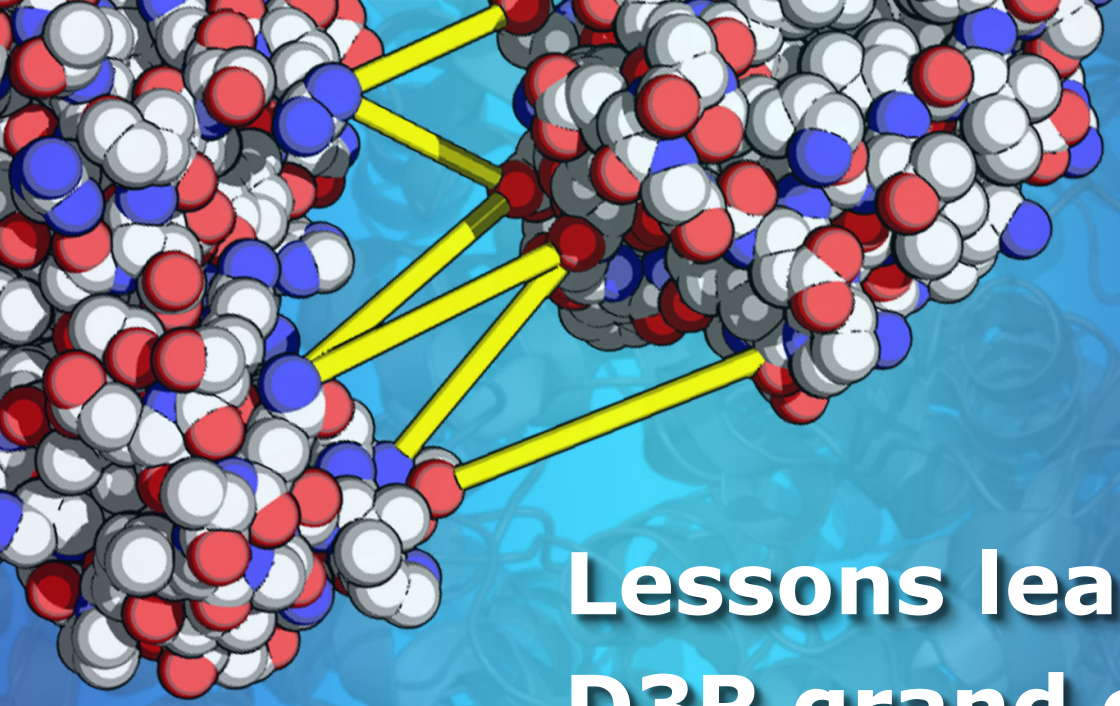
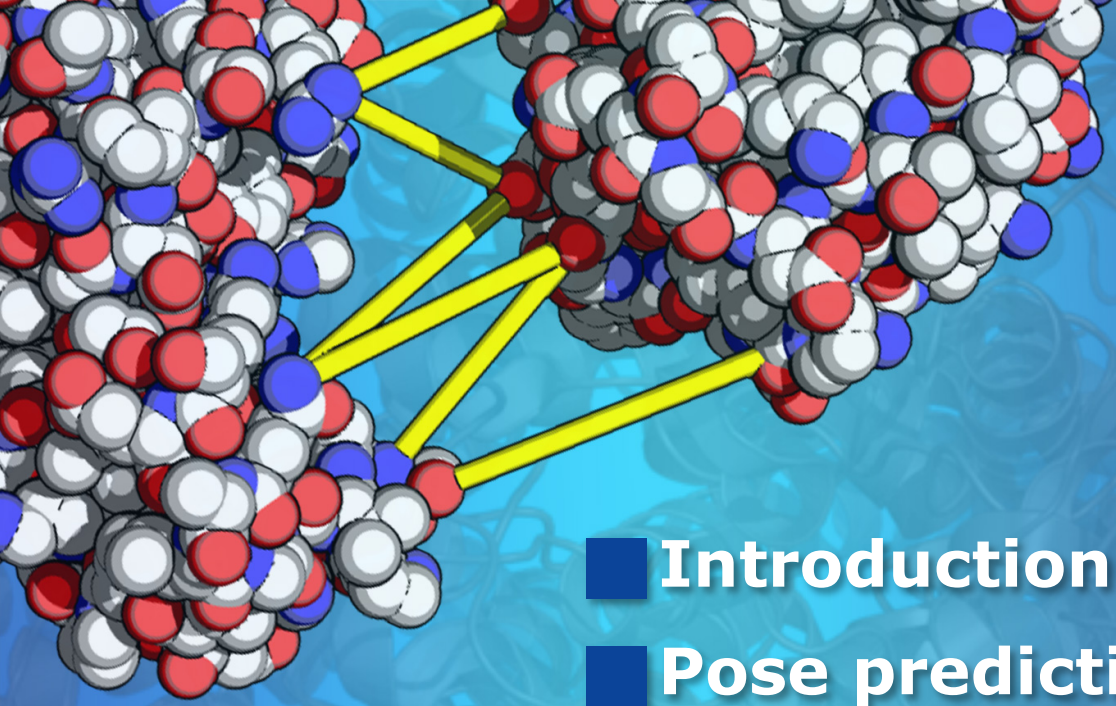


Lessons learned from two D3R grand challenges rounds

Prof. Alexandre M.J.J. Bonvin
Bijvoet Center for Biomolecular Research
Faculty of Science, Utrecht University
the Netherlands
a.m.j.j.bonvin@uu.nl

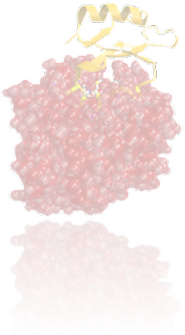
Utrecht Bioinformatics Center



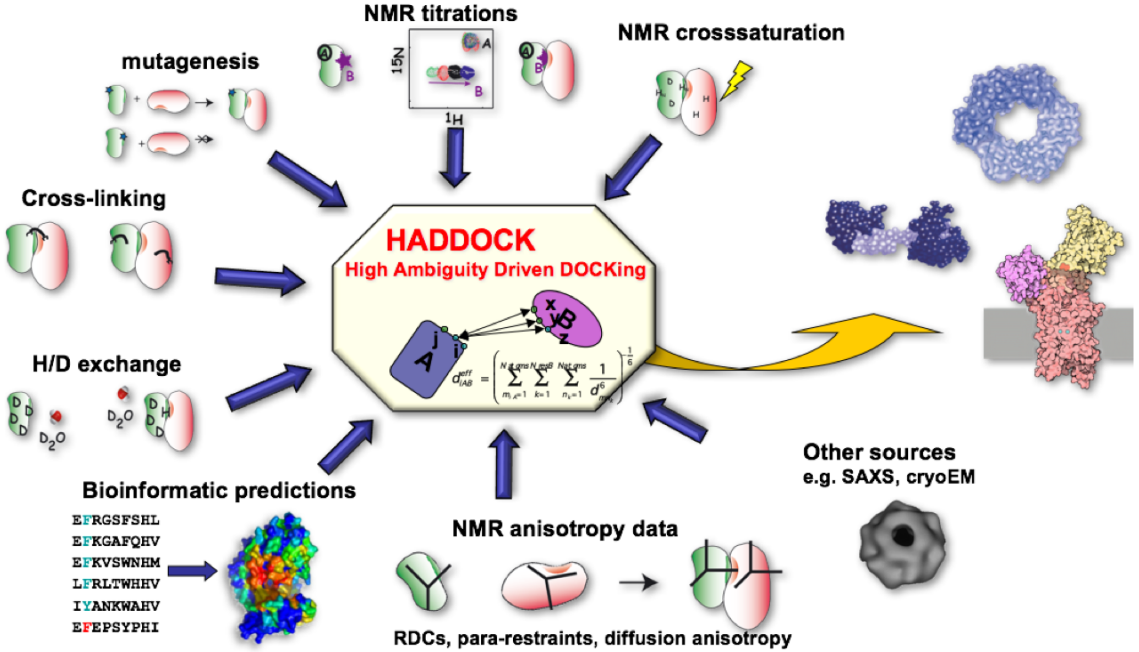
Overview

- Introduction
- Pose predictions
  - Lessons from GC2
  - Applying them to GC3
- Binding affinity/ranking
- Conclusions

HADDOCK:
An integrative modeling platform

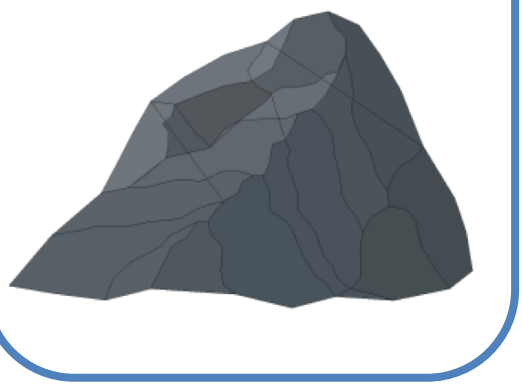


- Incorporates ambiguous and low-resolution data to aid the docking
- Capable of docking up to 6 molecules
- Symmetries can be leveraged
- Powerful algorithms to handle flexibility at the interface
- Final flexible refinement in explicit solvent
- One of the best performing software in CAPRI



Dominguez, Boelens & Bonvin. *JACS* 125, 173 (2003).

HADDOCK docking protocol

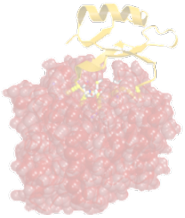


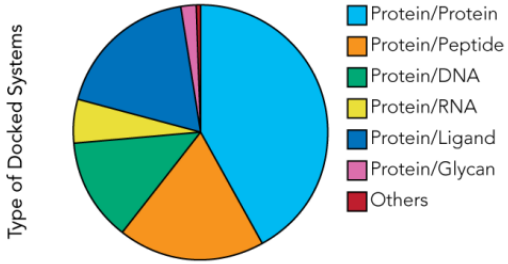
it0 **it1** **itw**

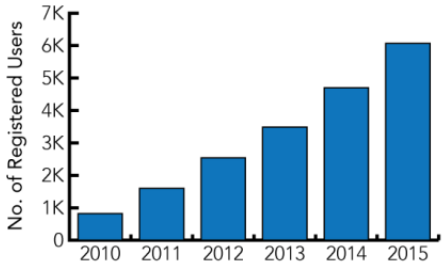
Succession of energy minimization and molecular dynamics protocols

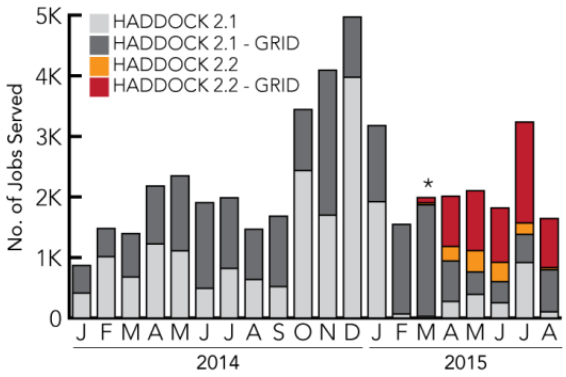
reminiscent of NMR structure calculations

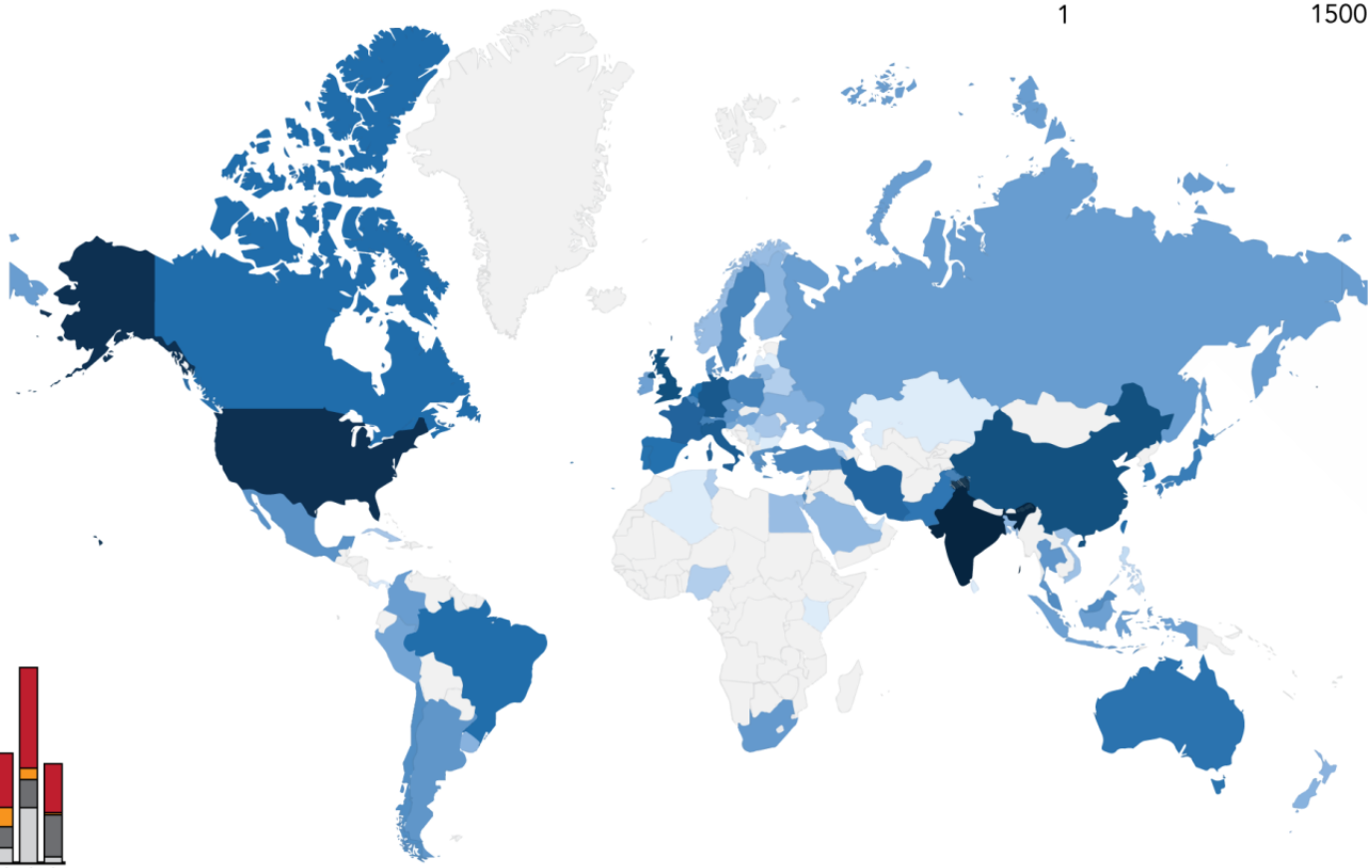
HADDOCK's user base



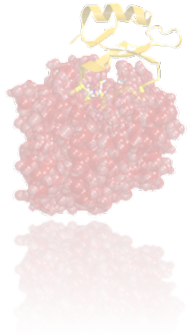






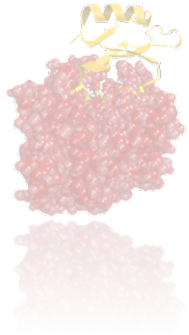


HADDOCK protein-ligand protocol



- Receptor binding site:
  - active during it0 to attract the ligand into the pocket.
  - passive for flexible stages to allow exploration by the ligand.
- Parameters:
  - **Protein: OPLS** (Jorgensen and Tirado-Rives 1988)
  - **Ligand: PRODRG** (Schüttelkopf and van Aalten 2004)
- No high-T search in semi-flexible refinement.
- Modified scoring: w_vdw_it0 = 1, w_elec_water = 0.1
- Pooled results from two runs in case of buried binding site:
  - Standard with ligand settings (as above)
  - Buried-site settings (k_inter=0.001, w_vdw_it0 = 0)

Unbound receptor docking results

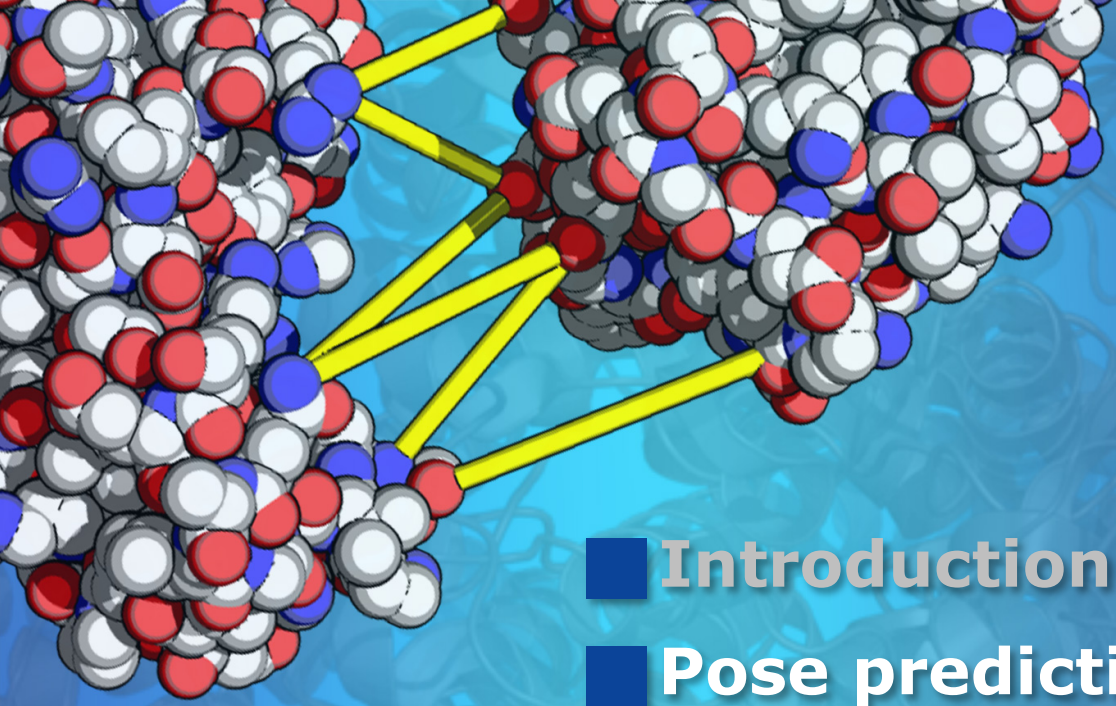


71 cases from the Astex NonNative Set [1]

10Å interface	Cluster quality water – combined analysis			
	Top 1	Top 2	Top 3	Top 5
#total	71	71	71	71
#success *	41	49	55	58
#success **	23	26	28	29
#success ***	2	2	2	2
%success *	**57,7**	**69,0**	**77,5**	**81,7**
%success **	32,4	36,6	39,4	40,8
%success ***	2,8	2,8	2,8	2,8

***: ligand-interface RMSD < 2.0Å**
****: ligand-interface RMSD < 1.0Å**
*****: ligand-interface RMSD < 0.5Å**

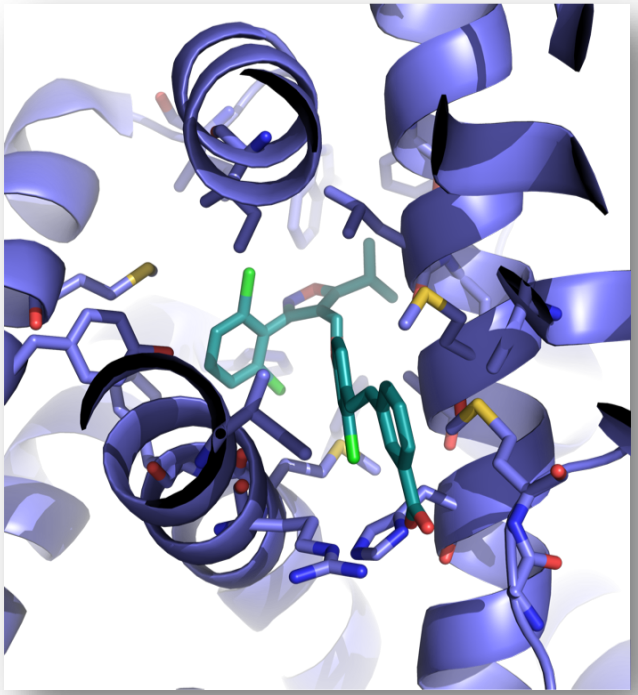
1. Verdonk et al. *J. Chem. Inf. Model.* 2008

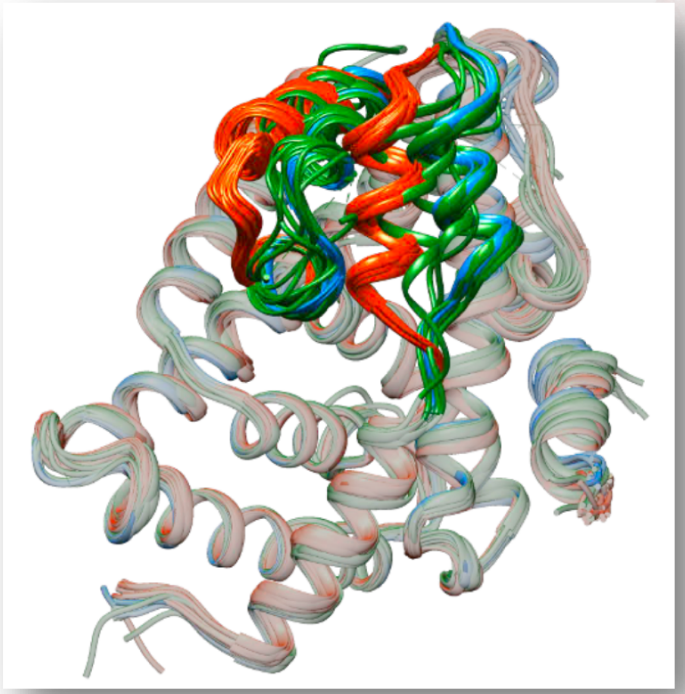


Overview

- Introduction
- Pose predictions
  - Lessons from GC2
  - Applying them to GC3
- Binding affinity/ranking
- Conclusions

GC2: The challenges



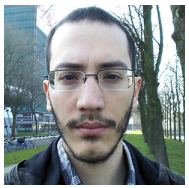


Zeid et al.

Deep and mostly hydrophobic pocket
Rather complex and diverse set of ligands (102 in total)

Flexible helices at the receptor proximal end

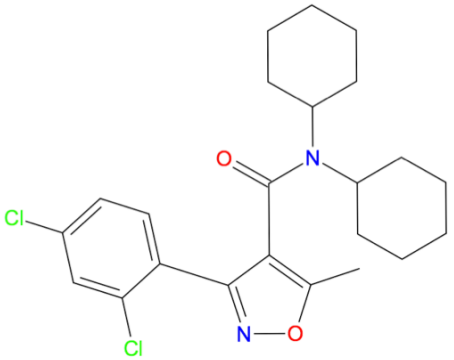
Zeynep Kurkcuoglu

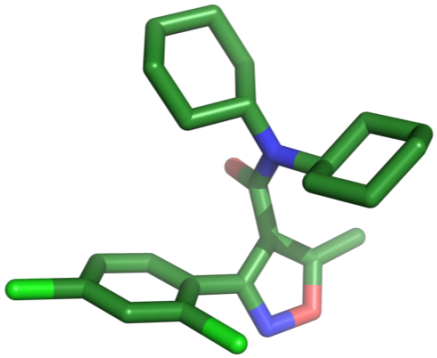


Panos Koukos

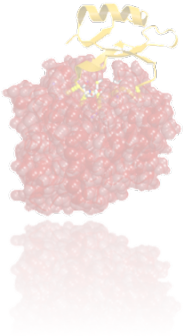
GC2: Ligand generation

- Generate 3D conformers from the SMILE strings (using the OpenEye Omega Toolkit)
- Cluster (hierarchically) the conformers and select representative structures
- Ensemble docking in HADDOCK

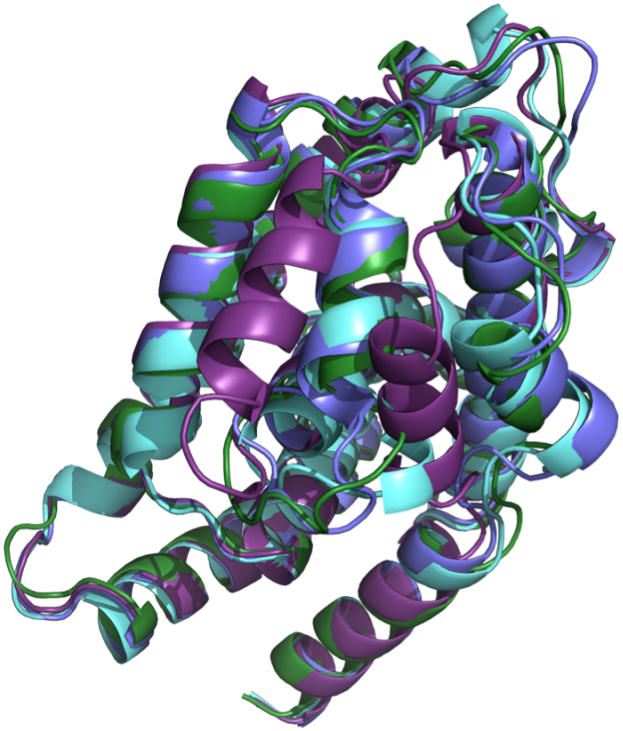




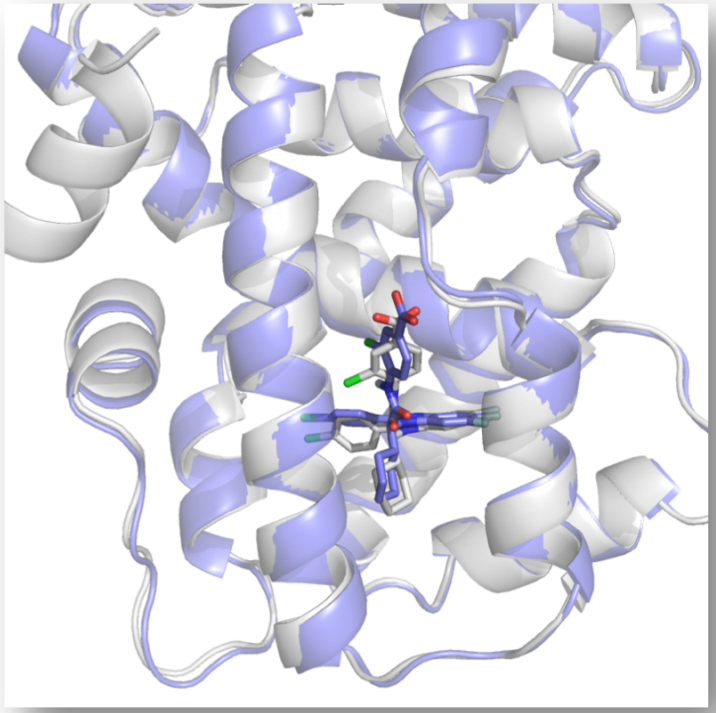
GC2: Receptor for docking



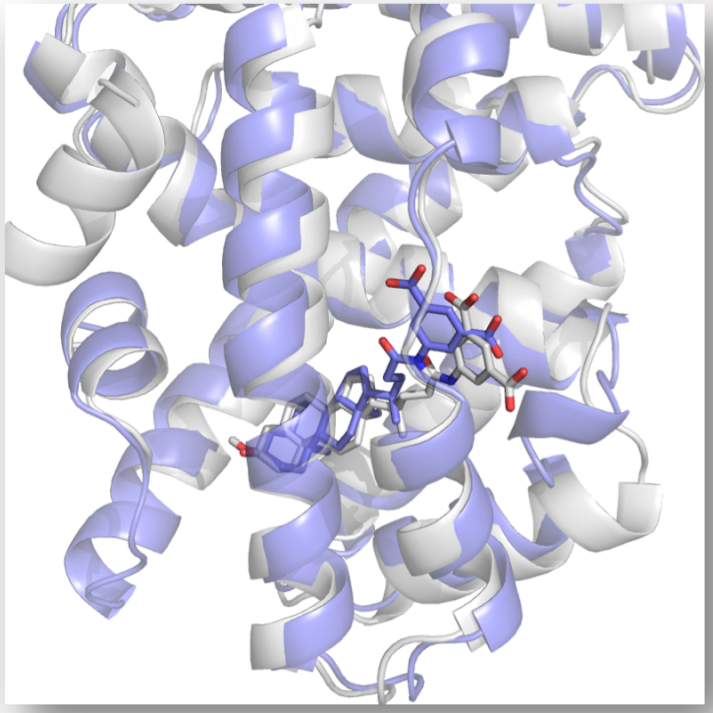
- BLAST Apo protein sequence against the PDB
- Remove problematic entries, cluster the proteins based on binding pocket backbone-RMSD
- Mutate as necessary, select representatives (4 for stage 1)
- Binding pocket defined as the union of all the residues within 5Å of all the ligands



Some successful predictions

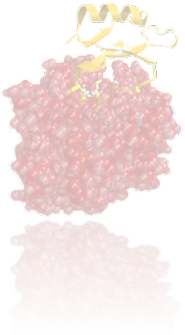


FXR-27: l-RMSD of 1.17Å



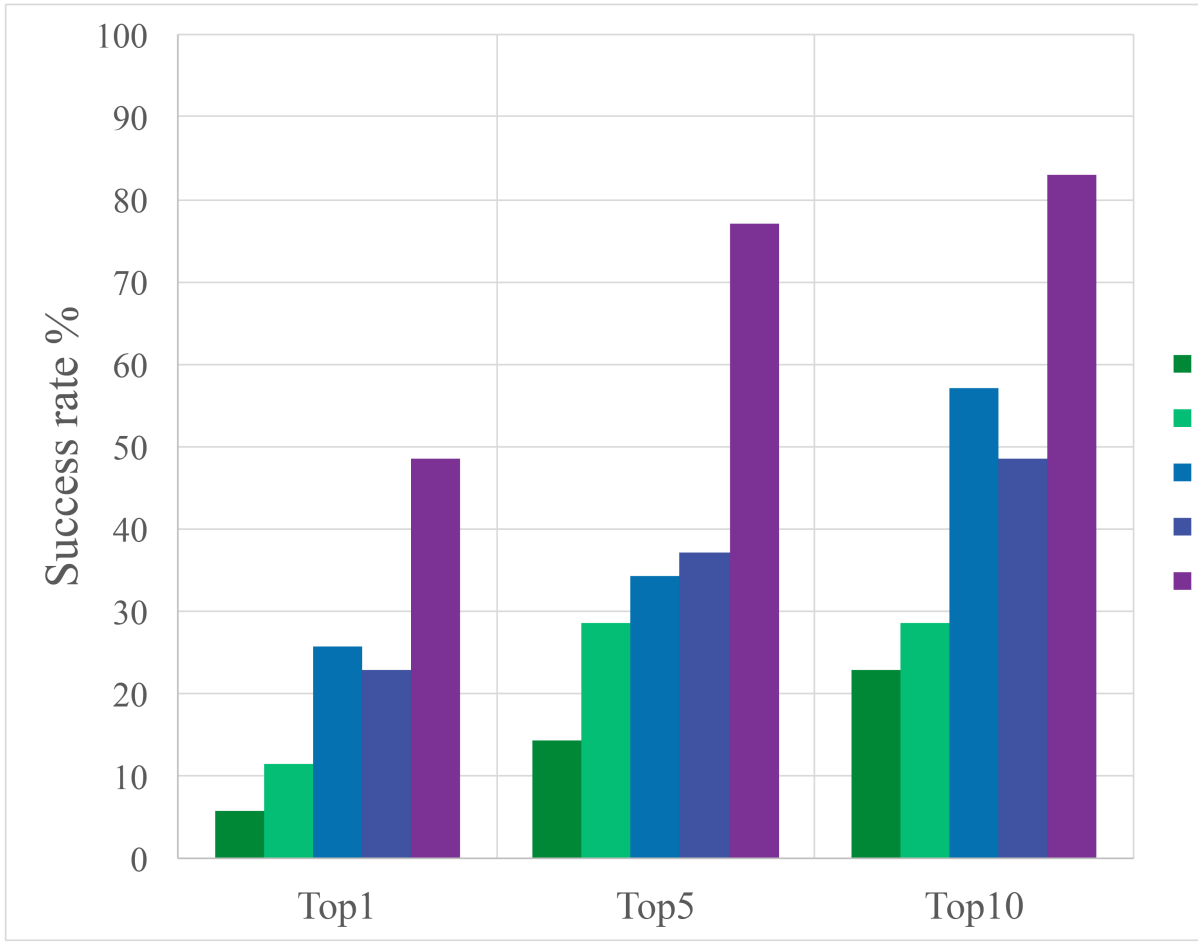
FXR-34: l-RMSD of 1.94Å

Stage1:
Pose prediction performance

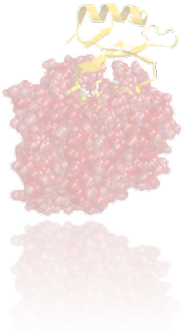


Average best RMSD from 5 poses per ligand

What limits our performance?



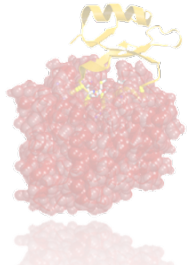
Improving our docking performance in Stage2

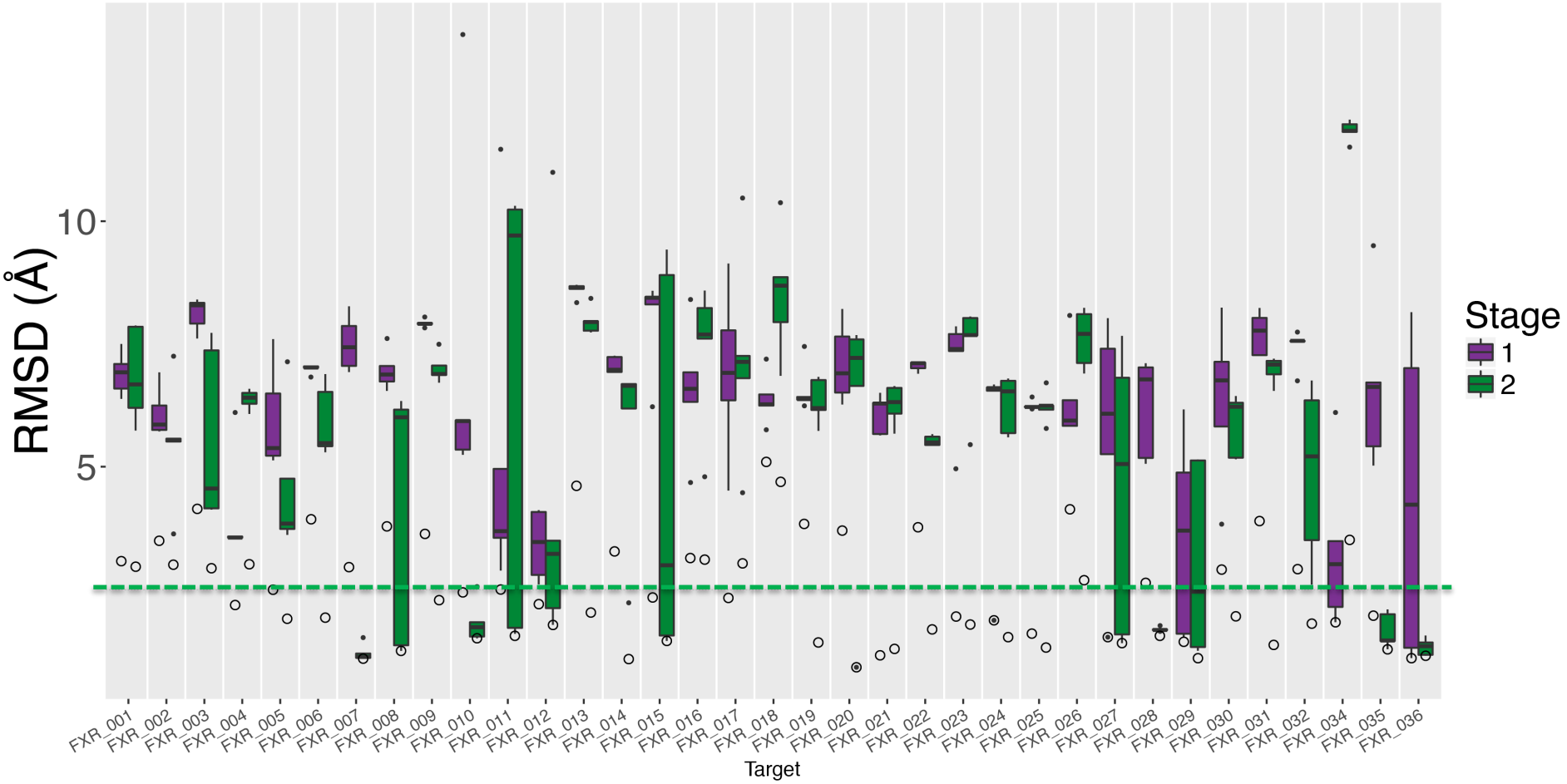


- **Tackling the receptor limiting factor:**
- Select receptor template based on the similarity of the target ligand to the template ligand
- Tanimoto coefficient

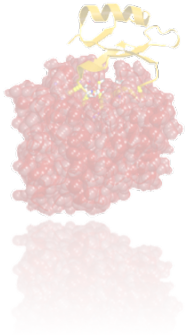
- **Tackling the ligand limiting factor:**
- Up-sample the major cluster if possible

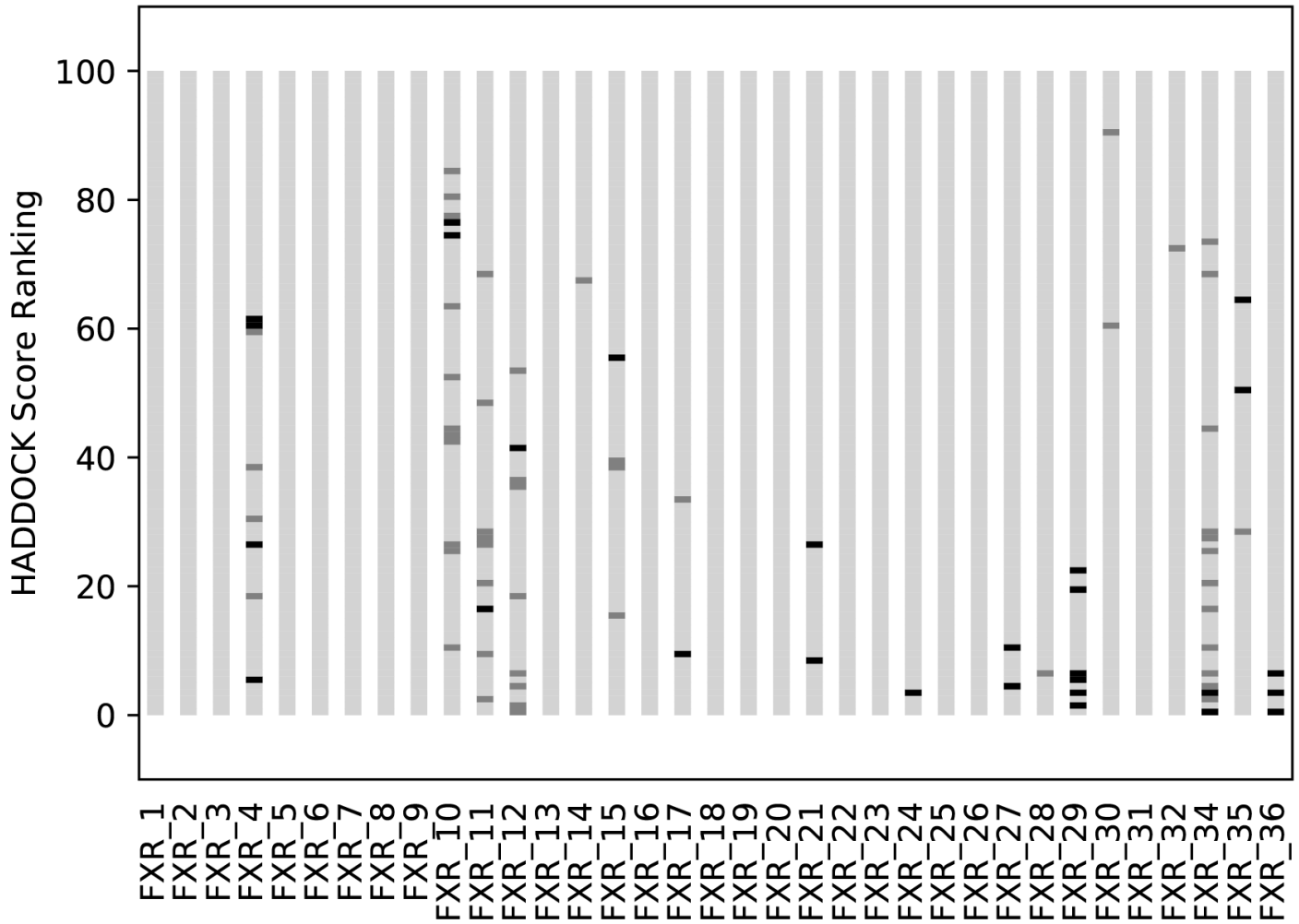
Stage1 vs Stage 2 performance





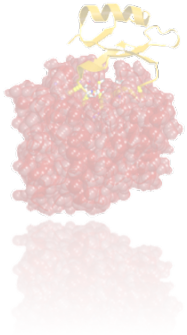
Impact on quality and scoring

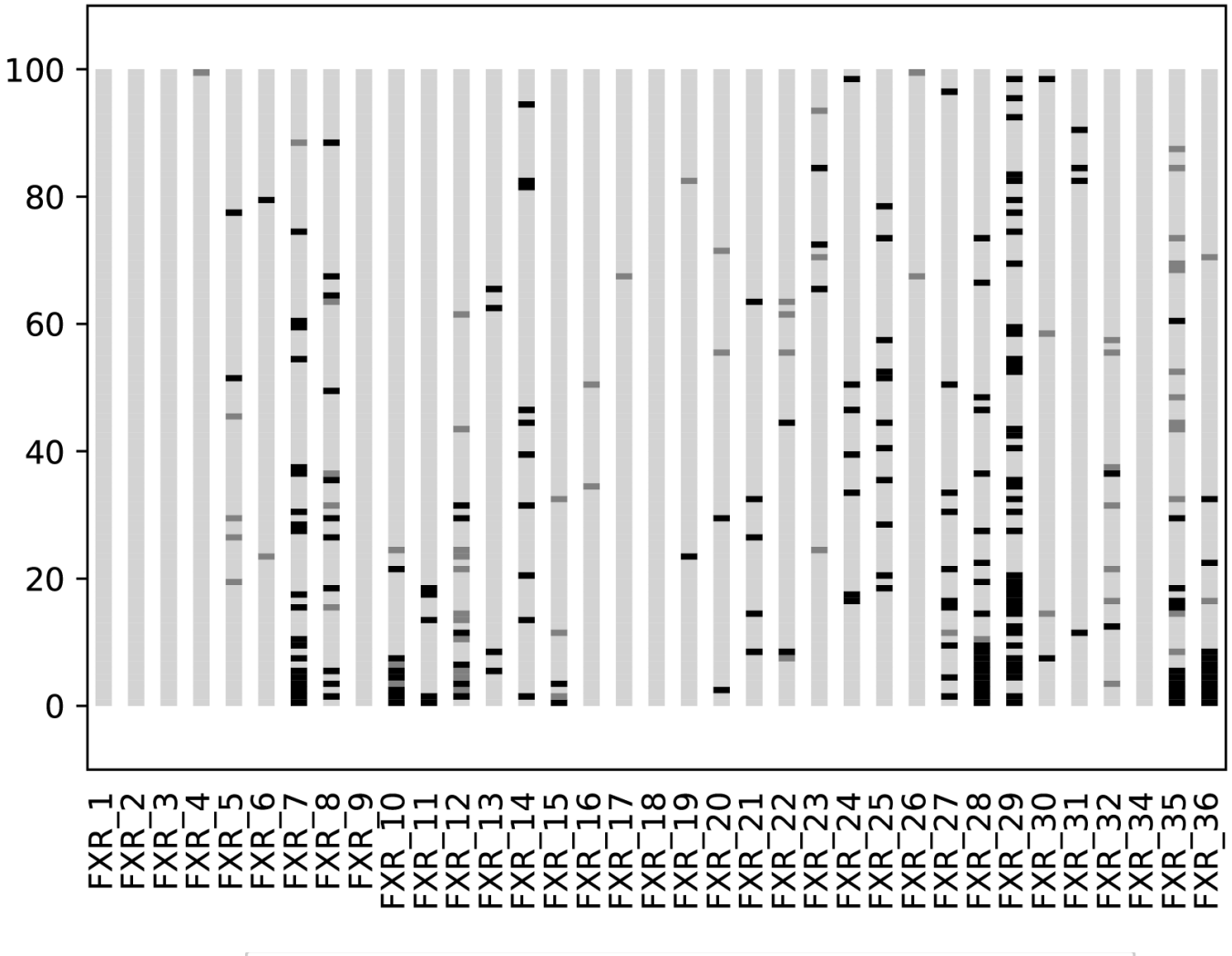




Stage 1

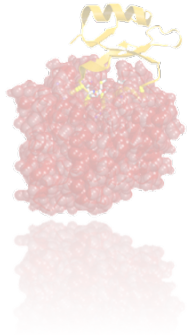
Impact on quality and scoring



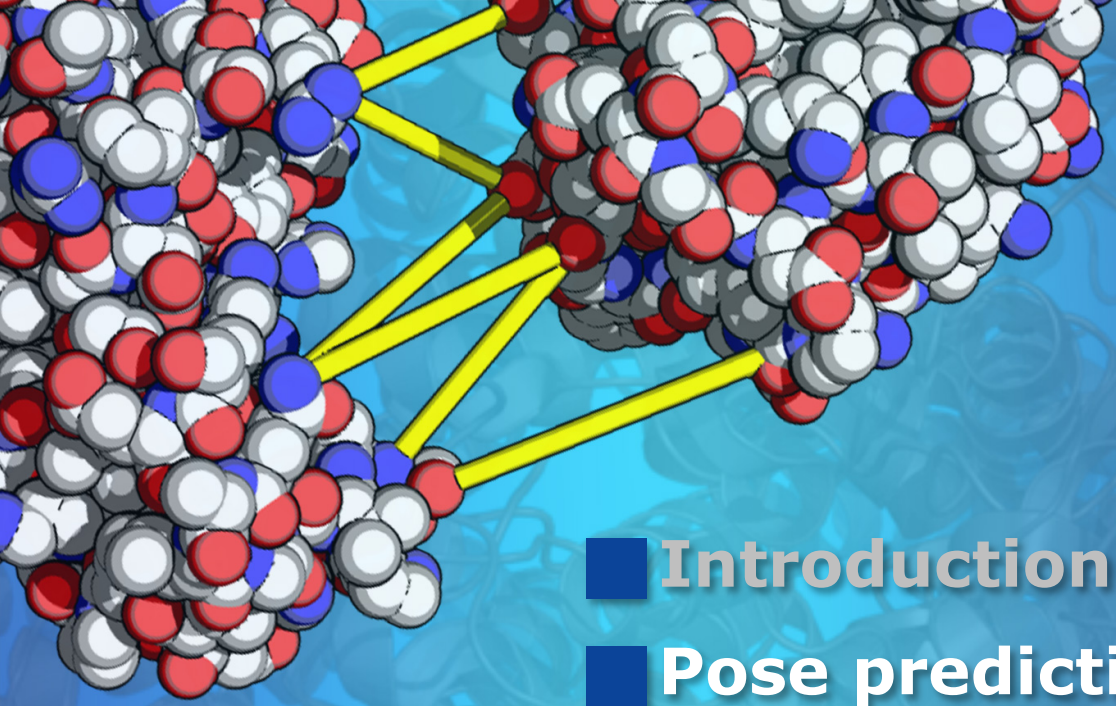


Stage 2

GC2: Main lessons



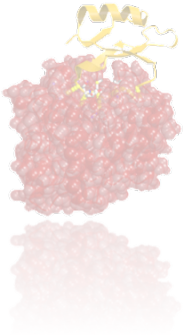
- **Lesson 1:** Smart choice of receptor conformation is crucial
- **Lesson 2:** Need to better select ligand conformations for docking
- **Lesson 3:** Our docking protocol starts from randomly rotated, separated conformations -> need for better starting conformations



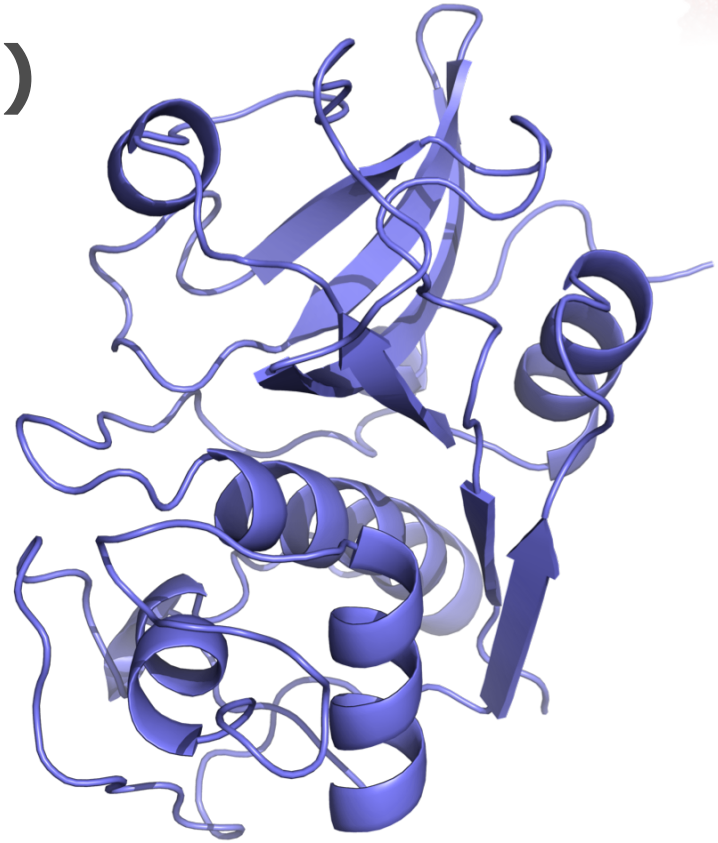
Overview

- Introduction
- Pose predictions
  - Lessons from GC2
  - Applying them to GC3
- Binding affinity/ranking
- Conclusions

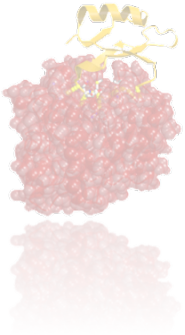
D3R 2017 Grand Challenge 3



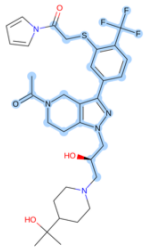
- **Protein (cathepsin) against 141 (24) small molecules**

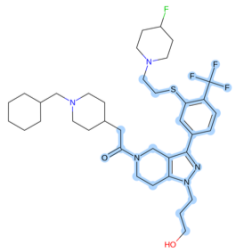


Selecting the receptor

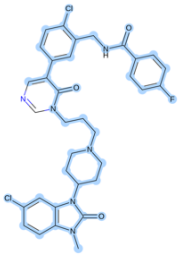


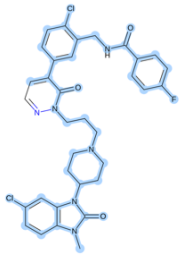
- **Identify templates with high sequence identity (>70%) to the target protein sequence with at least one bound ligand.**





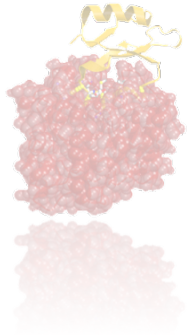
- **Compare the crystallographic ligand to the Cathepsin set using the Maximum Common Substructure (MCS)** (as implemented in ChemmineR)



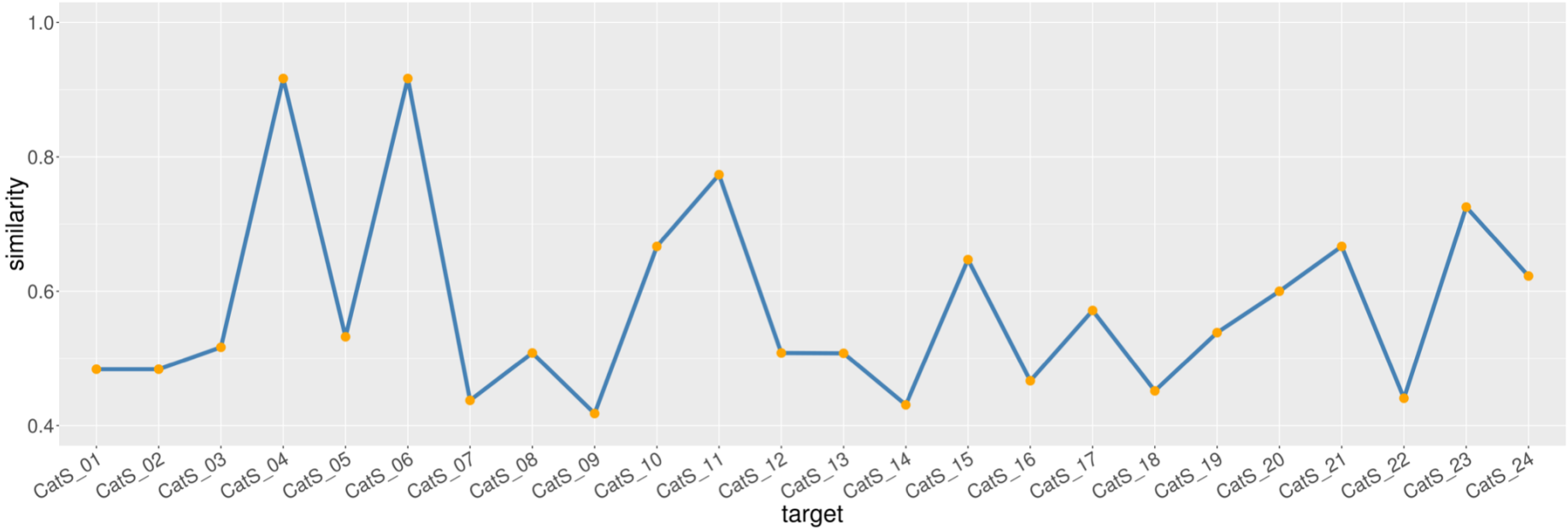


- **Select the receptor with the highest Tanimoto Coefficient (GC2 lesson 1)**

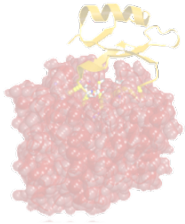
Selecting the receptor



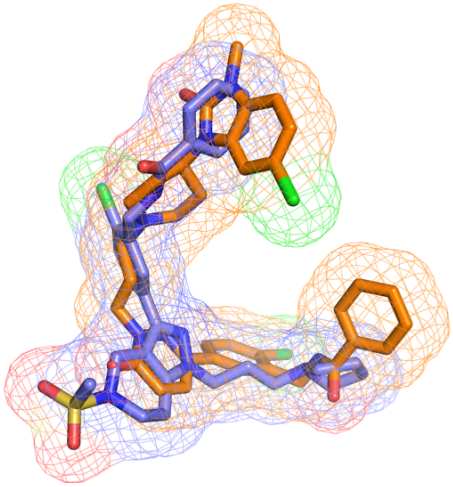
Ligand similarity of best template found



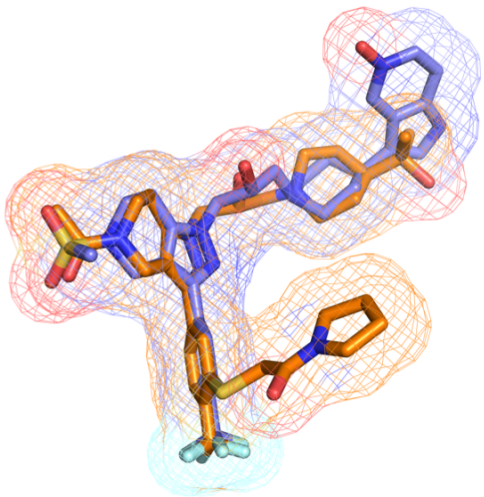
Ligand conformation



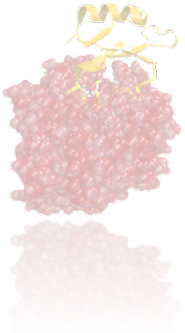
- **Compare similarity of crystallographic template ligands with generated conformers using shape and color tanimoto** (as implemented in OpenEye ROCS)



- **Select 10 conformers with the highest combined score for ensemble docking (GC2 Lesson 2)**



Starting conformations



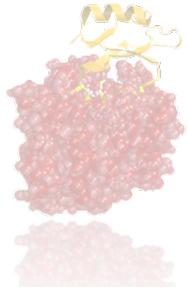
- **Superpose the selected conformers on the crystallographic ligands** (OpenEye shape-TK)

- **Refine using HADDOCK – only short minimization in 2nd stage and final refinement in explicit solvent**

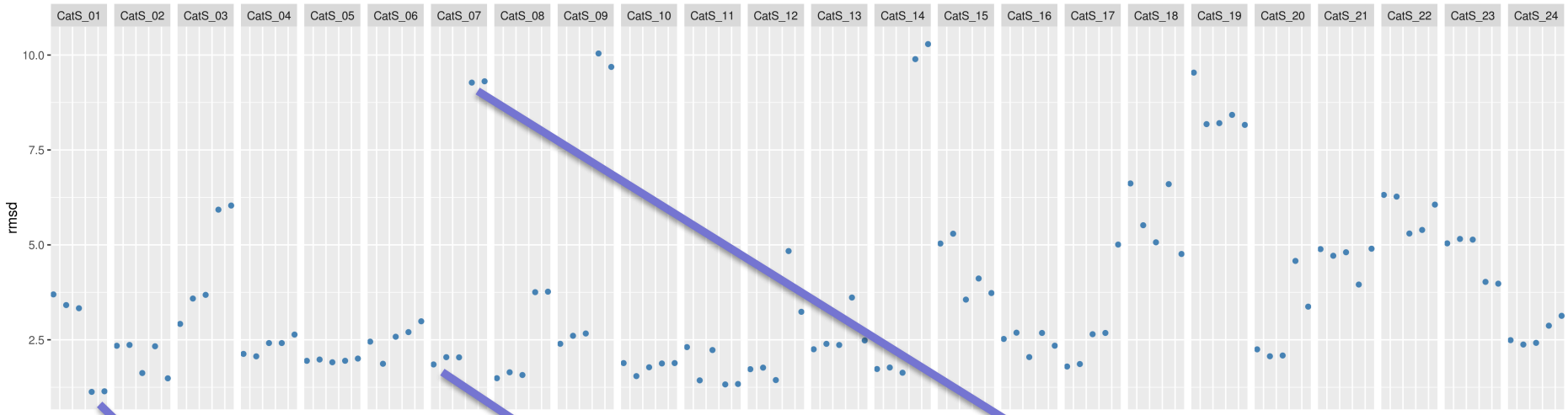
(GC2 Lesson 3)

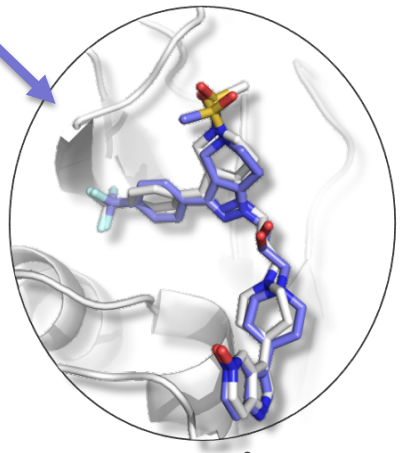
- For stage 1b, receptor, water molecules and other small molecules kept rigid in place (not much impact on performance)

Docking results

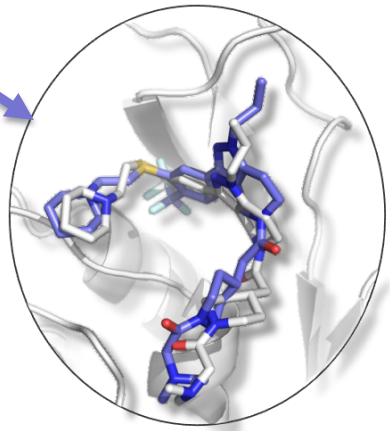


Stage 1 heavy-atom RMSD

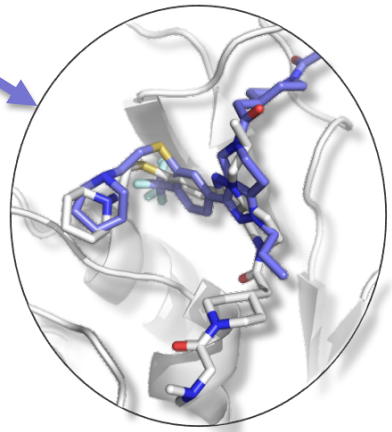




0.91Å

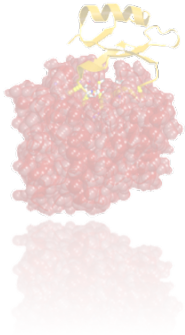


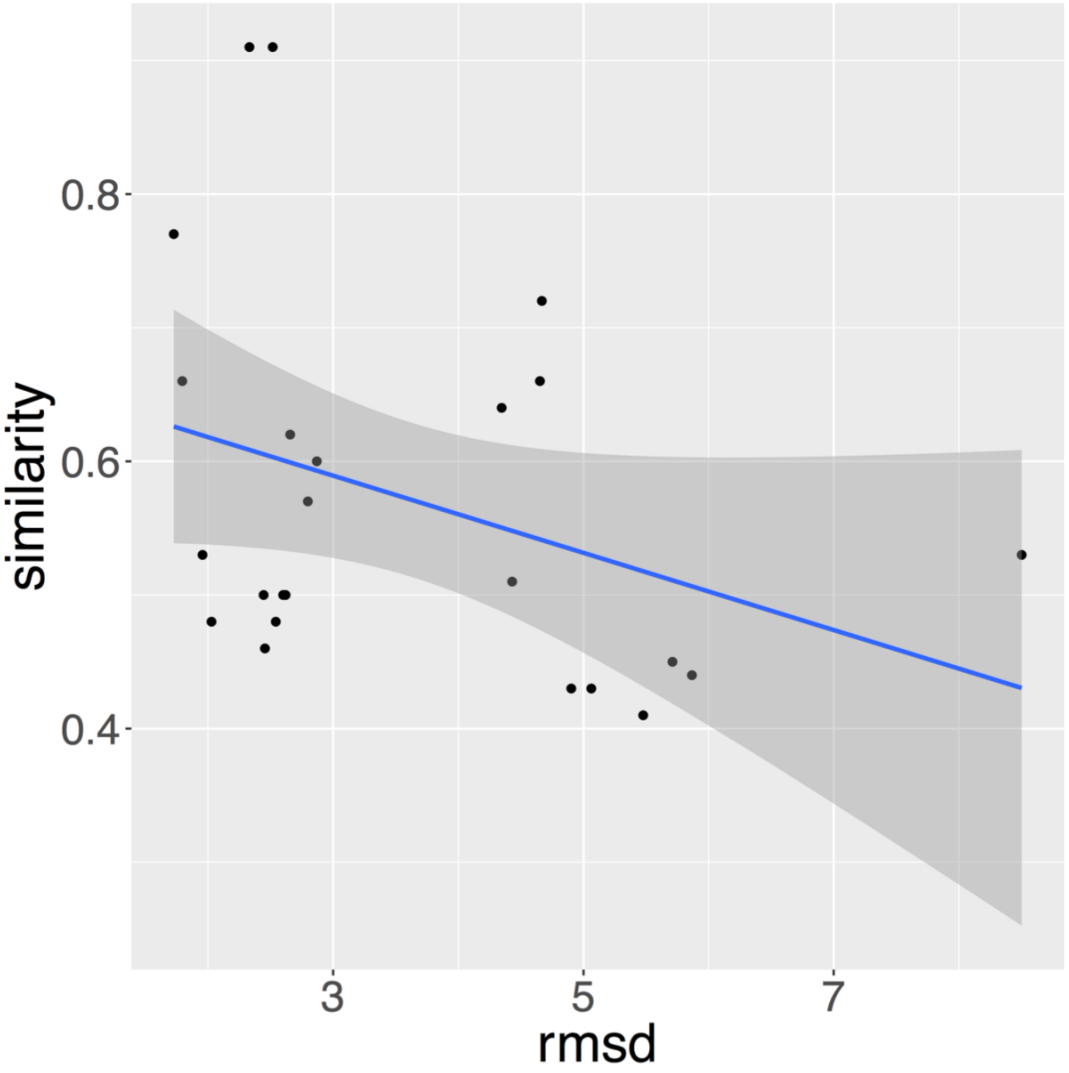
1.75Å



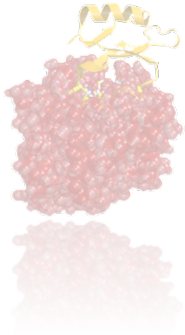
9.26Å

Impact of ligand similarity

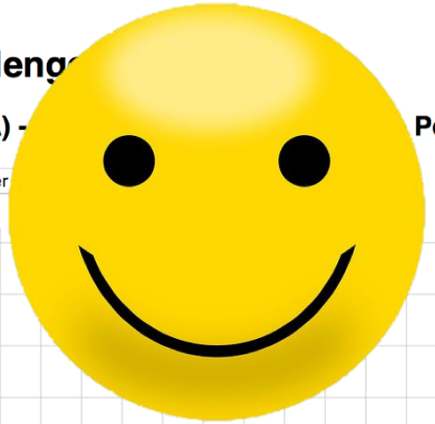


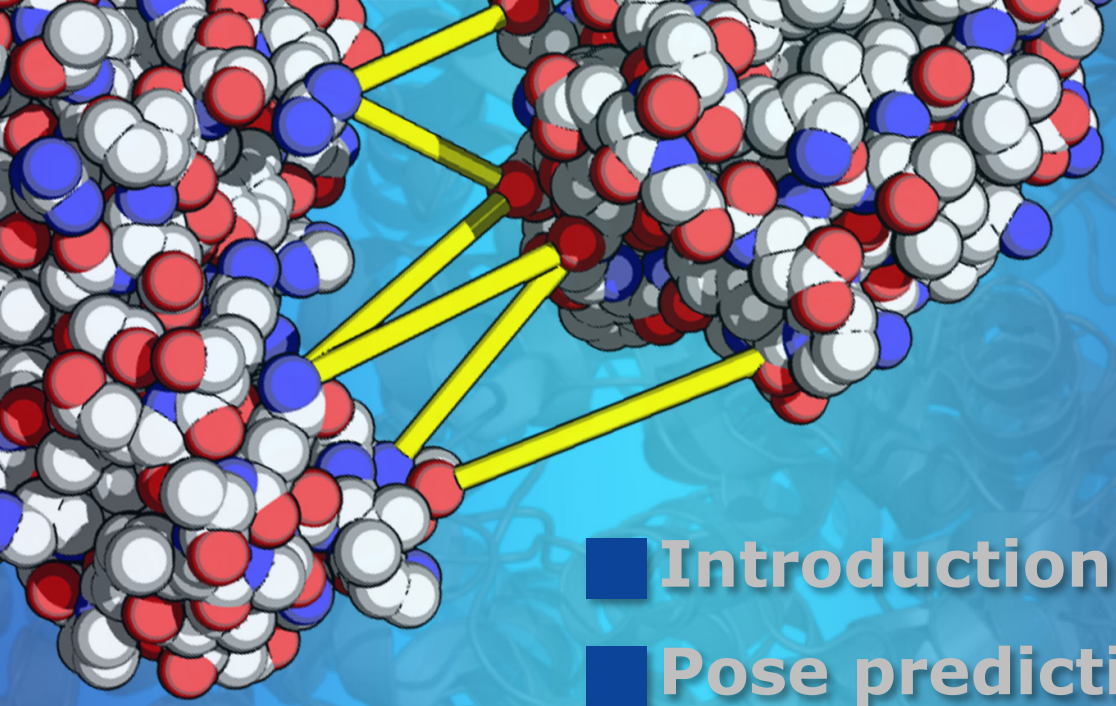


Lessons learned!



GC3 – Excellent performance

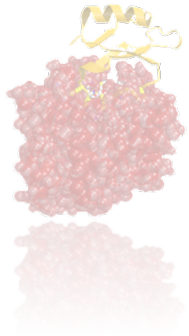




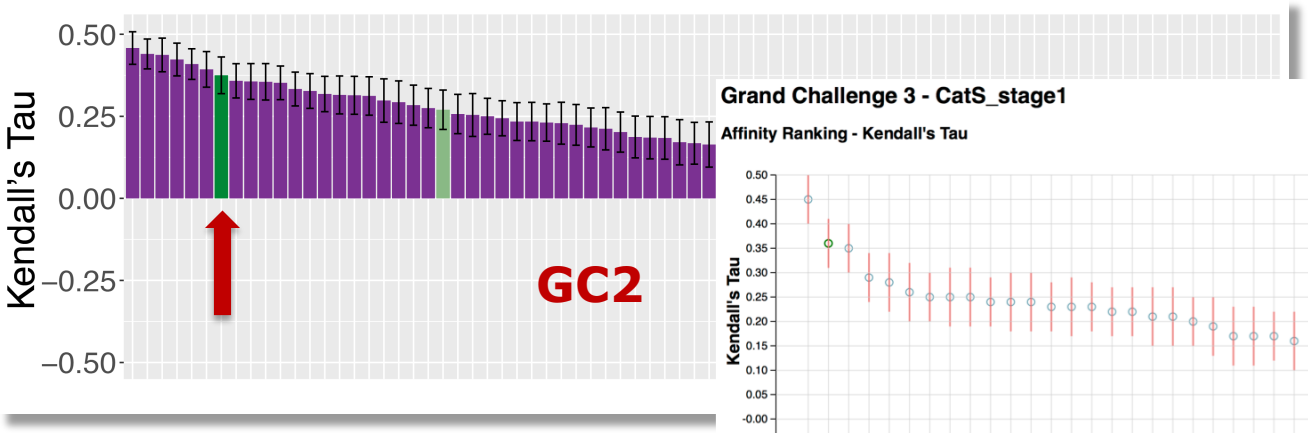
Overview

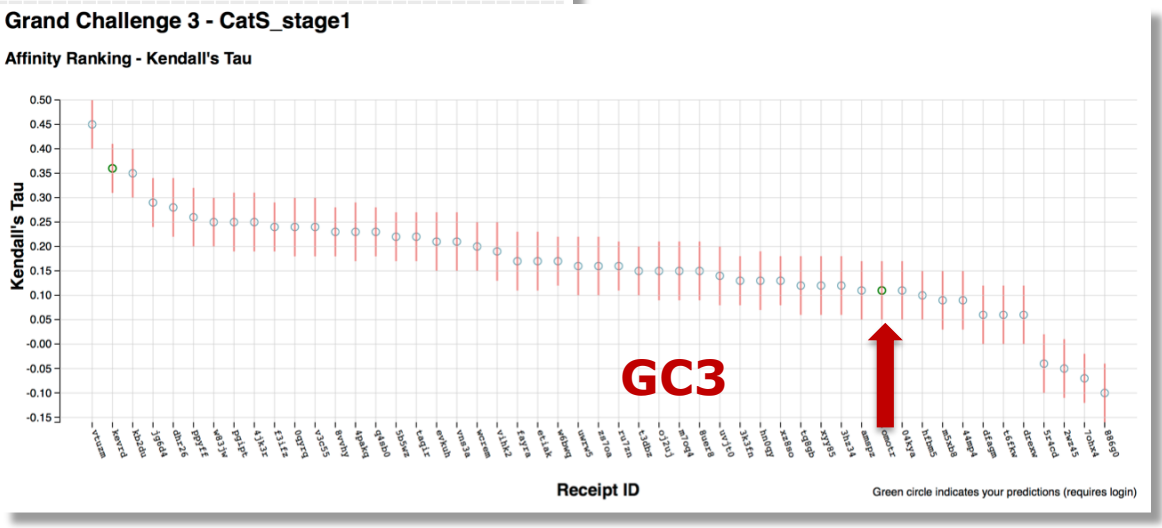
- Introduction
- Pose predictions
  - Lessons from GC2
  - Applying them to GC3
- **Binding affinity/ranking**
- Conclusions

Ligand binding affinity/ranking prediction



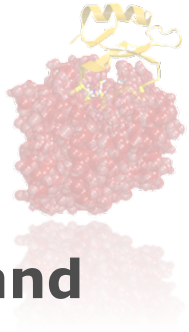
- **Nothing changed between GC2 and GC3**
- **Structure-based prediction using an atomic contact model** (see our GC2 paper)



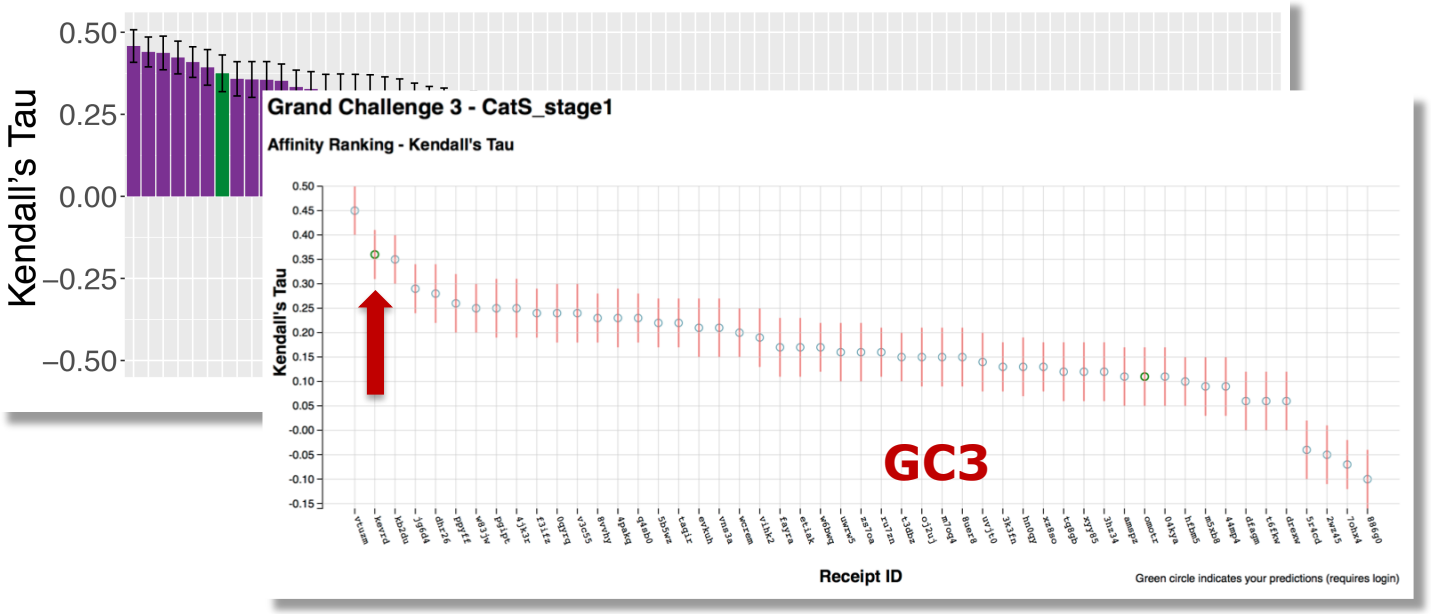


Kurkcuoglu et al. *J. Comp. Aid. Mol. Des.* 2017

Ligand binding affinity/ranking prediction

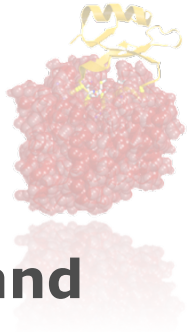


- **Ligand-based prediction using target-specific ligand similarity SVR model** (see our GC2 paper)

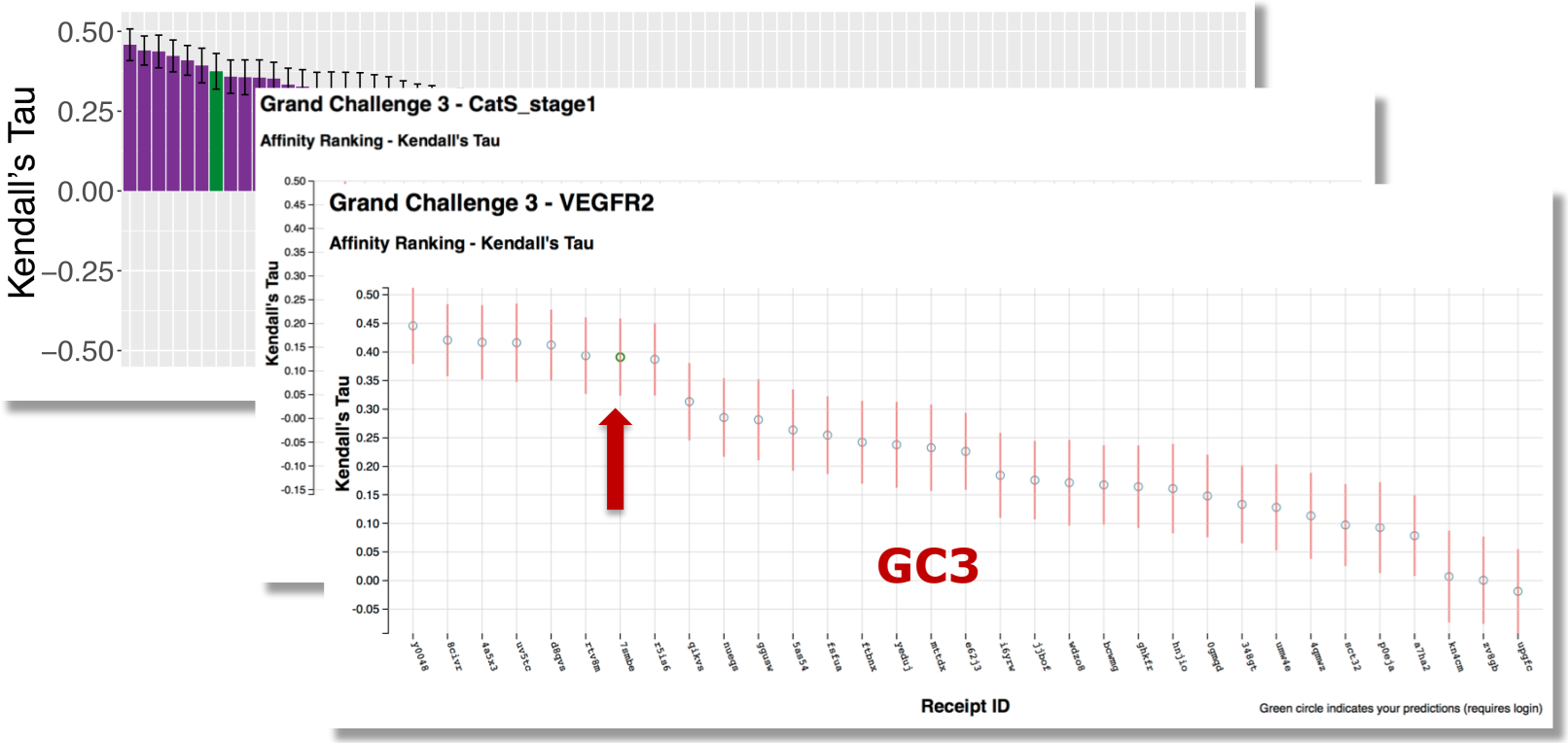


Kurkcuoglu et al. *J. Comp. Aid. Mol. Des.* 2017

Ligand binding affinity/ranking prediction

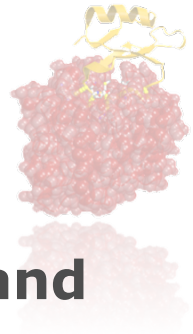


- **Ligand-based prediction using target-specific ligand similarity SVR model** (see our GC2 paper)

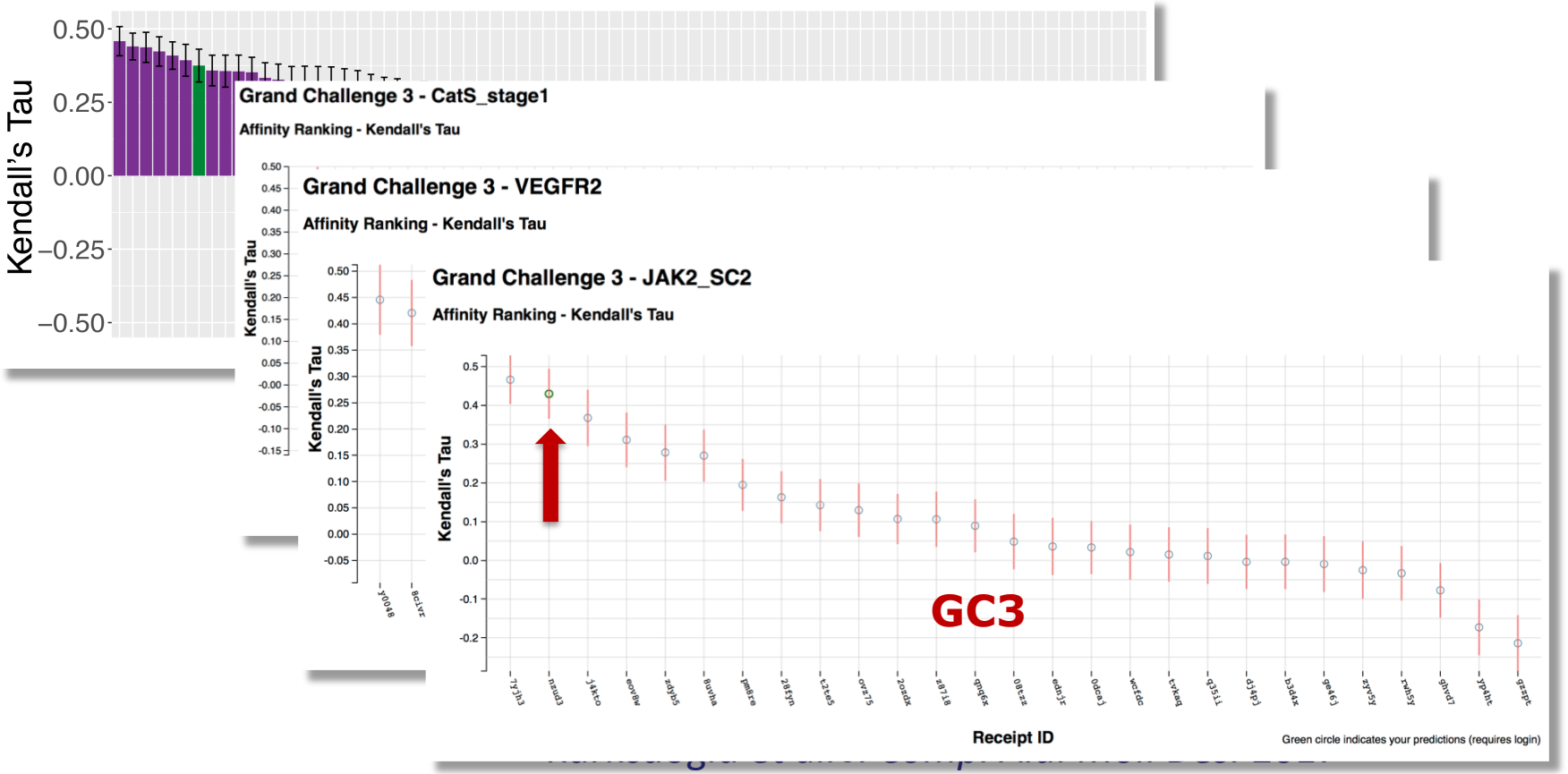


Kurkcuoglu et al. *J. Comp. Aid. Mol. Des.* 2017

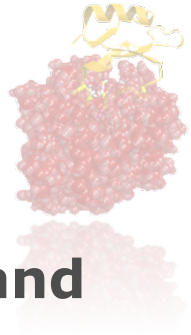
Ligand binding affinity/ranking prediction



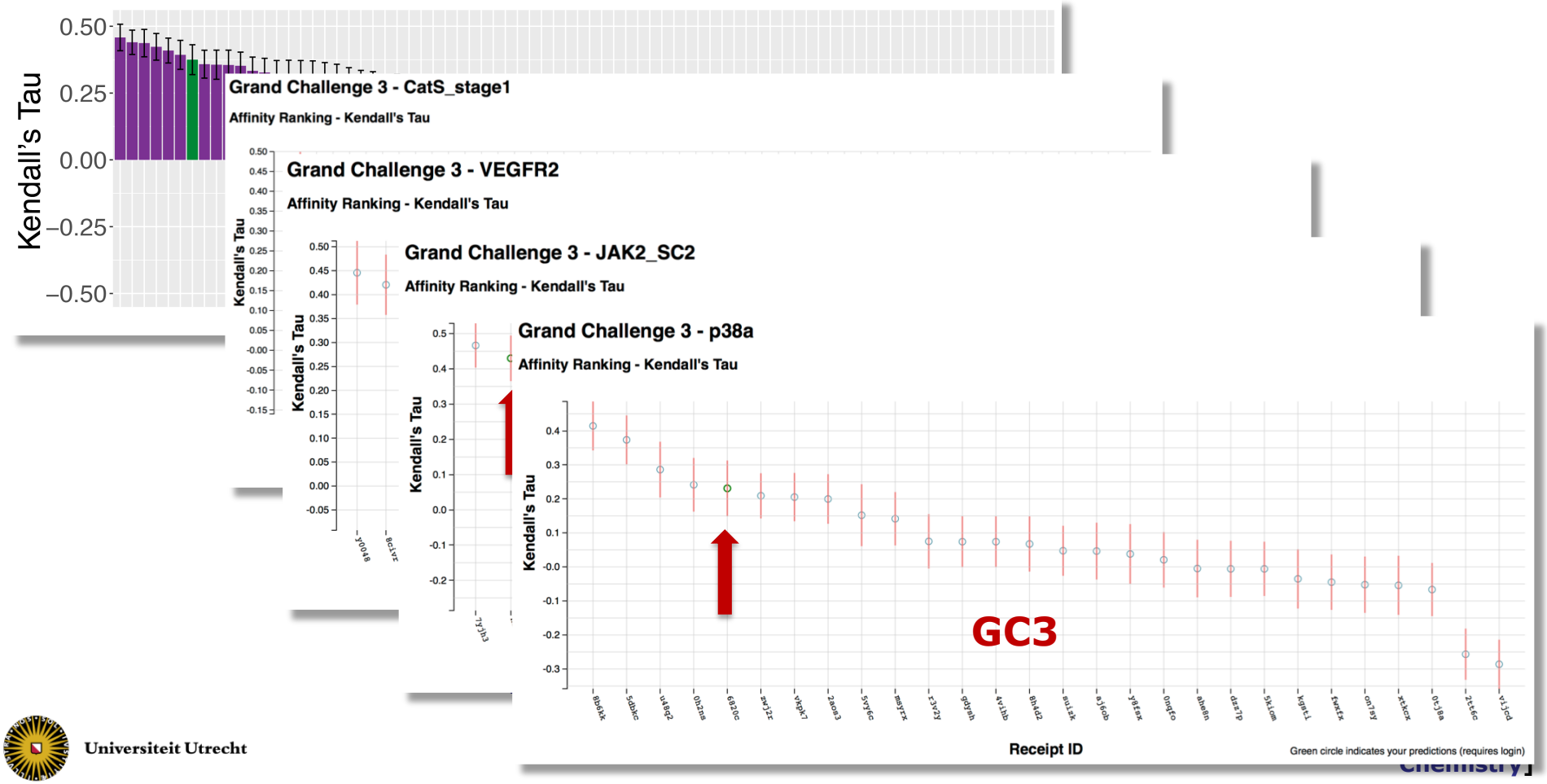
- **Ligand-based prediction using target-specific ligand similarity SVR model** (see our GC2 paper)

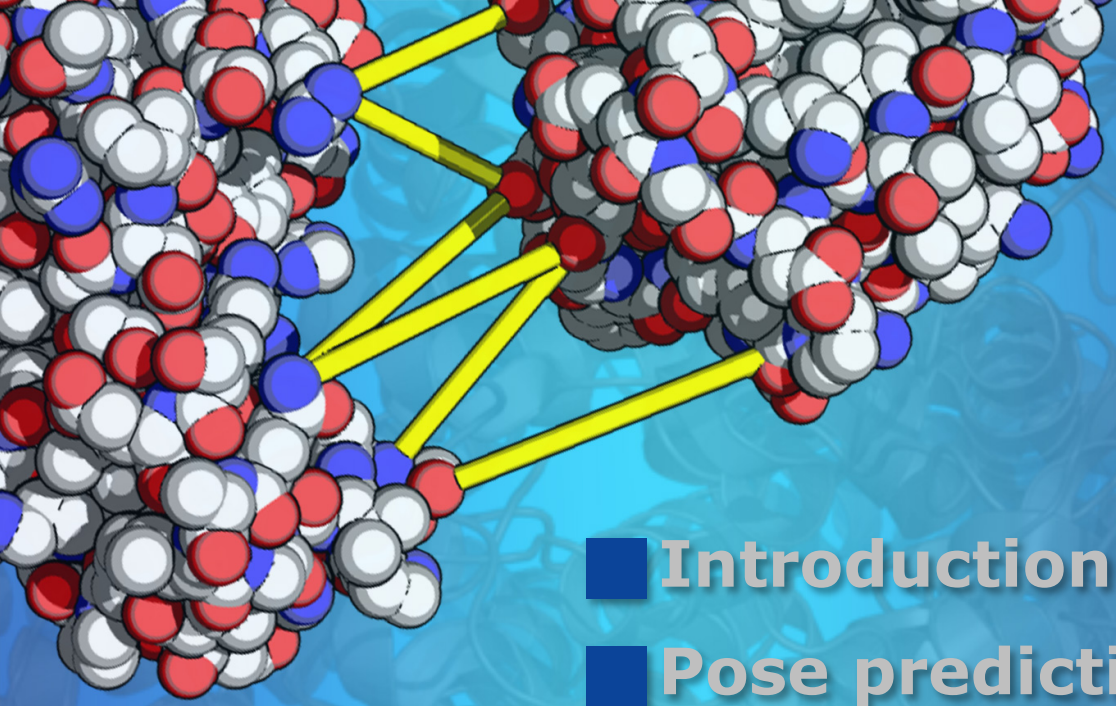


Ligand binding affinity/ranking prediction



- **Ligand-based prediction using target-specific ligand similarity SVR model** (see our GC2 paper)

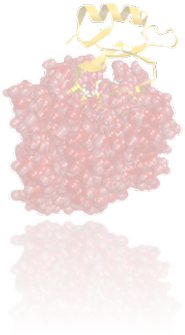




Overview

- Introduction
- Pose predictions
  - Lessons from GC2
  - Applying them to GC3
- Binding affinity/ranking
- Conclusions

Conclusions



- as catalyzer for learning from failures
- **Success factors (for us):**
  - **Smart selection of receptor and ligand conformations**
  - **Smart positioning of ligand in binding pocket**

Kurkcuoglu et al. *J. Comp. Aid. Mol. Des.* 2017

Acknowledgments:

the CSB group@UU

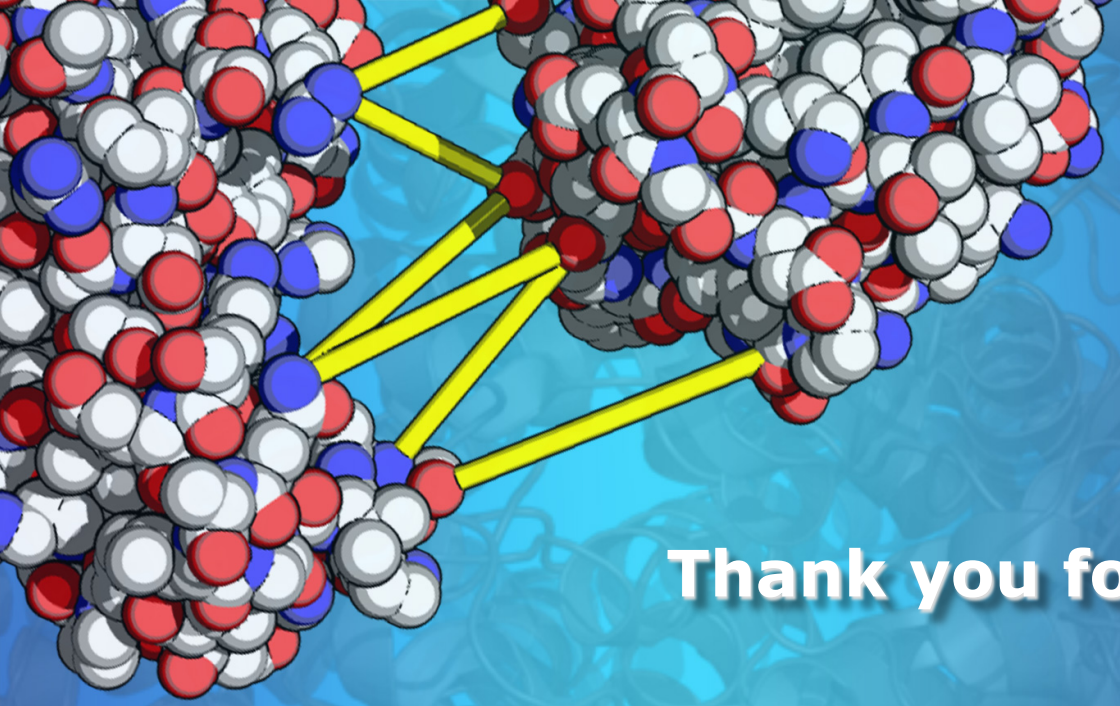
€€

NWO

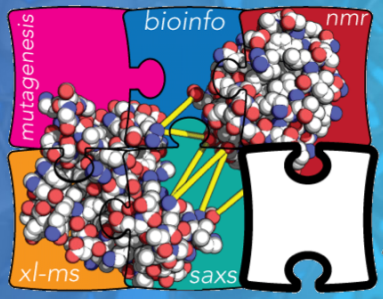
VICI
TOP-PUNT

WeNMR
West-Life
EGI-Engage
INDIGO-
Datacloud
BioExcel CoE
EOSC-Hub

And the D3R team for running the show!



Thank you for your attention!



HADDOCK
High-Ambiguity Driven Docking

HADDOCK online:

- **http://haddock.science.uu.nl**
- **http://bonvinlab.org/software**

Utrecht Bioinformatics Center