumner, E. Knab, C. M. Davis, and C. S. O’Hern, “Identifying the minimal sets of distance restraints for FRET-assisted protein structural modeling,” *Protein Science* (2024).

J. Sumner, G. Meng, N. Brandt, A. T. Grigas, A. Cordoba, M. D. Shattuck, and C. S. O”Hern, “Assessment of scoring functions for computational models of protein-protein interactions,” submitted to Journal of Computational Chemistry (2025).

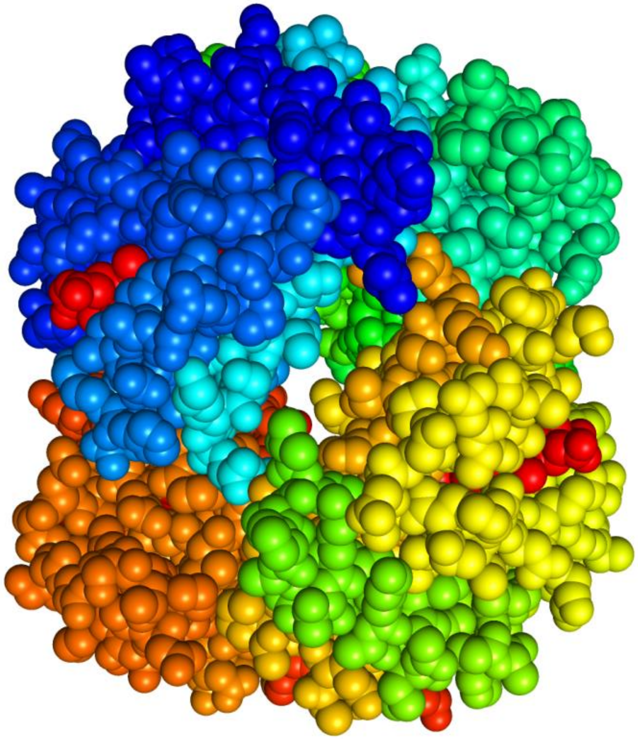
A. T. Grigas, Z. Liu, L. Regan, and C. S. O’Hern, “Corey packing of well-defined xray and NMR protein structures is the same,” Protein Science 31 (2022) e4373.

NIH T32 GM 008384
NIH T32 GM 008283

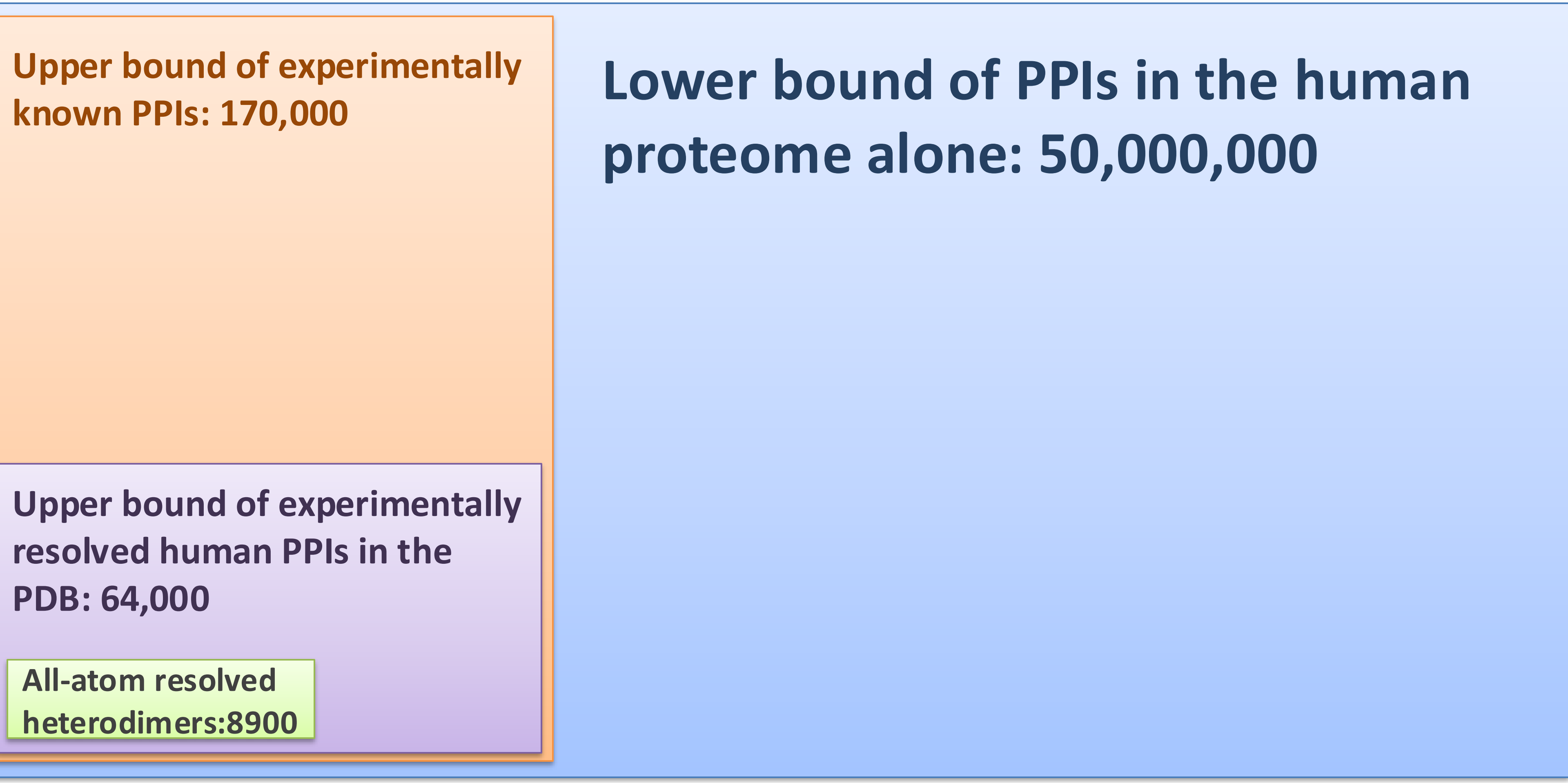
Protein-protein binding/interactions



SARS-CoV-2 spike protein with complexed antibodies



Hemoglobin – complex of 8 proteins – oxygen transporter



Lower bound of PPIs in the human proteome alone: 50,000,000

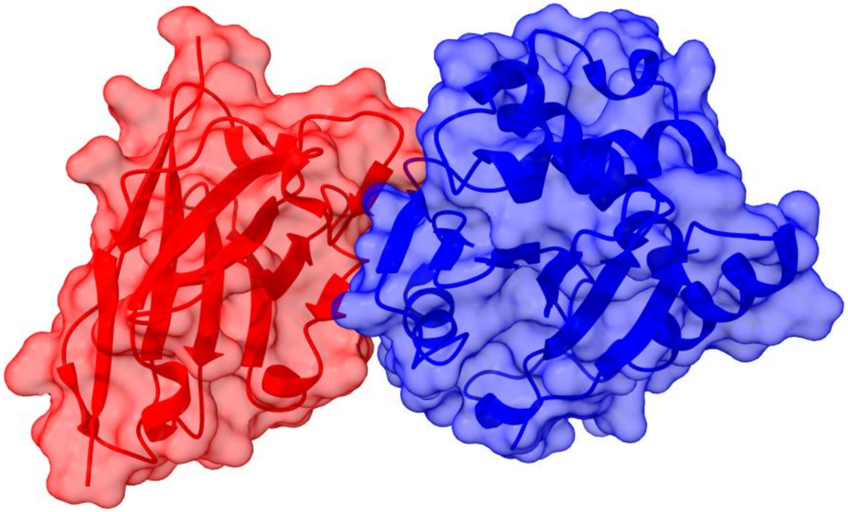
Upper bound of experimentally known PPIs: 170,000

Upper bound of experimentally resolved human PPIs in the PDB: 64,000

All-atom resolved heterodimers:8900

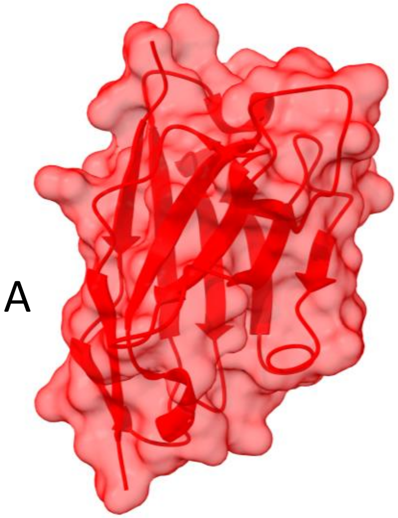
PDBID: 5sv3

heterodimer



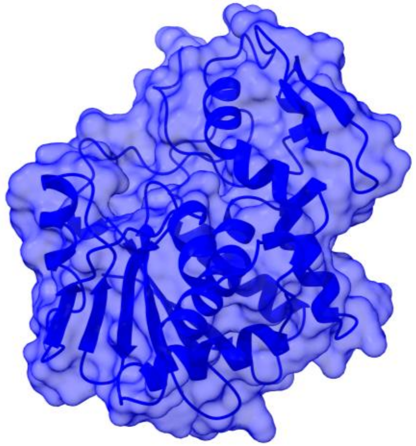
PDBID: 5sv4

Monomer A



PDBID: 1ift

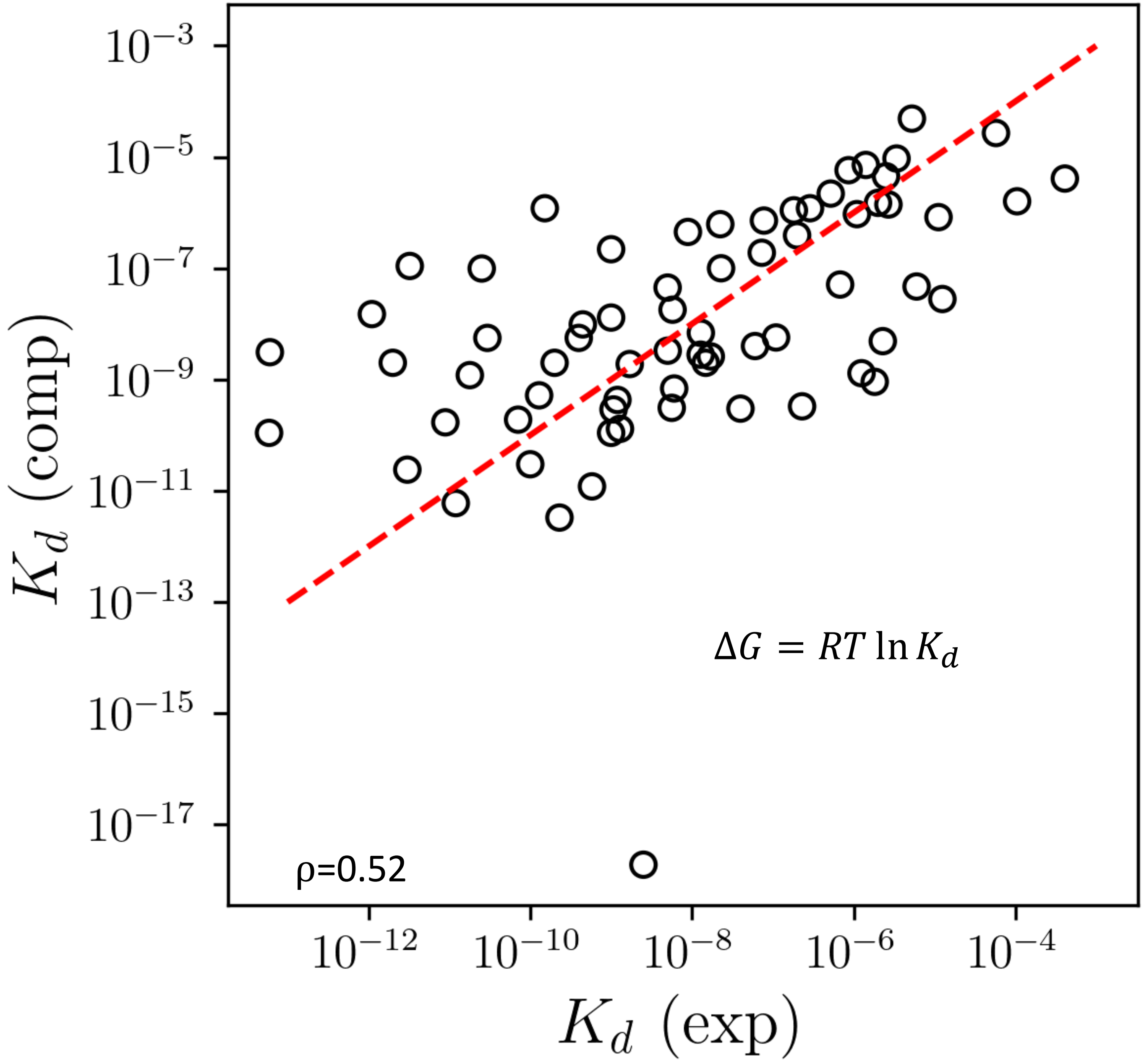
Monomer B



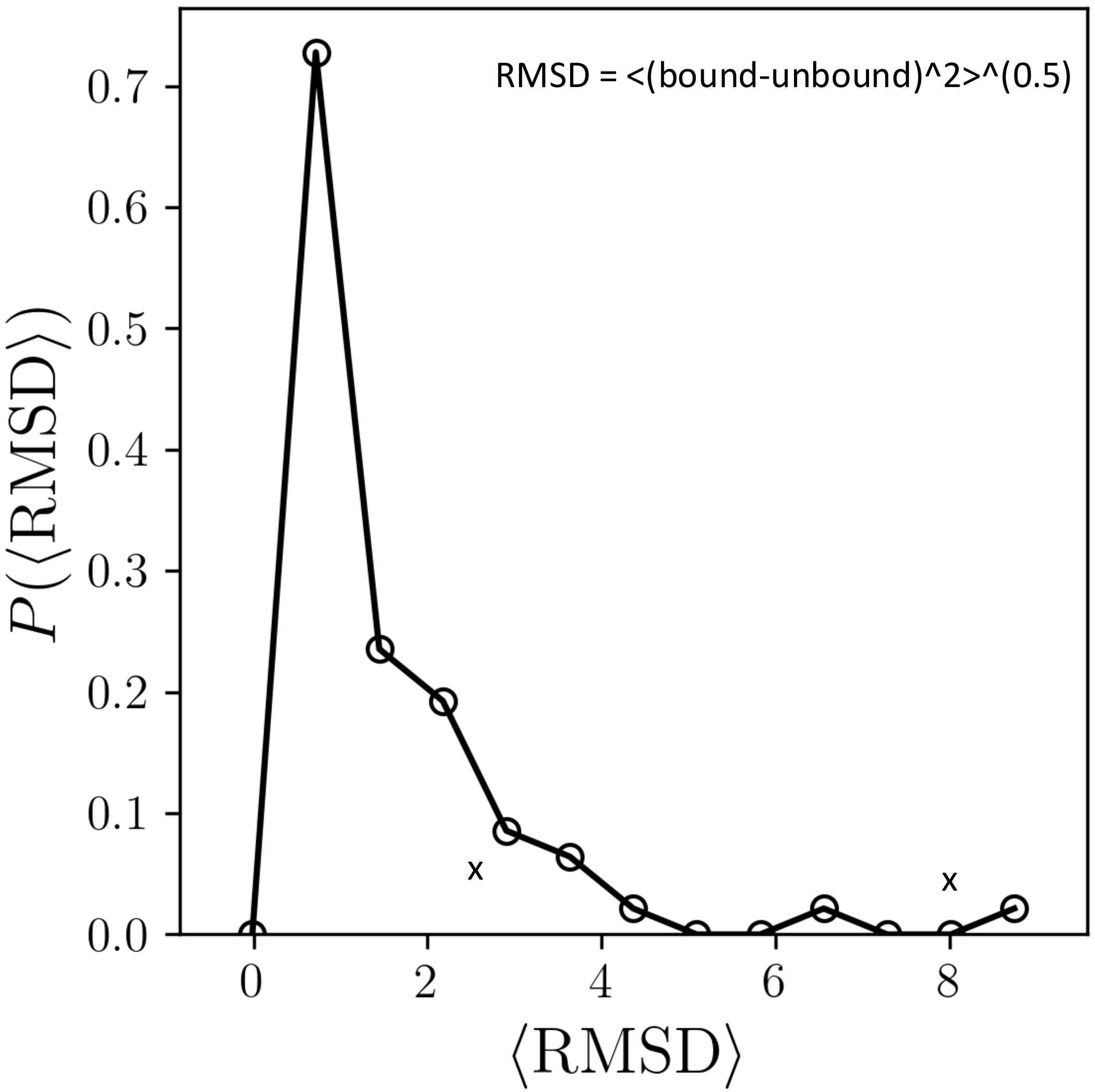
Possible Questions about Heterodimers

- Do proteins A and B bind? How strongly do proteins A and B bind?

- Where do proteins A and B bind and in what conformations?



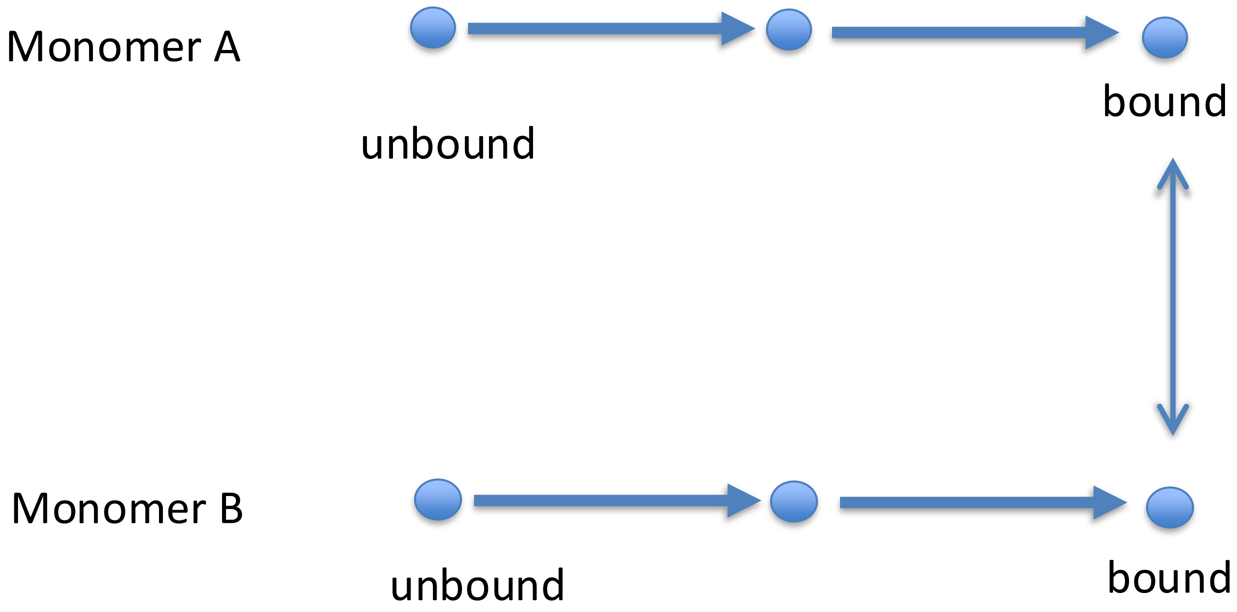
Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. Elife. 2015 Jul 20;4:e07454. doi: 10.7554/eLife.07454.


RMSD = <(bound-unbound)^2>^(0.5)
$P(\langle\mathrm{RMSD}\rangle)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
2
4
6
8
$\langle\mathrm{RMSD}\rangle$
x
x

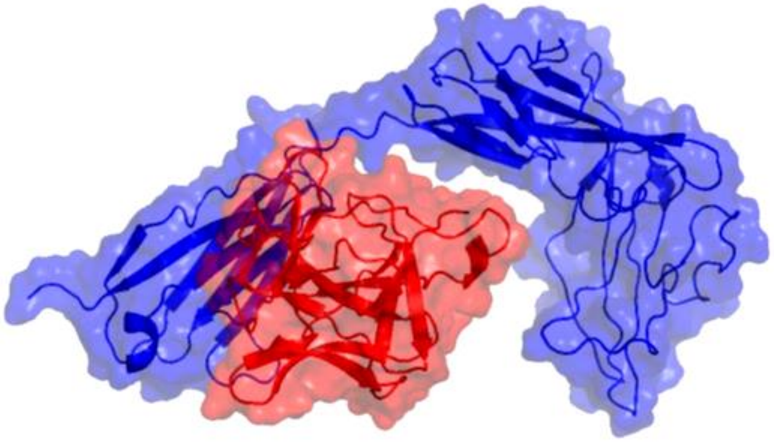
1acb

1ira

Sampling and Scoring

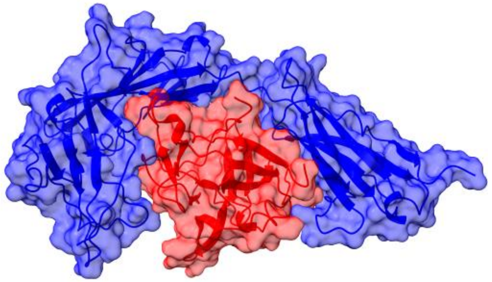


Sampling of Bound Forms

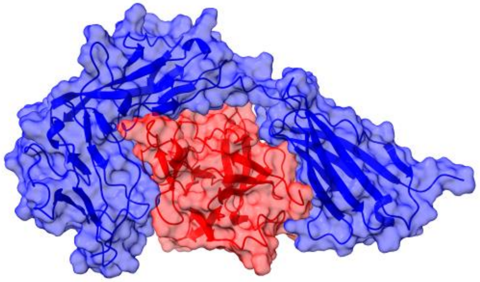


PDB: 1ITB

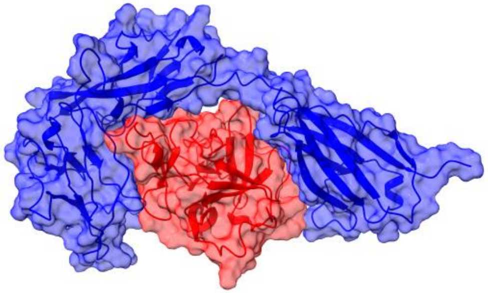
PPI decoy scoring using ground truth



Crystal structure
DockQ: 1.0

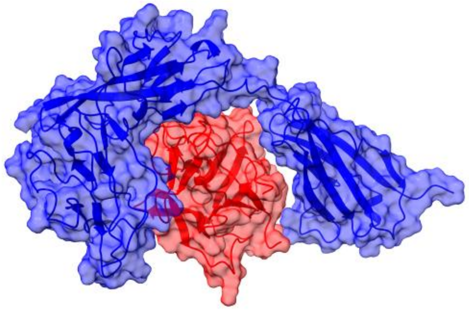


DockQ: 0.847
CAPRI: High

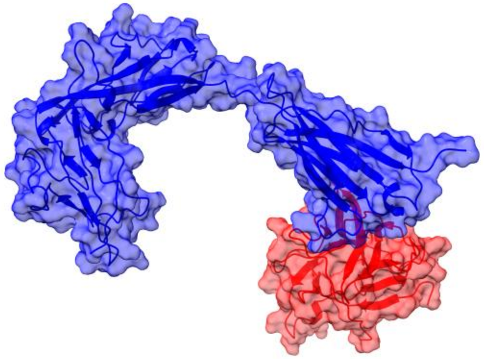


DockQ: 0.506
CAPRI: Medium

shown



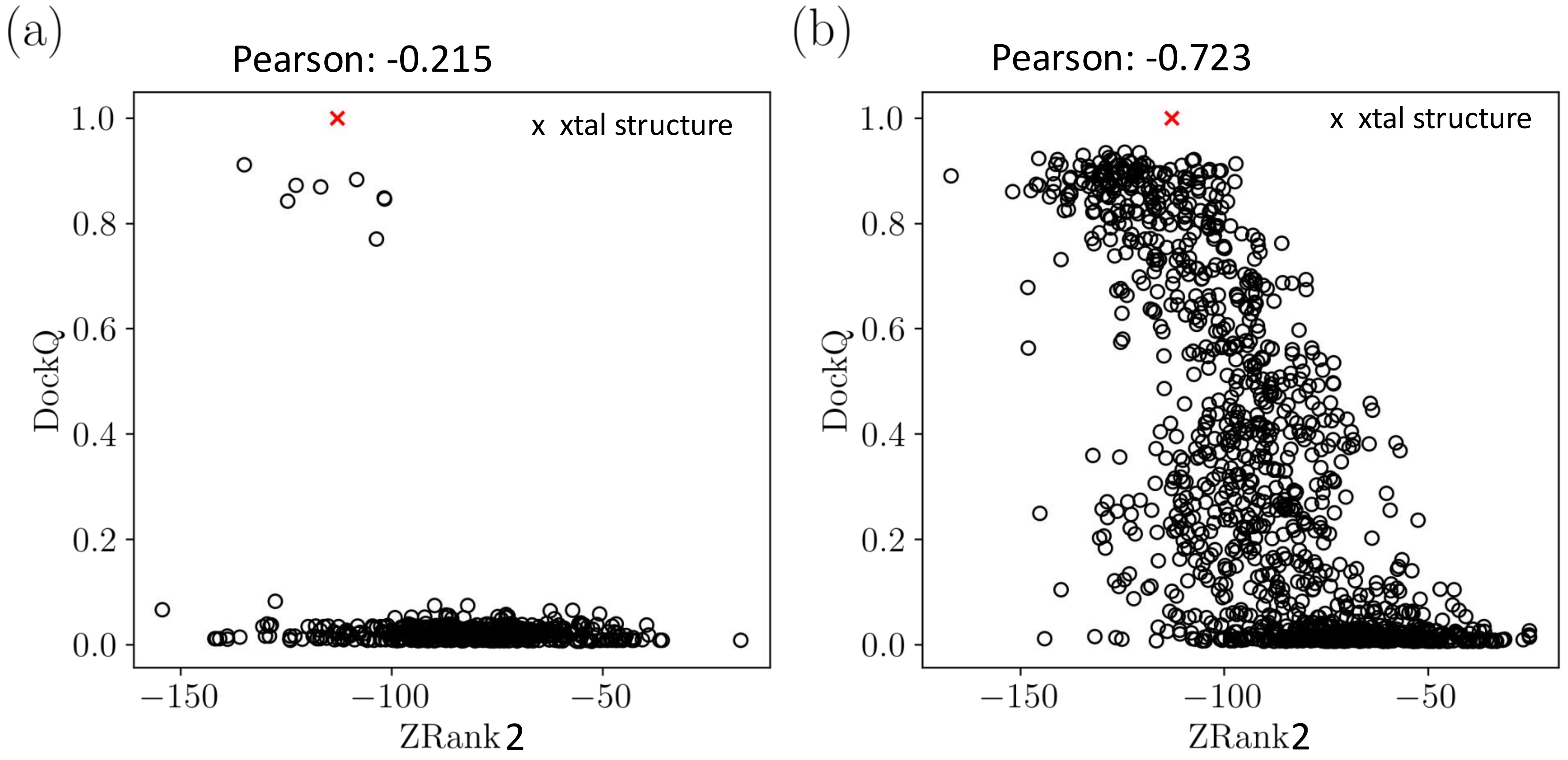
DockQ: 0.286
CAPRI: Acceptable



DockQ: 0.012
CAPRI: Incorrect

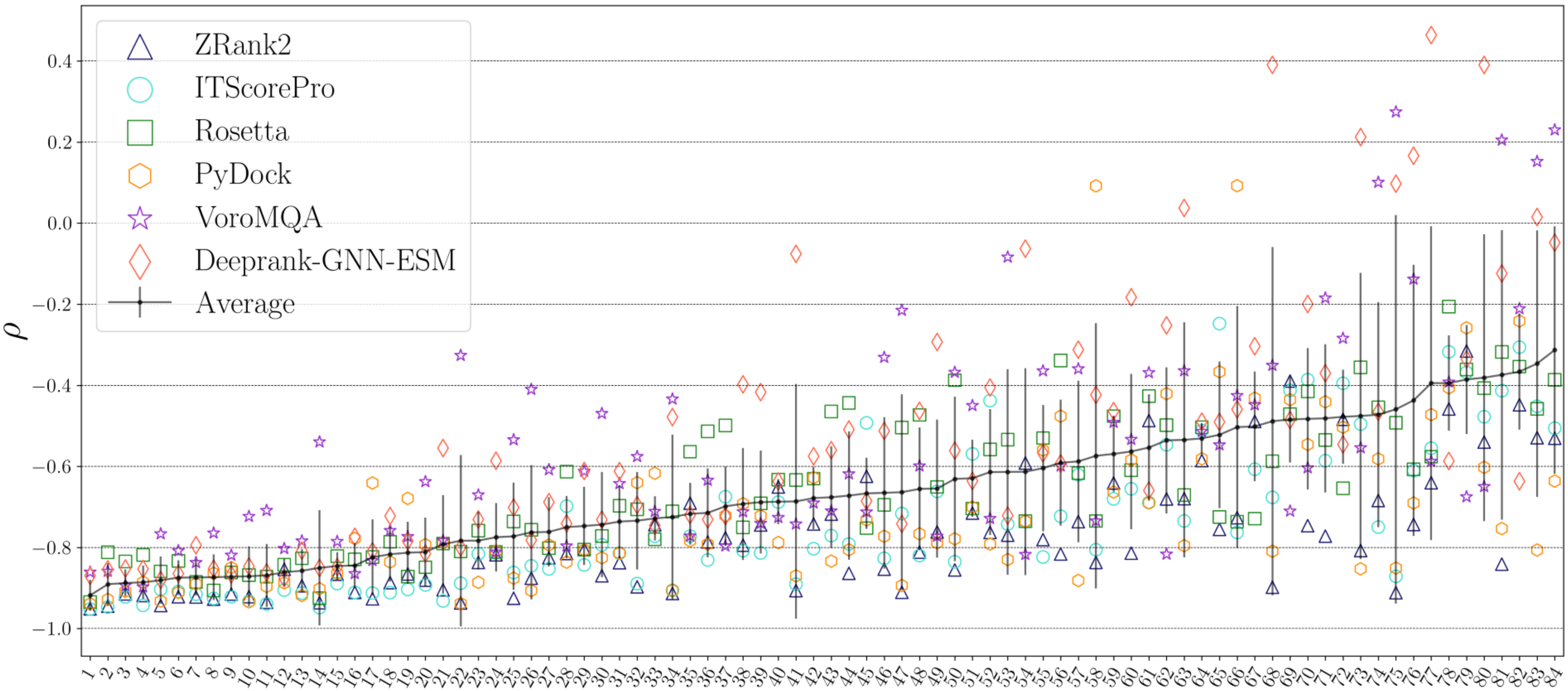
What is the performance of PPI scoring functions on models obtained from bound forms?

Correlation between ground truth and PPI scoring functions



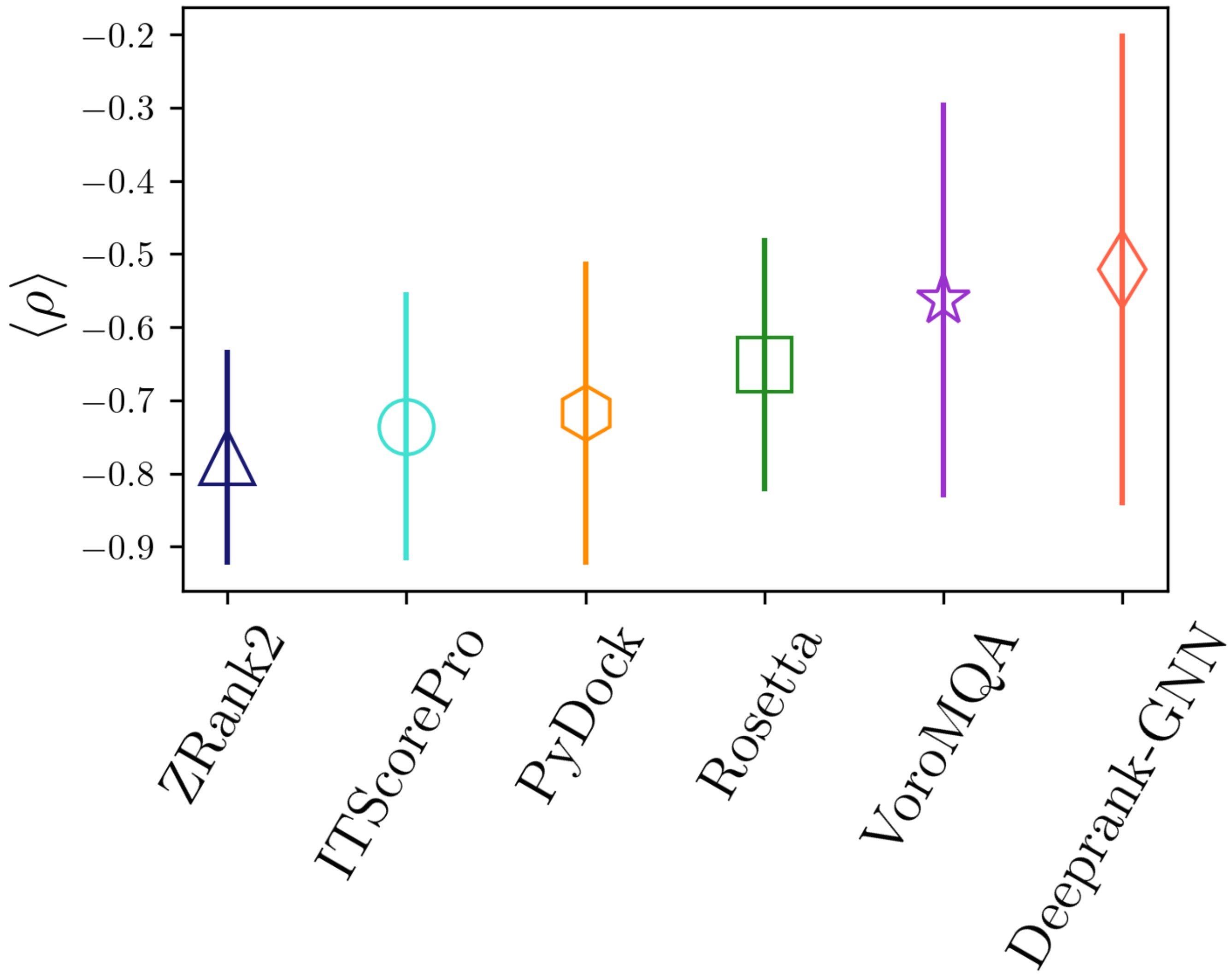
PDBID: 1XG2

Pearson correlation between DockQ and scoring function

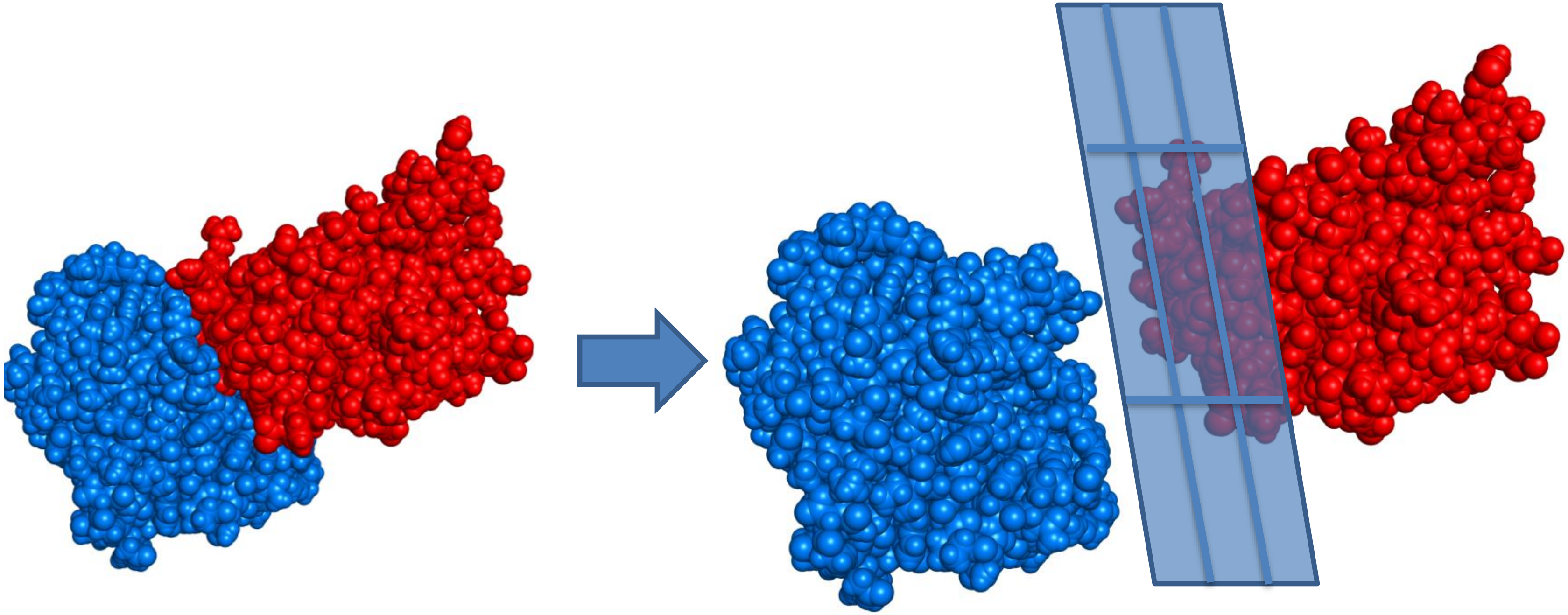


PPI target

Average correlations for state-of-the-art scoring functions



Physical Features

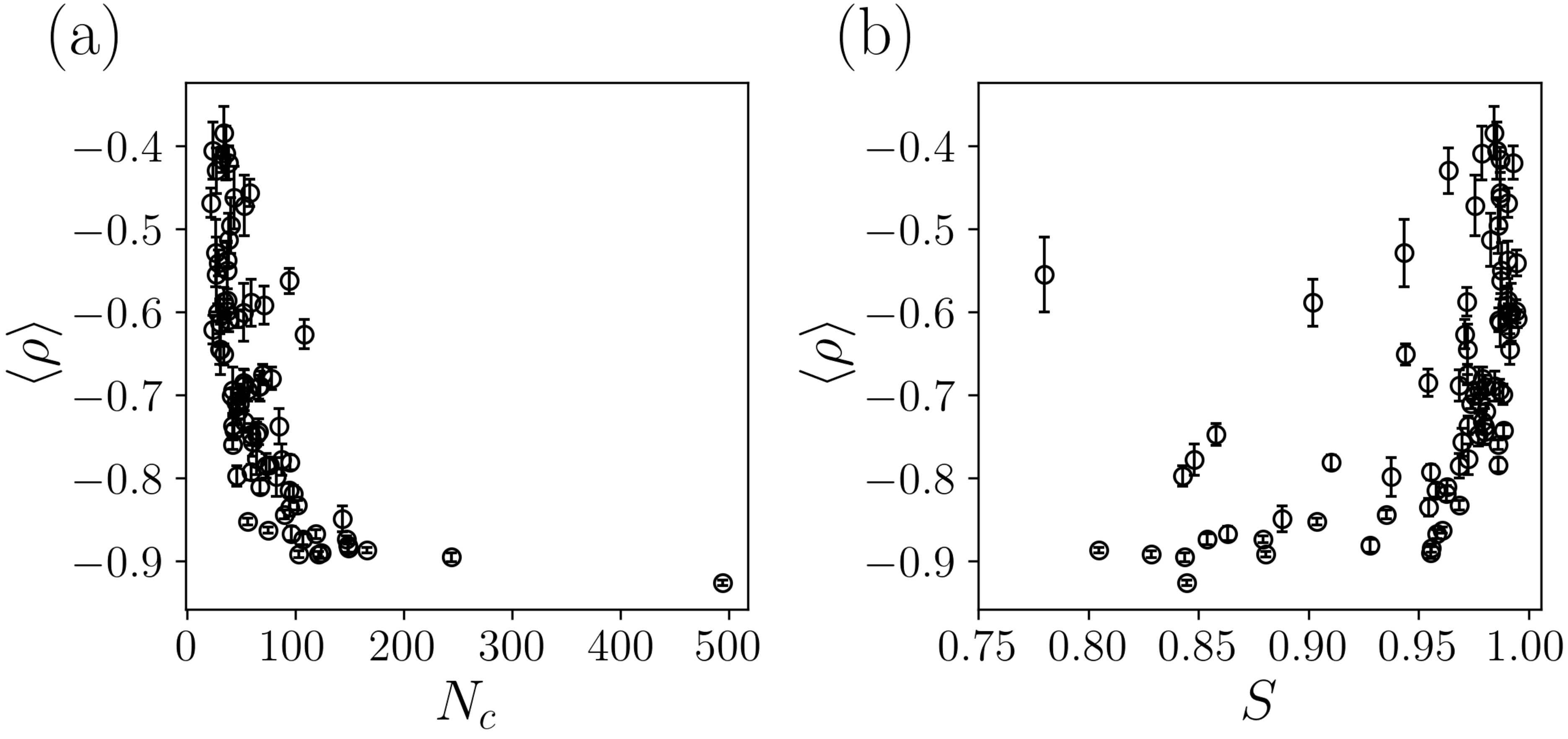


Intertwined score

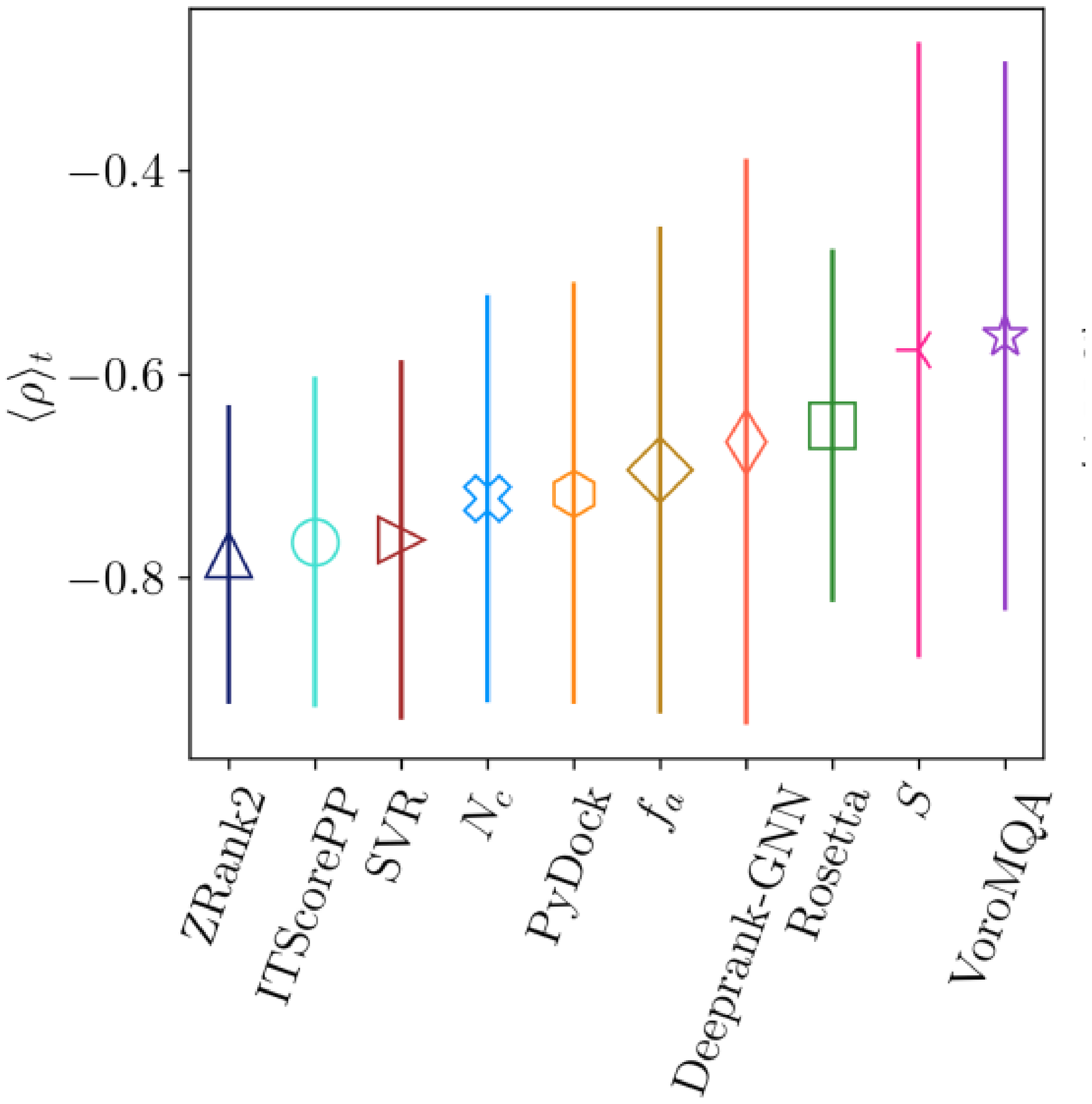
0.5 = intertwined interface

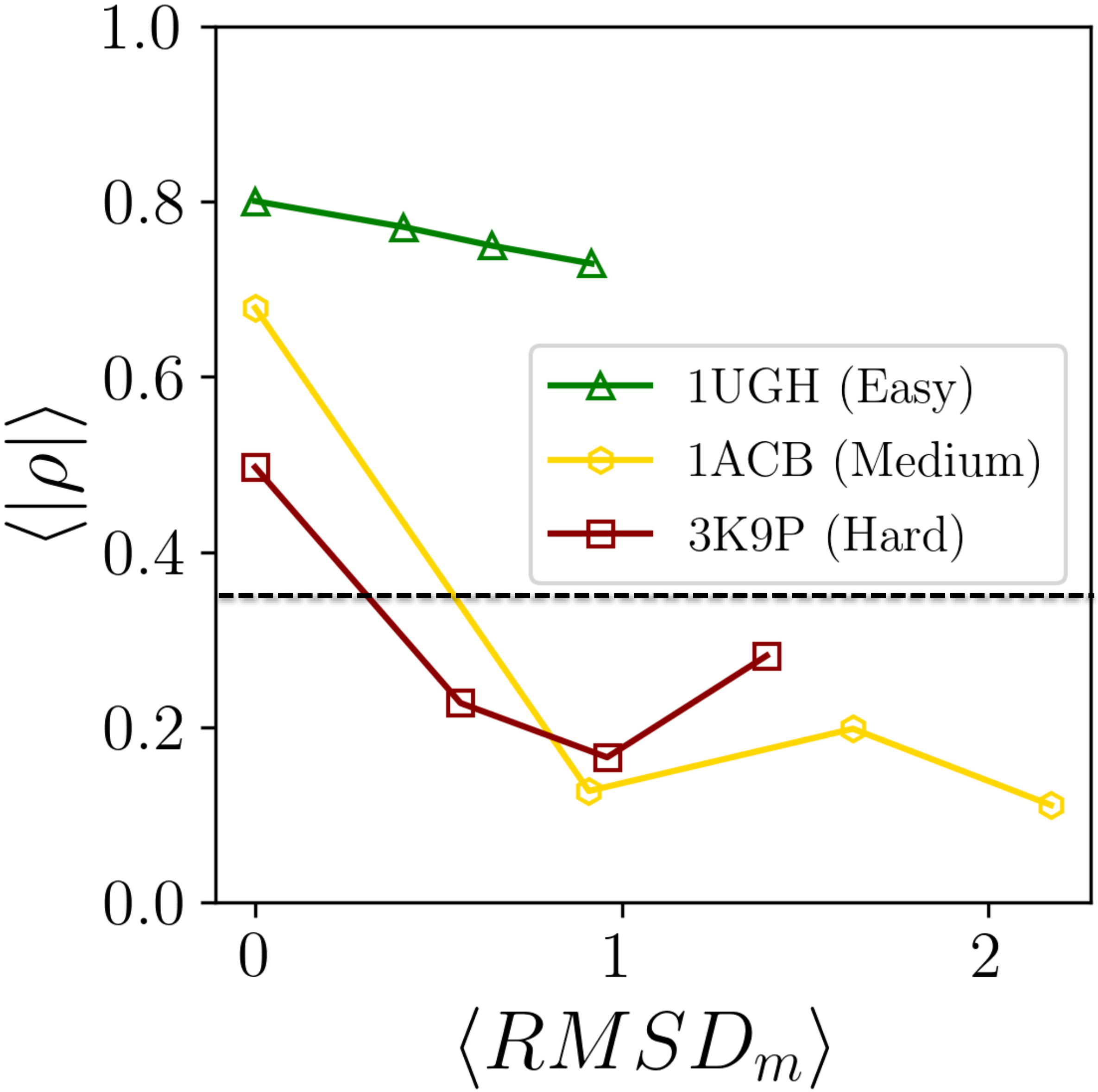
1.0 = completely flat interface

How do we improve PPI scoring functions: Identify important physical features of interface



What progress have we made in 20 years?



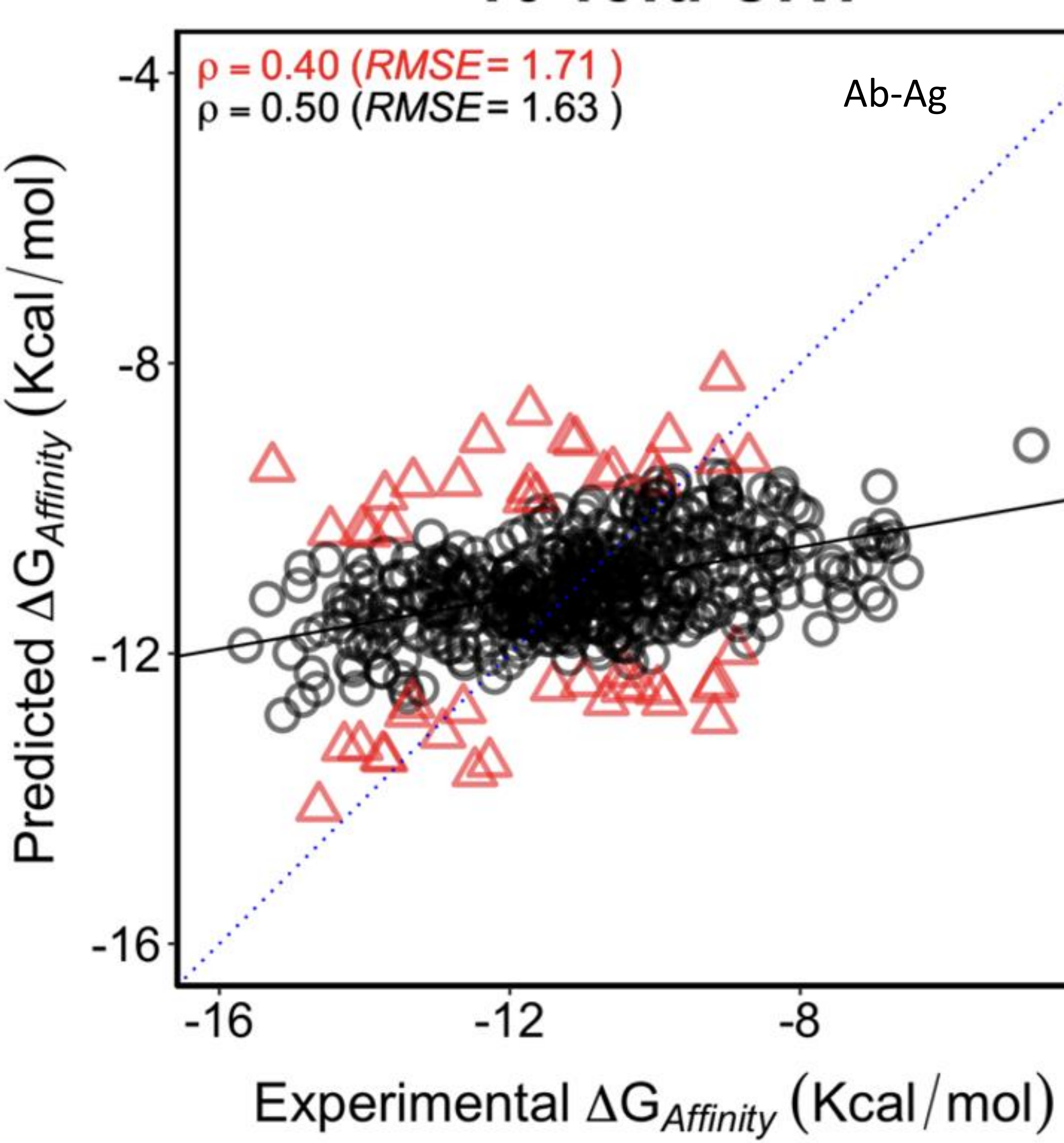

1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
$\langle|\rho|\rangle$
$\langle RMSD_m \rangle$
1UGH (Easy)
1ACB (Medium)
3K9P (Hard)

Conclusions and Future Directions

- PPI decoy scoring for even rigid-body docking of bound forms needs improvement. Previous studies of PPI scoring have mostly focused on identifying a single high-quality model, not correlation between PPI score and ground truth score.

- Promising physical features for scoring functions include flatness of interface and fraction of interface contacts. Develop new scoring function with physical features that achieves $|\rho| \sim 1$ for bound forms of all targets.

- What is the characteristic RMSD beyond which we cannot obtain large $|\rho|$? Develop methods to bring monomer conformations from unbound to bound forms, e..g low-frequency vibrational modes.



Yoochan Myung, Douglas E V Pires, David B Ascher, CSM-AB: graph-based antibody–antigen binding affinity prediction and docking scoring function, *Bioinformatics*, 38 (2022) 1141.