BINDING DECISIONS: PREDICTION OF PROTEIN/PEPTIDE INTERACTIONS

Joseph Crivelli
Meiler Group

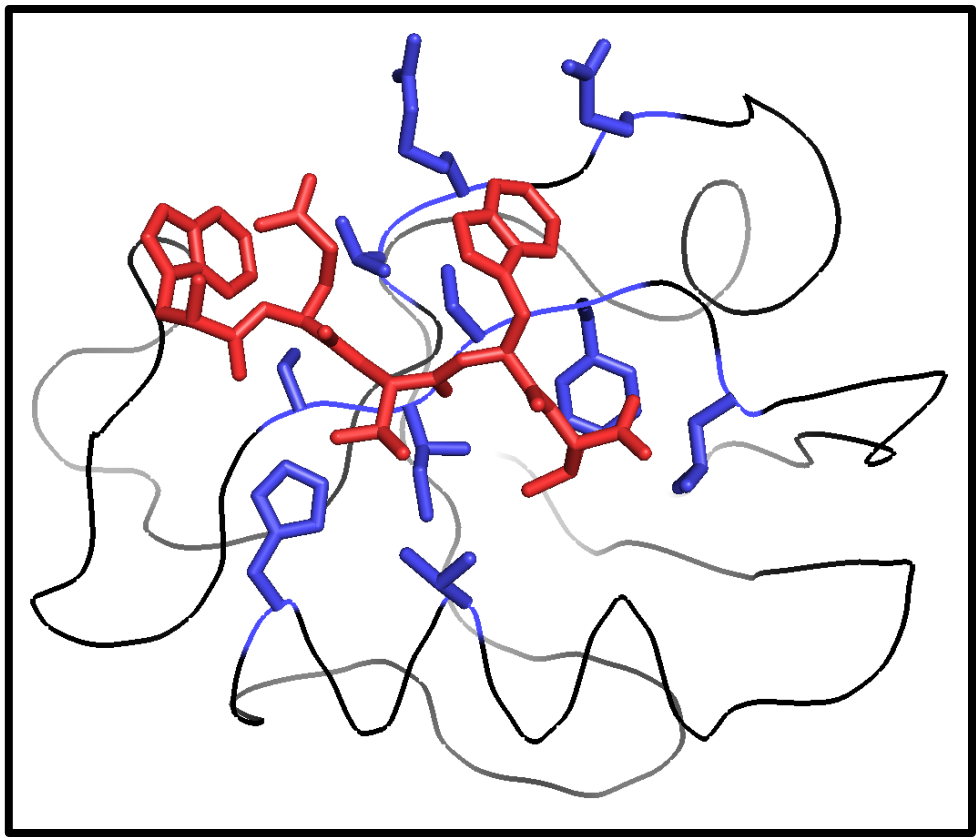
RosettaCon
08/05/2010

Objective

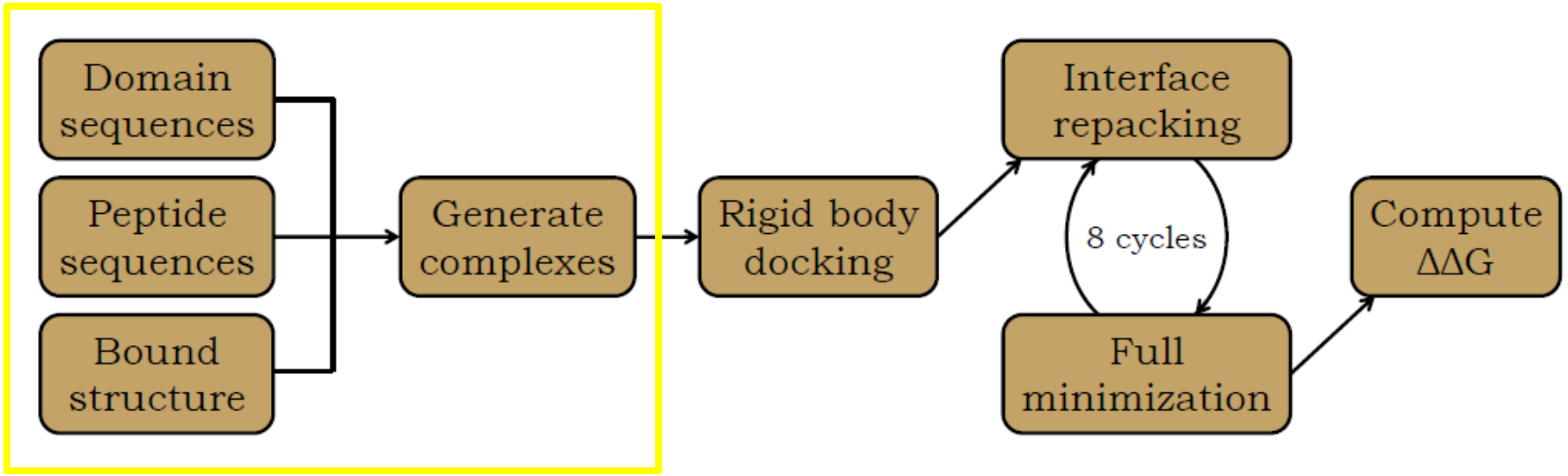
- Develop and test a protocol that accurately predicts
  - binding and non-binding events
  - binding preferences

PDZ single point mutant dataset

- Tonikian et al., 2008 phage display data
- Human Erbin (ERBB2IP-1) PDZ
- 91 mutations at 10 interface positions
- Unique phage derived peptides:
  - 2975 heptapeptides
  - 2156 hexapeptides
  - 934 pentapeptides
  - 328 tetrapeptides

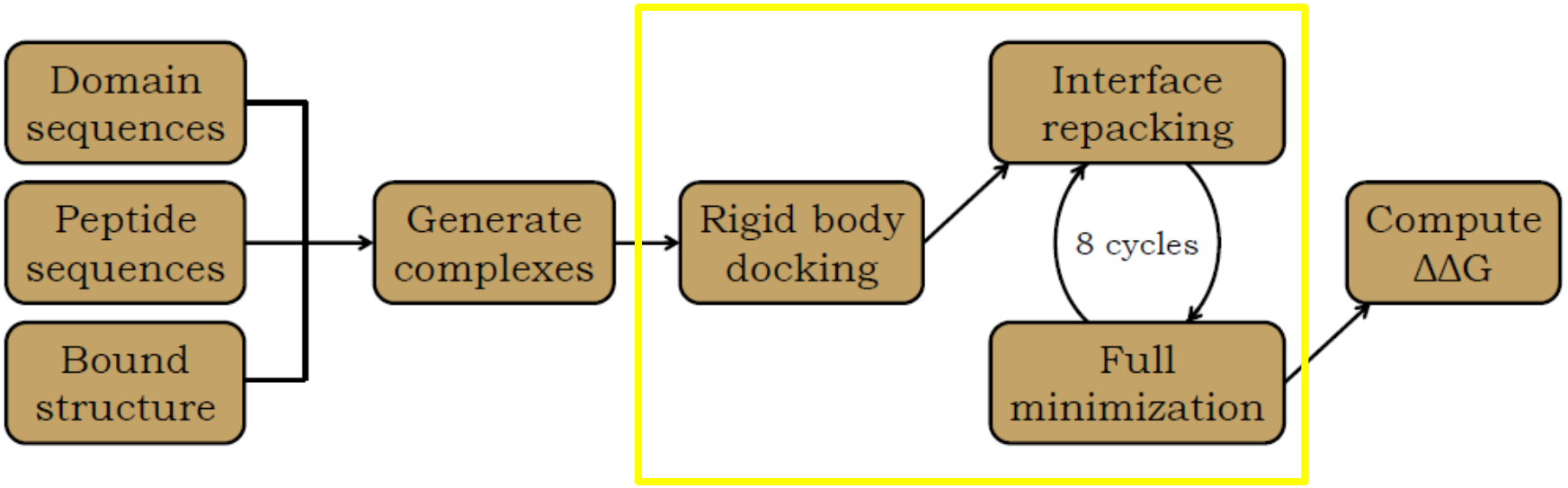


Protocol overview



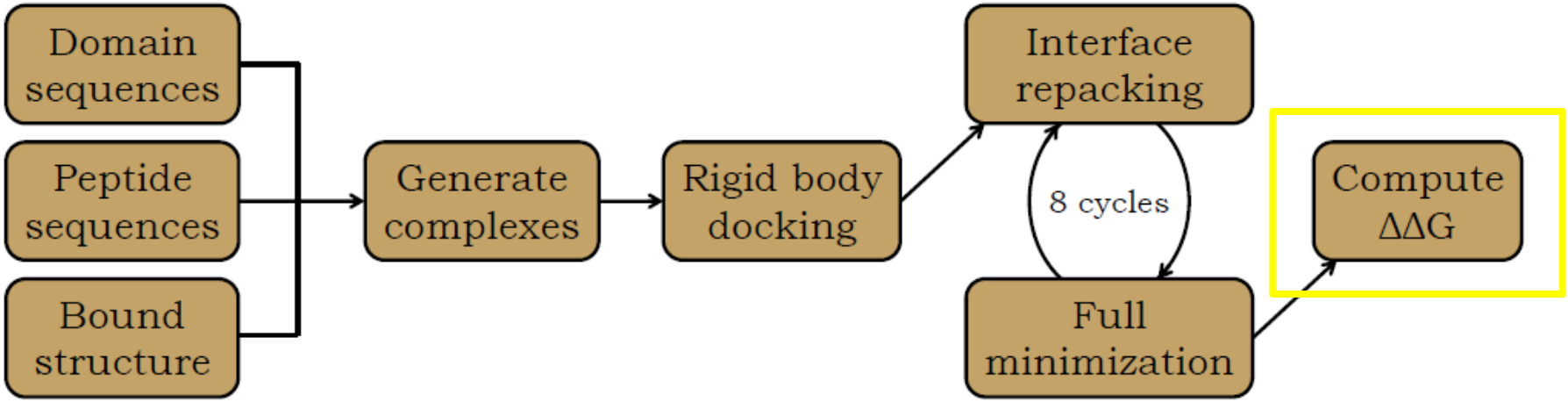
- Generated 29,900 complexes from 92 domain and 325 tetrapeptide sequences
- Threaded sequences onto relaxed Erbin PDZ crystal holo structure template (1.25 Å)

Protocol overview



- Small dock moves (S.D. 0.3 Å, 3.0°)
- Repulsive weight increased from 12.5% to 100% Score12 value during iterative repack/minimize

Protocol overview



- $\Delta\Delta G = \Delta G_{bound} - \Delta G_{unbound}$
- Interface residues repacked in the unbound state before scoring

Optimization of the *ΔΔG* function

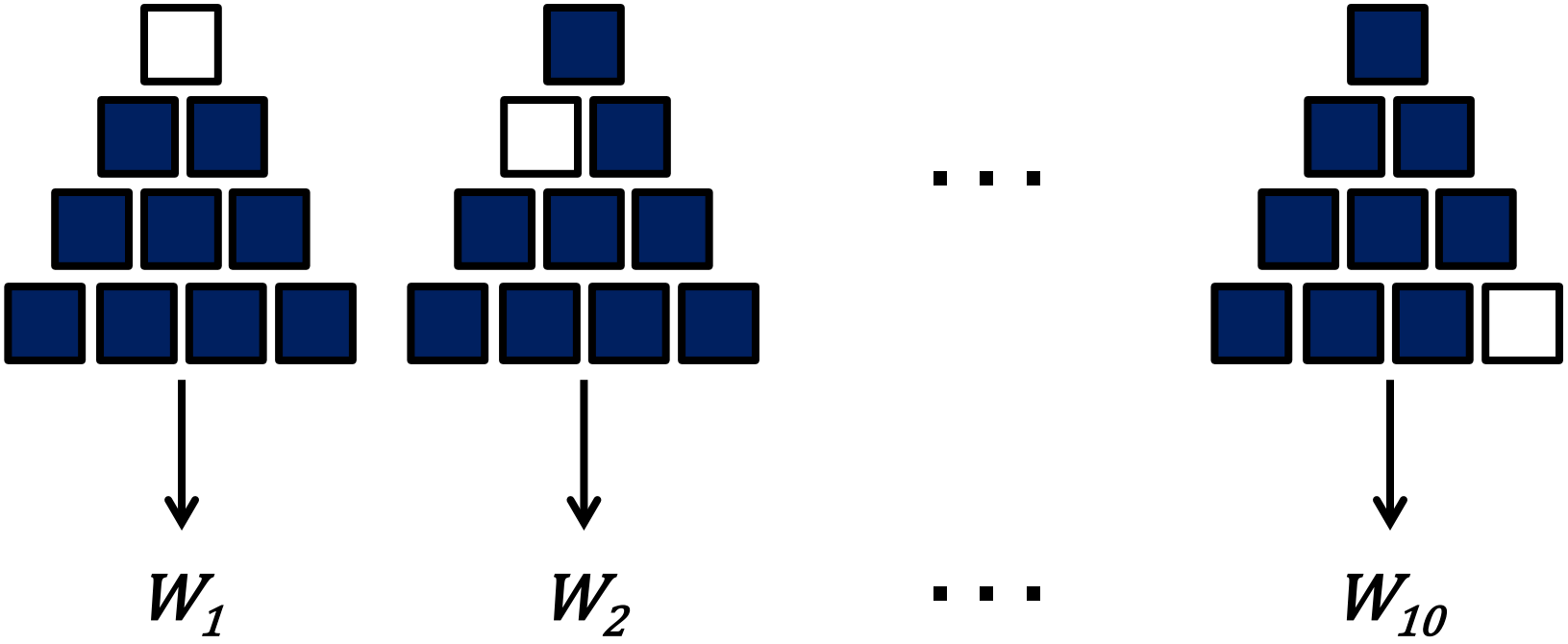
$$\Delta\Delta G_{total} = w_{atr}\Delta\Delta G_{atr} + w_{rep}\Delta\Delta G_{rep} + w_{sol}\Delta\Delta G_{sol} + w_{hbond\_bb\_bb}\Delta\Delta G_{hbond\_bb\_bb} + w_{hbond\_bb\_sc}\Delta\Delta G_{hbond\_bb\_sc} + w_{hbond\_sc\_sc}\Delta\Delta G_{hbond\_sc\_sc}$$

Complex	$\Delta\Delta G_{atr}$	$\Delta\Delta G_{rep}$	$\Delta\Delta G_{total}$	Binds?
A	-60	12	-48	0
B	-60	13	-47	0
C	-44	7	-37	1
D	-42	7	-35	1

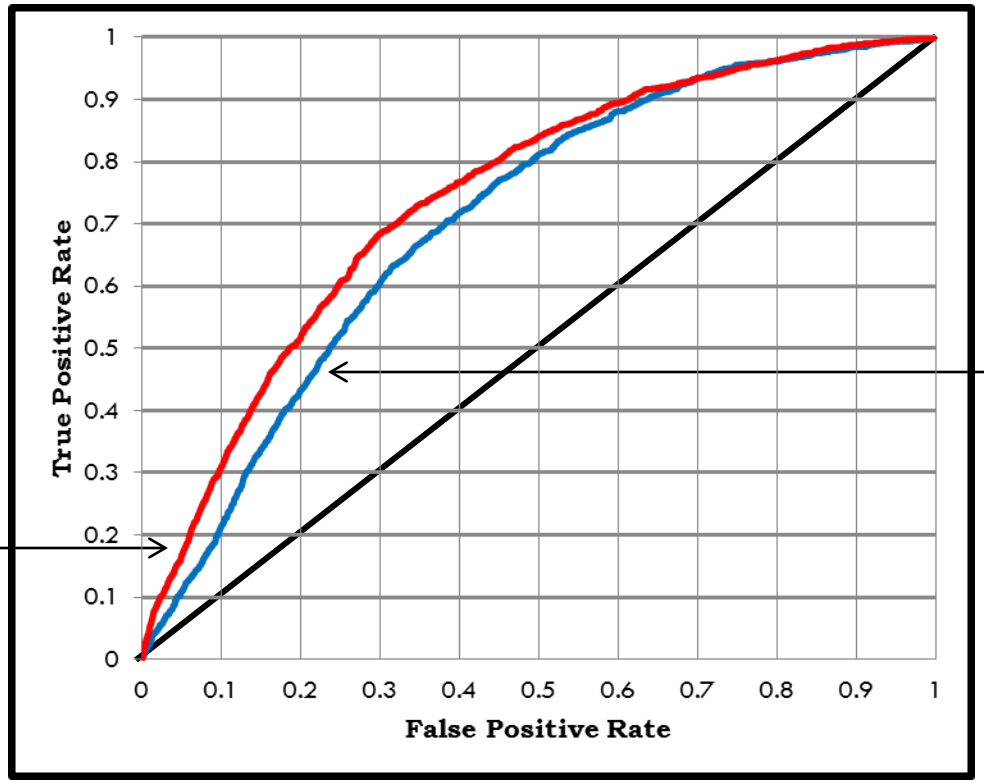
Complex	$\Delta\Delta G_{atr}$	$\Delta\Delta G_{rep}$	$\Delta\Delta G_{total}$	Binds?
C	-22	14	-8	1
D	-21	14	-7	1
A	-30	24	-6	0
B	-30	26	-4	0

10-fold cross-validation analysis

- Partitioned set of 92 domains into 10 subsets
- *ΔΔG* function weights are optimized based on 9 of 10 subsets in a round-robin setup
- Weights are averaged over all 10 steps



Interaction prediction is 24% better than random



PDZ weights:
AUC = 0.744

Score12 weights:
AUC = 0.707

Results

Optimized *ΔΔG* weights are stable

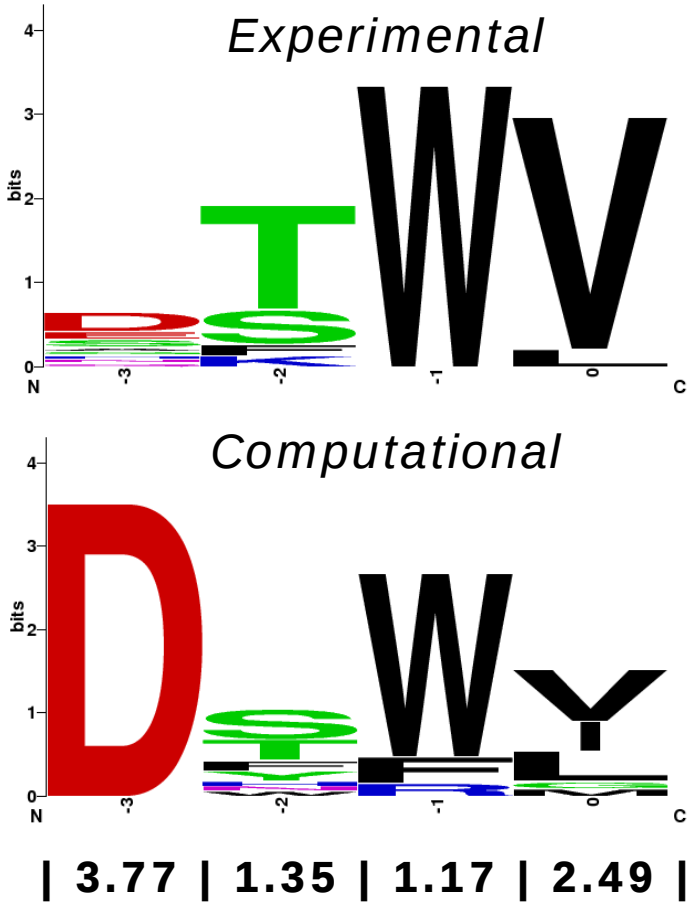
Scoring term	**PDZ weight**	**Score12 weight**
Lennard-Jones:		
attraction	0.228 ± 0.006	0.684
repulsion	0.125 ± 0.004	0.376
Solvation	0.294 ± 0.011	0.556
Hydrogen bonding:		
backbone-backbone	0.559 ± 0.034	1.000
backbone-side chain	0.258 ± 0.038	1.000
side chain-side chain	1.000 ± 0.031	0.940

*All weights are normalized

Comparing binding profiles

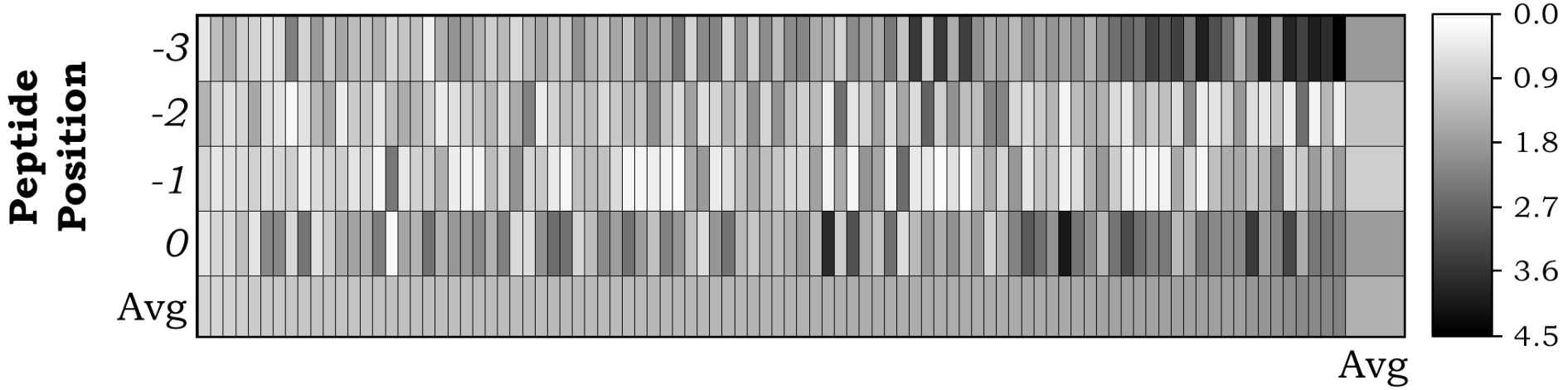
At peptide position k, given the number of bits of information b_k and a vector of amino acid frequencies $\boldsymbol{v_k}$, the distance between an experimental and a computational profile is

$$D_k := \left\| b_k^{exp} \boldsymbol{v_k^{exp}} - b_k^{comp} \boldsymbol{v_k^{comp}} \right\|_2.$$

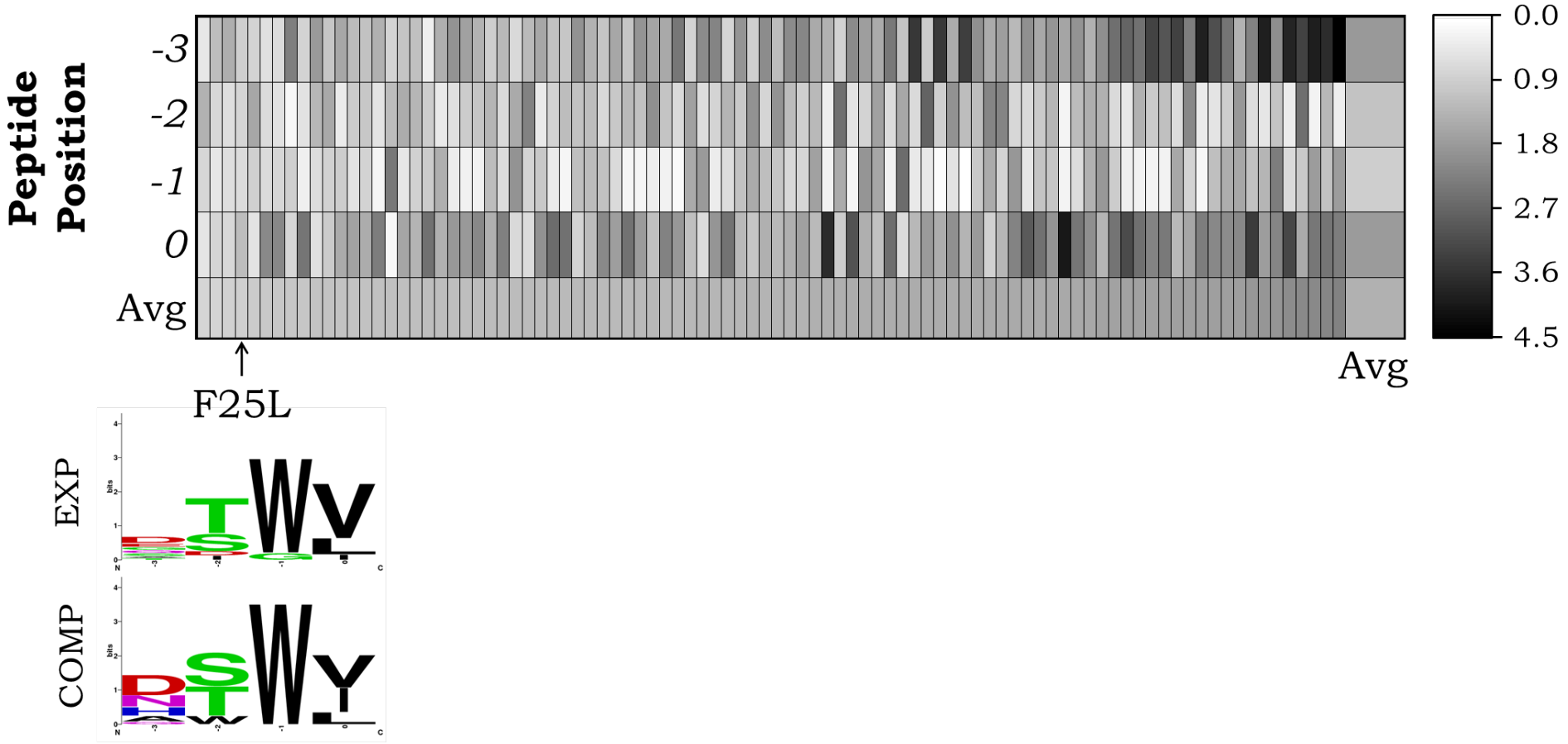


Results

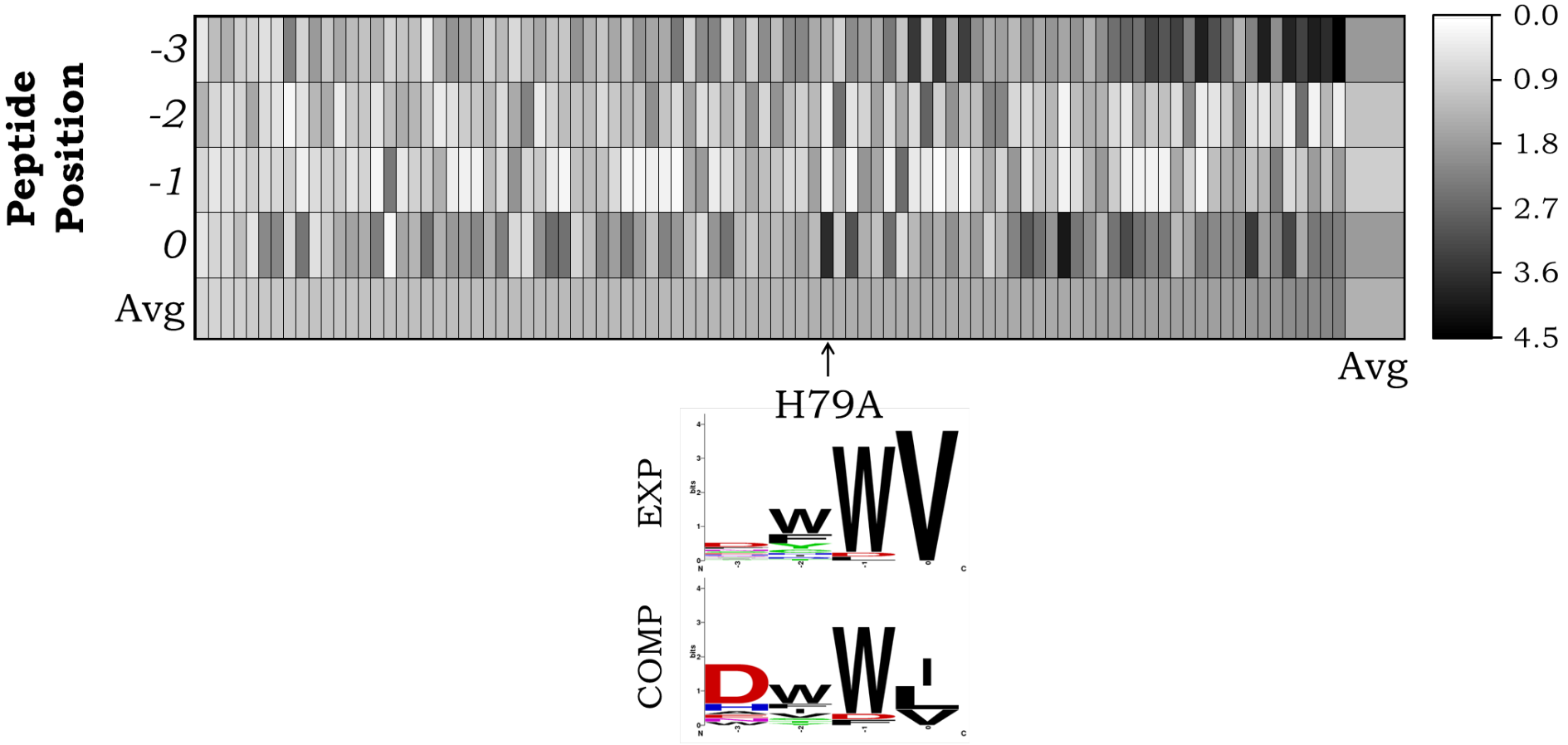
-1 & -2 positions are most accurately predicted



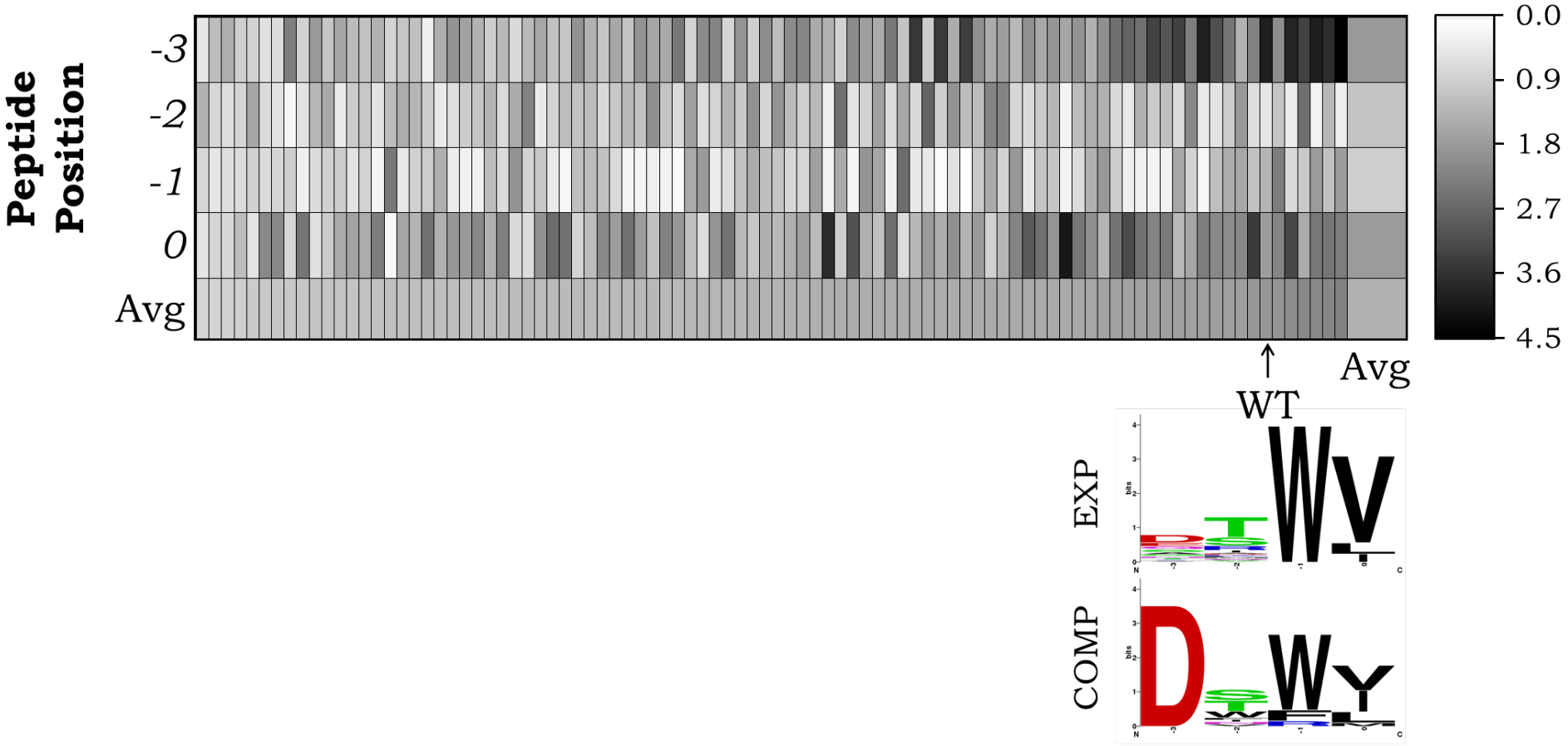
Best case scenario



Hydrophobes at 0 are often incorrectly predicted

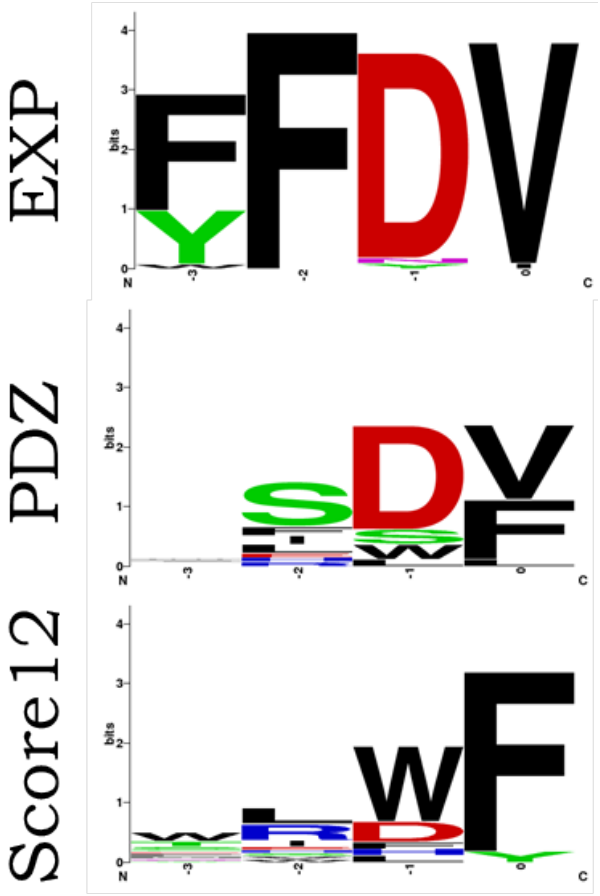


For WT, promiscuity at -3 unclear



Independent test on a homologous domain

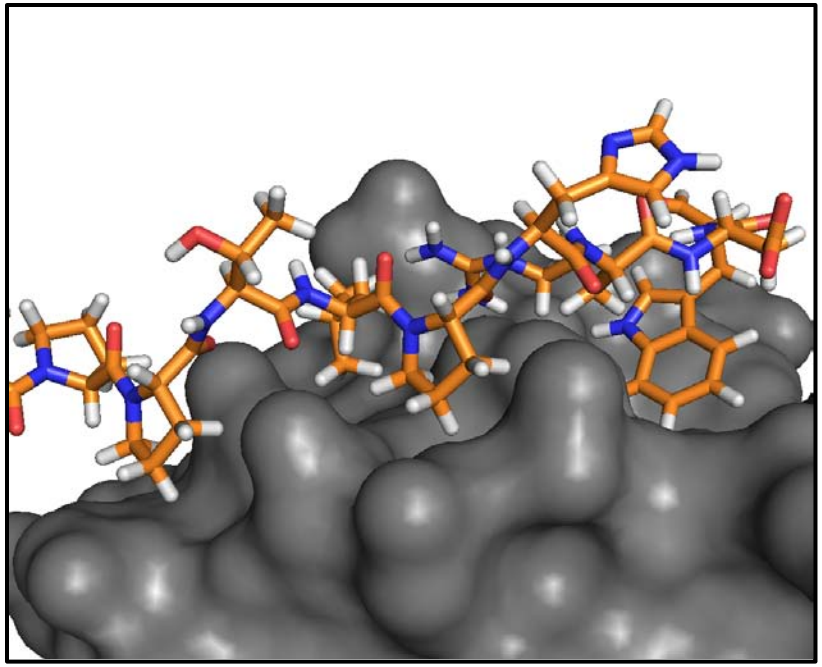
- Tonikian et al. (2008) phage display experiment on 54 wild type human PDZ domains
- Computational test case: CASK-1 PDZ
- Performed the interaction prediction protocol for 760 CASK PDZ protein/peptide complexes



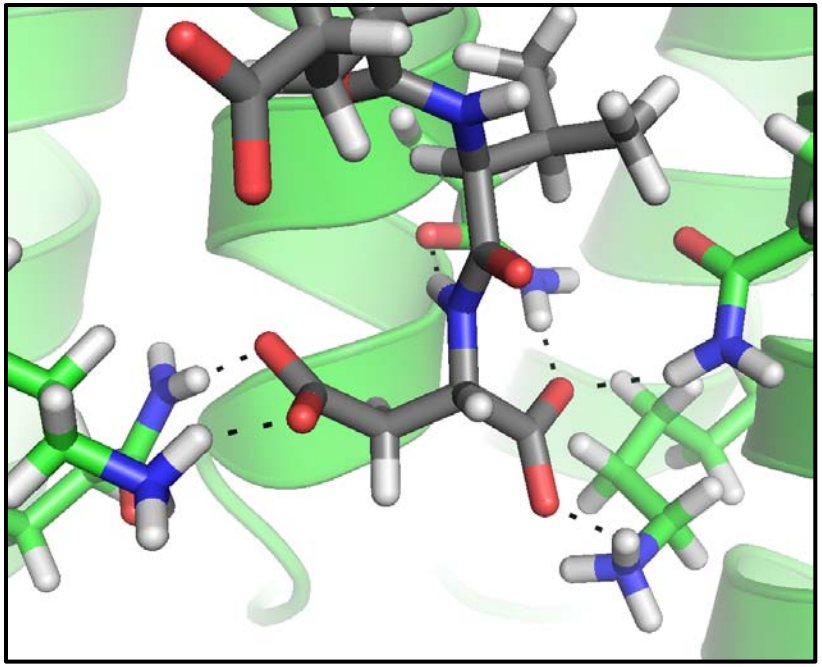
Expand & fine-tune the PDZ protocol

- Will repeat all experiments for 5, 6 and 7 amino acid peptide ligands
- Include peptide backbone flexibility while docking (FlexPepDock)
- Measure the impact of individual ROSETTA refinement steps (docking, repacking, minimization)
- Find the correlation between experimental and computational $\Delta\Delta G$ values

Other systems with canonical binding modes



SH3-domain/PPII-helix



TPR-domain/peptide

Acknowledgements







Arnold and Mabel BECKMAN FOUNDATION