Evaluation Overview of GC3

Neysa Nevins & Mill Lambert
Computational & Modeling Sciences
GlaxoSmithKline

do more
feel better
live longer

GSK Docking & Scoring Study (carried out 2002-2003)

J. Med. Chem. **2006**, *49*, 5912−5931

A Critical Assessment of Docking Programs and Scoring Functions

Gregory L. Warren,[*,†] C. Webster Andrews,[‡] Anna-Maria Capelli,[#] Brian Clarke,[||] Judith LaLonde,[†,§] Millard H. Lambert,[‡] Mika Lindvall,[⊥,●] Neysa Nevins,[†] Simon F. Semus,[†] Stefan Senger,[⊥] Giovanna Tedesco,[#] Ian D. Wall,[||] James M. Woolven,[⊥] Catherine E. Peishoff,[†] and Martha S. Head[†]

GlaxoSmithKline Pharmaceuticals, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, GlaxoSmithKline, Five Moore Drive, Research Triangle Park, North Carolina 27709, GlaxoSmithKline, Centre Via Alessandro, Fleming 4, 37135, Verona, Italy, GlaxoSmithKline, New Frontiers Science Park, Third Avenue, Harlow, Essex CM19 5AW, U.K., and GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, U.K.

Received April 17, 2005

Docking is a computational technique that samples conformations of small molecules in protein binding sites; scoring functions are used to assess which of these conformations best complements the protein binding site. An evaluation of 10 docking programs and 37 scoring functions was conducted against eight proteins of seven protein types for three tasks: binding mode prediction, virtual screening for lead identification, and rank-ordering by affinity for lead optimization. All of the docking programs were able to generate ligand conformations similar to crystallographically determined protein/ligand complex structures for at least one of the targets. However, scoring functions were less successful at distinguishing the crystallographic conformation from the set of docked poses. Docking programs identified active compounds from a pharmaceutically relevant pool of decoy compounds; however, no single program performed well for all of the targets. For prediction of compound affinity, none of the docking programs or scoring functions made a useful prediction of ligand binding affinity.

Overall conclusions of Warren, *et. al.*

Circa 2005

Docking is a productive technology

- Can predict binding modes correctly
  - select using the score, SAR, and intuition
  - predictions have successfully guided chemistry
- Can identify active compounds by virtual screening
  - enrichment from corporate collection or purchasable compounds

But…

- Cannot predict best target/algorithm pairing
- Cannot reliably rank order compounds by affinity*

*Extensive and immediate development is needed

ACS Natl Meeting 2011 Docking & Scoring Symposium

The Challenge

- Six pages of instructions from Greg Warren to participants
- Programs: DOCK (Brozell), DOCK 6.4 (Mukherjee), eHITS, FlexX/Hyde, FRED 3.0, Glide, Gold5.0/ChemPLP, ICM, LeadFinder, LibDock, MOE, Surflex-Dock
  - Pre-work: run program/algorithm in best way possible against two public data sets
    - 85 structures in Astex data set* for binding mode prediction
    - 40 DUD** v2 targets (using the matched active and decoys sets) for virtual screening (e.g. run HSP90 ligands against EGFR)
- Each participant was asked to spend 25 min of their 40 min symposium talk reporting their data for binding mode prediction & virtual screening efficiency exercises above

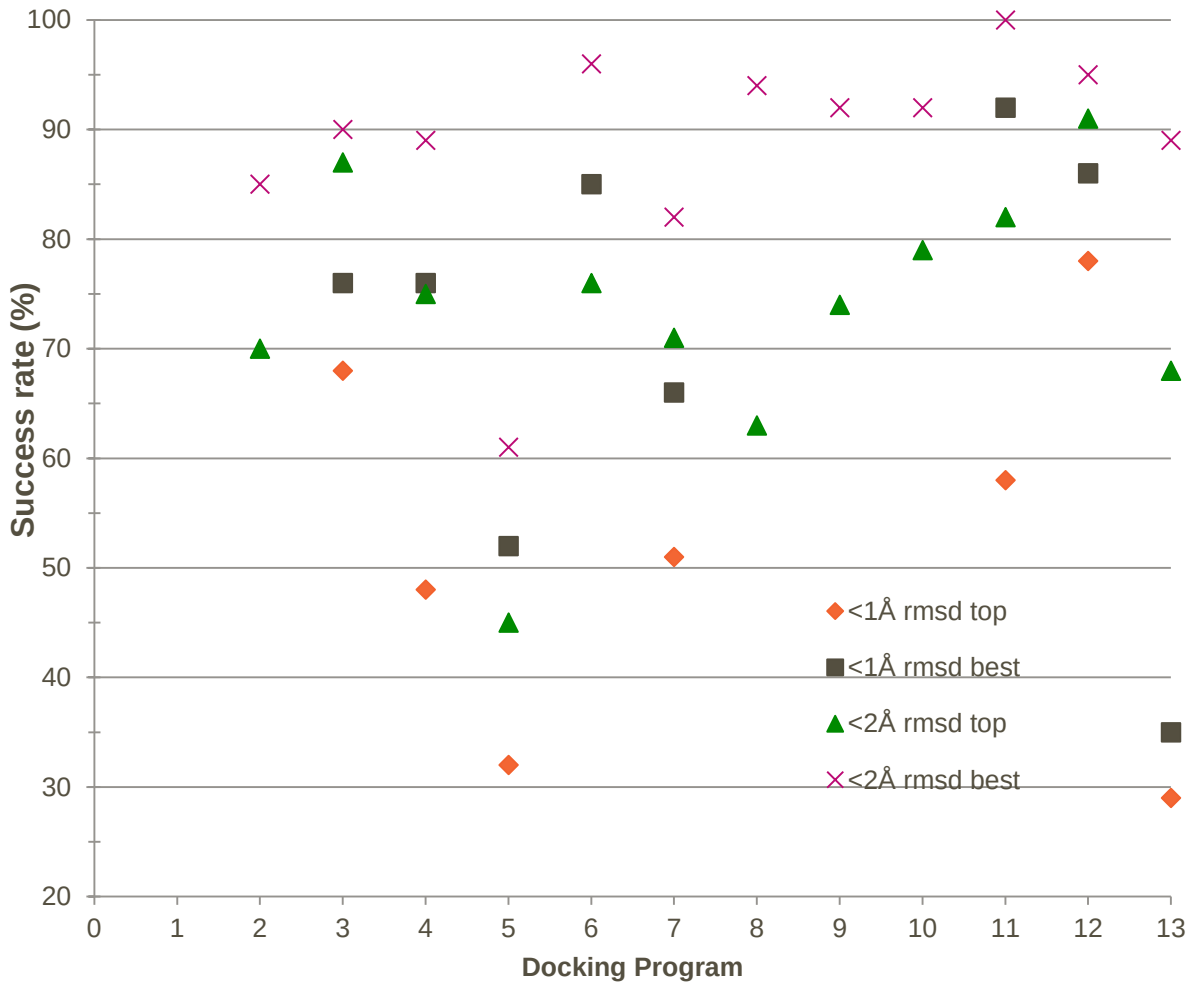
- It was our hope that this exercise would give audience sense of each algorithm's optimum performance since developers carried out calculations using optimum settings

* M. J. Hartshorn, …, C. W. Murray, Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance, *J. Med. Chem.* **2007,** *50,* 726-741.

** Huang, Shoichet and Irwin, *J. Med. Chem.* **2006**, 49(23), 6789-6801.

Pose prediction circa 2010

Prospective Docking Exercise on Astex data set

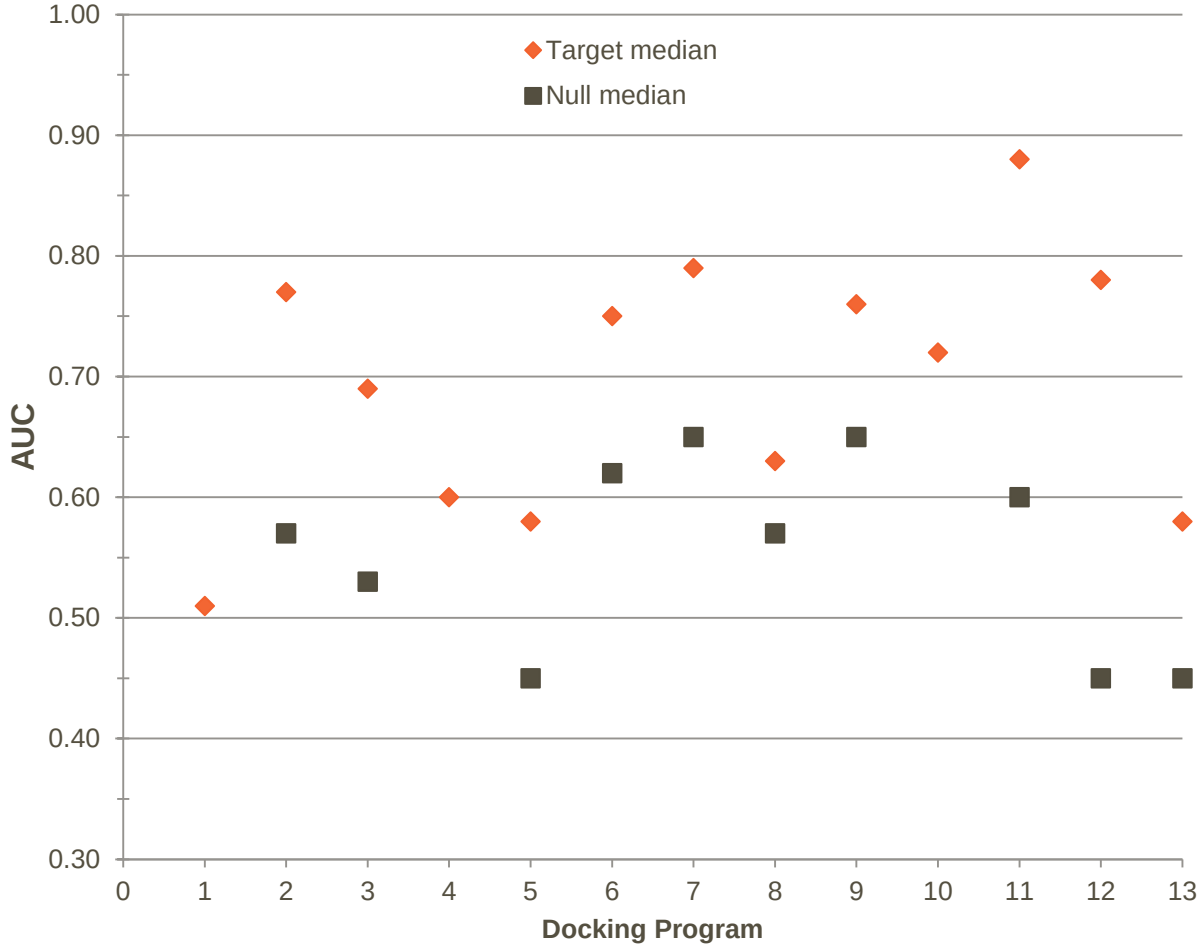


Nine of 11 docking programs find crystallographic pose 70% of the time

241st ACS meeting
March 2011,
Anaheim, CA

Virtual screening

Prospective Virtual Screening Exercise on DUD data set



Docking ligands from one target into another still provides enrichment!

241st ACS meeting
March 2011,
Anaheim, CA

Lessons learned (and re-learned) in Anaheim

241st ACS meeting, March 2011

- Data Sets
  - Deviations from PDB formatting cause issues (missing TER, HETATM vs ATOM, etc)
  - Be mindful of significant conformational differences and critical missing residues when selecting proteins for set, *e.g.* kinase DFG-in vs DFG-out, activation loop residue
  - Refining protein structures (ensuring quality) improves results
  - Ensure actives are actives, decoys are decoys
  - Alternate conformations don't seem to matter (for this experiment, still small # cases)
  - Scripts/preparation can fail
- Charges can heavily influence outcome, non-charge dependent codes more flexible?
- Optimizing H-bond network improves results and potentially reduces cross-docking issues
- Binding site optimization within experimental error improves results for cognate docking
- Force field based scores + extra terms (training) – tend to have higher %success predicting binding modes, virtual screening AUC's
- Problem with RMSD metric: highly symmetric ligands (*e.g.* 1jje 8 Å RMSD)
- Null hypothesis added insight; issue with thrombin/FXa cross-reactivity
- Early enrichment metrics have no statistical meaning!
  - However, real-life virtual docking often involves selecting only the first 0.1-1% of hits for testing in assay
- Are errors in data the "boat anchor dragging down scoring functions"?

Review of issues by folks attending D&S symposium

Not necessarily new, some programs now address – *i.e.* incremental improvements

- "H-bonding network" not optimized to ligand (issue with parameterized scoring fxns)
- Ionization state issues
- Missing water
- Solvation not treated properly
- Errors in ligand conformation
- Ligand and protein entropy
- Metals

241st ACS meeting
March 2011,
Anaheim, CA

CSAR

2008-2014

Organized four exercises

1. 2010 Benchmark (CSAR-HiQ set)
2. 2012 CSAR Exercise (Cdk2 , Cdk2-CyclinA, LpxC) (Michigan), Urokinase (Abbott), Chk1 (Abbott), ERK2 (Vertex)
3. 2013 Benchmark Exercise (Protein design in collab w/ David Baker, Univ of Washington)
4. 2014 Benchmark Exercise (fXa, Syk, TrmD donated by GlaxoSmithKline)

Lessons Learned*

- Lesson 1: Good Crystal Structures Are Hard to Find
- Lesson 2: Several Metrics Are Needed for Assessing Docking and Scoring
- Lesson 3: Embrace Statistics, Error Bars, and Confidence Intervals
- Lesson 4: Making a Good Data Set Is a Difficult Multi-optimization Process
- Lesson 5: Please Stop Using FXa as a Model System

*Heather Carlson, “Lessons Learned over Four Benchmark Exercises from the Community Structure–Activity Resource”, *J. Chem. Inf. Model.* **2016**, 56(6), 951-954.

Protein-ligand crystal set quality recommendations

REVIEWS

Drug Discovery Today • Volume 17, Numbers 23–24 • December 2012

ELSEVIER

All structure-based drug design predictions are dependent on the quality of the protein–ligand structure(s) used. This review discusses methods for assessing the quality of this critical data.

Essential considerations for using protein–ligand structures in drug discovery

Reviews•

Gregory L. Warren[1], Thanh D. Do[2,3], Brian P. Kelley[1], Anthony Nicholls[1] and Stephen D. Warren[2]

[1] OpenEye Scientific Software, Inc., 9 Bisbee Court Suite D, Santa Fe, NM 87508, USA
[2] Department of Chemistry & Biochemistry, Gonzaga University, 502 E Boone Ave, Spokane, WA 99258, USA

Protein–ligand structures are the core data required for structure-based drug design (SBDD). Understanding the error present in this data is essential to the successful development of SBDD tools. Methods for assessing protein–ligand structure quality and a new set of identification criteria are presented here. When these criteria were applied to a set of 728 structures previously used to validate molecular docking software, only 17% were found to be acceptable. Structures were re-refined to maintain internal consistency in the comparison and assessment of the quality criteria. This process resulted in Iridium, a highly trustworthy protein–ligand structure database to be used for development and validation of structure-based design tools for drug discovery.

Gregory L. Warren received his bachelor's degree in chemistry and biology from Walla Walla University (1986) and his PhD in biochemistry from Massachusetts Institute of Technology (1994). He worked as a postdoctoral fellow in the laboratory of Axel Brunger as part of the development team for the Crystallography & NMR System (CNS) refinement suite. Dr Warren worked for eight years as a molecular modeler at GlaxoSmithKline Pharmaceuticals before moving to OpenEye Scientific Software, Inc., in Santa Fe, NM, where his work includes structure-based design and X-ray crystallography applications.

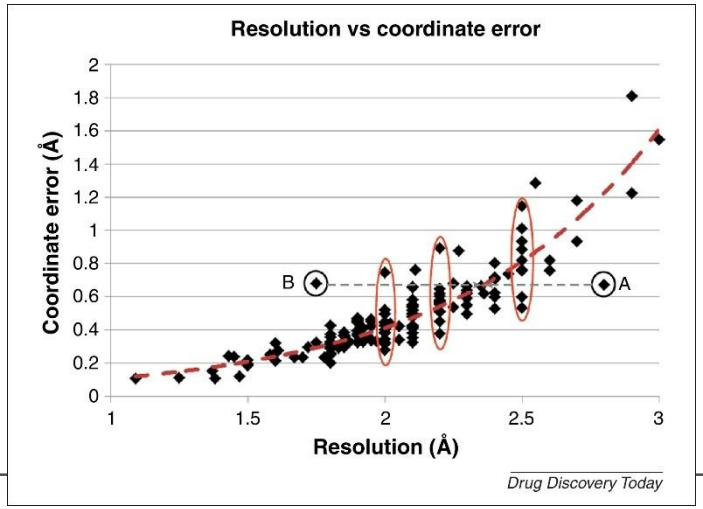
Anthony Nicholls received his PhD in biophysics from Florida State University in 1988. He worked as a post-doctoral fellow and later research associate with Barry Honig at Columbia University for seven years, developing the programs DelPhi and GRASP. In 1997 he founded OpenEye Scientific Software and remains the CEO and president of that company.

Warren recommendations for selecting high quality protein-ligand structures

- Check bond-order, tautomerization, ionization state, hydrogen positions to ensure consistency with binding site interactions
- Inspect electron density
  - Does the ligand have complete density, *e.g.* are there missing functional groups?
  - Are symmetry elements contacting ligand?
- Note difference between <u>resolution</u> (completeness of data) vs diffraction-component precision (Cruickshank) index or <u>DPI</u> (quality of data)

$$coordinate\ error = \frac{2.22 Rfree \sqrt{N_i}^3 \sqrt{V_a}}{n_{obs}^{5/6}}$$

$\mathbf{N_i}$ is number of heavy atoms w/ occupancy of 1, $\mathbf{V_a}$ is volume of asymmetric unit cell, $\mathbf{n_{obs}}$ is # of non-R_{free} reflections used during refinement

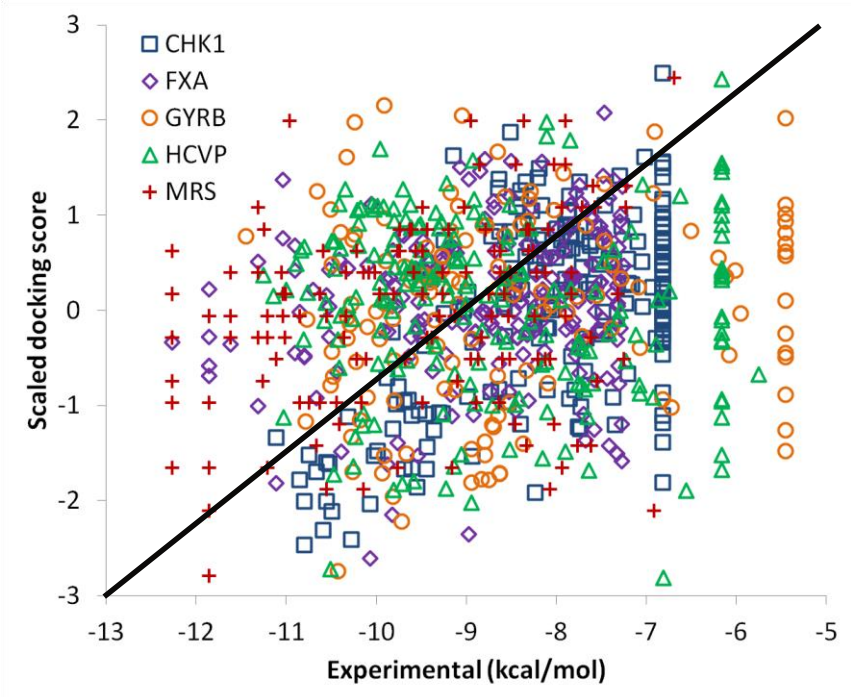


Warren, Do, Kelley, Nicholls, Warren, *Drug Disc Today*, **2012**, 17, 1270-1281

Now back to scoring

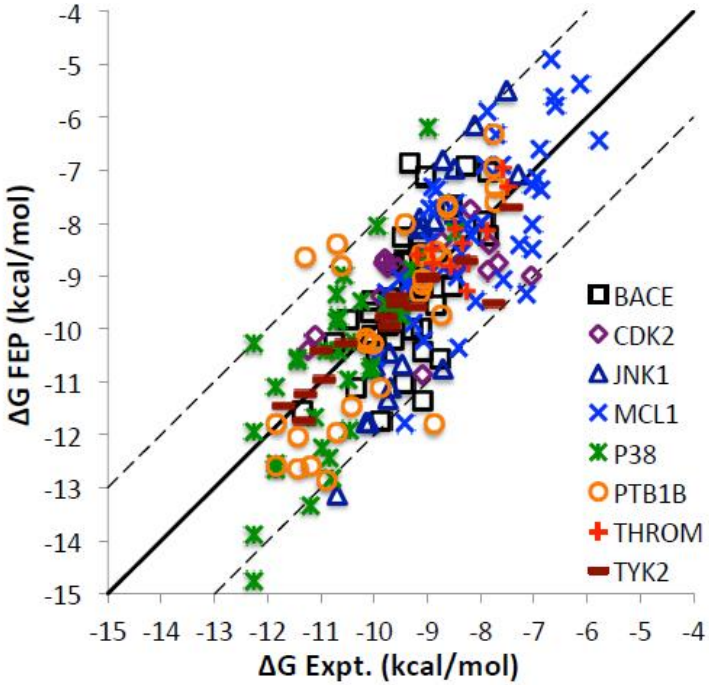
Circa 2015

Docking vs Experiment



Adapted from Warren (2006) *J. Med. Chem.*

FEP vs Experiment

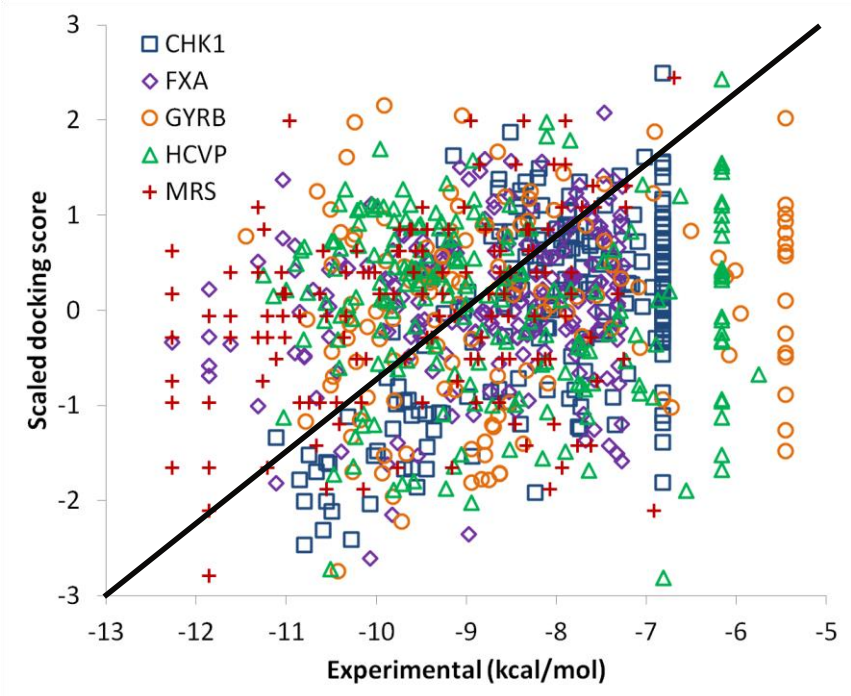


Wang (2015) *JACS*

Free energy perturbation (FEP) to the rescue?

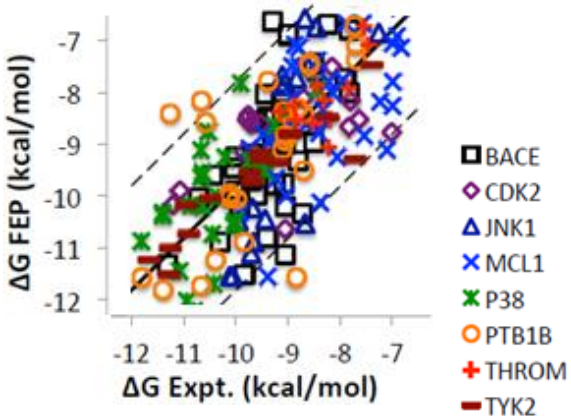
Showing signal but still ±2 kcal/mol*

Docking vs Experiment



Adapted from Warren (2006) *J. Med. Chem.*

FEP vs Experiment



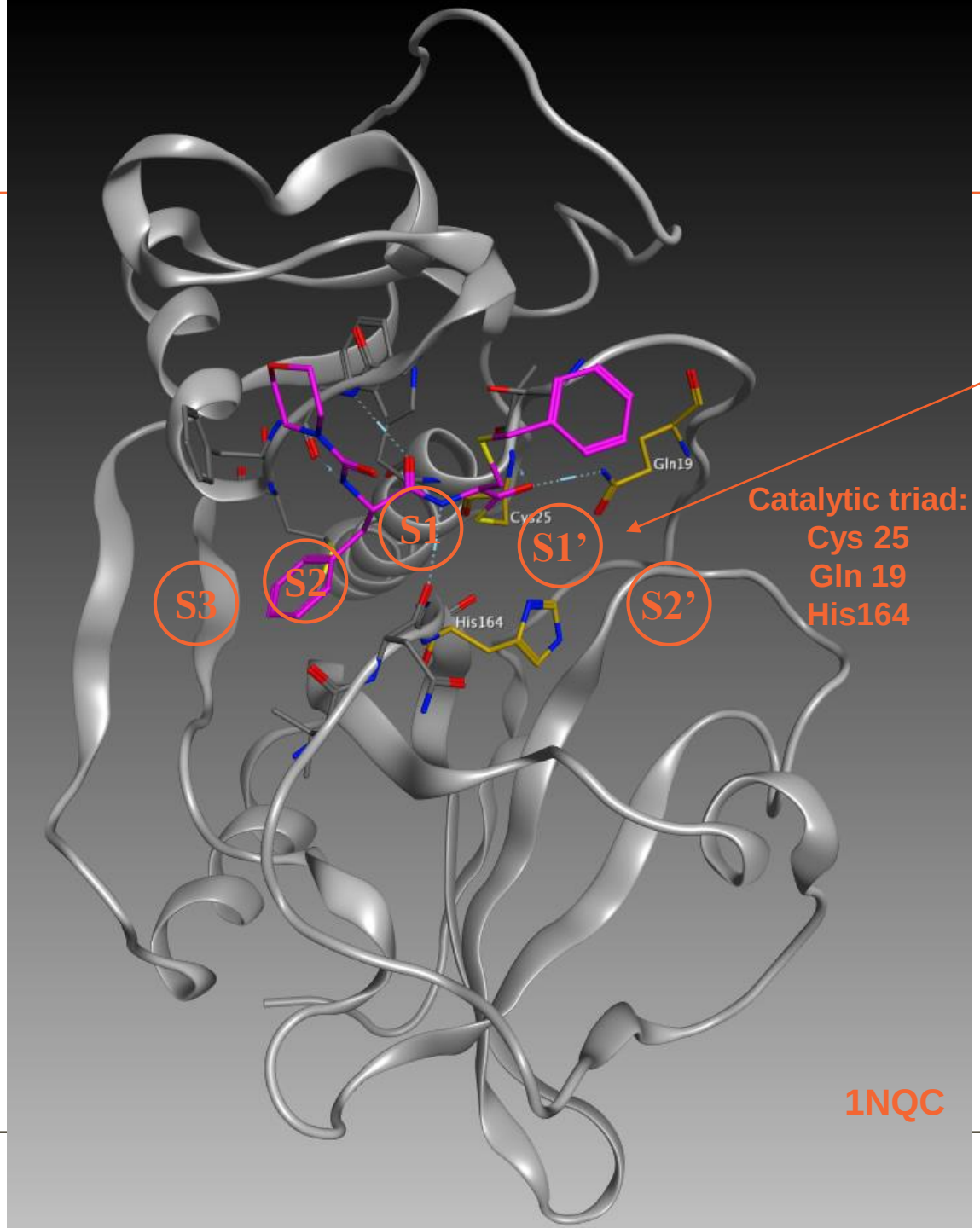
Adapted from Wang (2015) *JACS*
(cut two highest and one lowest affinity points)

*Doesn't look quite as good when looking at narrow range of data, e.g. pIC50 5-8

Cathepsin S

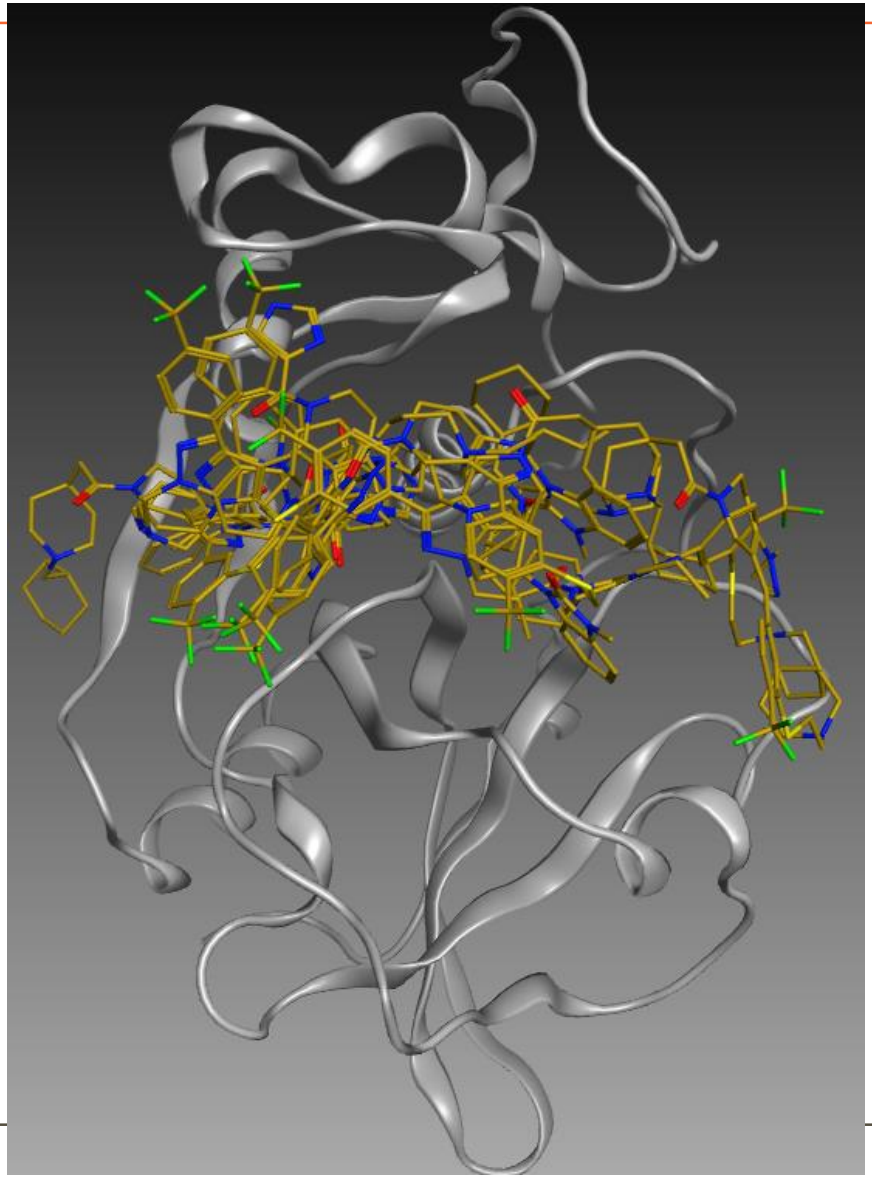
w/ covalent ligand

- Cysteine protease - catalytic activity occurs at conserved cysteine residue
- Active site nomenclature: S1, S2, S3… and S1', S2'…
  - each sub-site occupied by amino acid residue of peptide to be degraded
  - catalytic site at S1



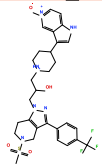
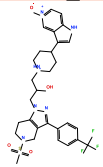
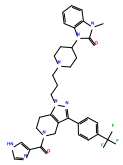
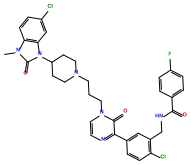
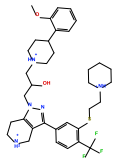
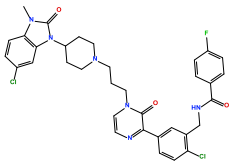
Peptide binding site

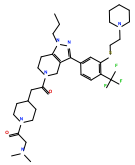
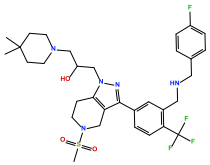
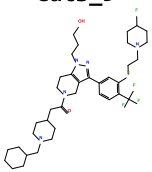
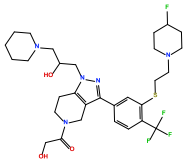
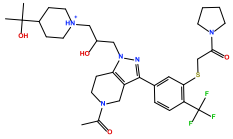
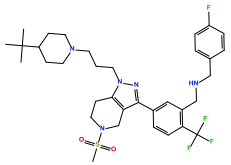
Docking large flexible ligands into flat, (almost) featureless protein binding sites

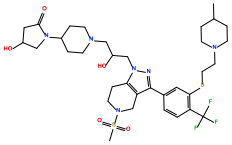
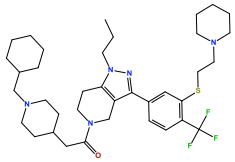
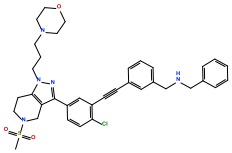
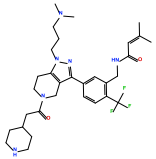
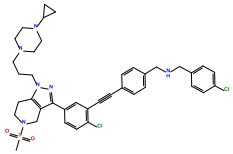
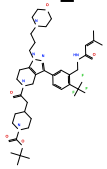


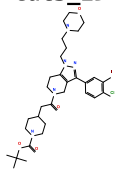
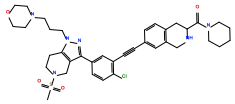
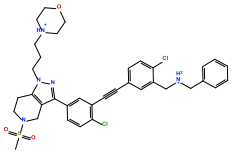
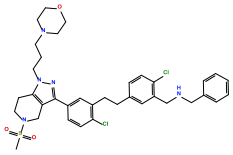
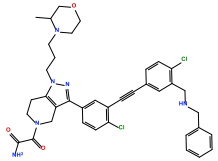
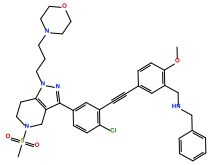
Closer look at Cat S pose challenge ligand set

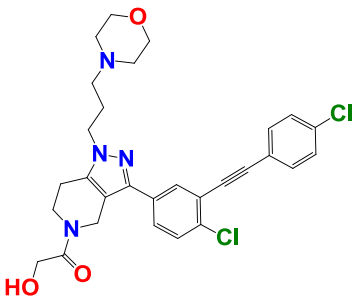
MW: 589 – 743; Heavy Atom Count: 42-52; # Rotatable bonds: 7-13

CatS_1	CatS_2	CatS_3	CatS_4	CatS_5	CatS_6
					
Sim: 0.72 MW: 620	Sim: 0.72 MW: 620	Sim: 0.77 MW: 633	Sim: 0.58 MW: 664	Sim: 0.72 MW: 661	Sim: 0.58 MW: 664
RotB: 7 HAC: 43	RotB: 7 HAC: 43	RotB: 8 HAC: 46	RotB: 8 HAC: 46	RotB: 11 HAC: 46	RotB: 8 HAC: 46

CatS_7	CatS_8	CatS_9	CatS_10	CatS_11	CatS_12
					
Sim: 0.75 MW: 663	Sim: 0.75 MW: 652	Sim: 0.74 MW: 708	Sim: 0.78 MW: 628	Sim: 0.76 MW: 653	Sim: 0.75 MW: 664
RotB: 13 HAC: 46	RotB: 10 HAC: 45	RotB: 13 HAC: 49	RotB: 11 HAC: 43	RotB: 11 HAC: 45	RotB: 10 HAC: 46

CatS_13	CatS_14	CatS_15	CatS_16	CatS_17	CatS_18
					
Sim: 0.73 MW: 743	Sim: 0.75 MW: 674	Sim: 0.85 MW: 658	Sim: 0.80 MW: 589	Sim: 0.83 MW: 732	Sim: 0.80 MW: 731
RotB: 11 HAC: 50	RotB: 12 HAC: 47	RotB: 12 HAC: 46	RotB: 11 HAC: 42	RotB: 13 HAC: 50	RotB: 12 HAC: 52

CatS_19	CatS_20	CatS_21	CatS_22	CatS_23	CatS_24
					
Sim: 0.83 MW: 712	Sim: 0.81 MW: 705	Sim: 0.84 MW: 695	Sim: 0.79 MW: 697	Sim: 0.92 MW: 700	Sim: 0.82 MW: 688
RotB: 9 HAC: 42	RotB: 10 HAC: 49	RotB: 12 HAC: 47	RotB: 13 HAC: 47	RotB: 13 HAC: 49	RotB: 13 HAC: 48

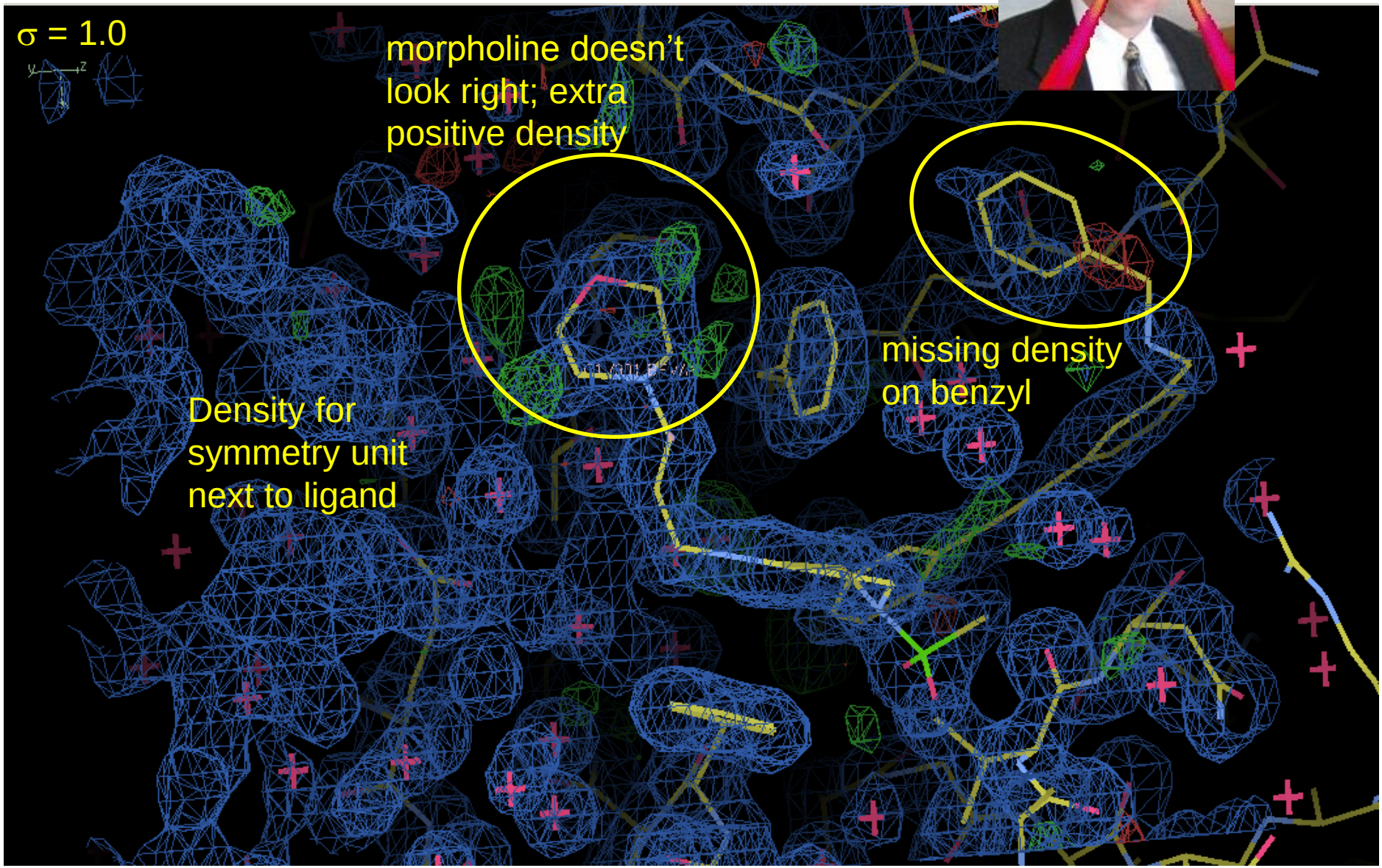


3IEJ*
Sim†: 1.00
RB: 9
MW: 553
HAC: 38

*Deposited 2009
†Sim = Tanimoto similarity

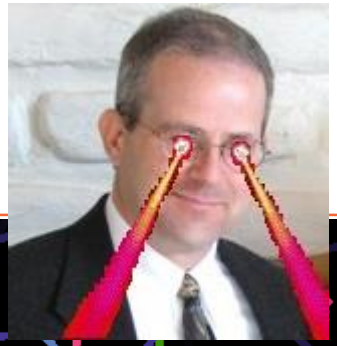
1: Missing density on ligand

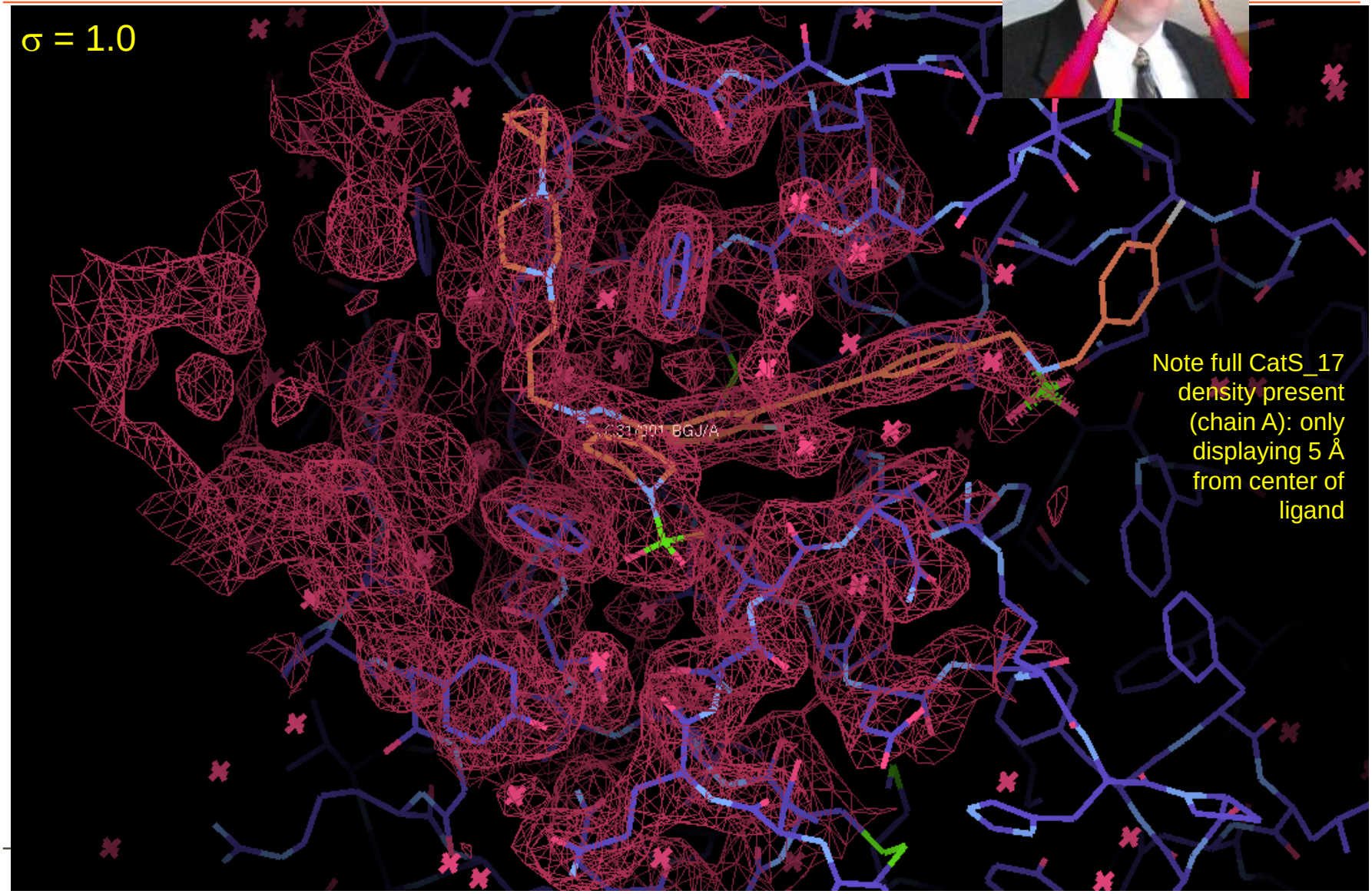
5QC8 chain A (ligand ID CatS_15)



2: Symmetry elements interaction with ligand

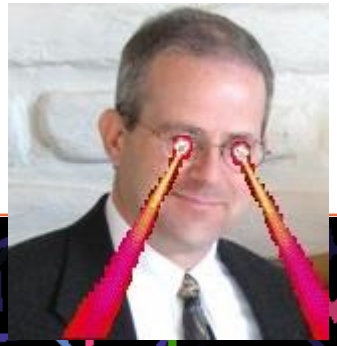
5QCA chain A: density near ligand CatS_17

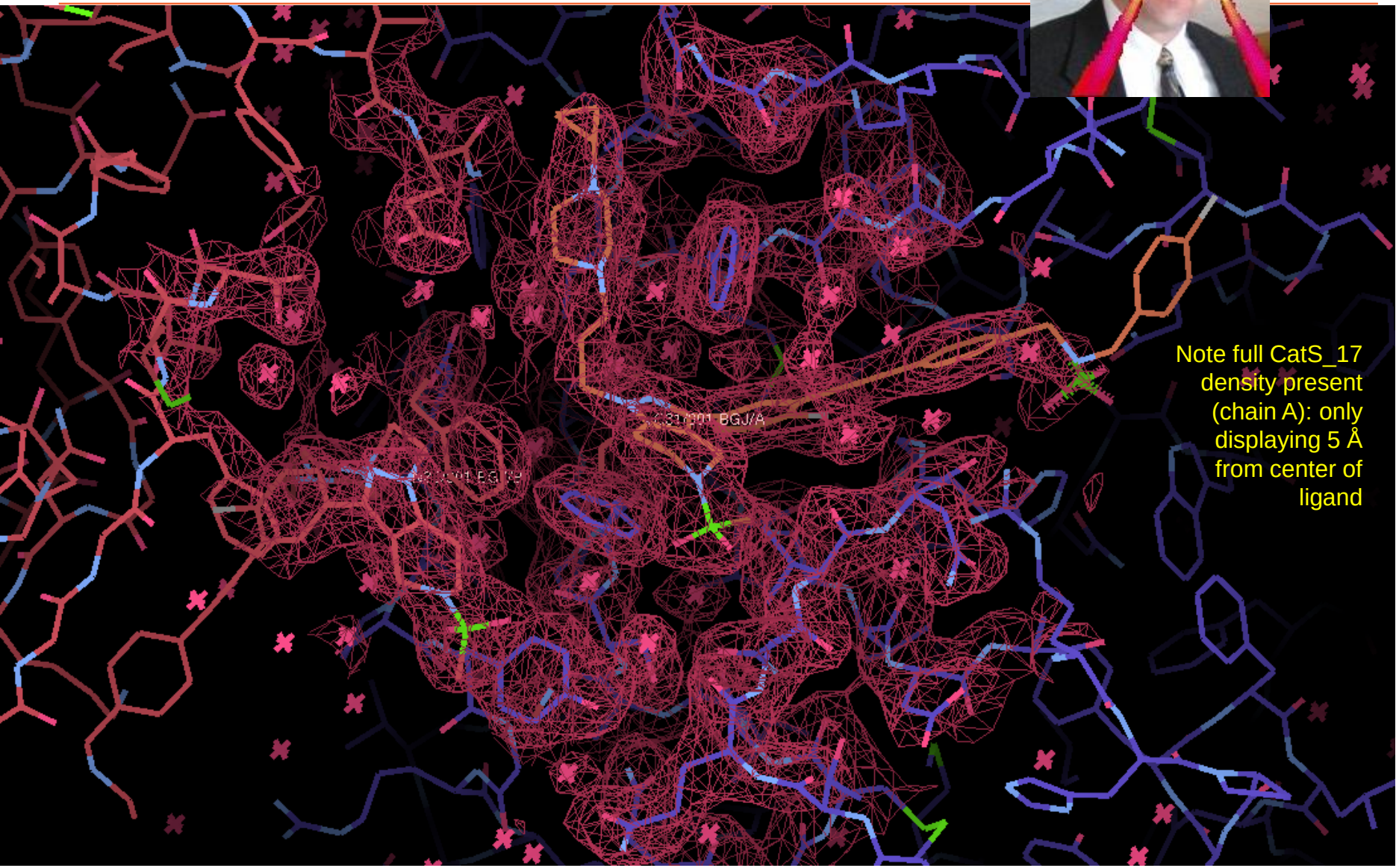




2: Symmetry elements interaction with ligand

5QCA chain A: density near ligand due to symmetry element





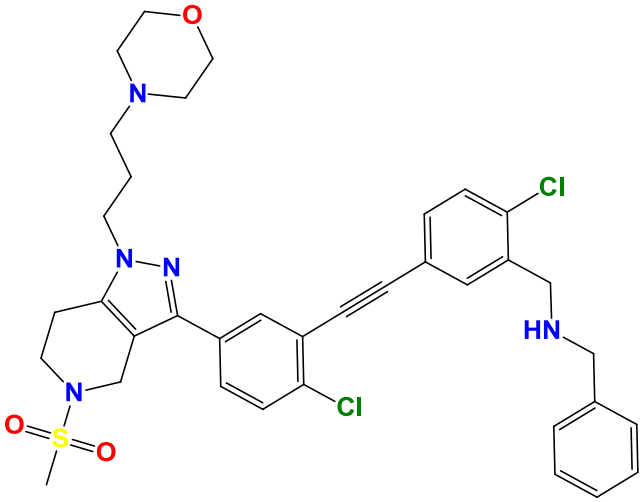
Note full CatS_17 density present (chain A): only displaying 5 Å from center of ligand

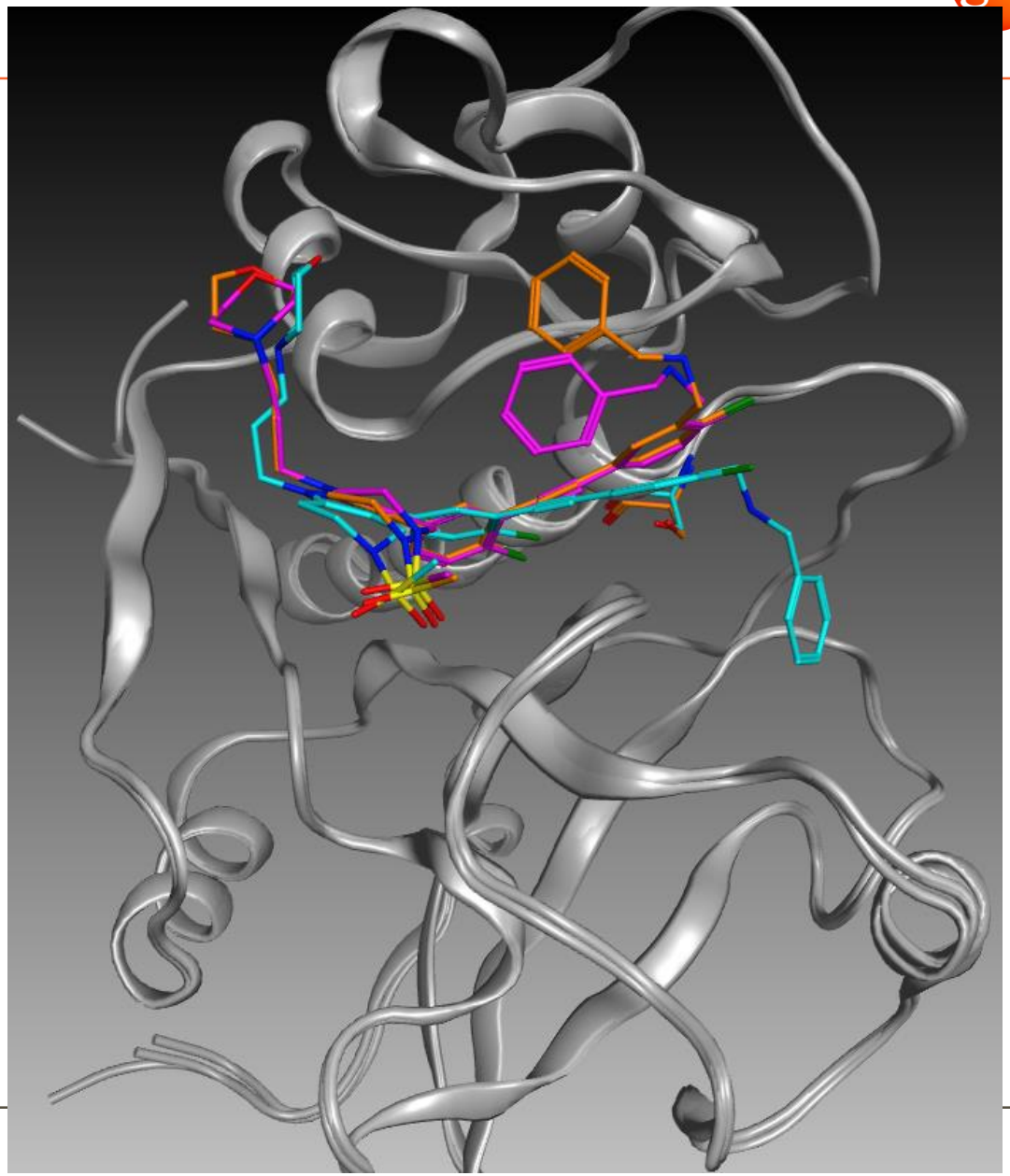
3: Multiple crystallographic ligand conformers for ID CatS_21

5QCE chain A

5QCE chain B

5QCF chain A





D3R Grand Challenge 3

Our Observations and Comments

Cathepsin S pose predictions
Kinase energy predictions

Subchallenges in Grand Challenge 3

challenge	target	ic50s	xrays	type	submissions	ic50-range nM
968-1	CatS	0	24	pose	52	
968-2	CatS	136	0	ligand-based	11	3.0-8520
968-3	CatS	136	0	structure-based	43	3.0-8520
968-4	CatS	33	0	free energy	35	
972-1	CatS	0	24	pose	47	
1009-2	CatS	136	0	ligand-based	9	3.0-8520
1009-3	CatS	136	0	structure-based	72	3.0-8520
1009-4	CatS	33	0	free energy	34	
969-2	JAK2-SC2	89	0	ligand-based	3	0.66-10000
969-3	JAK2-SC2	89	0	structure-based	28	0.66-10000
965-2	p38a	72	0	ligand-based	3	0.23-10000
965-3	p38a	72	0	structure-based	26	0.23-10000
966-2	VEGFR	85	0	ligand-based	3	0.62-10000
966-3	VEGFR	85	0	structure-based	31	0.62-10000
970-3	JAK2-SC3	17	0	structure-based	18	53.0-10000
970-4	JAK2-SC3	17	0	free energy	7	53.0-10000
967-3	TIE2	18	0	structure-based	18	3.4-10000
967-4	TIE2	4	0	free energy1	7	200.0-10000
967-5	TIE2	6	0	free energy2	7	3.4- 3200
971-3	ABL1	12	0	mutagenesis	11	49.0-10000
		1176	**24**		**465**	

Grand Challenge 3

1. Cathepsin S docking and scoring
   - Stage 1a: predict crystallographic poses of 24 ligands & predict affinities/rankings for 136 ligands and/or absolute or relative binding affinities for the designated free energy subset of 33 compounds
   - Stage 1b: repeat pose prediction exercise from stage 1a as a self-docking challenge
   - Stage 2: repeat affinity prediction exercise from stage 1a with crystal structures now available.
2. Kinase Selectivity: JAK2, p38a, VEGFR2
   - Predict affinities/rankings for set of ligands given FASTA sequences of targets and SMILES of ligands
3. Kinase Activity Cliff: JAK2
   - An affinity ranking/scoring & free energy challenge designed to test ability of current methods to detect large changes in affinity due to small changes in chemical structure.
   - Dataset comprises 17 congeneric compounds with Kd values for the kinase JAK2
4. Kinase Activity Cliff: TIE2
   - Predict affinities, or affinity rankings, for 18 ligands and/or predict the absolute or relative binding affinities for two designated free energy subsets of 4 and 6 compounds.
5. ABL1 mutations: predict affinities/rankings for all mutants for each of two ligands
   - ABL1(F317I), ABL1(F317L), ABL1(H396P), ABL1(Q252H), and ABL1(T315I)

1. Janssen Pharmaceuticals

2-5. Structural Genomics Consortium group at the University of North Carolina at Chapel Hill (SGC-UNC)
Drewry DH, Wells CI, Andrews DM, et al (2017) Progress towards a public chemogenomic set for protein kinases and a call for contributions. PLOS ONE 12:e0181585

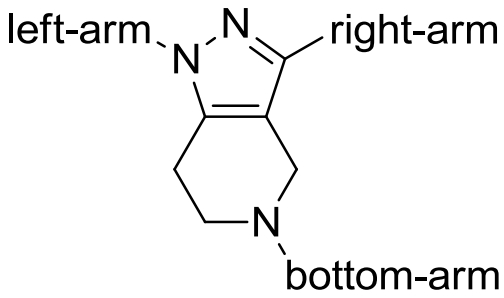
Cathepsin S (CatS) Stage 1a:

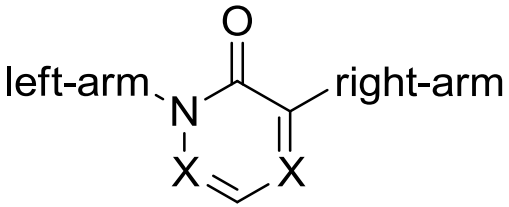
D3R provides 141 compounds (SMILES),
and coordinates for 2 crystal structures (minus inhibitor)

Participants dock the compounds and predict the affinities

D3R takes the submissions,
compares 24 docking predictions with crystal structures,
compares 136 affinity predictions with measured IC50s
compares 33 affinity predictions for free energy subset

Two chemical series for Cathepsin S

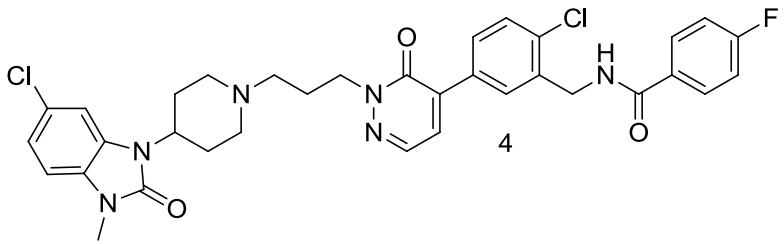




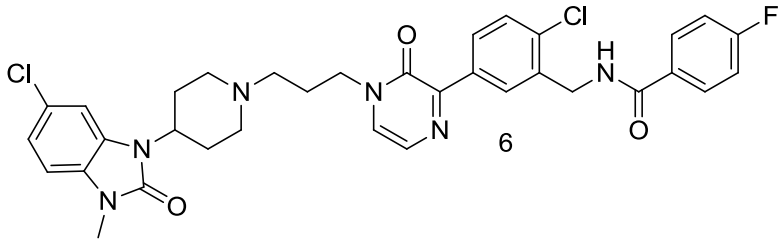
tetrahydropyrido-pyrazole
134 measured IC50s
22 crystal structures

pyridinone-like
2 measured IC50s
2 crystal structures

Only two pyridinones, and they are similar to each other

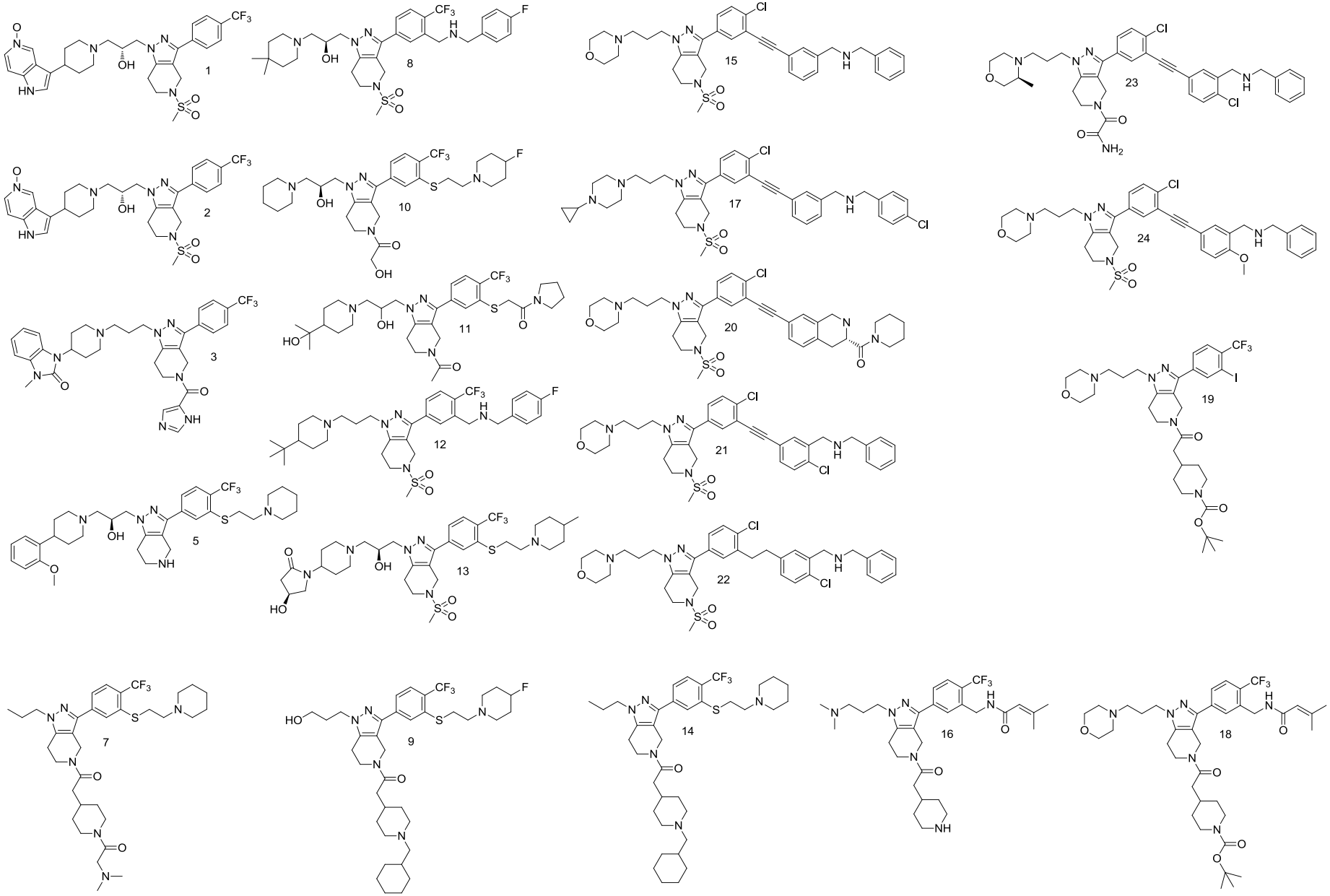


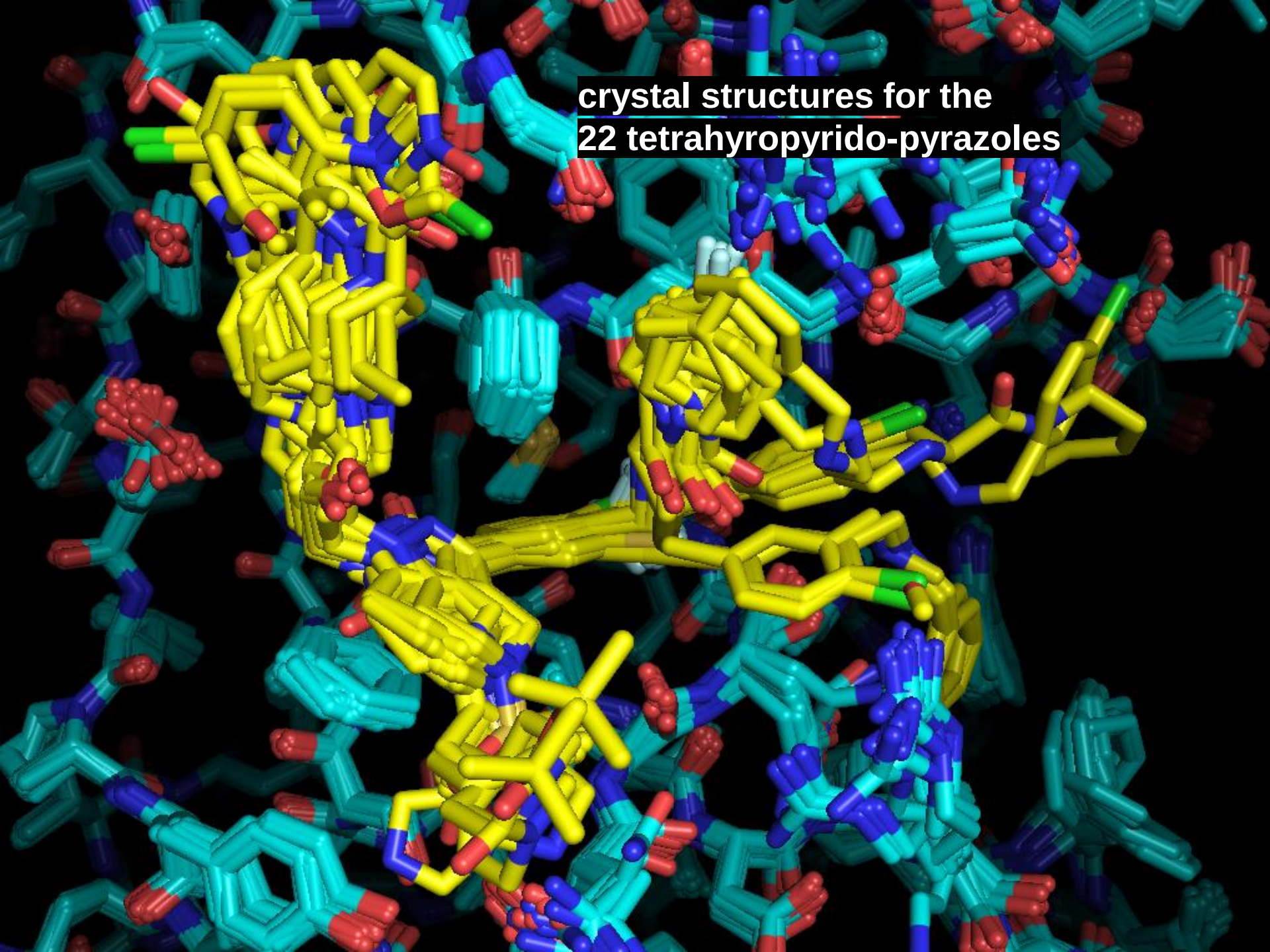
4

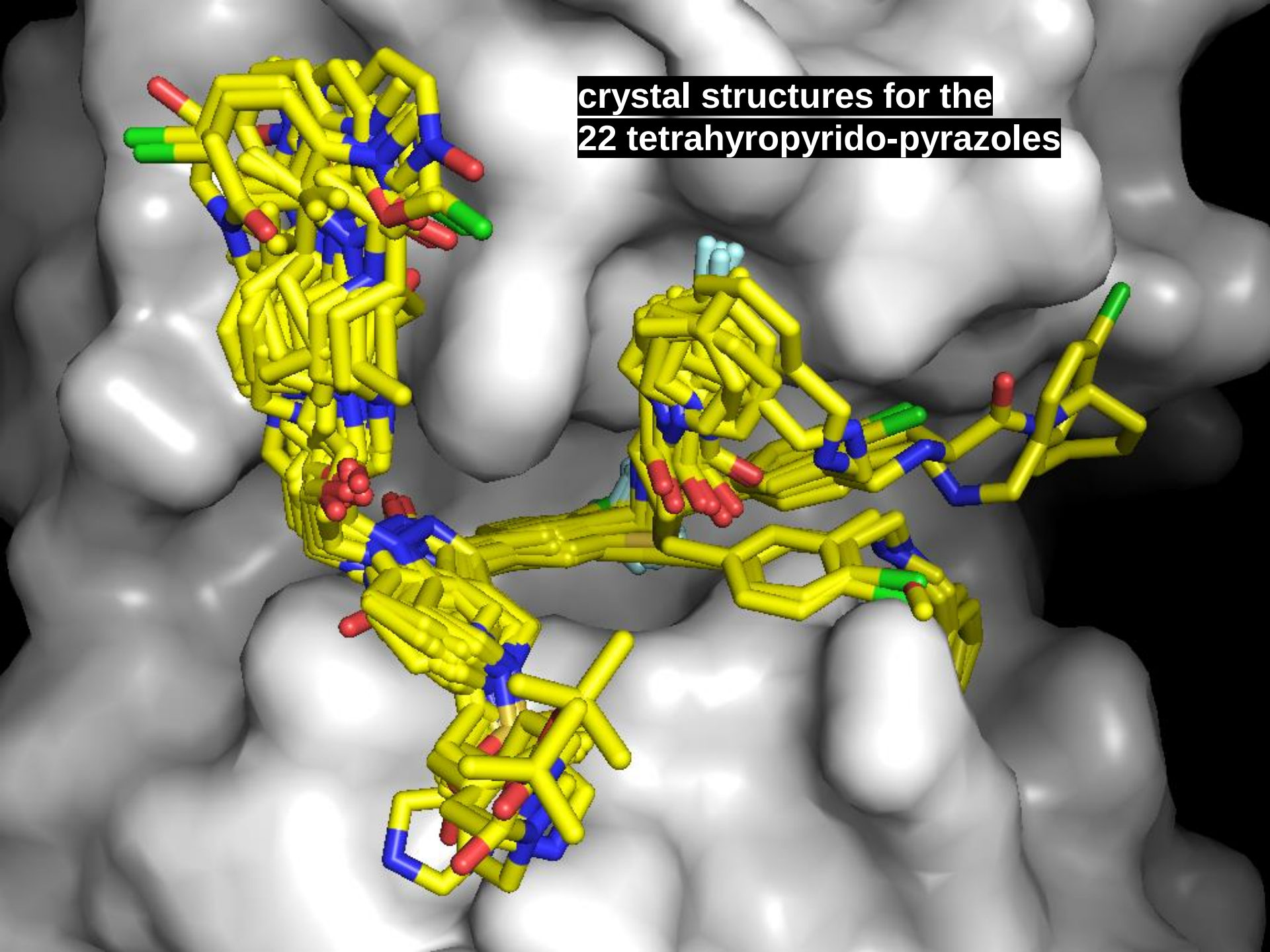


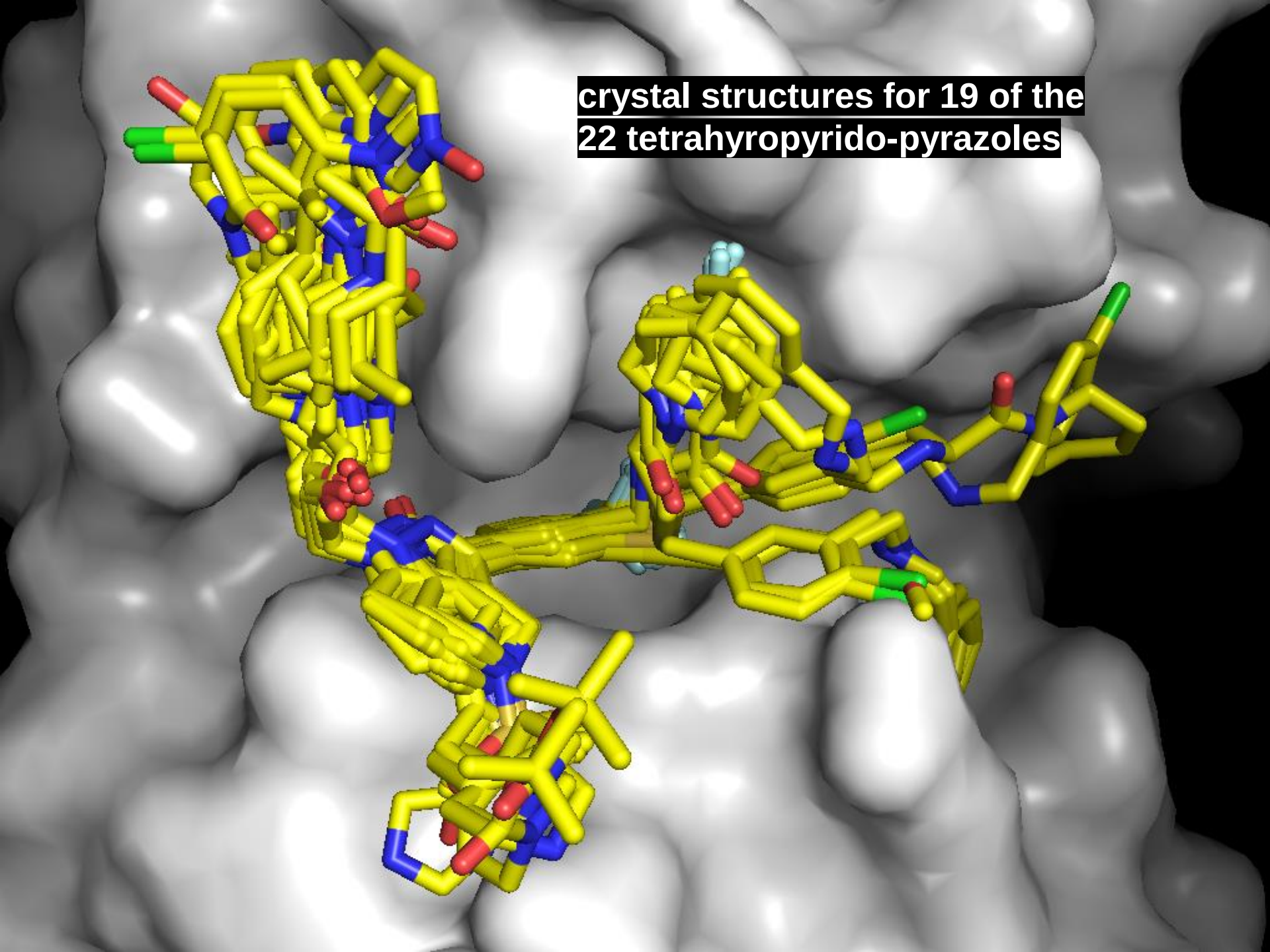
6

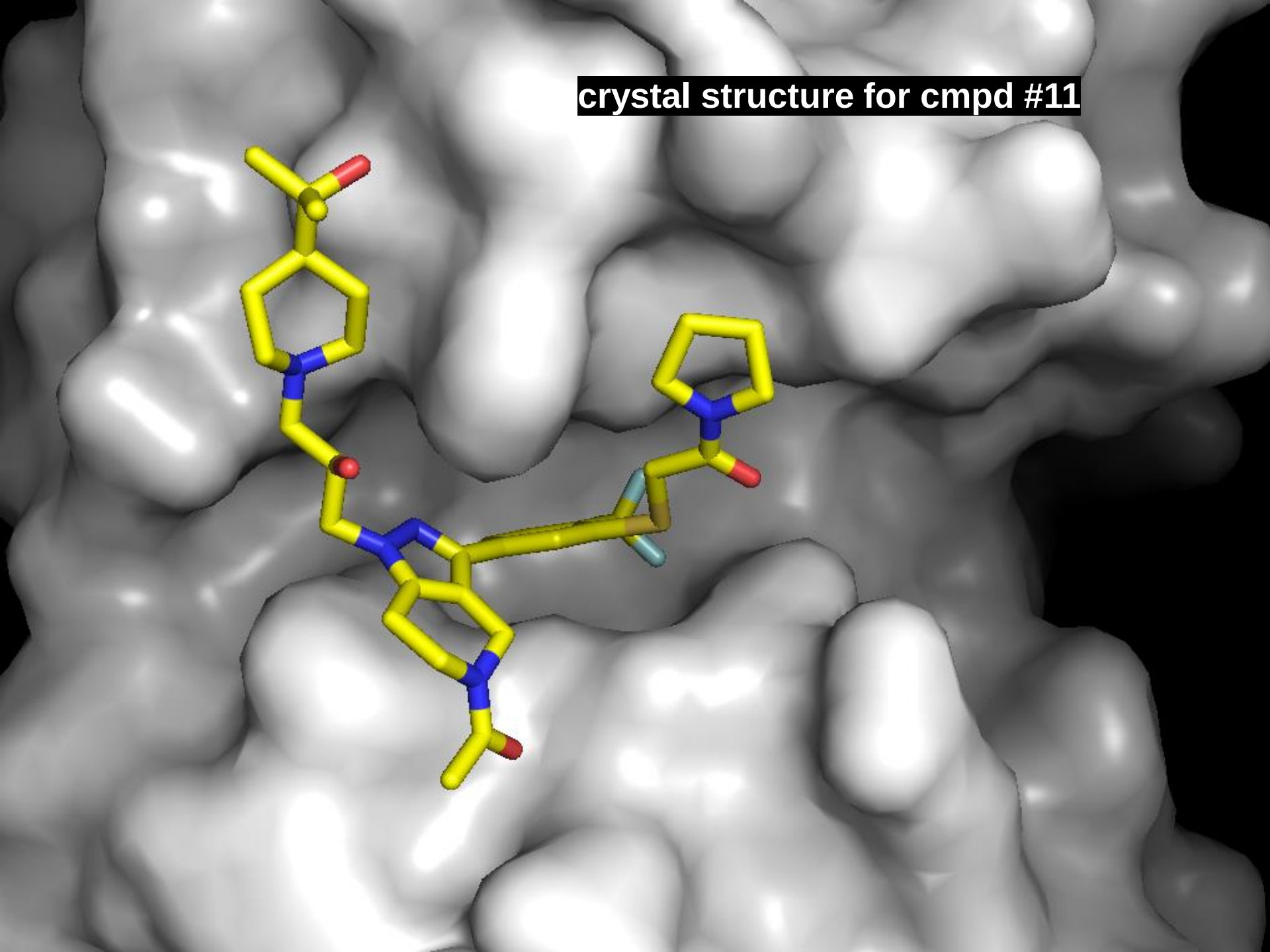
22 tetrahydropyrido-pyrazoles with newly refined crystal structures

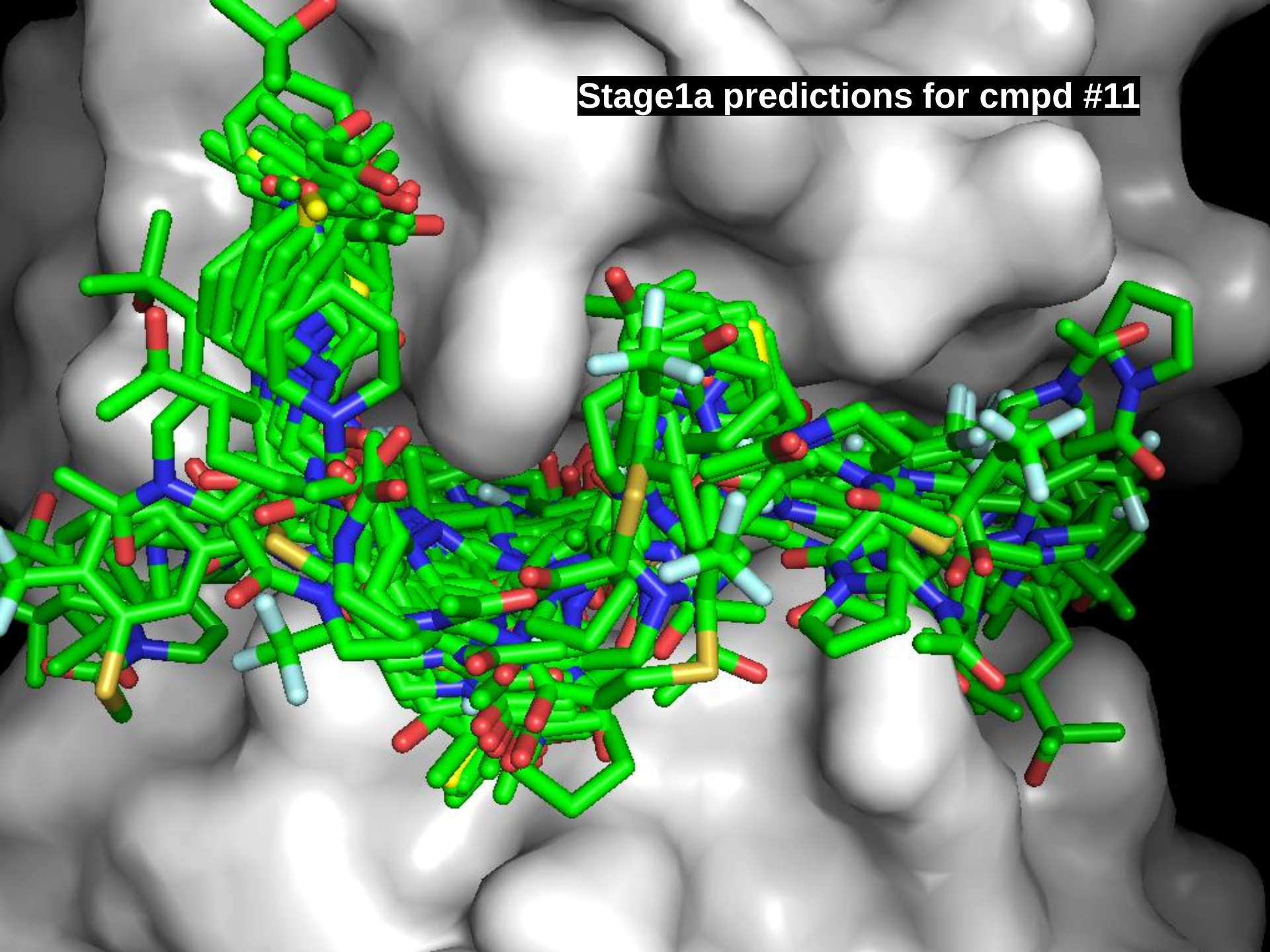



crystal structures for the
22 tetrahyropyrido-pyrazoles


crystal structures for the
22 tetrahyropyrido-pyrazoles


crystal structures for 19 of the
22 tetrahyropyrido-pyrazoles


crystal structure for cmpd #11


Stage1a predictions for cmpd #11

RMSD calculations

Thanks to Conor Parks and Pat Walters for RMSD calculations

Conor and Pat (separately) calculated the RMS-deviation of each pose from the X-ray.

1. Each used Maximum Common Subgraph algorithms, ignoring bond order.
2. Protein superimposed into standard orientation prior to RMSD calcn.
3. No adjustment for poor density, alternate conformations in X-rays.

Pose1 is the pose that the participant identified as “best”
Pose2 is the second best
Pose3 is third best, etc

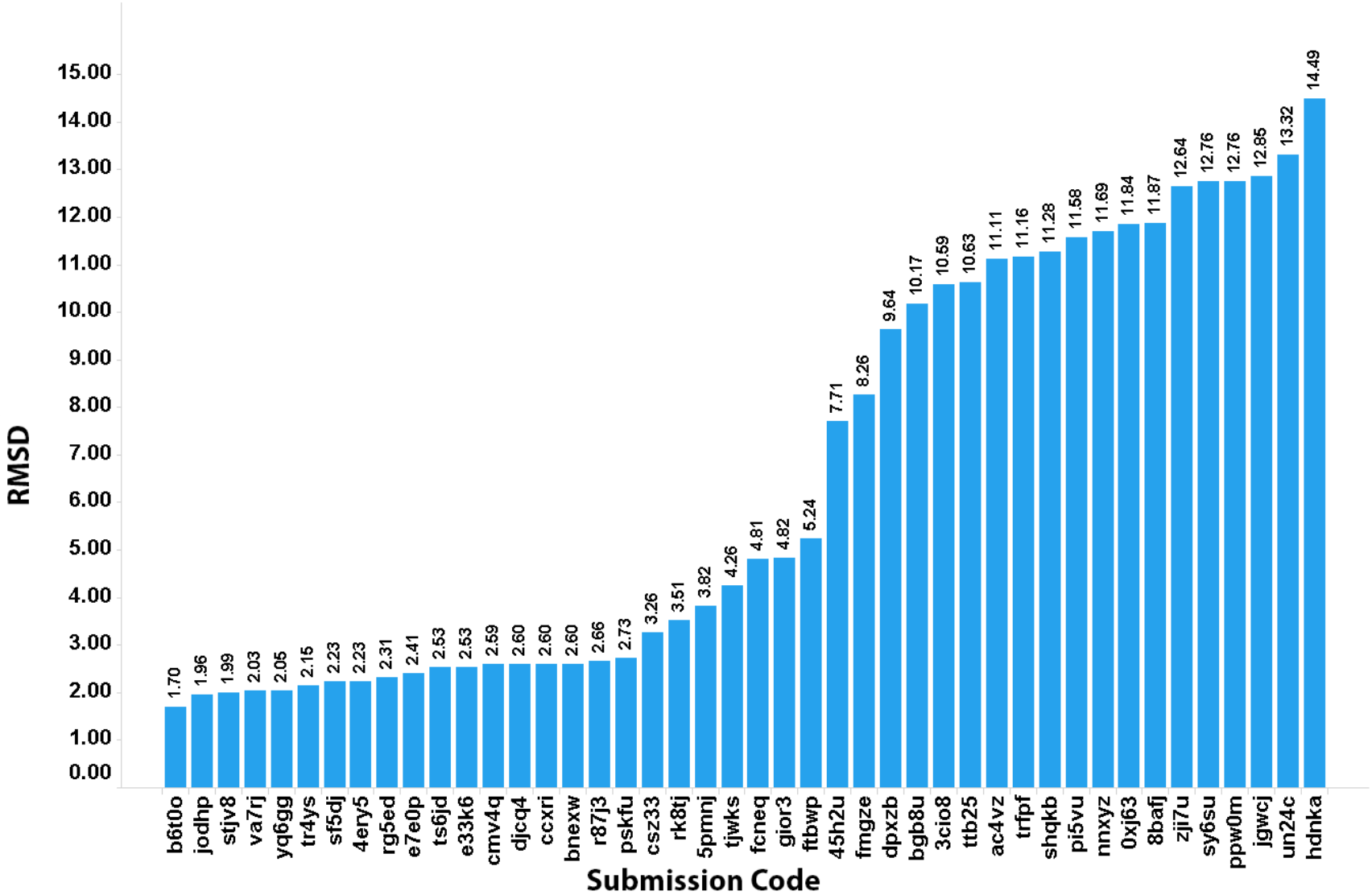
We are reporting RMS-deviations for Pose1 only.

Each submission will have 24 Pose1 RMSD values, for comparison of Pose1 from predictions for the 24 compounds with crystal structures.

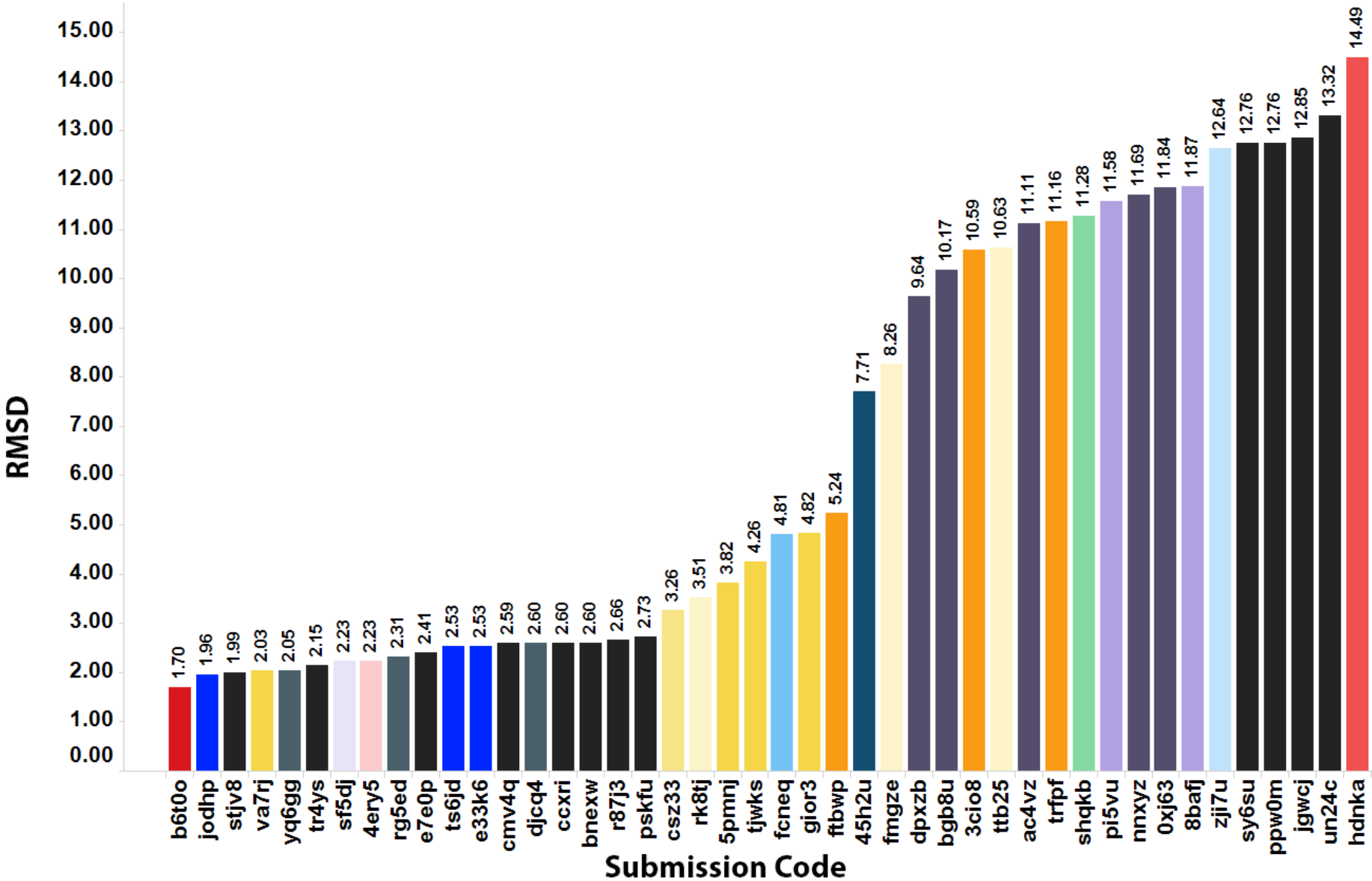
The “Median Pose1 RMSD” is the Median of these 24 values.

Median Pose 1 RMSD for CatS stage1a (median across 24 predicted structures)

Thanks to Conor Parks and Pat Walters for RMSD calculations



CatS Stage1a Pose RMSD, colored by submitting group



Affinity (potency) correlation coefficients

Thanks to Zied Gaieb and Pat Walters for Calculating Correlation Coefficients

Each submission predicts the affinity for 136 cmpds.

Zied and Pat calculated various correlation coefficients:

1. Pearson
2. Spearman rank-order correlation coefficient (non-parametric)
3. Kendall (non-parametric)

We are using the Spearman as the accuracy measure for affinity (potency).

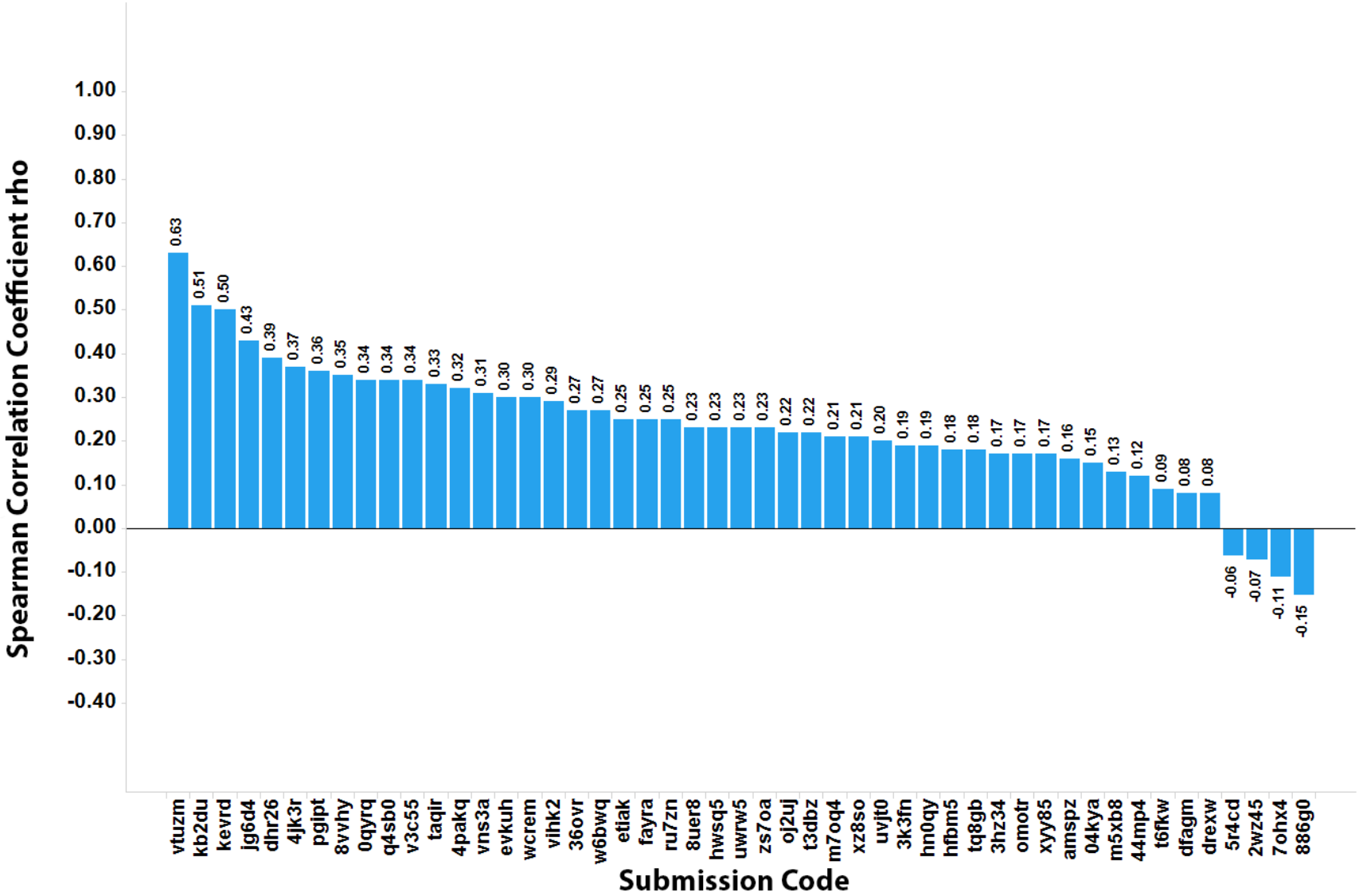
Affinity prediction, 136 compounds, structure based

Stage 1a 968-3: 43 submissions, ranging from good (ρ=0.63) to poor (ρ=-0.12)

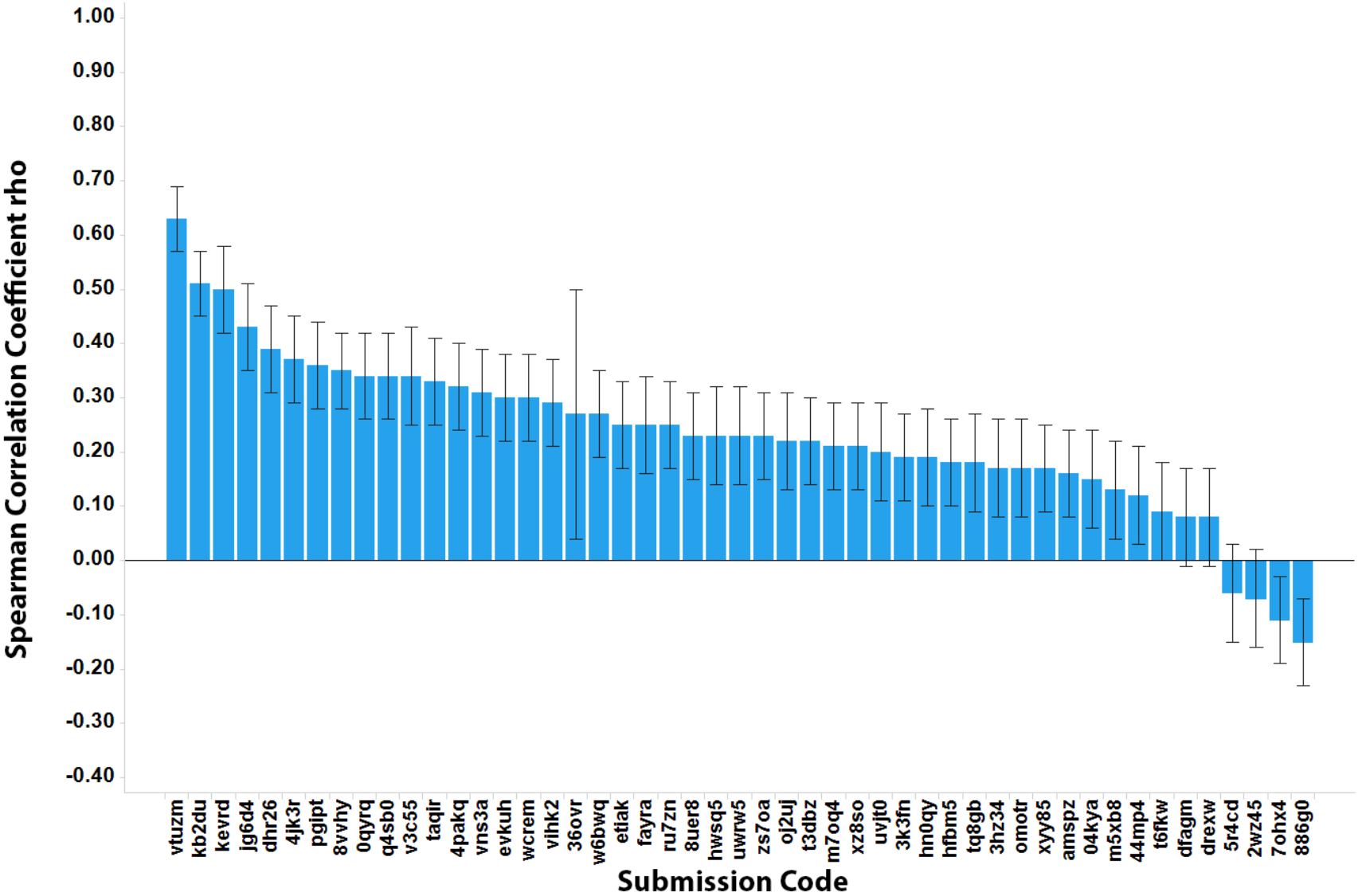
submissio	Num Inhibs	Kendalls Tau	Kendalls Tau Error	Spearman's Rho	Spearman's Rho Error
vtuzm	136	0.45	0.05	0.63	0.06
jg6d4	136	0.29	0.05	0.43	0.08
dhr26	136	0.28	0.06	0.39	0.08
4jk3r	136	0.25	0.06	0.37	0.08
w83jw	136	0.25	0.05	0.37	0.08
ppyff	136	0.26	0.06	0.37	0.08
pgipt	136	0.25	0.06	0.36	0.08
f3ifz	136	0.24	0.05	0.36	0.08
8vvhy	136	0.23	0.05	0.35	0.07
v3c55	136	0.24	0.06	0.34	0.09
q4sb0	136	0.23	0.05	0.34	0.08
0qyrq	136	0.24	0.06	0.34	0.08
taqir	136	0.22	0.05	0.33	0.08
4pakq	136	0.23	0.06	0.32	0.08
5b5wz	136	0.22	0.05	0.32	0.08
vns3a	136	0.21	0.06	0.31	0.08
wcrem	136	0.20	0.05	0.30	0.08
evkuh	136	0.21	0.06	0.30	0.08
vihk2	136	0.19	0.06	0.29	0.08
w6bwq	136	0.17	0.05	0.27	0.08
36ovr	19	0.18	0.16	0.27	0.23
ru7zn	136	0.16	0.05	0.25	0.08
etiak	136	0.17	0.06	0.25	0.08
fayra	136	0.17	0.06	0.25	0.09
zs7oa	136	0.16	0.06	0.23	0.08
uwrw5	136	0.16	0.06	0.23	0.09
8uer8	136	0.15	0.06	0.23	0.08
oj2uj	136	0.15	0.06	0.22	0.09
t3dbz	136	0.15	0.05	0.22	0.08
xz8so	136	0.13	0.05	0.21	0.08
m7oq4	136	0.15	0.06	0.21	0.08
uvjt0	136	0.14	0.06	0.20	0.09
hn0qy	136	0.13	0.06	0.19	0.09
3k3fn	136	0.13	0.05	0.19	0.08
hfbm5	136	0.10	0.05	0.18	0.08
tq8gb	136	0.12	0.06	0.18	0.09
omotr	136	0.11	0.06	0.17	0.09
3hz34	136	0.12	0.06	0.17	0.09
xyy85	136	0.12	0.06	0.17	0.08
04kya	136	0.11	0.06	0.15	0.09
44mp4	136	0.09	0.06	0.12	0.09
5r4cd	136	-0.04	0.06	-0.06	0.09
72yx7	135	-0.07	0.06	-0.12	0.09

CatS stage1a correlation with potency, str- and lig-based

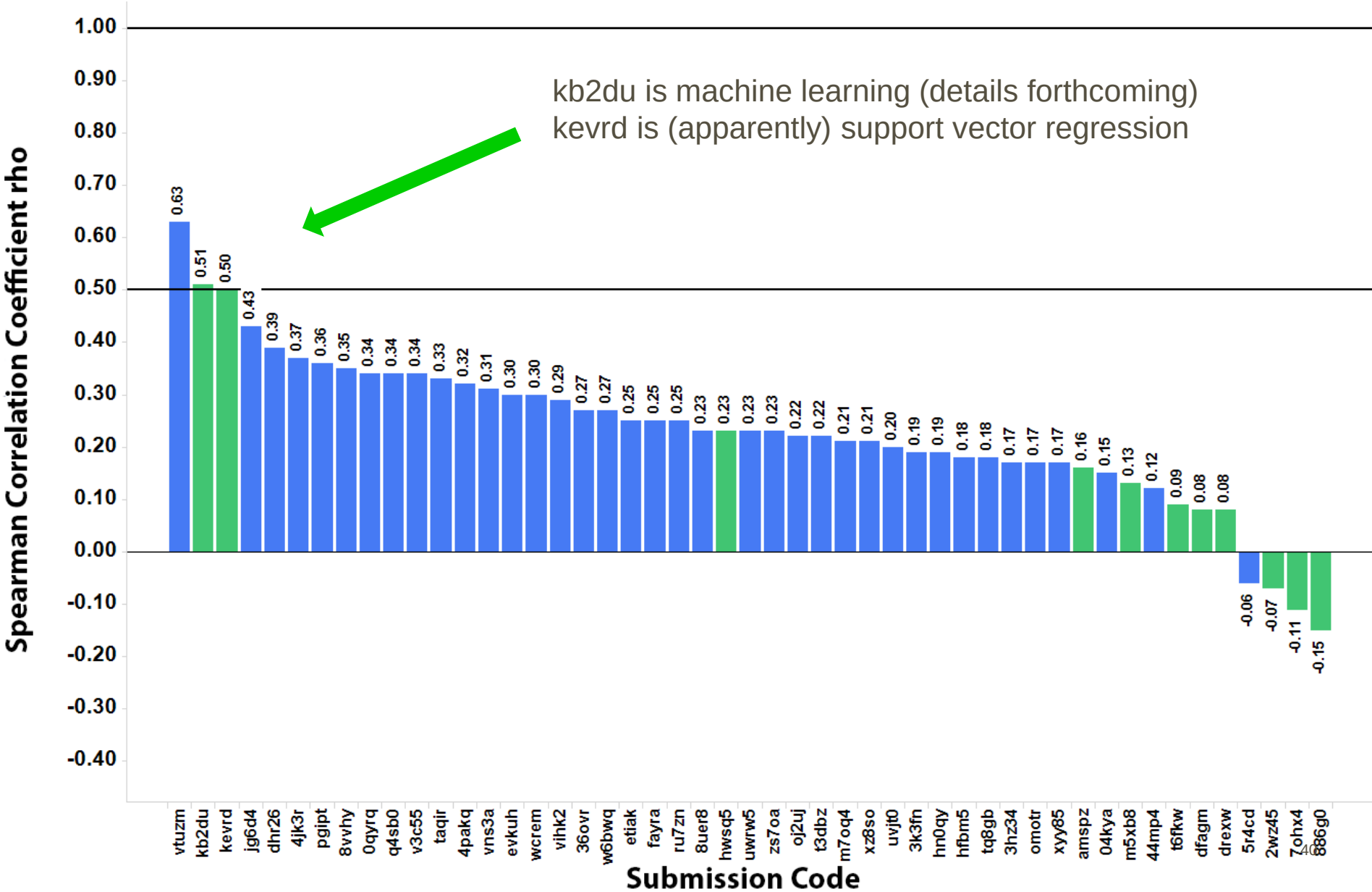
Thanks to Zied Gaieb and Pat Walters for Calculating Correlation Coefficients



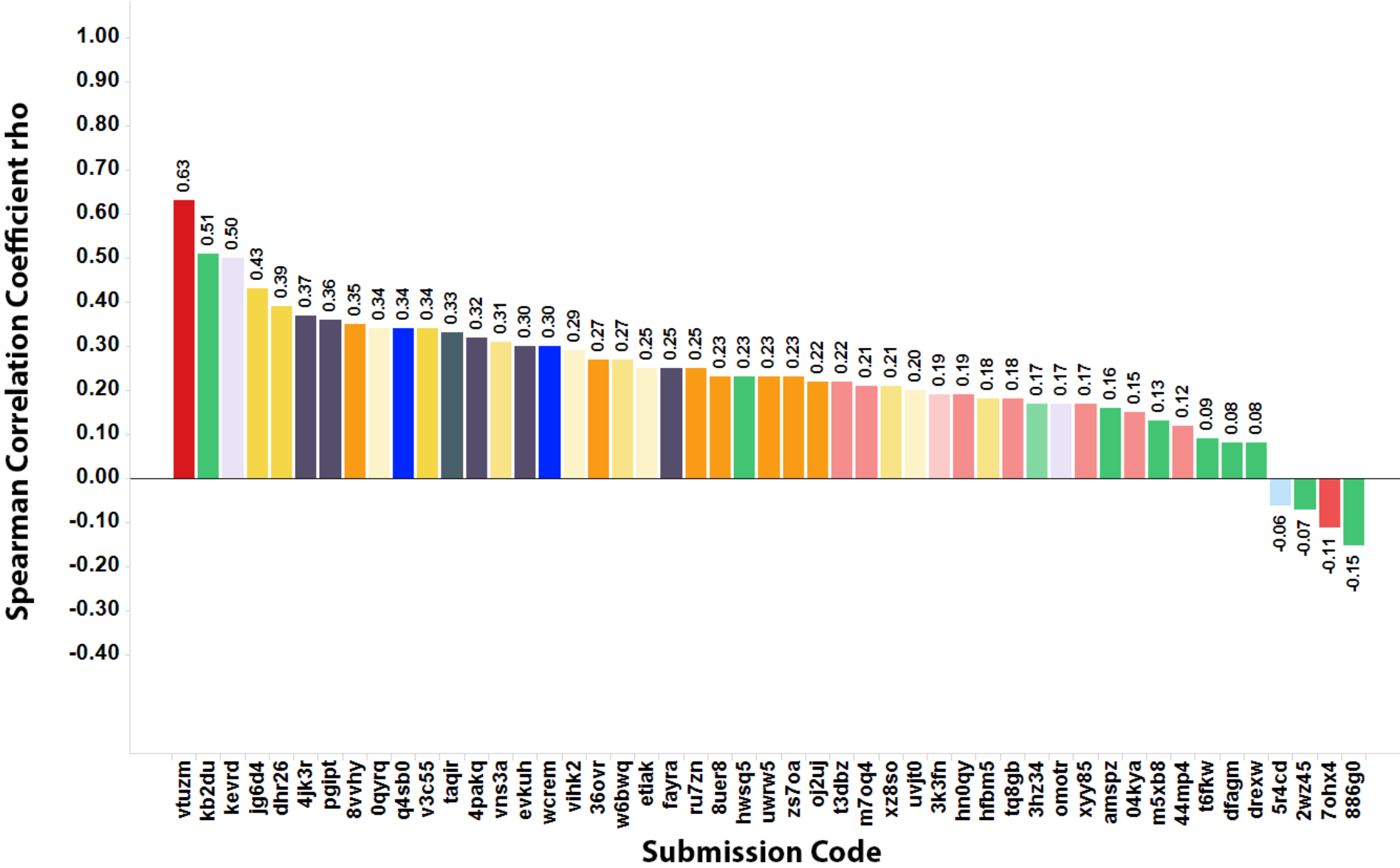
CatS stage1a correlations, showing error bars

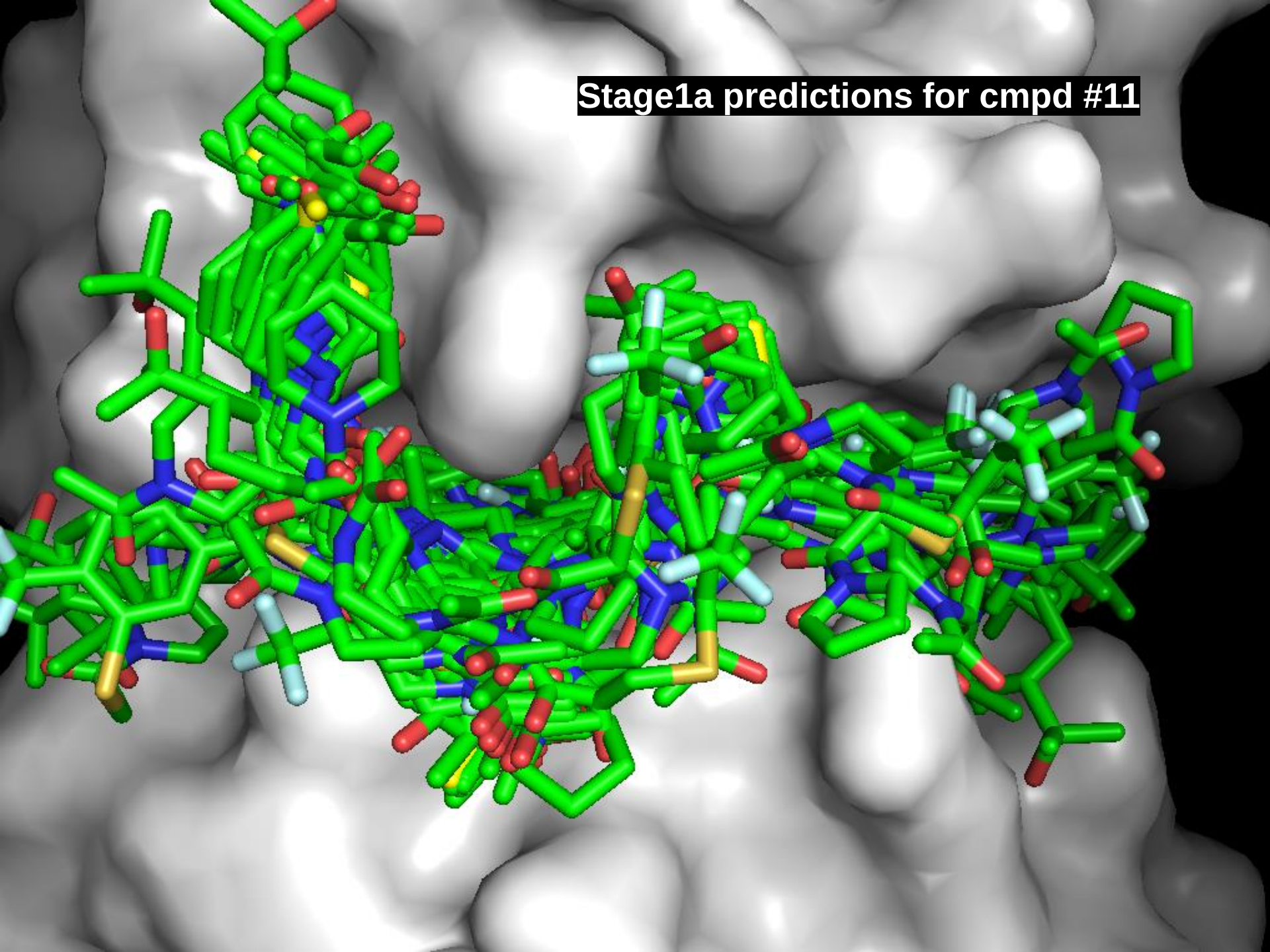


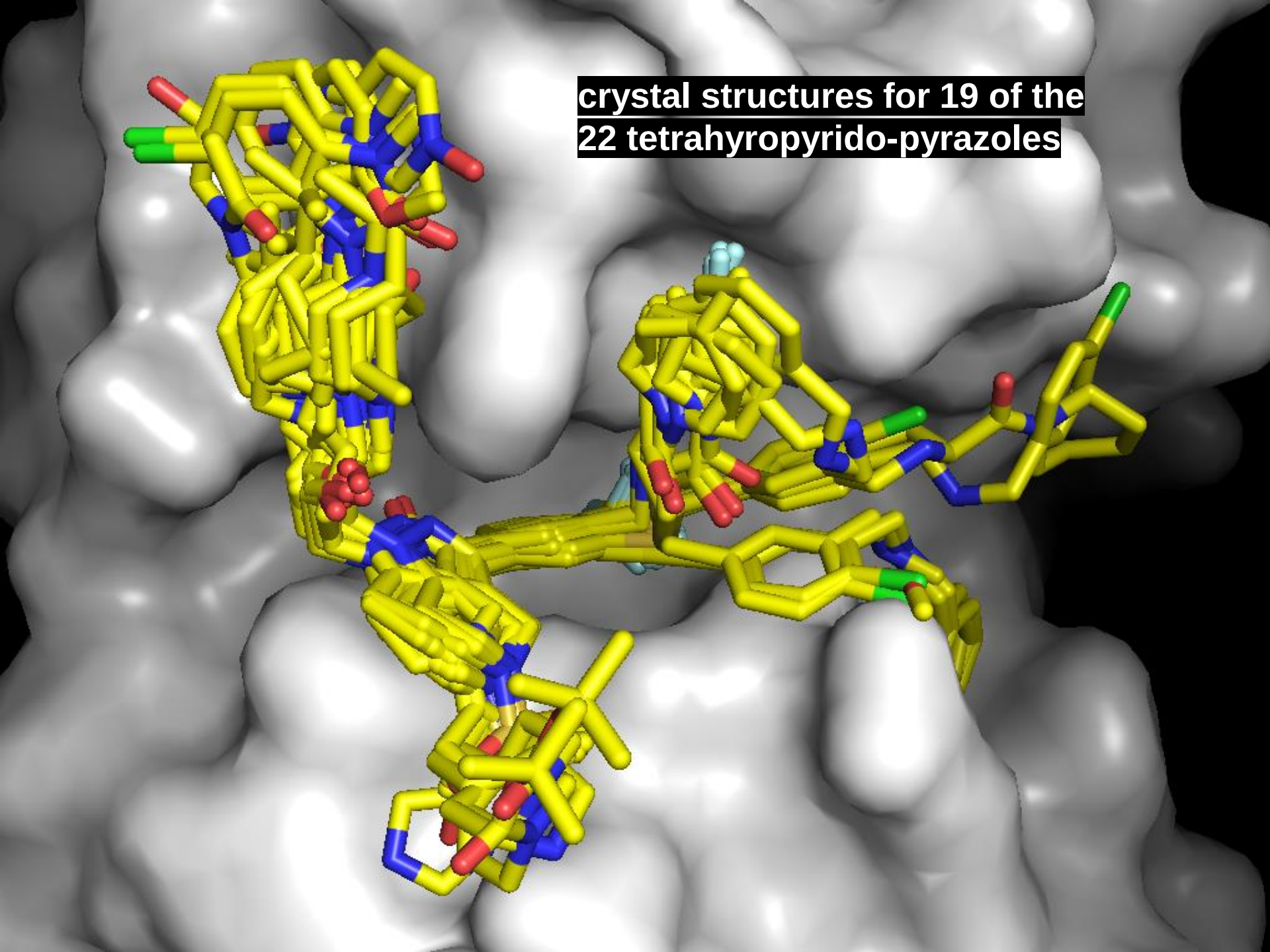
CatS stage1a, blue=structure-based, green=ligand-based

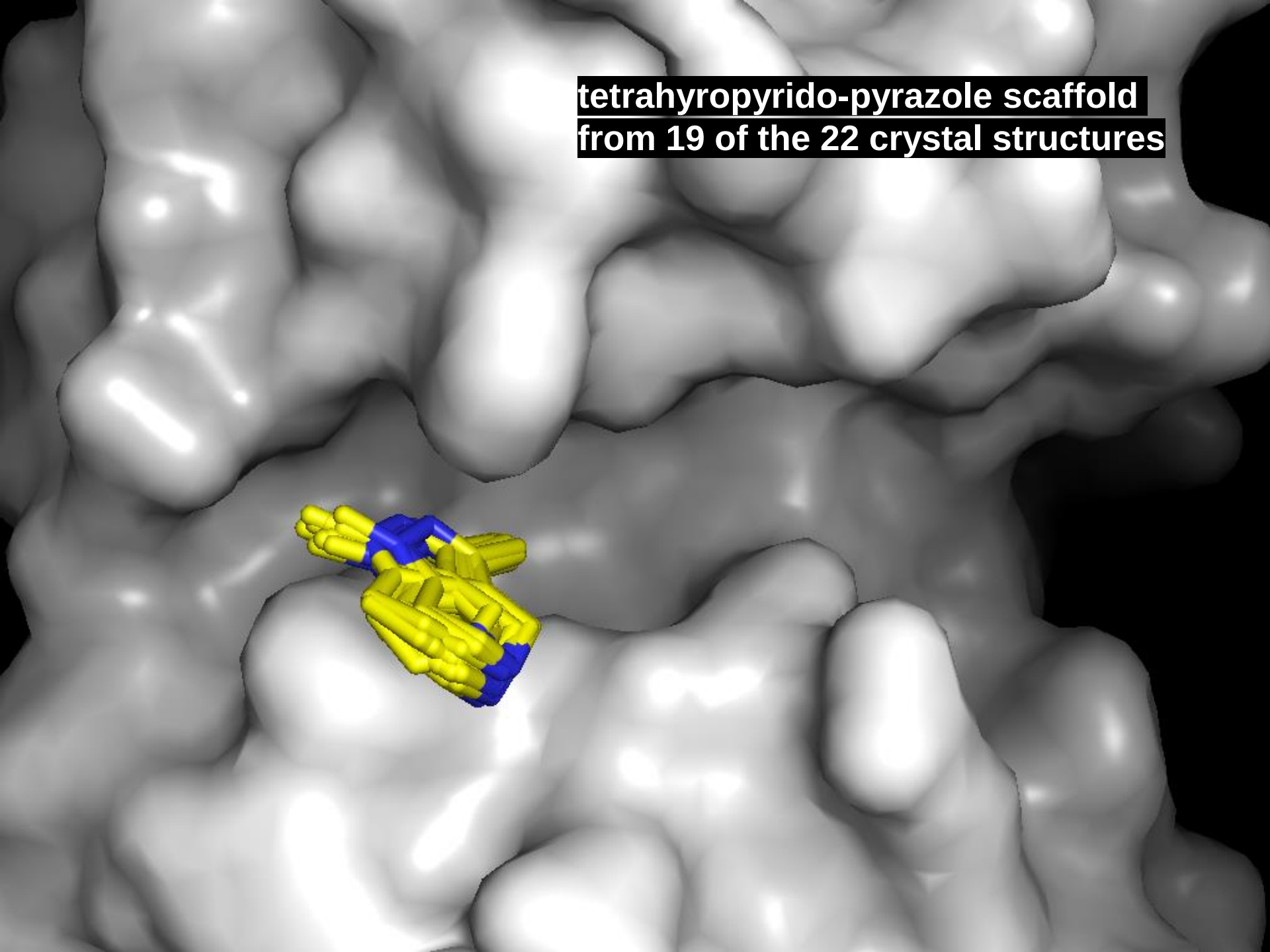


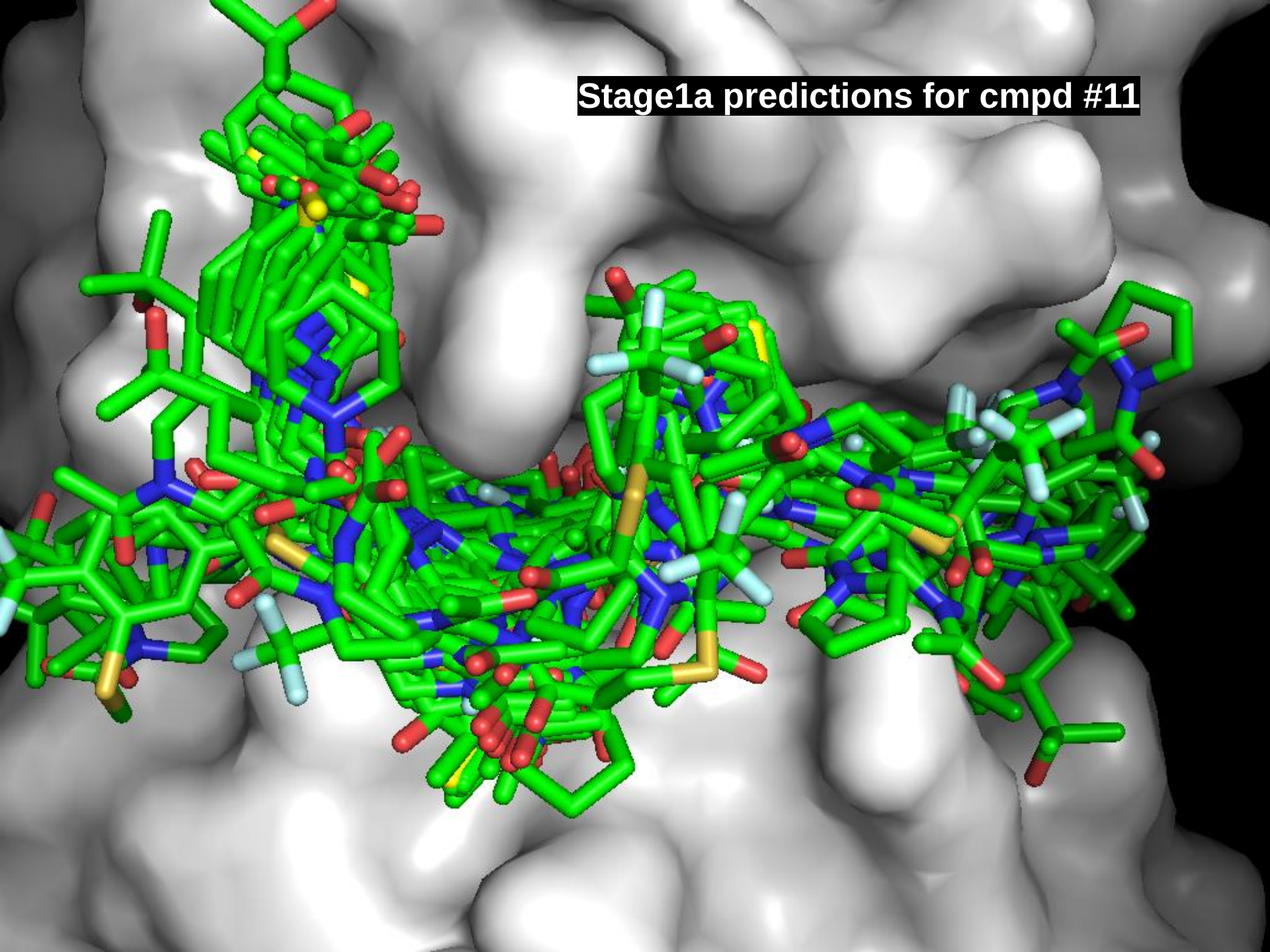
CatS stage1a potency, colored by submitting group

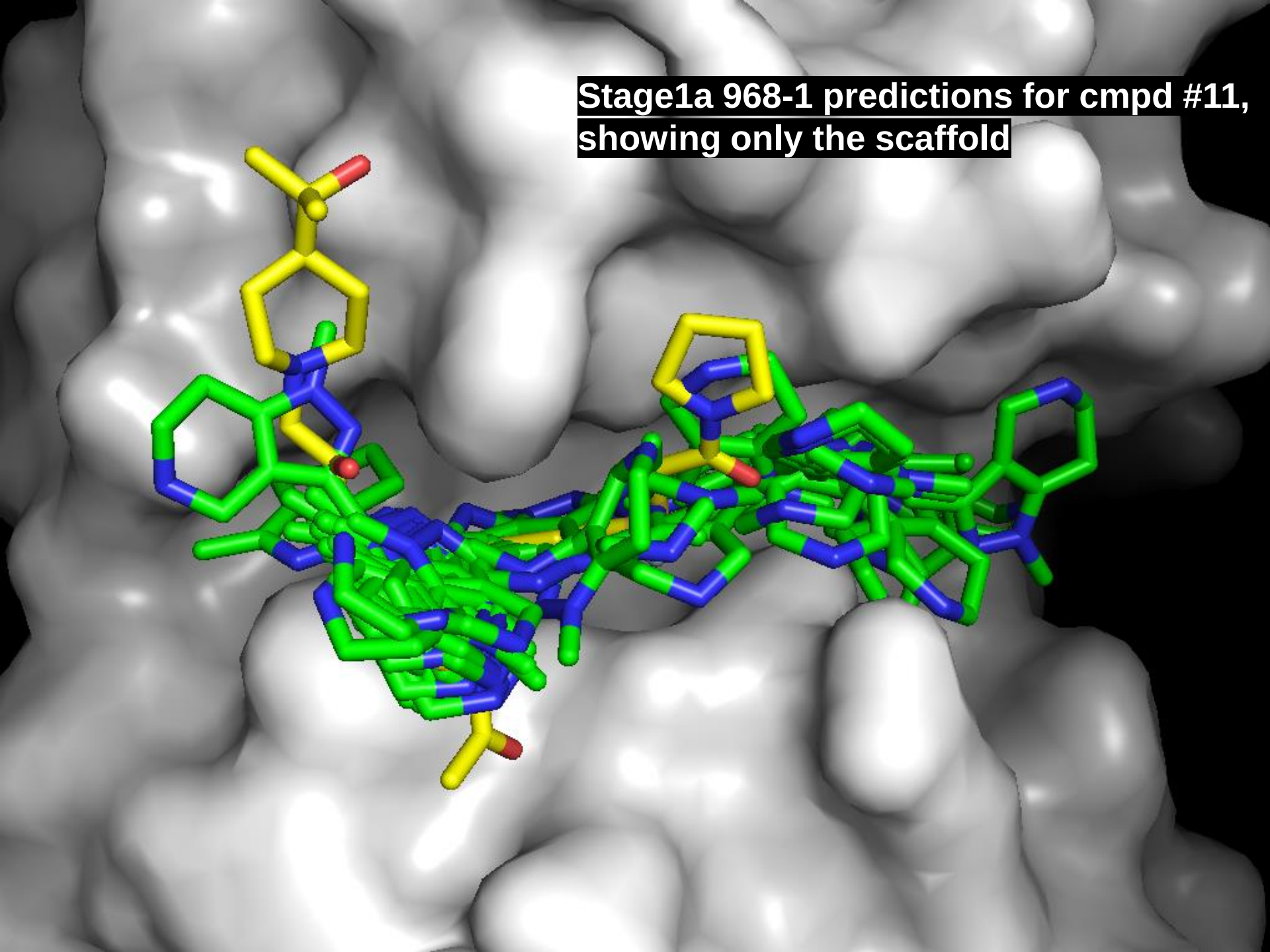



Stage1a predictions for cmpd #11


crystal structures for 19 of the
22 tetrahyropyrido-pyrazoles


tetrahyropyrido-pyrazole scaffold
from 19 of the 22 crystal structures

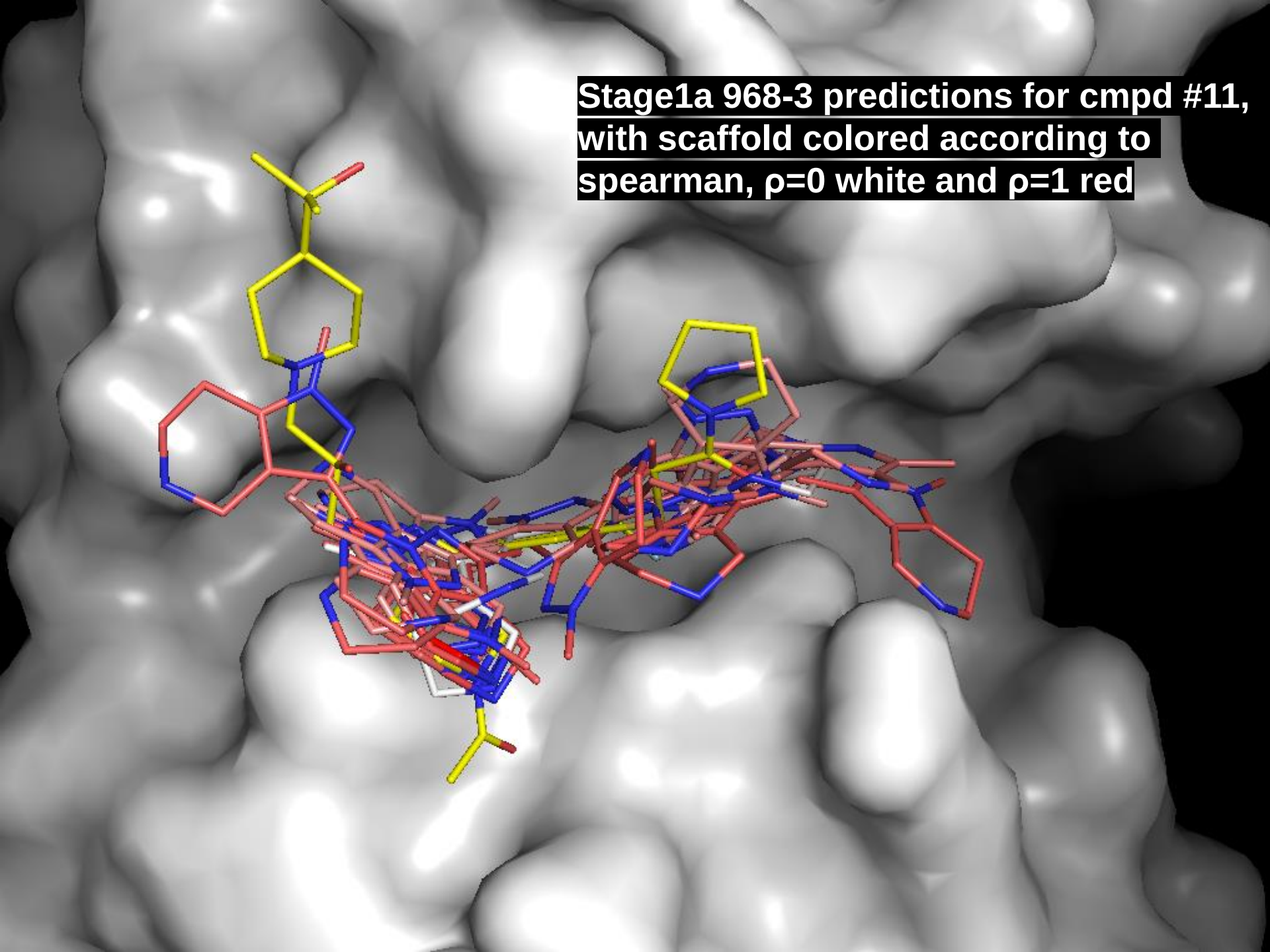

Stage1a predictions for cmpd #11

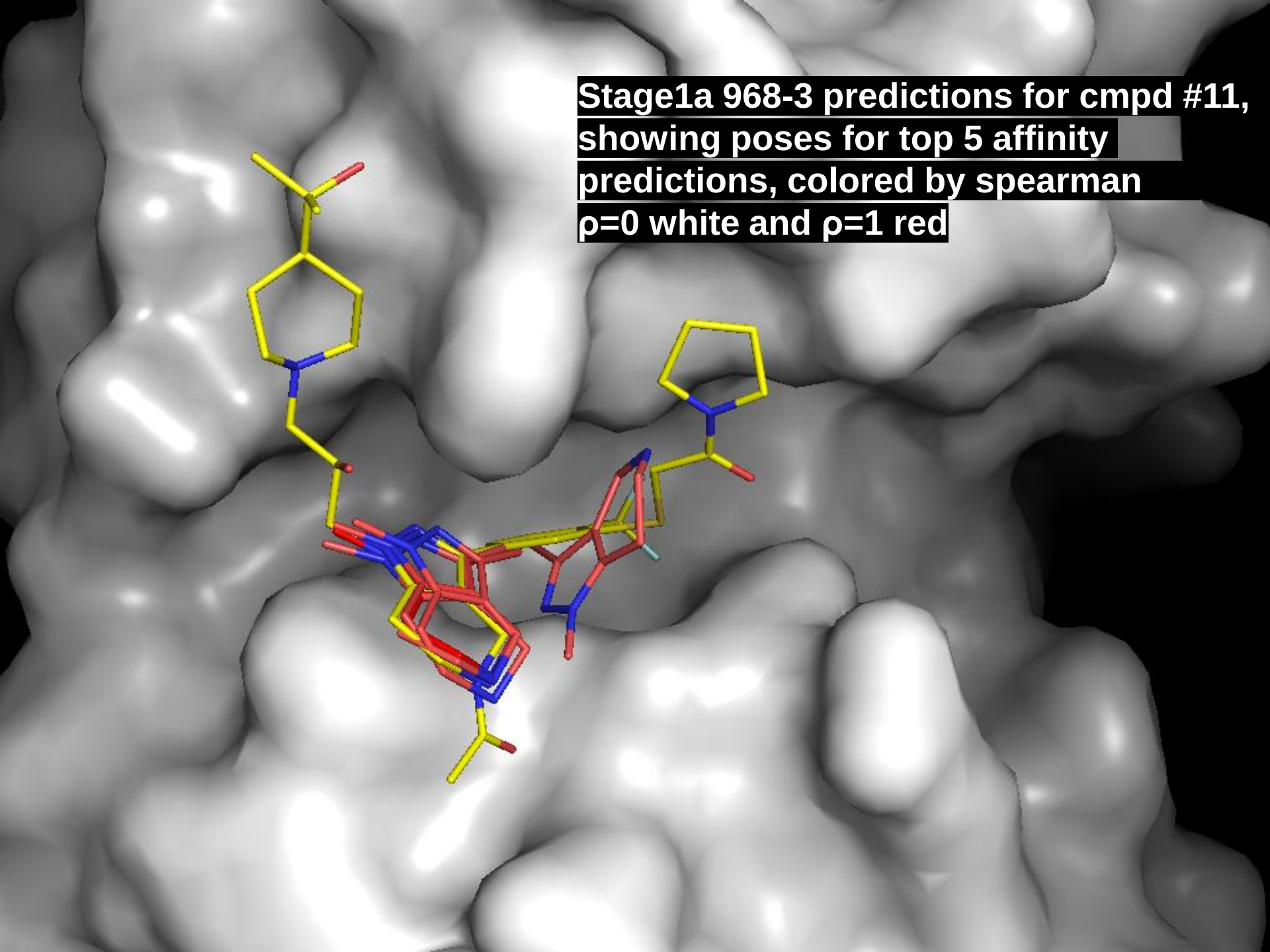

Stage1a 968-1 predictions for cmpd #11,
showing only the scaffold

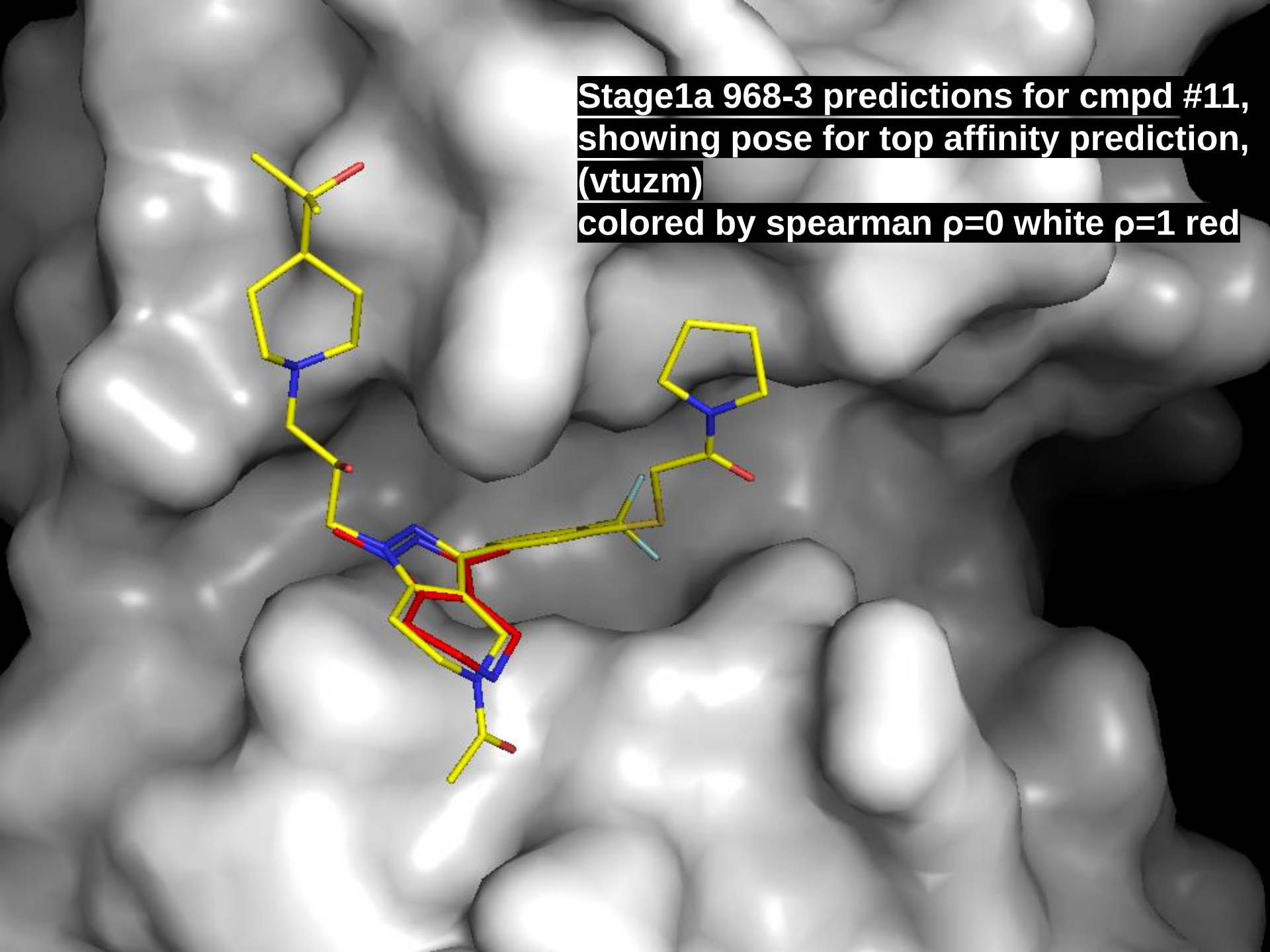
Affinity prediction, 136 compounds, structure based

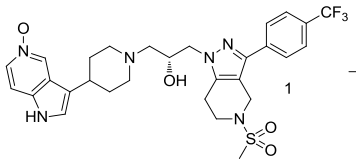
Stage 1a 968-3: 43 submissions, ranging from good (ρ=0.63) to poor (ρ=-0.12)

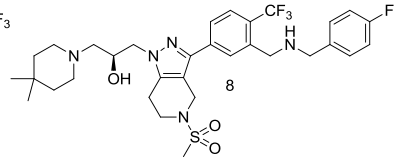
submissio	Num Inhibs	Kendalls Tau	Kendalls Tau Error	Spearman's Rho	Spearman's Rho Error
vtuzm	136	0.45	0.05	0.63	0.06
jg6d4	136	0.29	0.05	0.43	0.08
dhr26	136	0.28	0.06	0.39	0.08
4jk3r	136	0.25	0.06	0.37	0.08
w83jw	136	0.25	0.05	0.37	0.08
ppyff	136	0.26	0.06	0.37	0.08
pgipt	136	0.25	0.06	0.36	0.08
f3ifz	136	0.24	0.05	0.36	0.08
8vvhy	136	0.23	0.05	0.35	0.07
v3c55	136	0.24	0.06	0.34	0.09
q4sb0	136	0.23	0.05	0.34	0.08
0qyrq	136	0.24	0.06	0.34	0.08
taqir	136	0.22	0.05	0.33	0.08
4pakq	136	0.23	0.06	0.32	0.08
5b5wz	136	0.22	0.05	0.32	0.08
vns3a	136	0.21	0.06	0.31	0.08
wcrem	136	0.20	0.05	0.30	0.08
evkuh	136	0.21	0.06	0.30	0.08
vihk2	136	0.19	0.06	0.29	0.08
w6bwq	136	0.17	0.05	0.27	0.08
36ovr	19	0.18	0.16	0.27	0.23
ru7zn	136	0.16	0.05	0.25	0.08
etiak	136	0.17	0.06	0.25	0.08
fayra	136	0.17	0.06	0.25	0.09
zs7oa	136	0.16	0.06	0.23	0.08
uwrw5	136	0.16	0.06	0.23	0.09
8uer8	136	0.15	0.06	0.23	0.08
oj2uj	136	0.15	0.06	0.22	0.09
t3dbz	136	0.15	0.05	0.22	0.08
xz8so	136	0.13	0.05	0.21	0.08
m7oq4	136	0.15	0.06	0.21	0.08
uvjt0	136	0.14	0.06	0.20	0.09
hn0qy	136	0.13	0.06	0.19	0.09
3k3fn	136	0.13	0.05	0.19	0.08
hfbm5	136	0.10	0.05	0.18	0.08
tq8gb	136	0.12	0.06	0.18	0.09
omotr	136	0.11	0.06	0.17	0.09
3hz34	136	0.12	0.06	0.17	0.09
xyy85	136	0.12	0.06	0.17	0.08
04kya	136	0.11	0.06	0.15	0.09
44mp4	136	0.09	0.06	0.12	0.09
5r4cd	136	-0.04	0.06	-0.06	0.09
72yx7	135	-0.07	0.06	-0.12	0.09

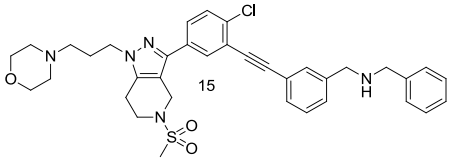

Stage1a 968-3 predictions for cmpd #11,
with scaffold colored according to
spearman, ρ=0 white and ρ=1 red

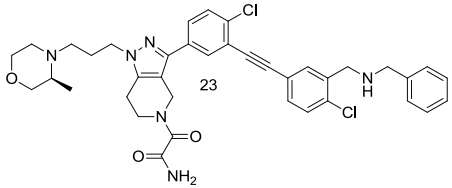

Stage1a 968-3 predictions for cmpd #11,
showing poses for top 5 affinity
predictions, colored by spearman
ρ=0 white and ρ=1 red

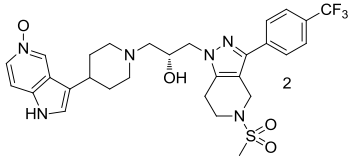

Stage1a 968-3 predictions for cmpd #11,
showing pose for top affinity prediction,
(vtuzm)
colored by spearman ρ=0 white ρ=1 red

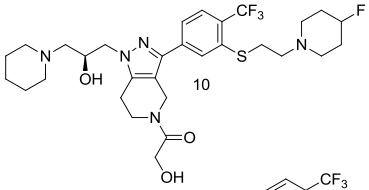

1

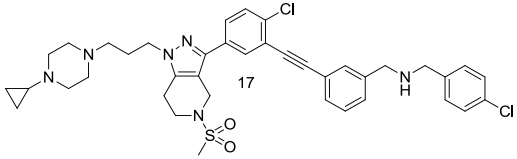

8

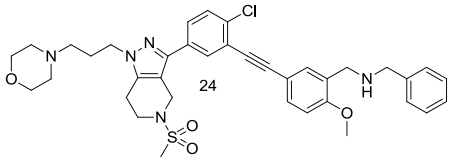

15

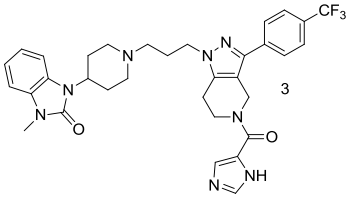

23

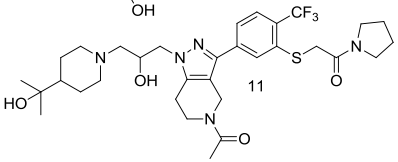

2

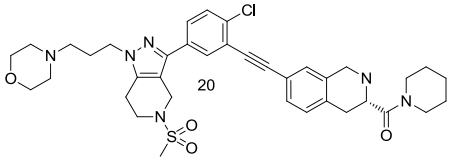

10

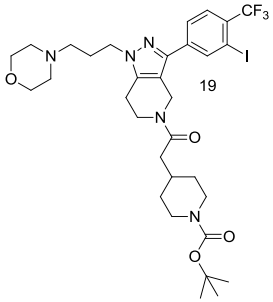

17

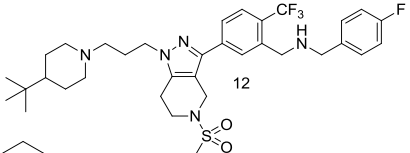

24

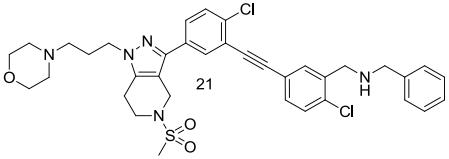

3

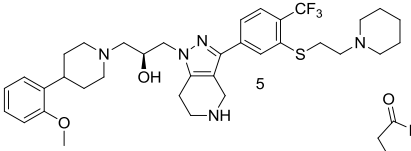

11

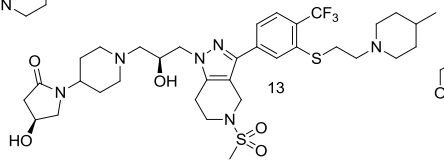

20

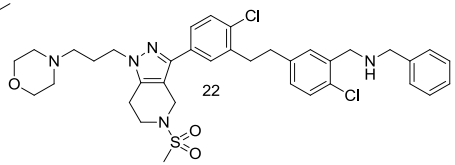

19

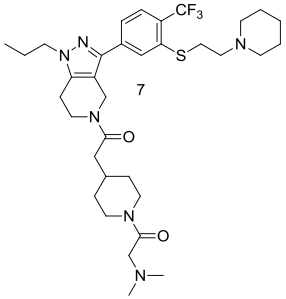

12

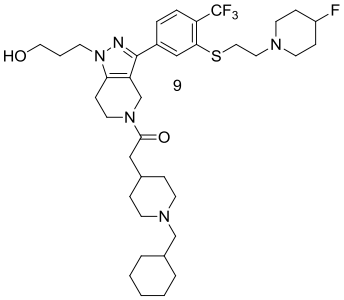

21

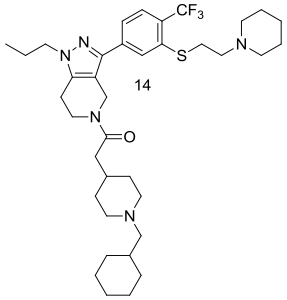

5

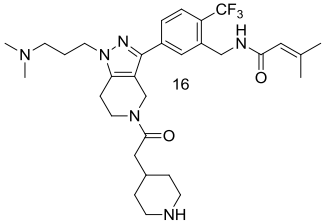

13

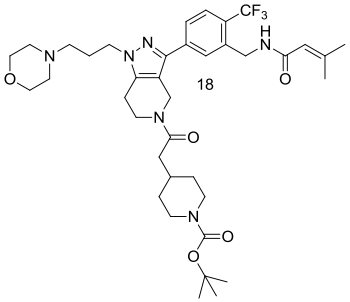

22

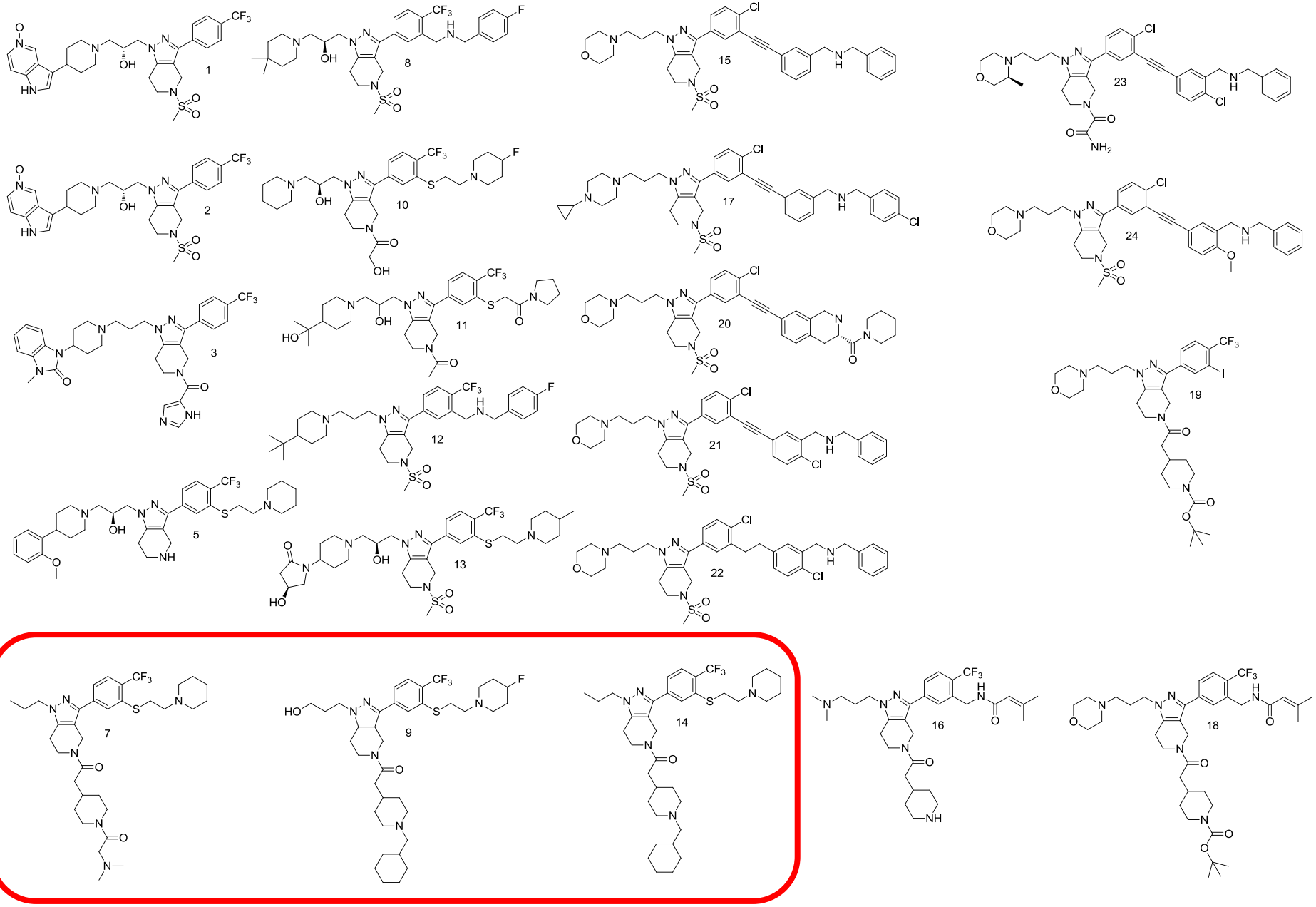

7

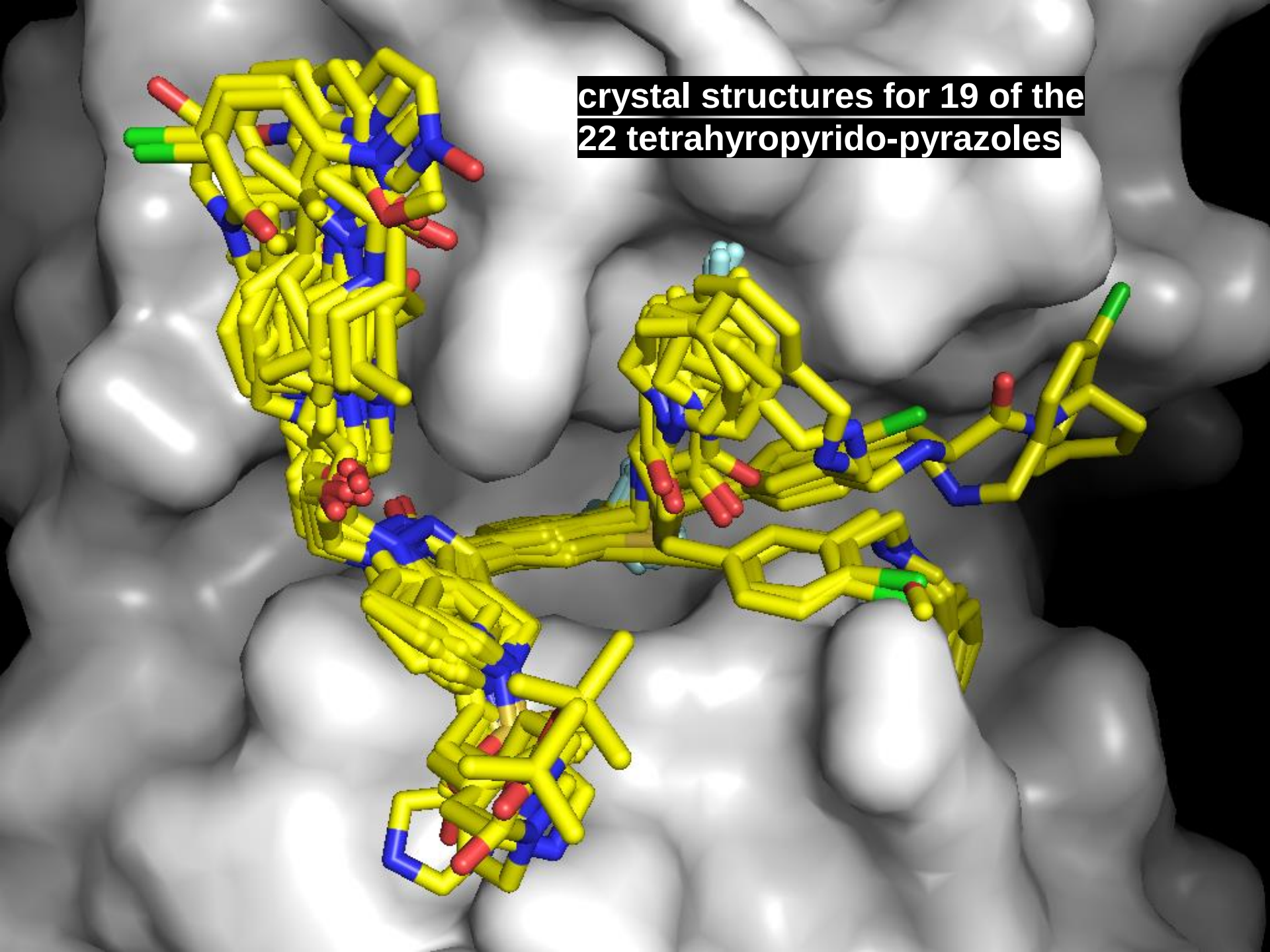

9

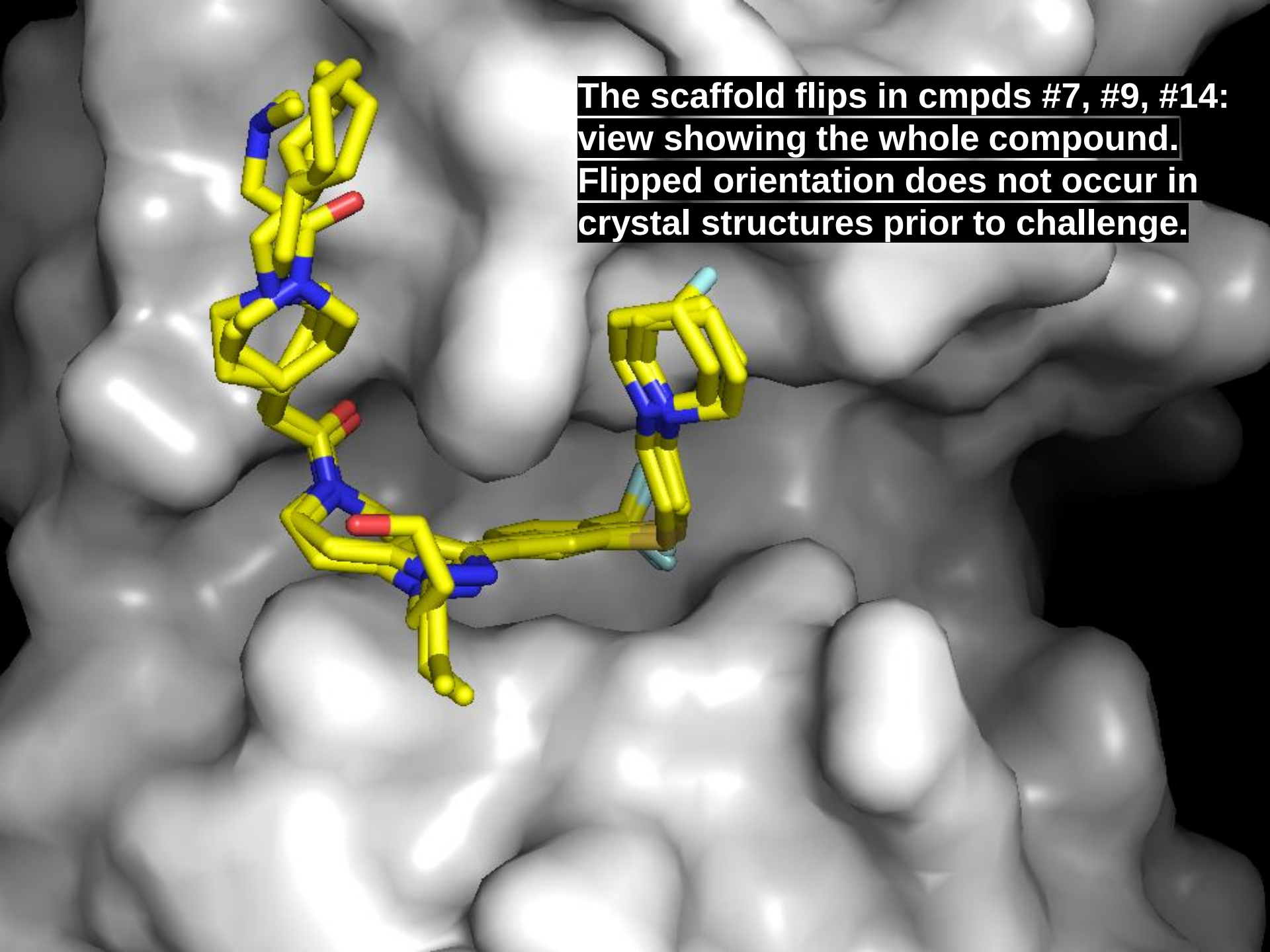

14

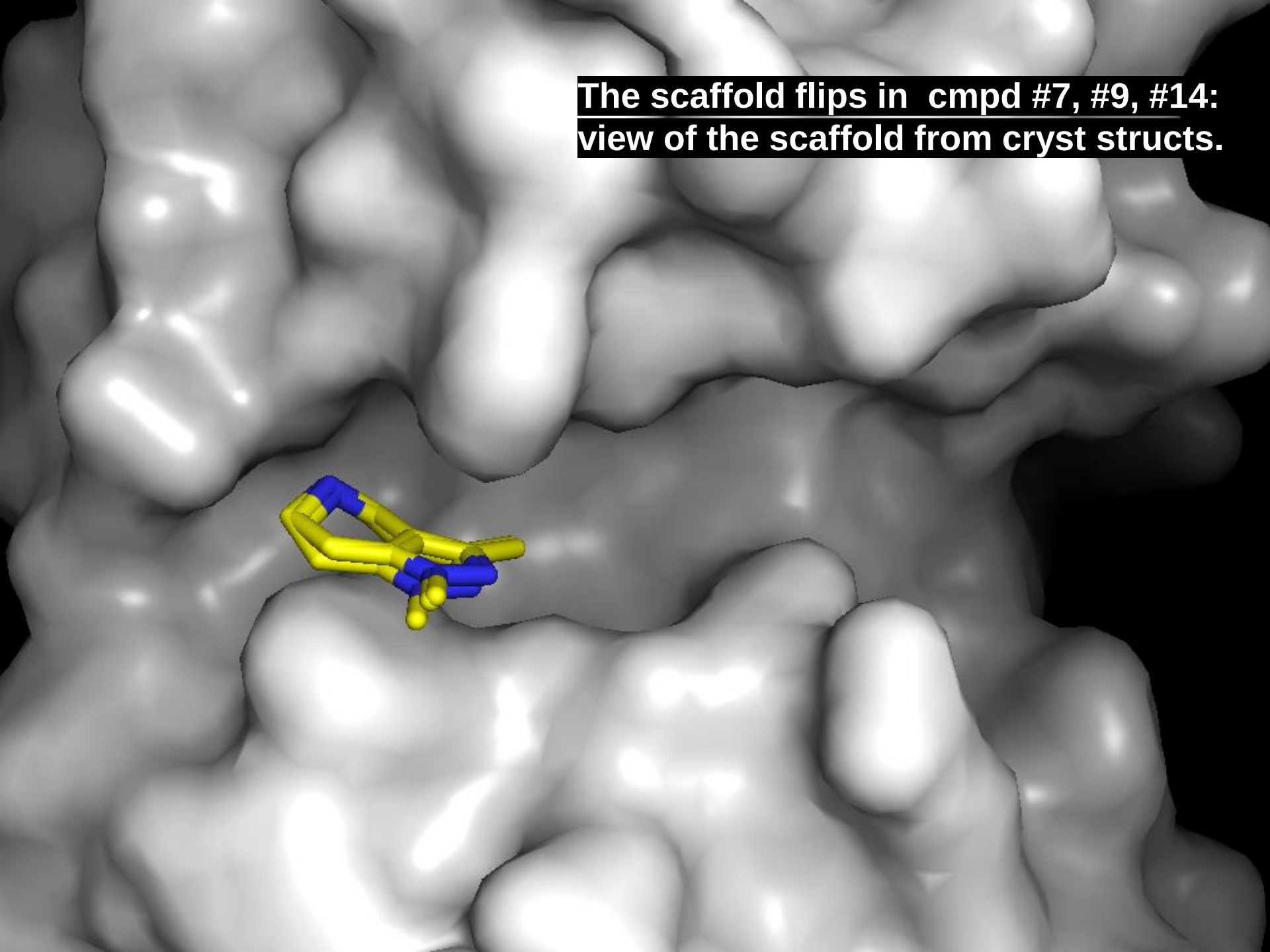

16

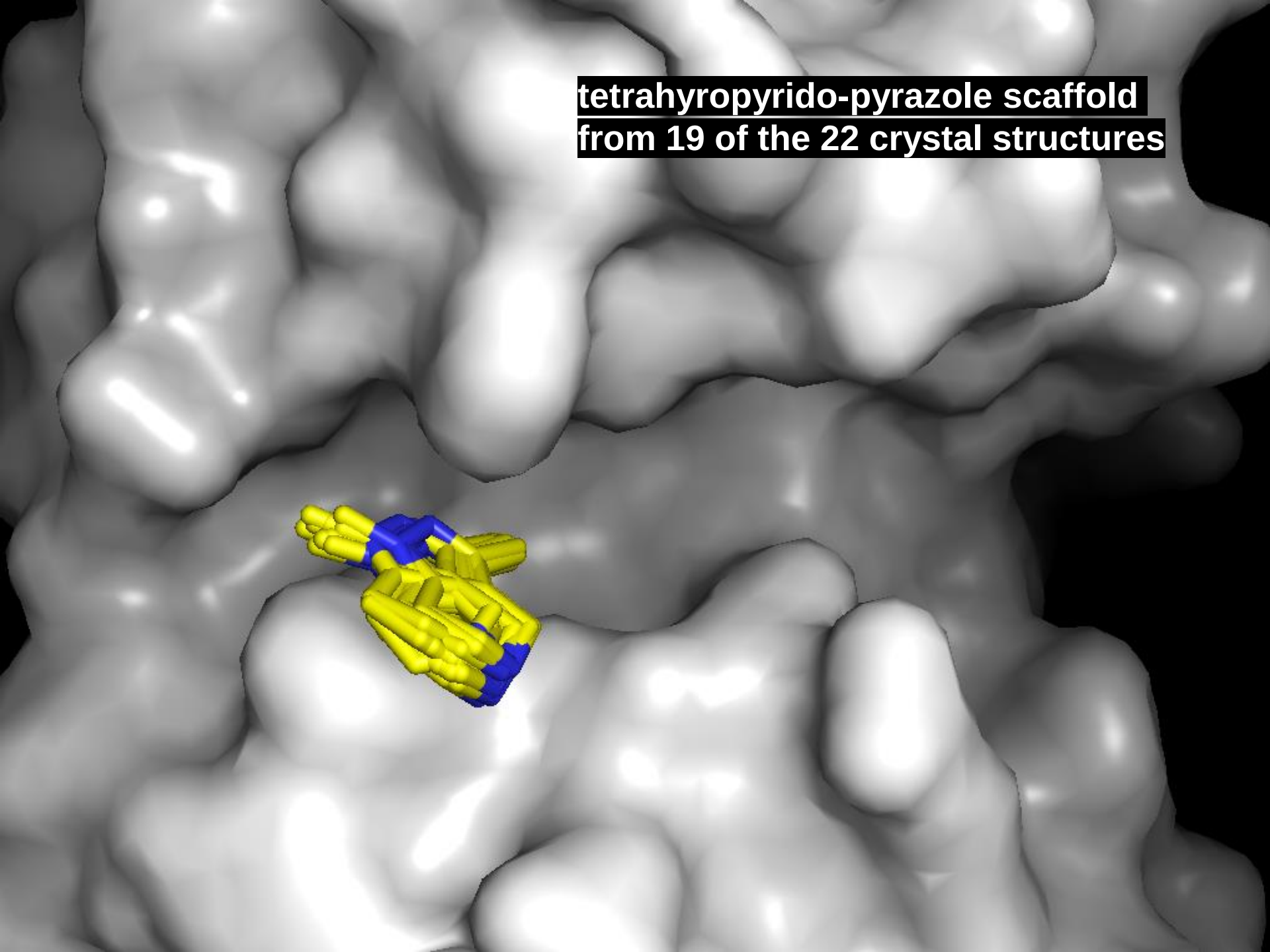

18

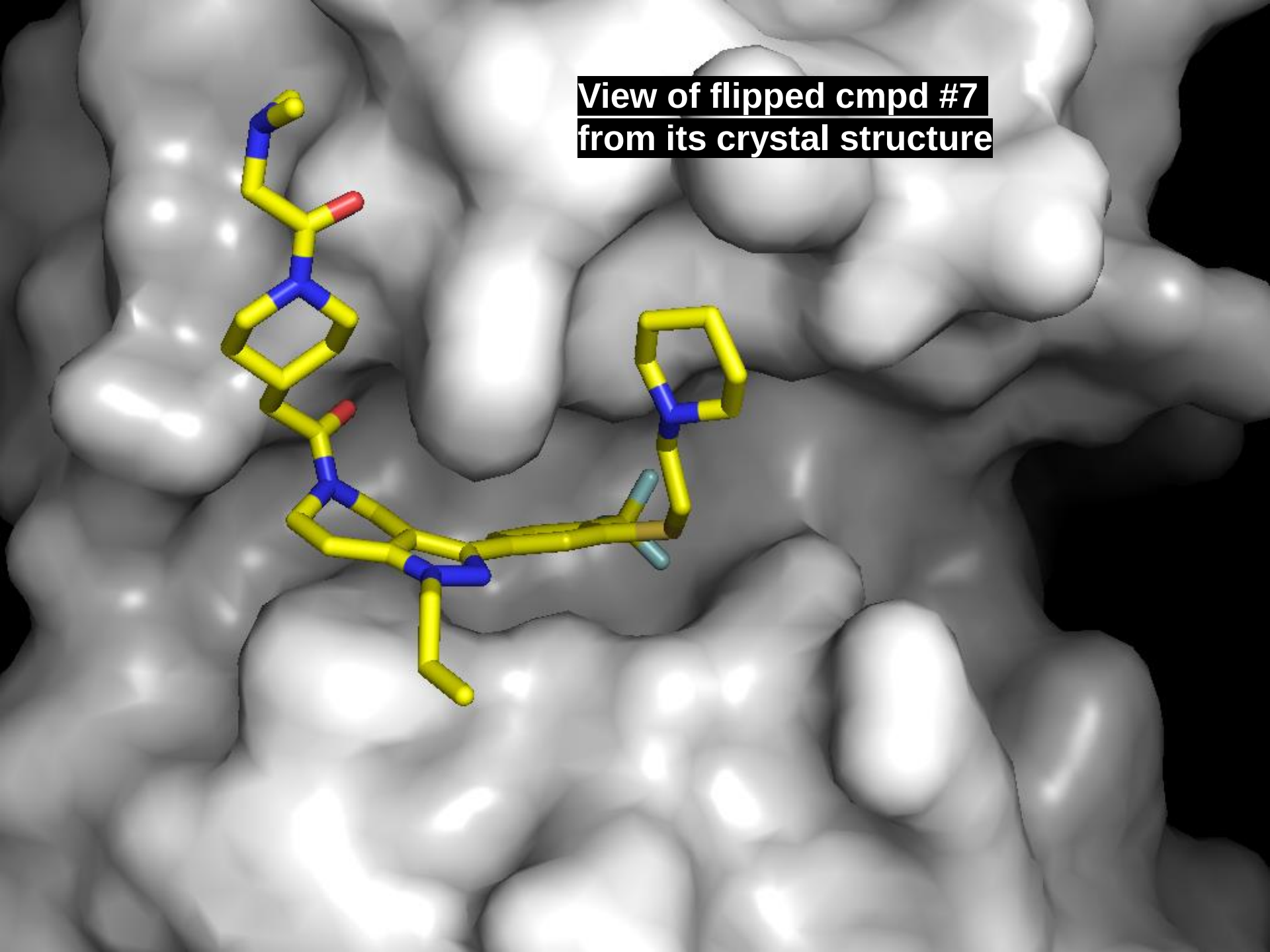

1
2
3
5
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

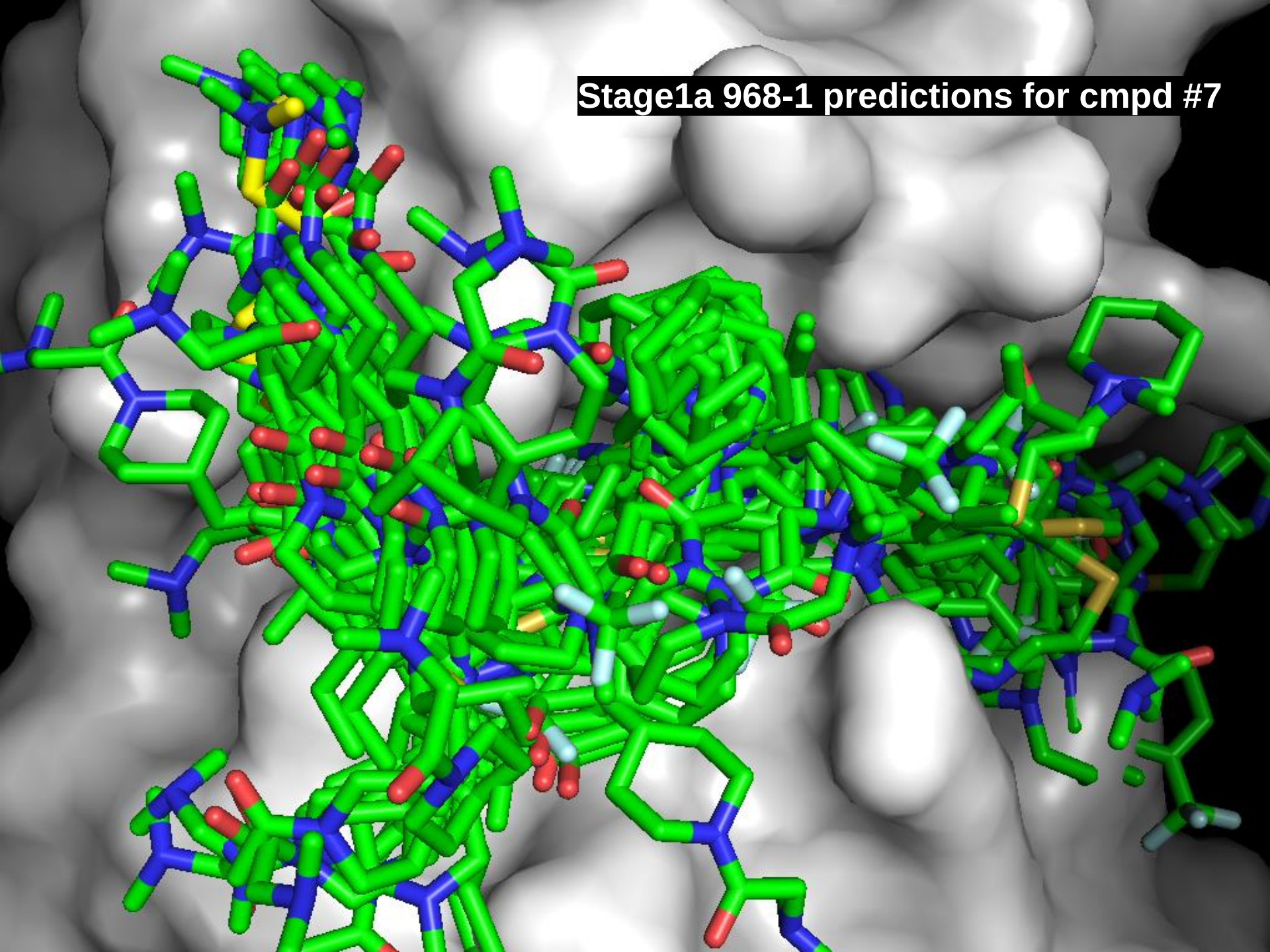

crystal structures for 19 of the
22 tetrahyropyrido-pyrazoles

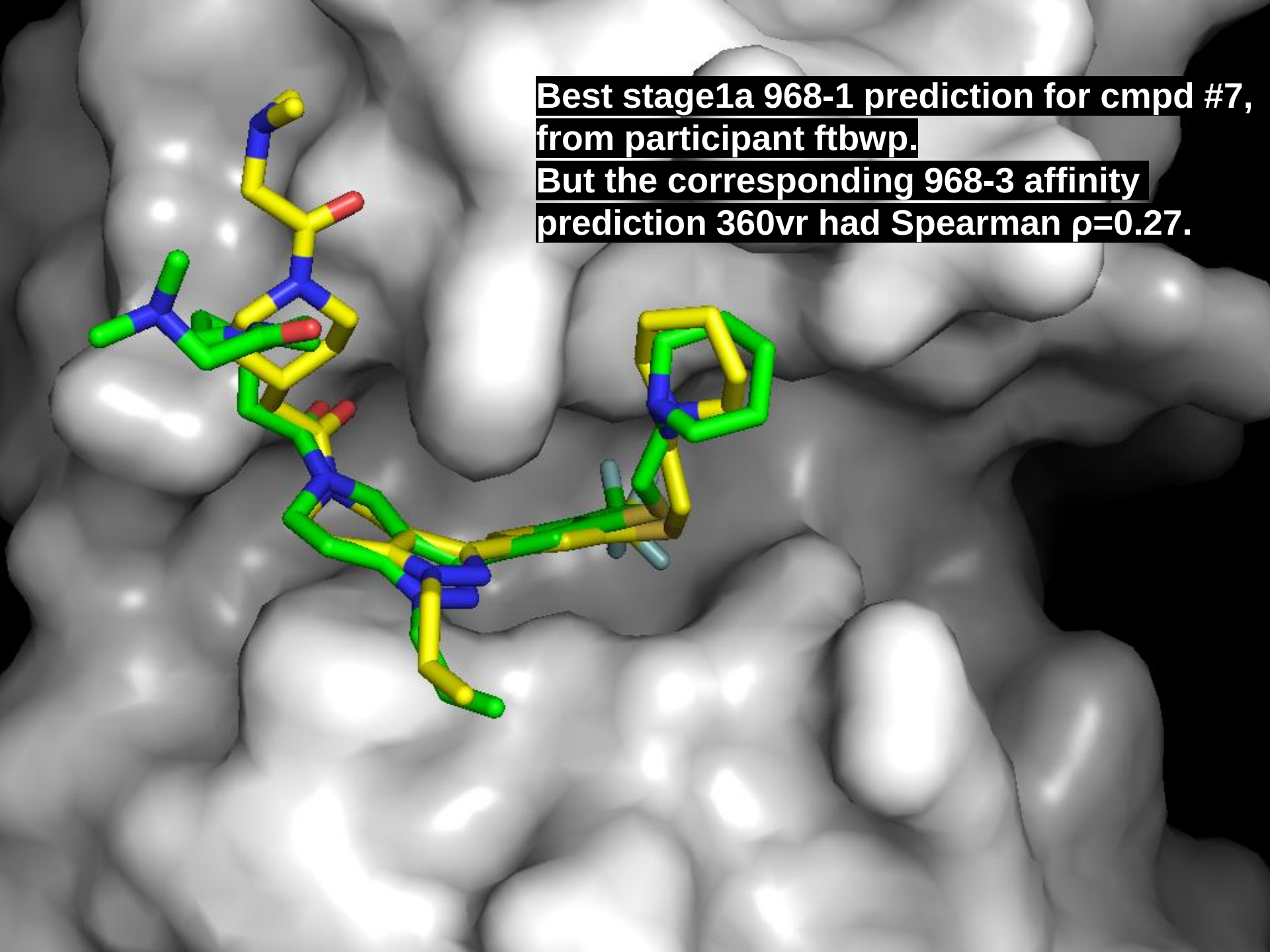

The scaffold flips in cmpds #7, #9, #14:
view showing the whole compound.
Flipped orientation does not occur in
crystal structures prior to challenge.

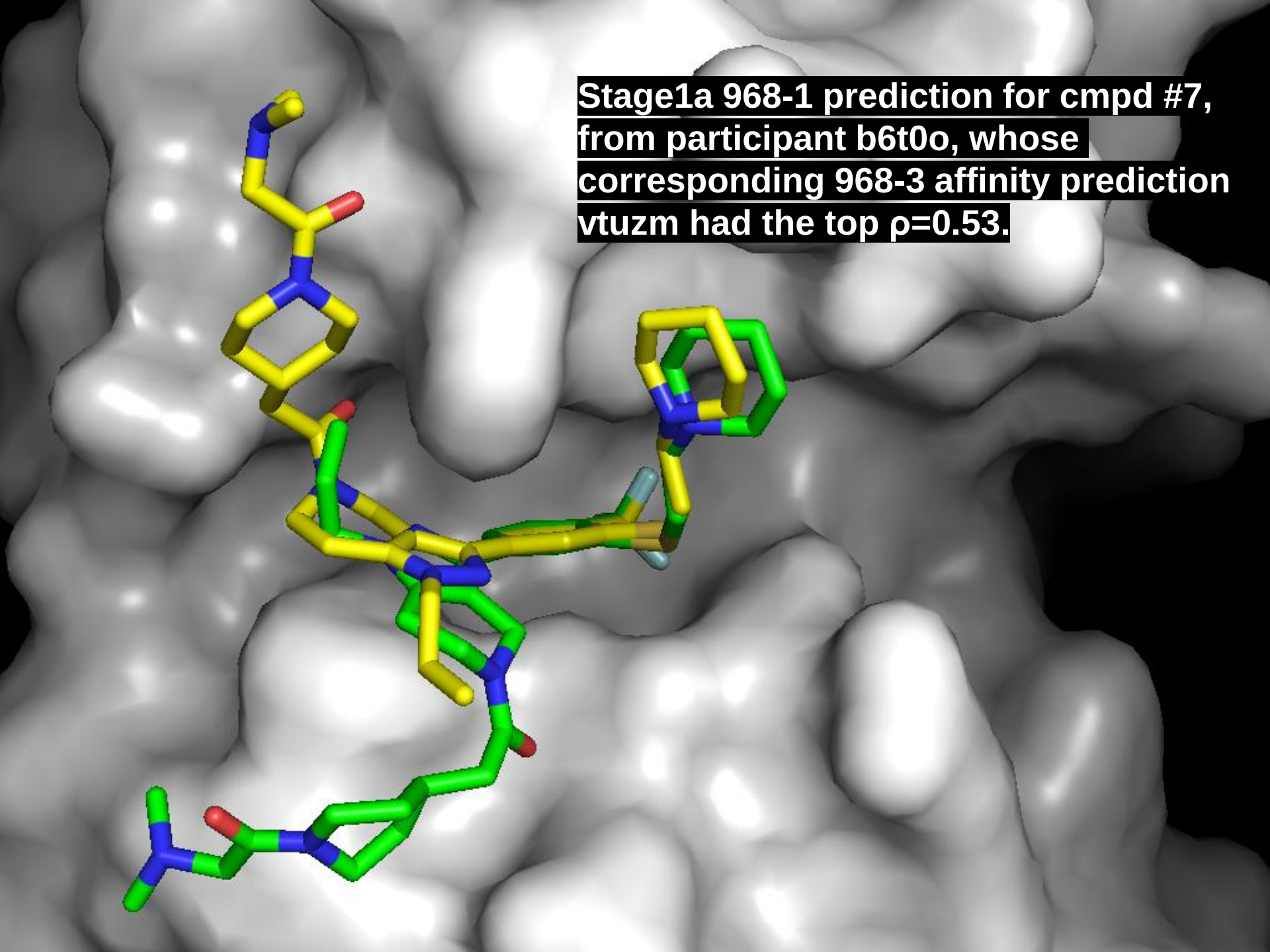

The scaffold flips in cmpd #7, #9, #14:
view of the scaffold from cryst structs.

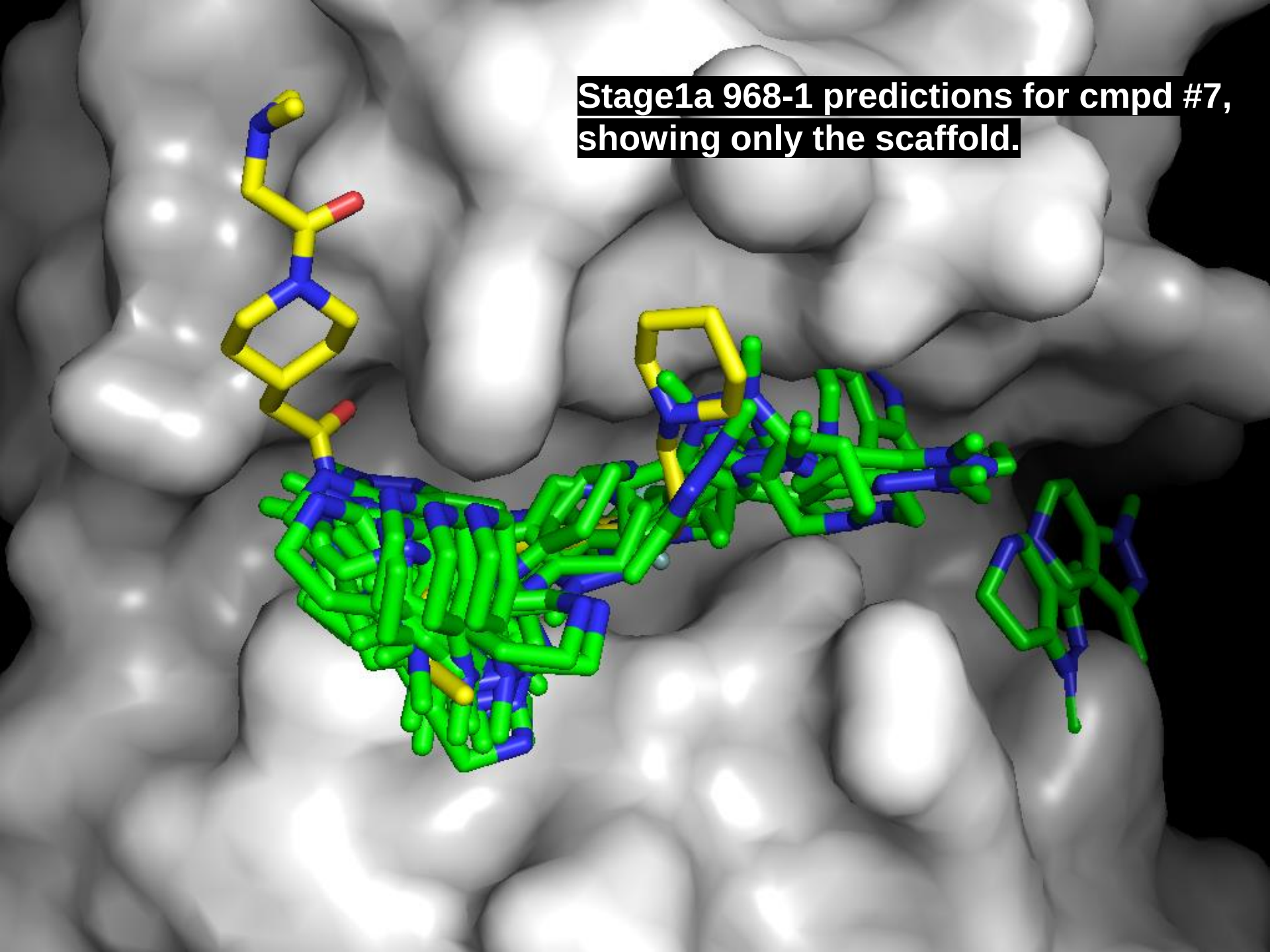

tetrahyropyrido-pyrazole scaffold
from 19 of the 22 crystal structures

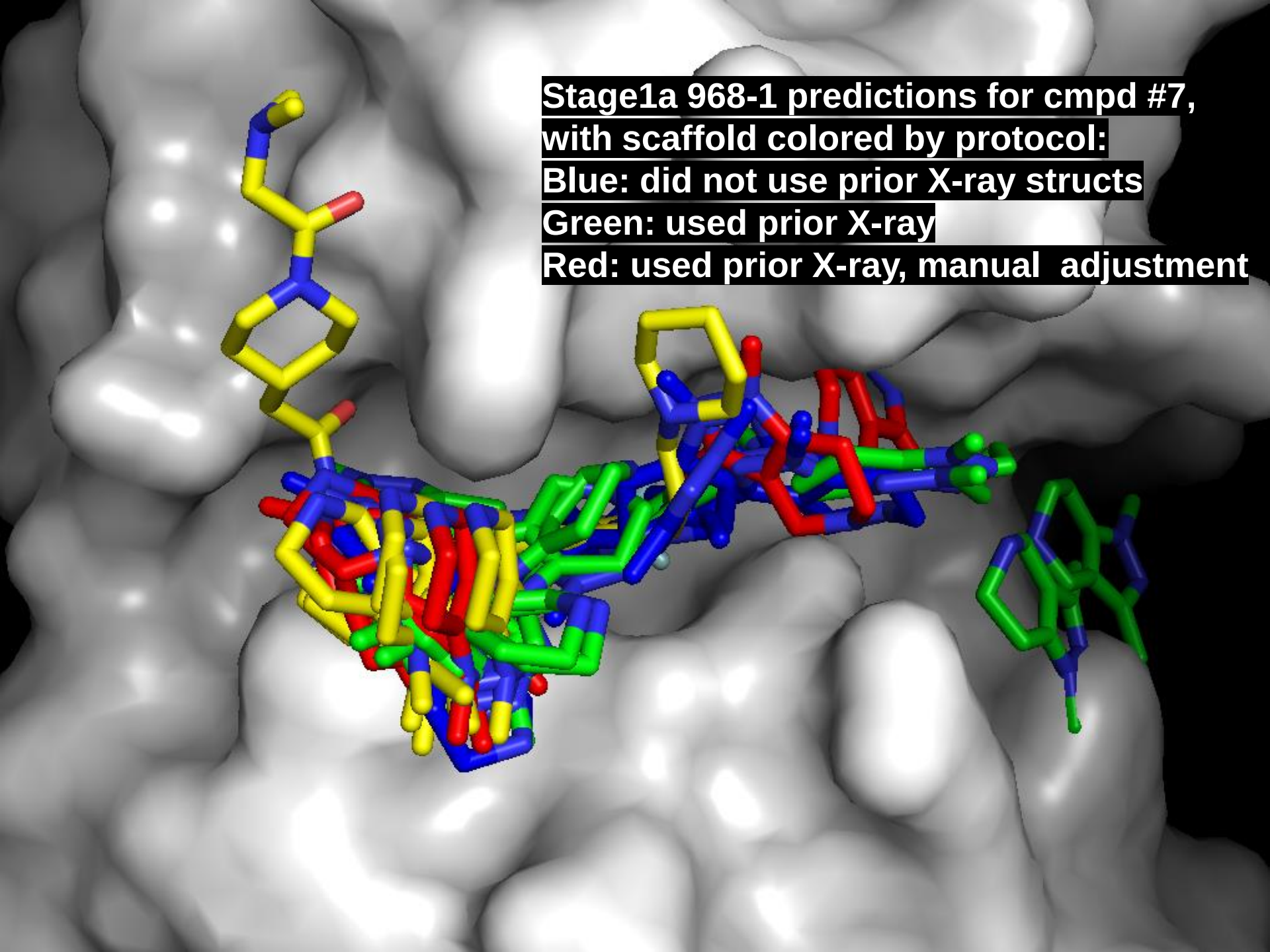

View of flipped cmpd #7
from its crystal structure


Stage1a 968-1 predictions for cmpd #7


Best stage1a 968-1 prediction for cmpd #7,
from participant ftbwp.
But the corresponding 968-3 affinity
prediction 360vr had Spearman ρ=0.27.


Stage1a 968-1 prediction for cmpd #7, from participant b6t0o, whose corresponding 968-3 affinity prediction vtuzm had the top ρ=0.53.


Stage1a 968-1 predictions for cmpd #7,
showing only the scaffold.


Stage1a 968-1 predictions for cmpd #7,
with scaffold colored by protocol:
Blue: did not use prior X-ray structs
Green: used prior X-ray
Red: used prior X-ray, manual adjustment

CatS Stage 1b:

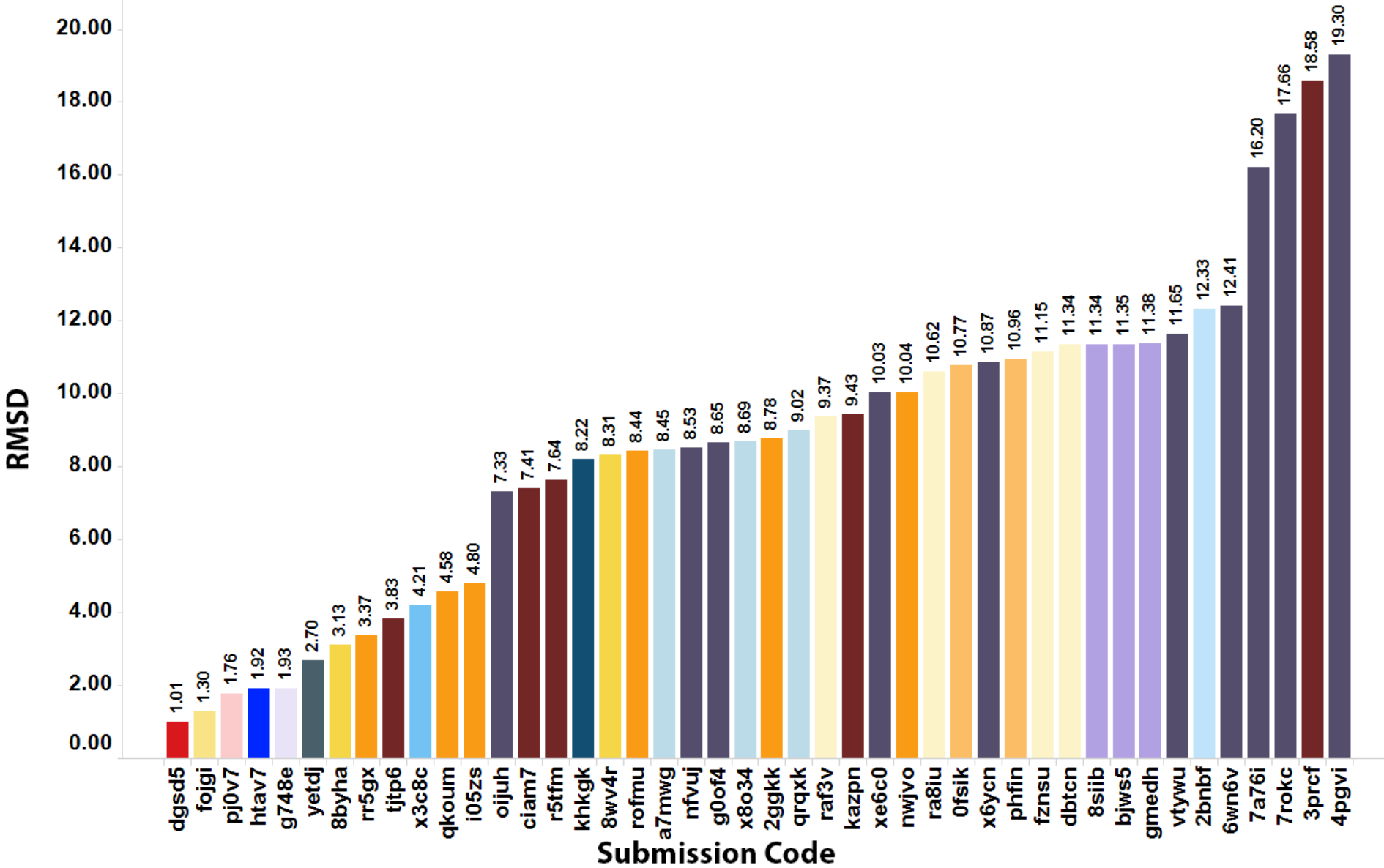
D3R releases coordinatess for all 24 crystal structures (minus inhibitor)

Participants dock the compounds again
(no need to consider protein flexibility)

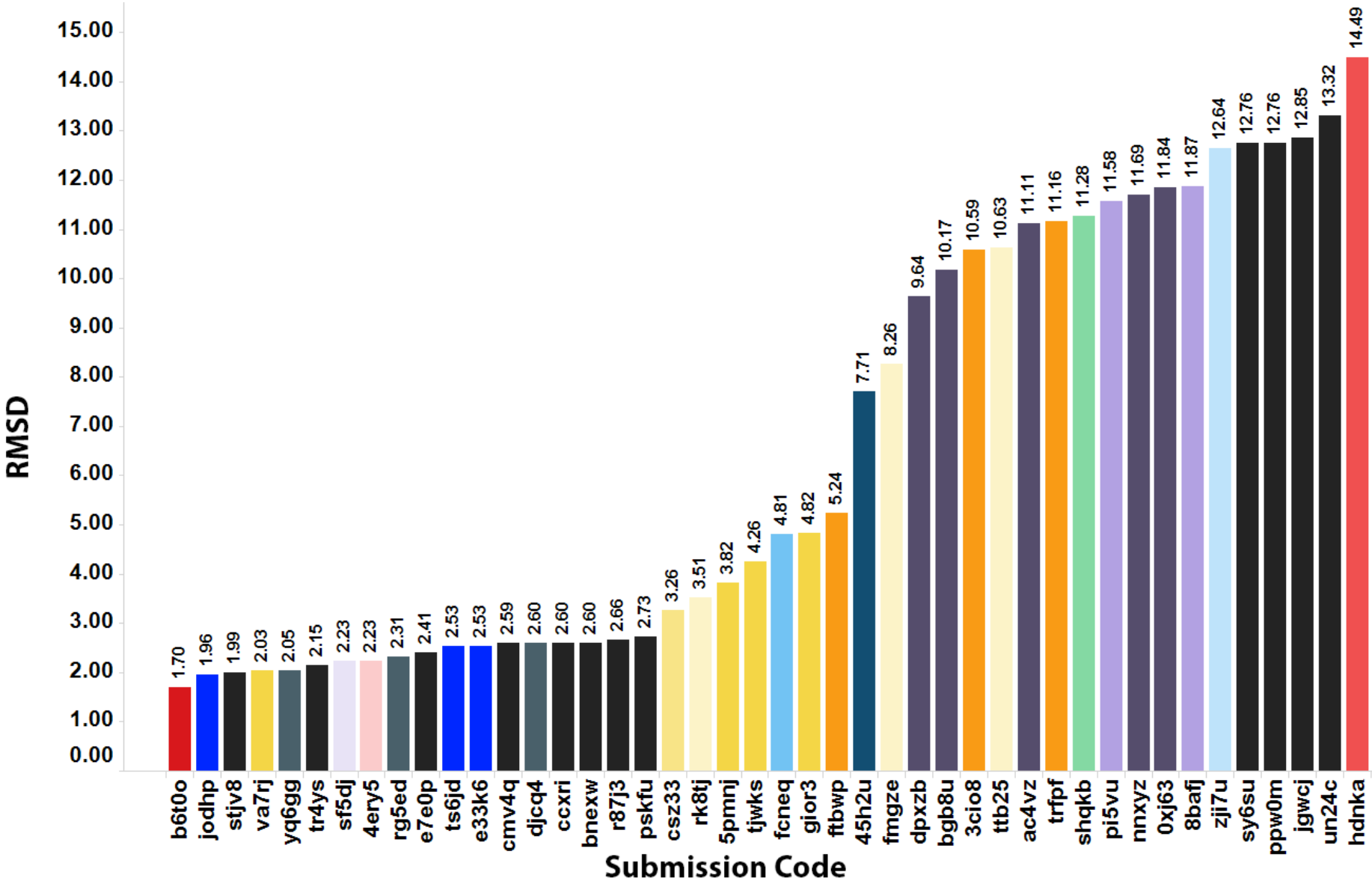
D3R takes the submissions,
compares 24 docking predictions with crystal structures

CatS Stage1b Pose RMSD

Thanks to Conor Parks and Pat Walters for RMSD calculations



CatS Stage1a Pose RMSD, colored by submitting group



CatS Stage 2:

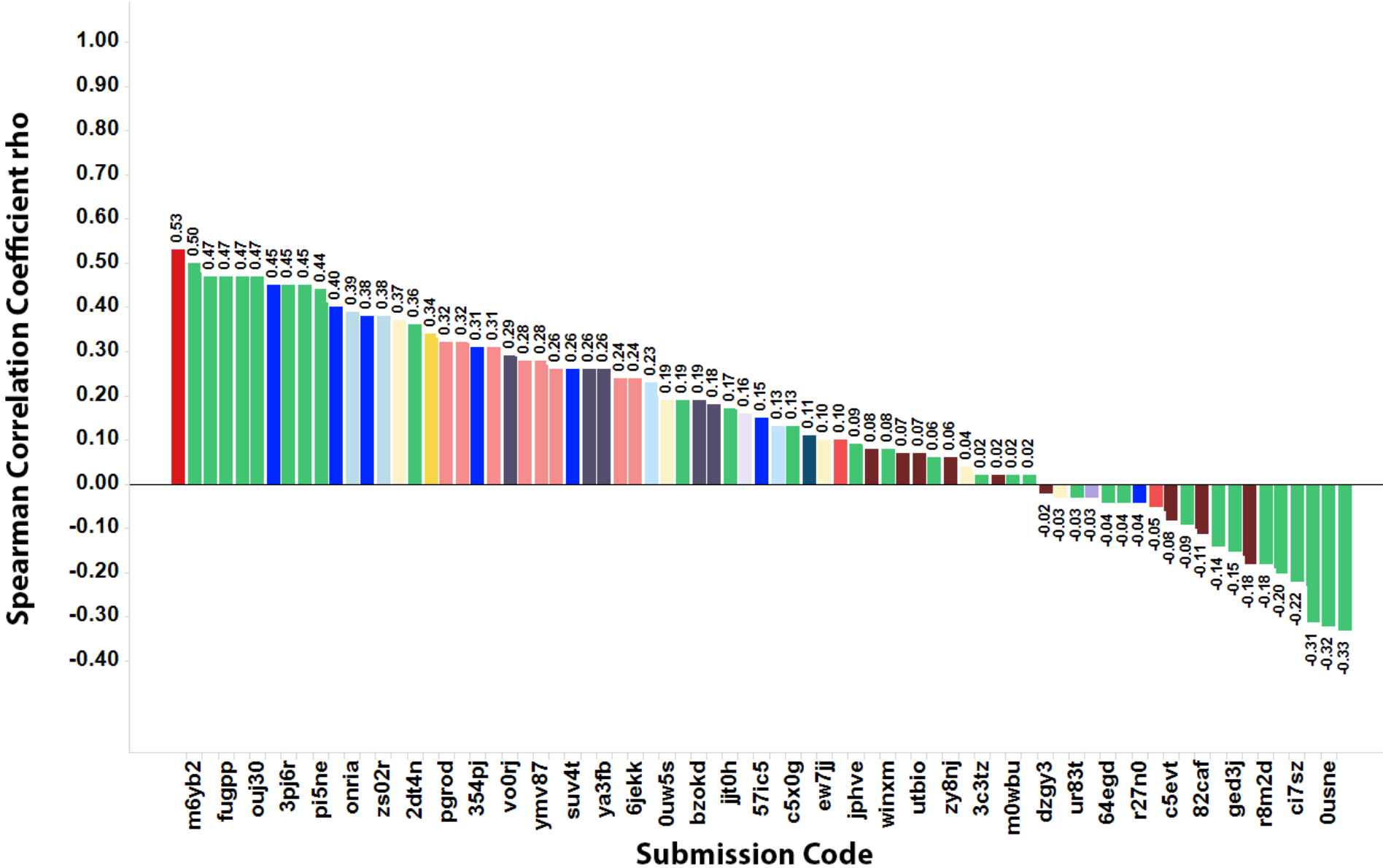
D3R releases all 24 crystal structures (including inhibitor)

Participants run scoring calculations again
(no need for docking)

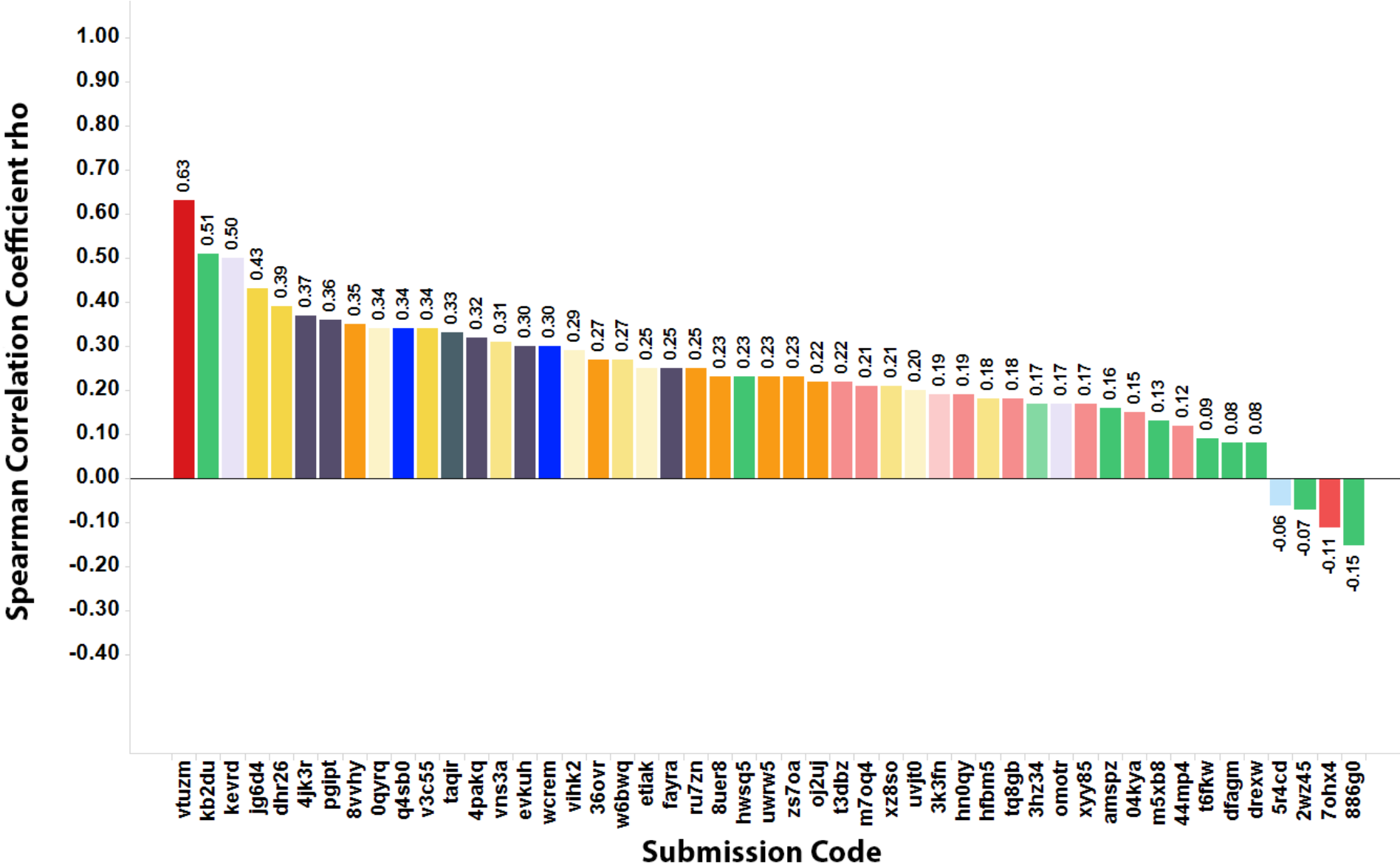
D3R takes the submissions,
compares 136 predicted affinities with measured IC50s
Compares 33 predicted affinities for free energy subset

CatS stage2 potency, colored by submitting group

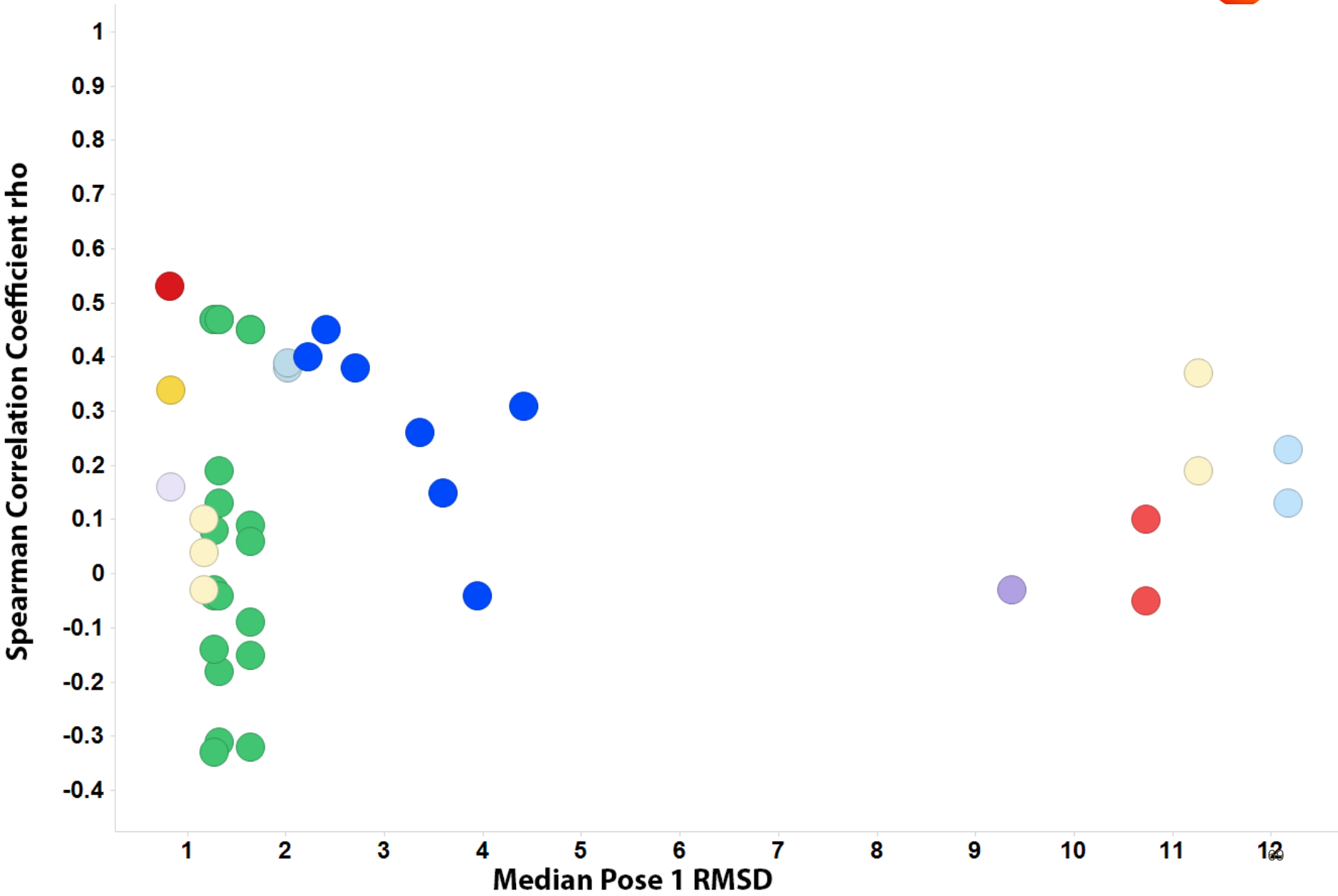
Some groups submitted many predictions



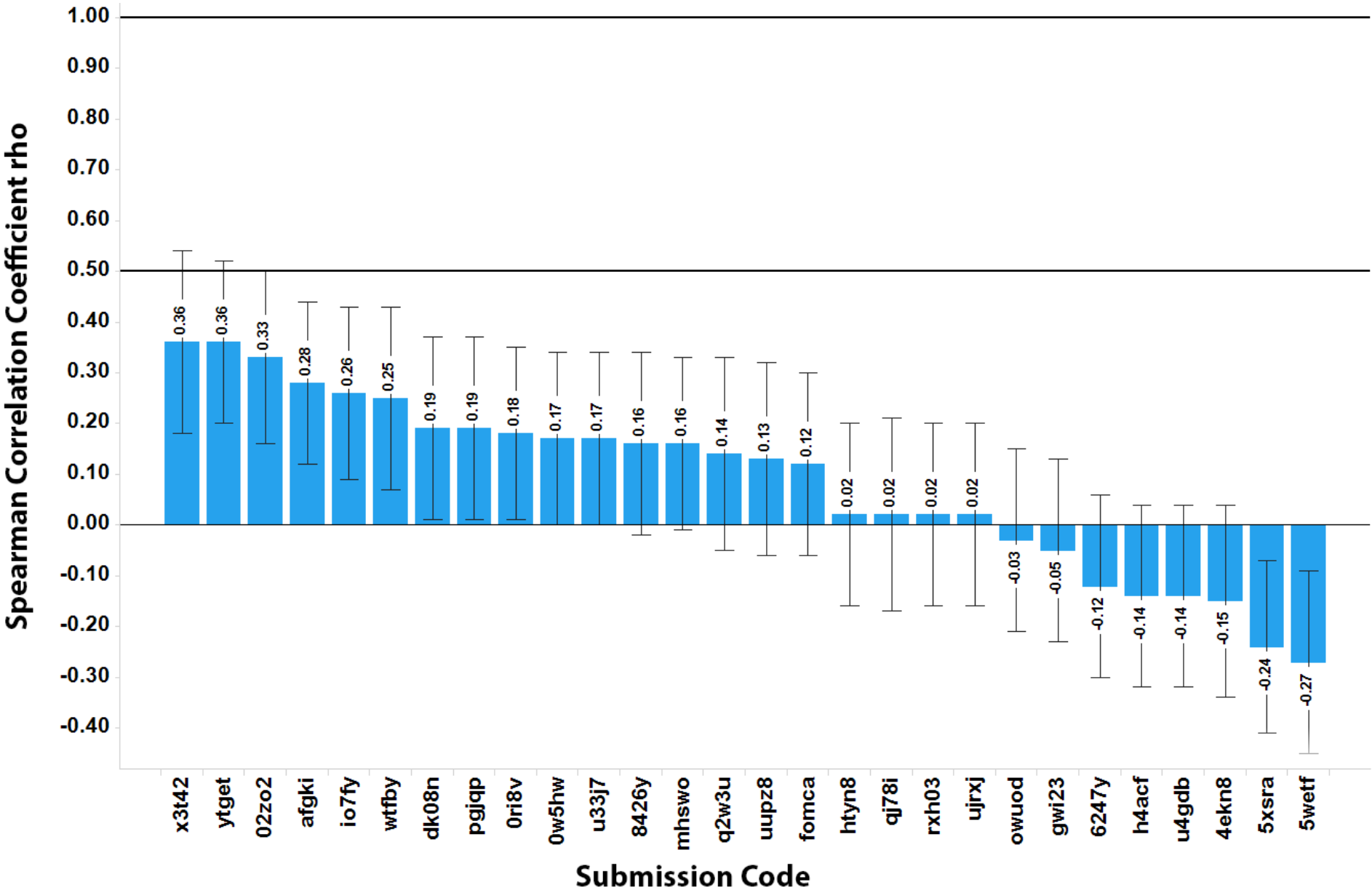
CatS stage1a potency, colored by submitting group



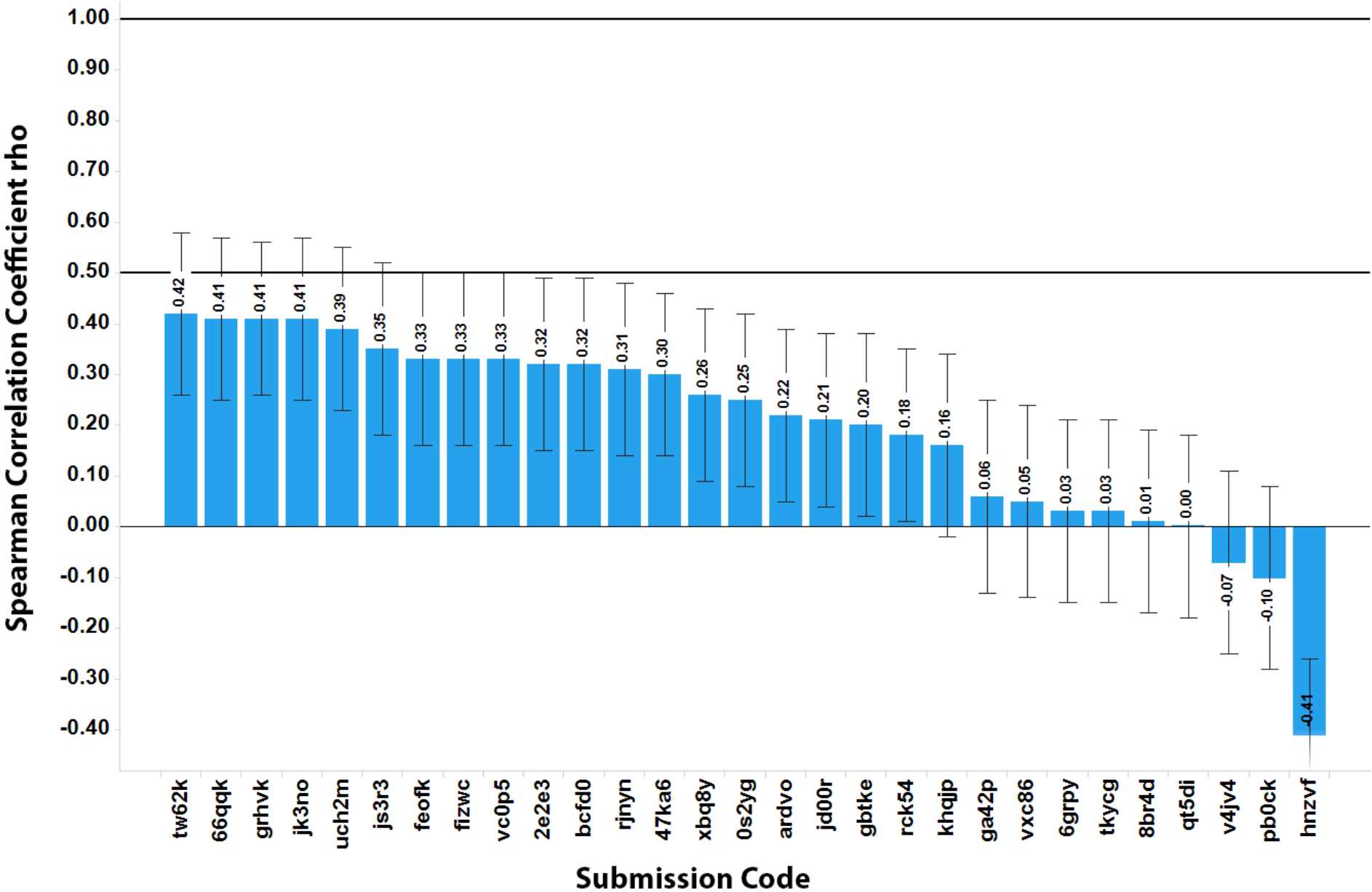
Potency correlation vs RMSD for Stage2 predictions



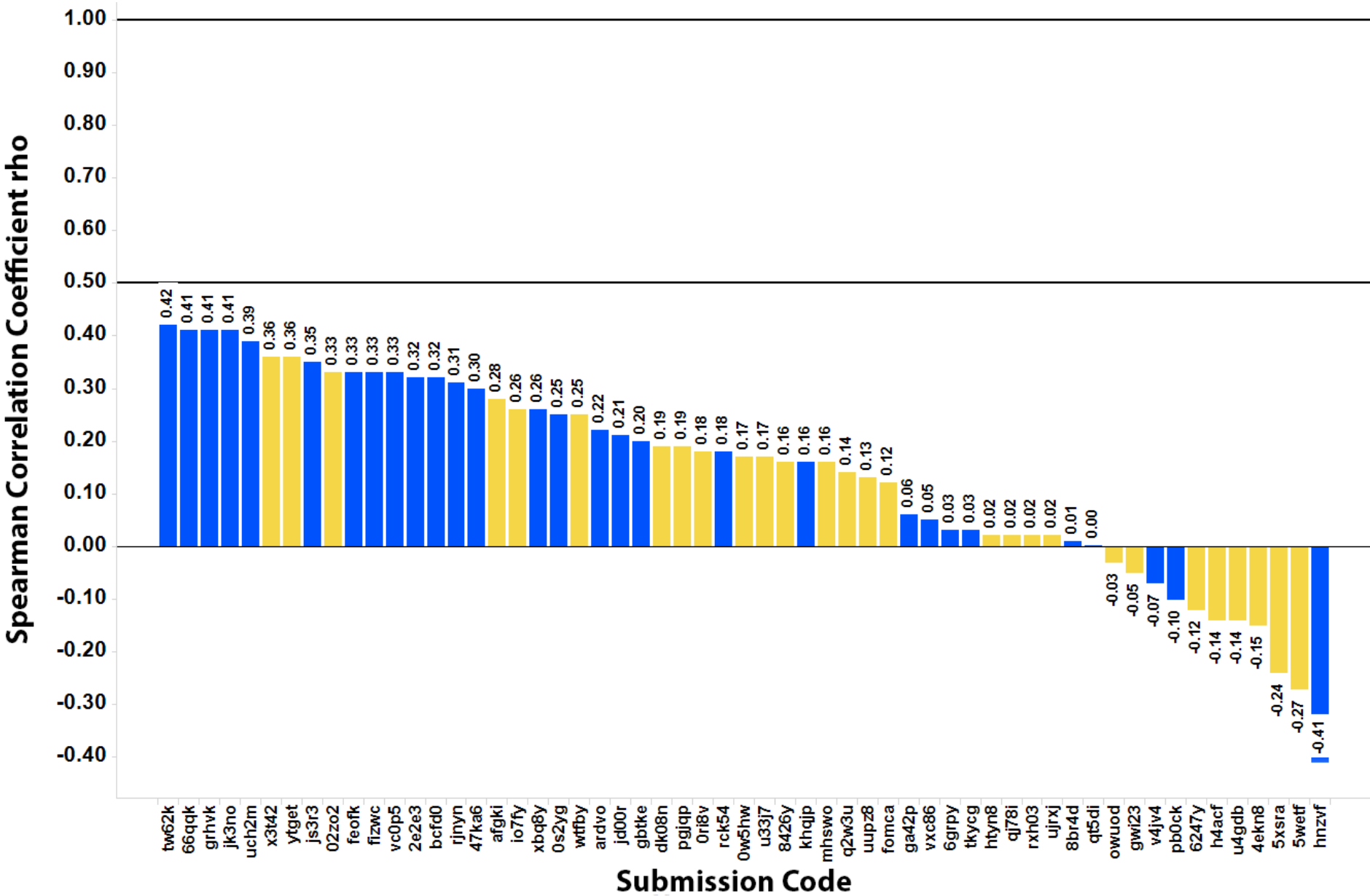
CatS stage1a Free Energy Subset (33 of 136 cmpds)



CatS stage2 Free Energy Subset (33 of 136 cmpds)

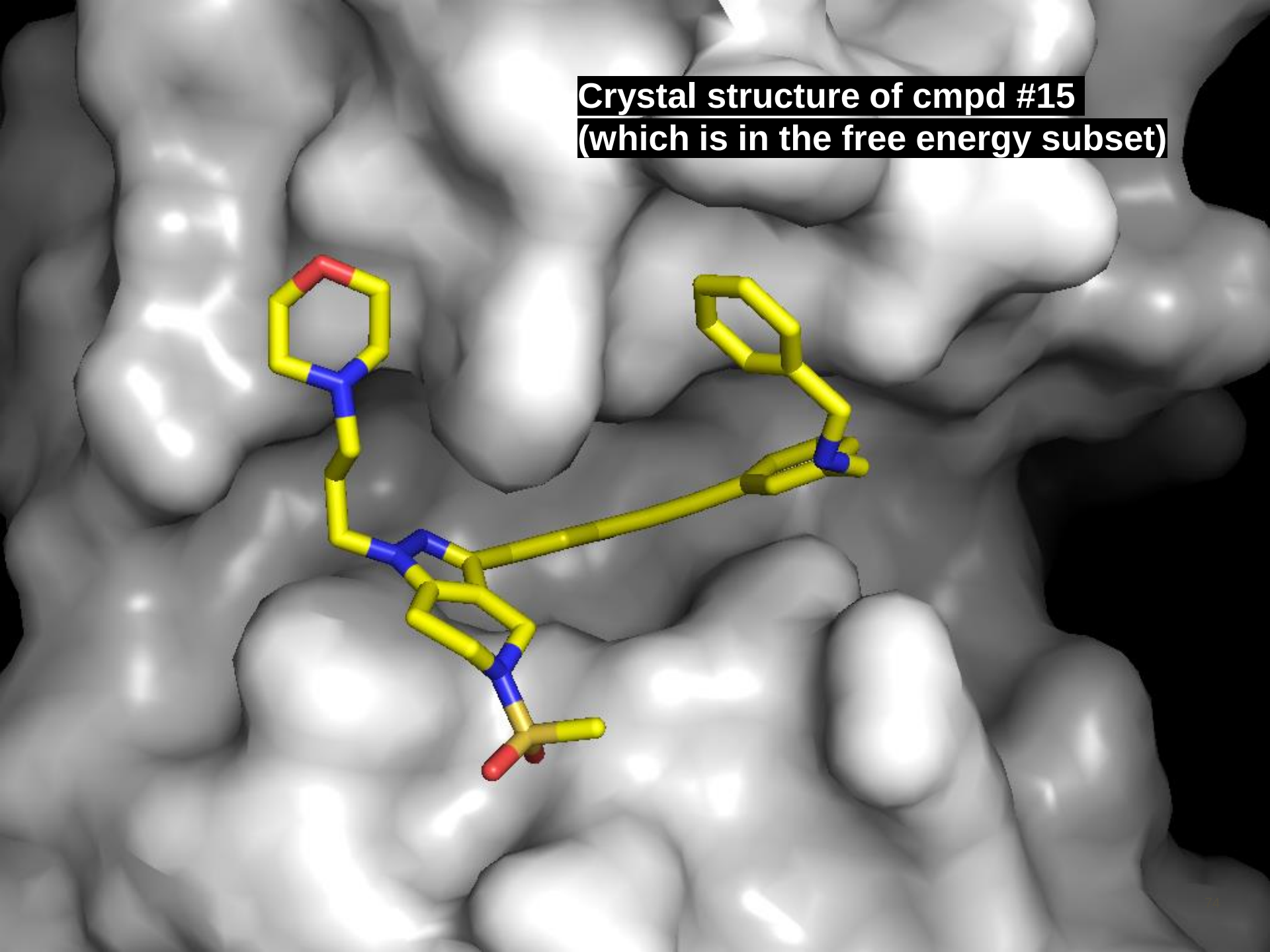


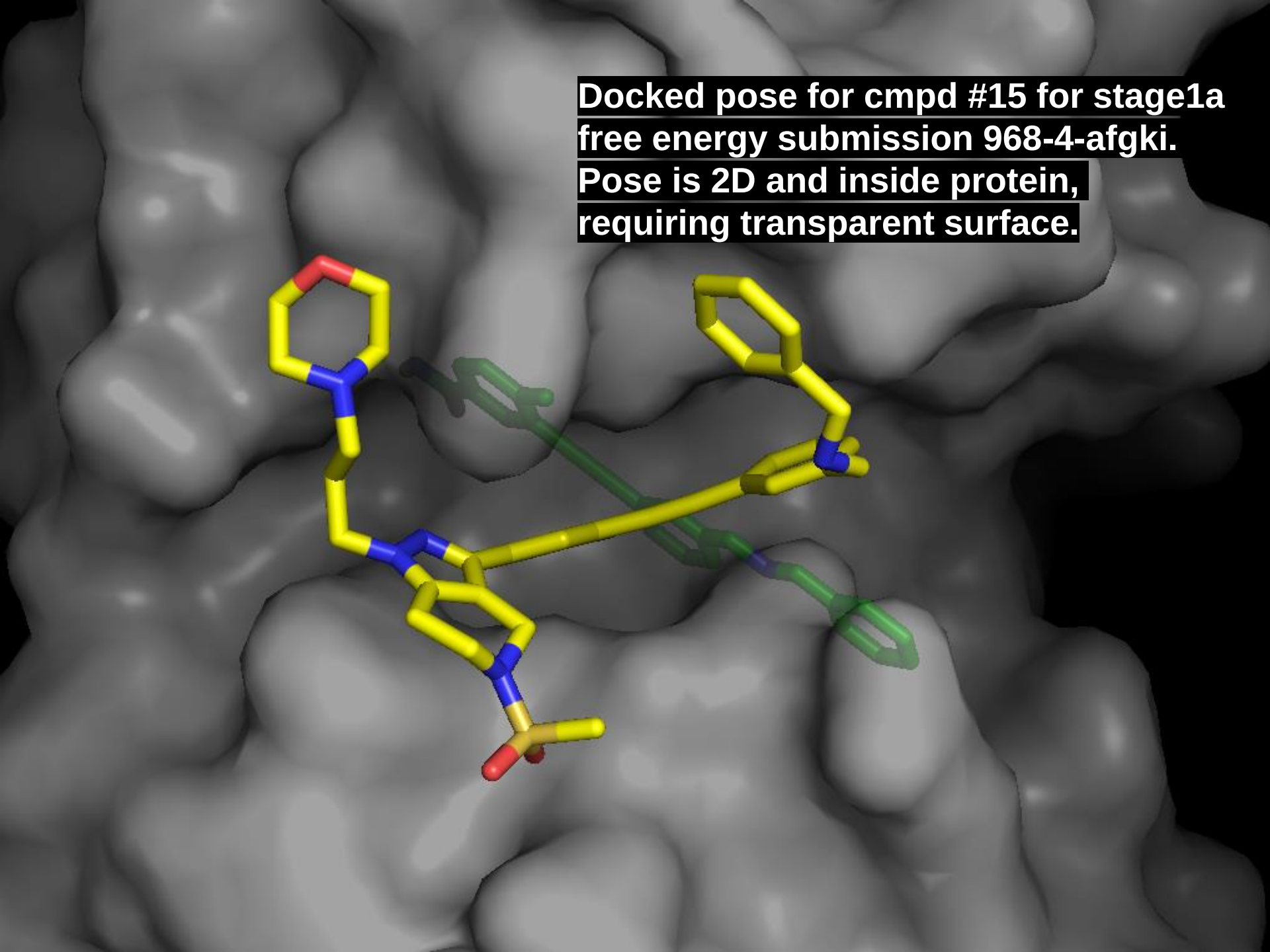
Free Energy subset, yellow=stage1a, blue=stage2



Top-scoring methods in the CatS Free Energy Subset

code	correl	state	method
tw62k	0.42	stage2	Machine learning 2D only no receptor info
66qqk	0.41	stage2	Machine learning 2D and 3D conformation no receptor info
grhvk	**0.41**	**stage2**	**gave journal reference but no explanation in text**
jk3no	0.41	stage2	Machine learning 2D and 3D conformation no receptor info
uch2m	0.39	stage2	Machine learning 2D only no receptor info
x3t42	0.36	stage1a	MM/PBSA on 100 docked poses
ytget	0.36	stage1a	Machine learning 2D only no receptor info
js3r3	**0.35**	**stage2**	**FEP**
02zo2	0.33	stage1a	knowledge-based scoring function
feofk	0.33	stage2	Machine learning 2D and 3D conformation no receptor info
afgki	0.28	stage1a	Machine learning 2D only no receptor info
io7fy	0.26	stage1a	MM/PBSA on 100 docked poses
wtfby	0.25	stage1a	MM/PBSA on 1500 snapshots from MD


Crystal structure of cmpd #15
(which is in the free energy subset)


Docked pose for cmpd #15 for stage1a
free energy submission 968-4-afgki.
Pose is 2D and inside protein,
requiring transparent surface.

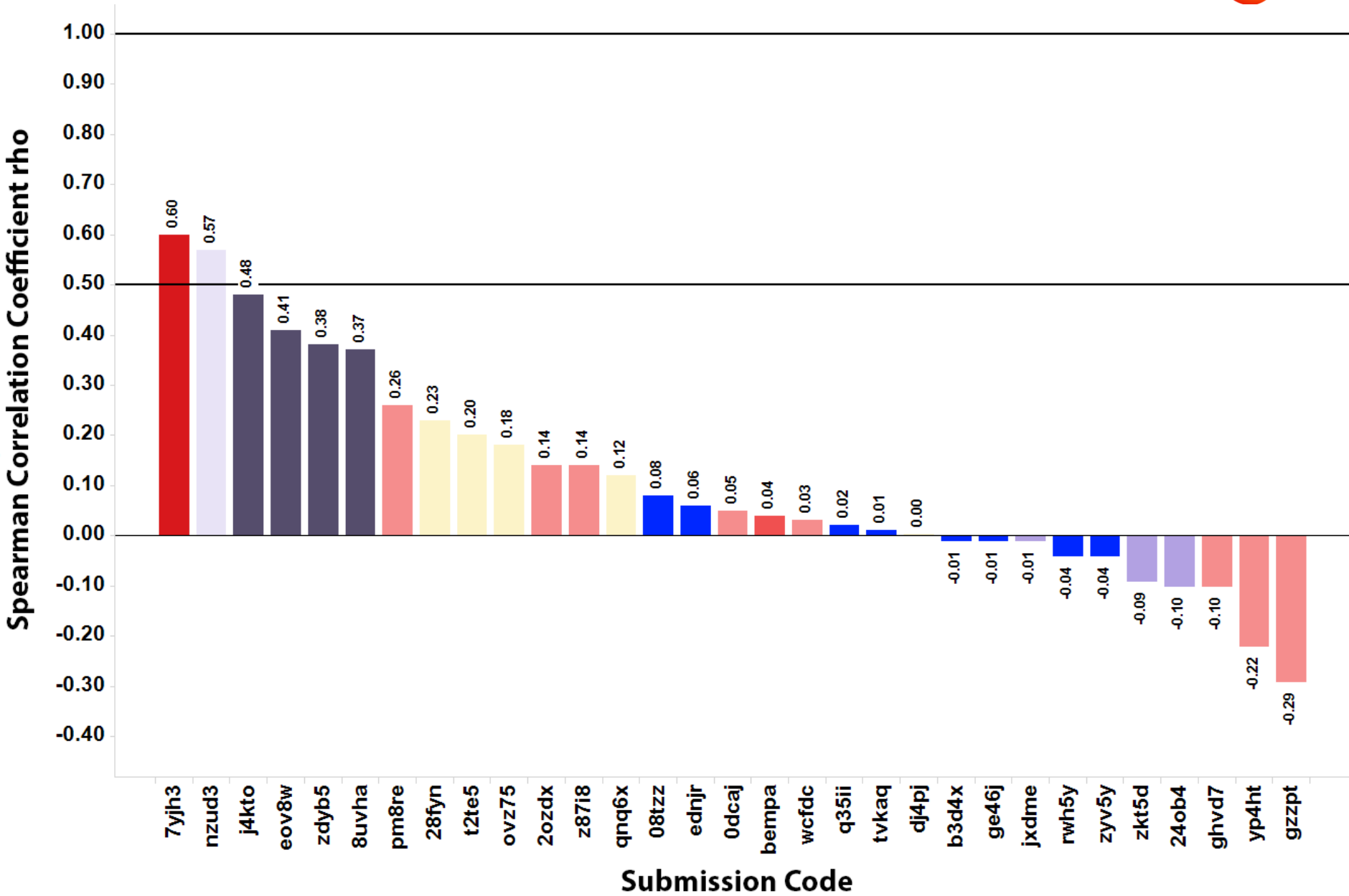
Top-scoring methods in the CatS Free Energy Subset

code	correl	state	method
tw62k	0.42	stage2	Machine learning 2D only no receptor info
66qqk	0.41	stage2	Machine learning 2D and 3D conformation no receptor info
grhvk	0.41	stage2	gave journal reference but no explanation in text
jk3no	0.41	stage2	Machine learning 2D and 3D conformation no receptor info
uch2m	0.39	stage2	Machine learning 2D only no receptor info
x3t42	0.36	stage1a	MM/PBSA on 100 docked poses
ytget	0.36	stage1a	Machine learning 2D only no receptor info
js3r3	0.35	stage2	FEP
02zo2	0.33	stage1a	knowledge-based scoring function
feofk	0.33	stage2	Machine learning 2D and 3D conformation no receptor info
afgki	**0.28**	**stage1a**	**Machine learning 2D only no receptor info**
io7fy	0.26	stage1a	MM/PBSA on 100 docked poses
wtfby	0.25	stage1a	MM/PBSA on 1500 snapshots from MD

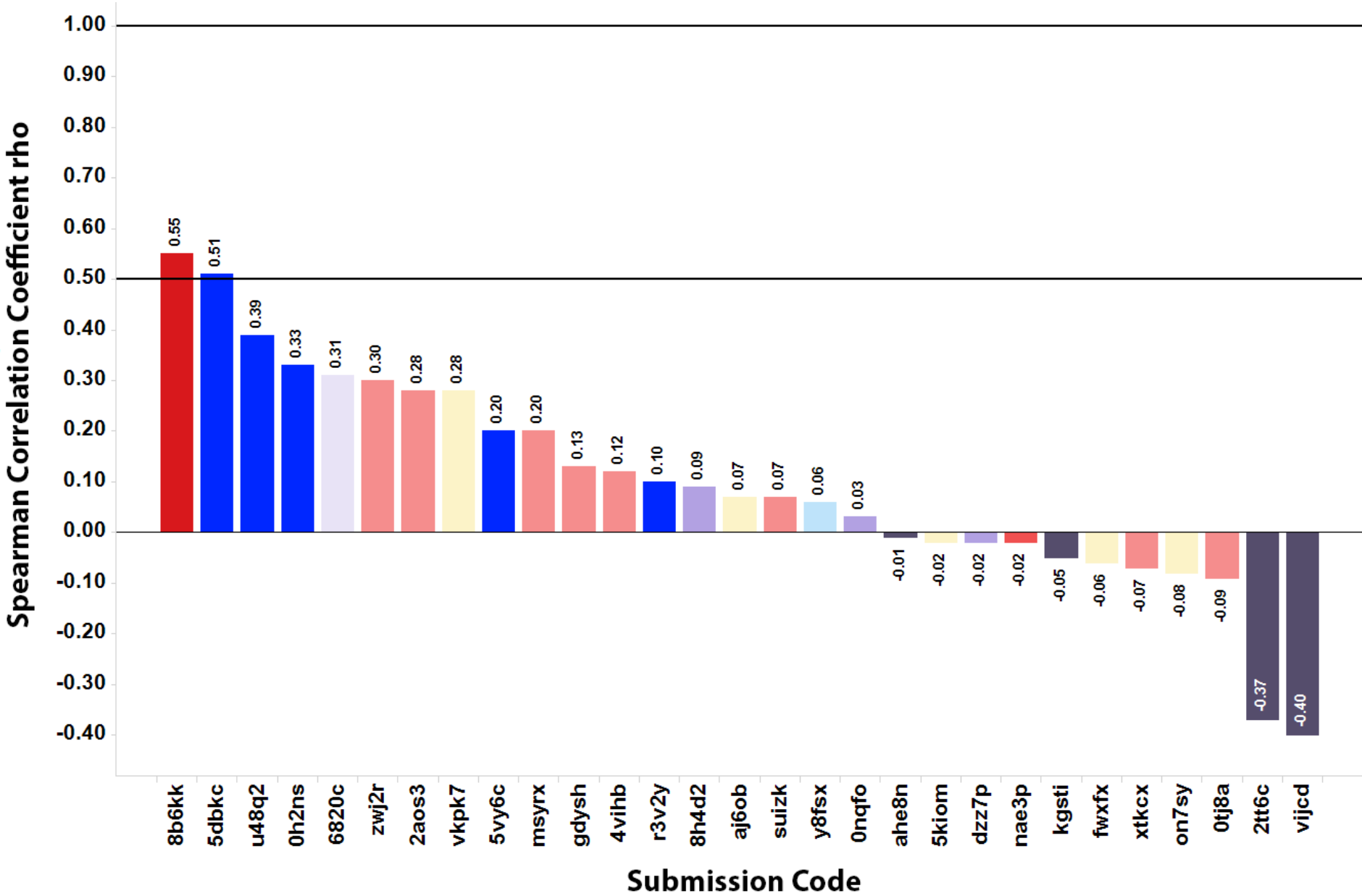
Subchallenges in Grand Challenge 3

challenge	target	ic50s	xrays	type	submissions	ic50-range nM
968-1	CatS	0	24	pose	52	
968-2	CatS	136	0	ligand-based	11	3.0-8520
968-3	CatS	136	0	structure-based	43	3.0-8520
968-4	CatS	33	0	free energy	35	
972-1	CatS	0	24	pose	47	
1009-2	CatS	136	0	ligand-based	9	3.0-8520
1009-3	CatS	136	0	structure-based	72	3.0-8520
1009-4	CatS	33	0	free energy	34	
969-2	**JAK2-SC2**	**89**	**0**	**ligand-based**	**3**	**0.66-10000**
969-3	**JAK2-SC2**	**89**	**0**	**structure-based**	**28**	**0.66-10000**
965-2	**p38a**	**72**	**0**	**ligand-based**	**3**	**0.23-10000**
965-3	**p38a**	**72**	**0**	**structure-based**	**26**	**0.23-10000**
966-2	**VEGFR**	**85**	**0**	**ligand-based**	**3**	**0.62-10000**
966-3	**VEGFR**	**85**	**0**	**structure-based**	**31**	**0.62-10000**
970-3	JAK2-SC3	17	0	structure-based	18	53.0-10000
970-4	JAK2-SC3	17	0	free energy	7	53.0-10000
967-3	TIE2	18	0	structure-based	18	3.4-10000
967-4	TIE2	4	0	free energy1	7	200.0-10000
967-5	TIE2	6	0	free energy2	7	3.4- 3200
971-3	ABL1	12	0	mutagenesis	11	49.0-10000
		1176	**24**		**465**	

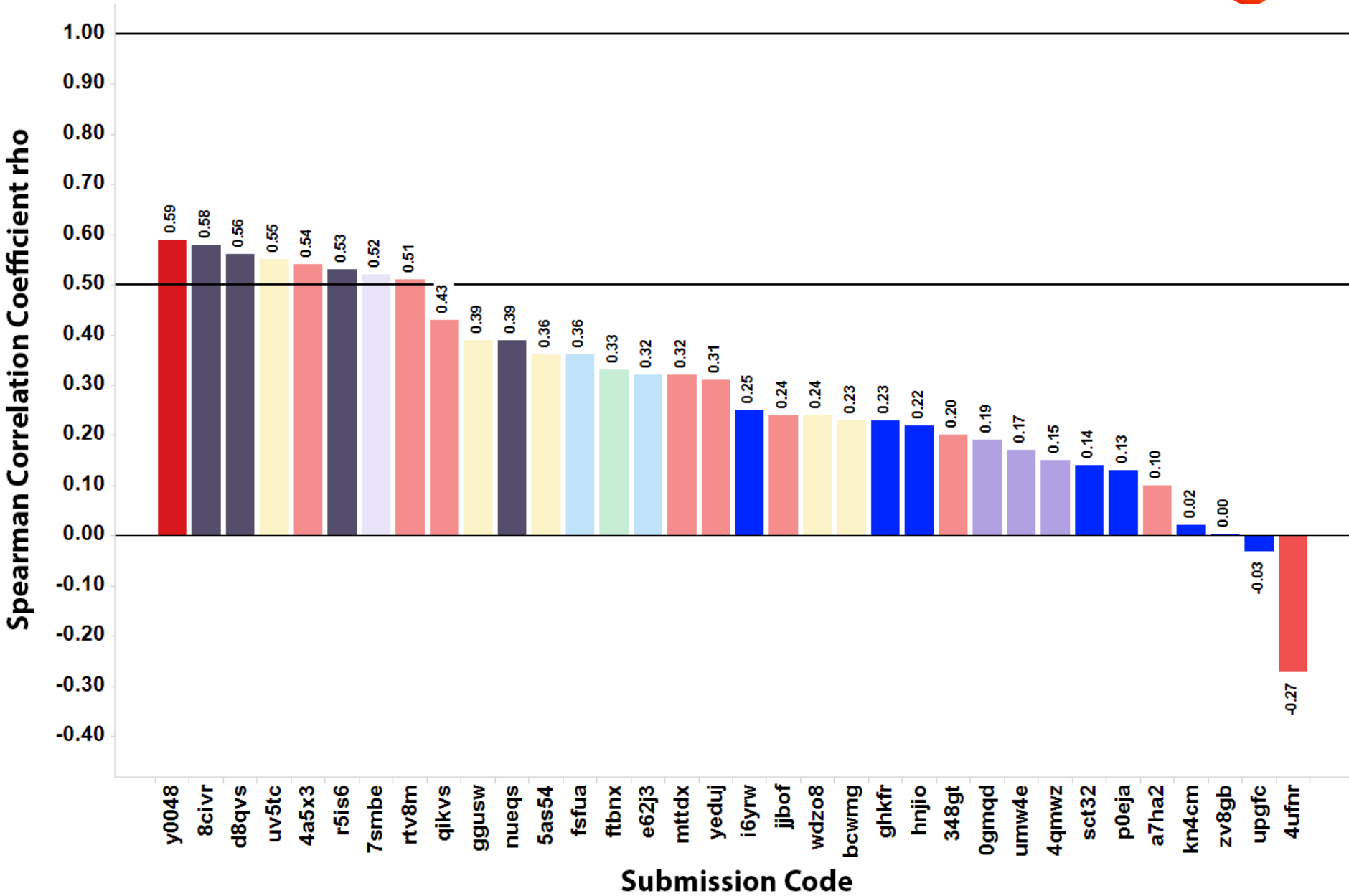
Correlations for SC2 JAK2 affinities, color by submitter



Correlations for SC2 p38 affinities, color by submitter



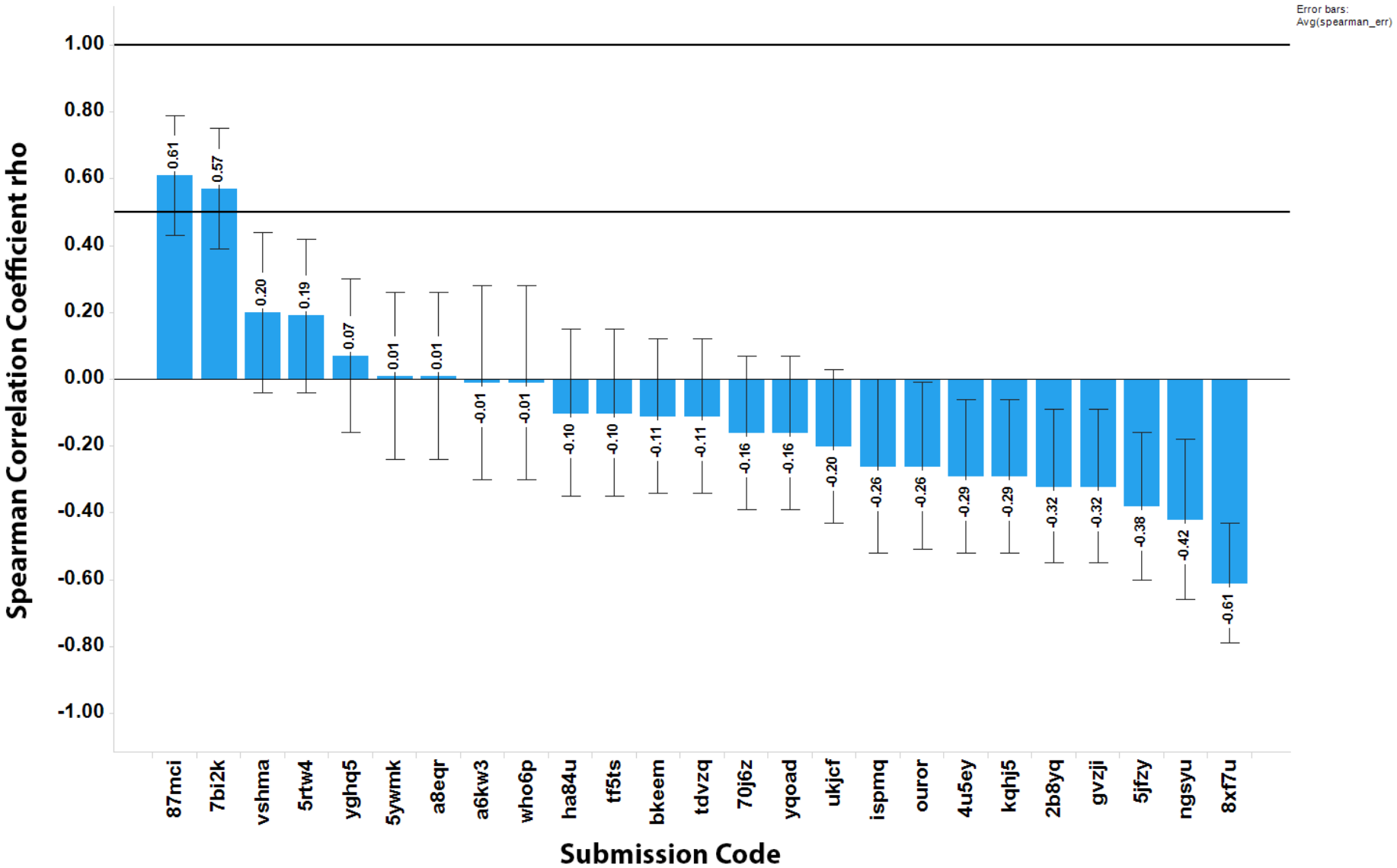
Correlations for SC2 VEGFR affinities, color by submitter



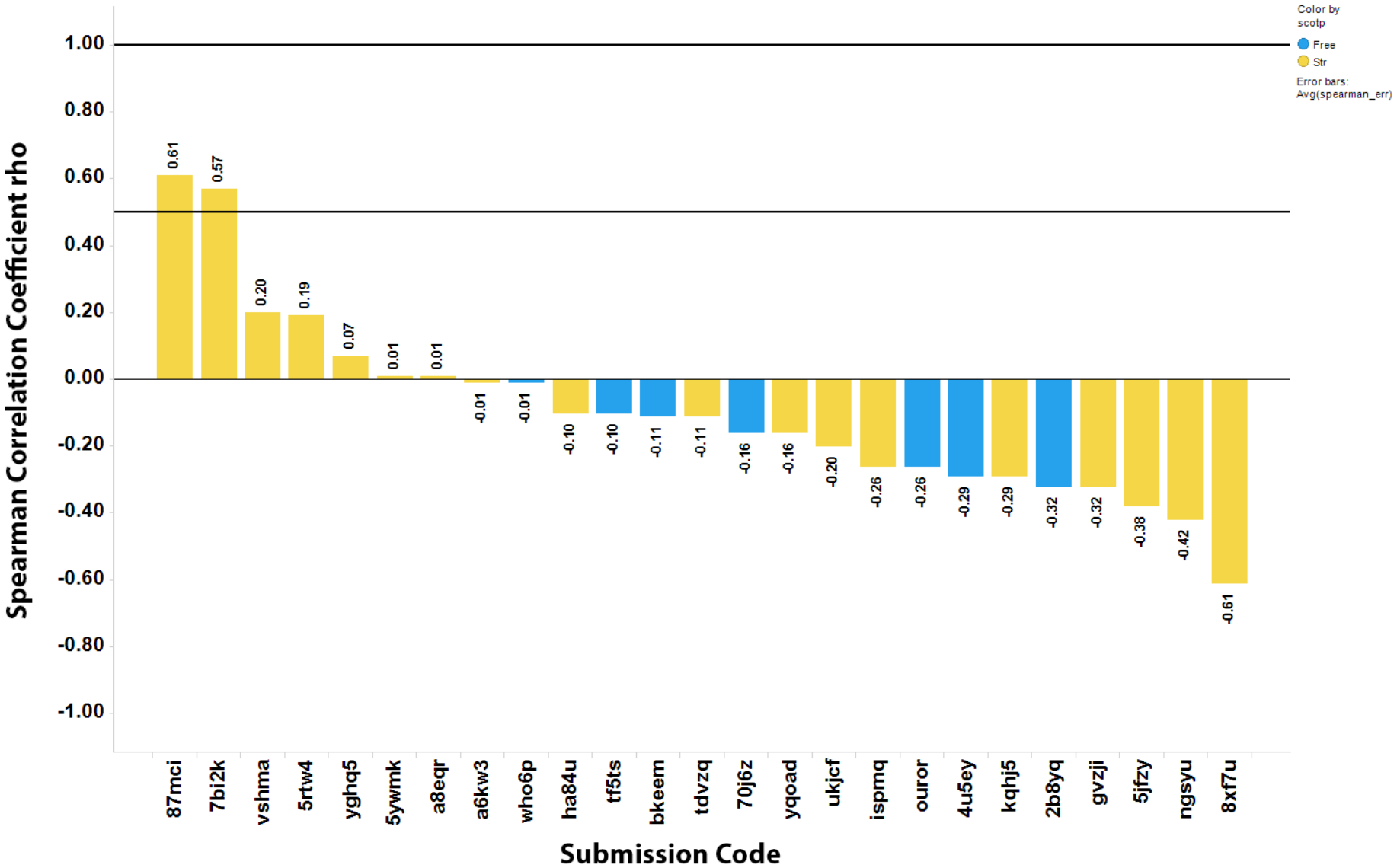
Subchallenges in Grand Challenge 3

challenge	target	ic50s	xrays	type	submissions	ic50-range nM
968-1	CatS	0	24	pose	52	
968-2	CatS	136	0	ligand-based	11	3.0-8520
968-3	CatS	136	0	structure-based	43	3.0-8520
968-4	CatS	33	0	free energy	35	
972-1	CatS	0	24	pose	47	
1009-2	CatS	136	0	ligand-based	9	3.0-8520
1009-3	CatS	136	0	structure-based	72	3.0-8520
1009-4	CatS	33	0	free energy	34	
969-2	JAK2-SC2	89	0	ligand-based	3	0.66-10000
969-3	JAK2-SC2	89	0	structure-based	28	0.66-10000
965-2	p38a	72	0	ligand-based	3	0.23-10000
965-3	p38a	72	0	structure-based	26	0.23-10000
966-2	VEGFR	85	0	ligand-based	3	0.62-10000
966-3	VEGFR	85	0	structure-based	31	0.62-10000
970-3	**JAK2-SC3**	**17**	**0**	**structure-based**	**18**	**53.0-10000**
970-4	**JAK2-SC3**	**17**	**0**	**free energy**	**7**	**53.0-10000**
967-3	TIE2	18	0	structure-based	18	3.4-10000
967-4	TIE2	4	0	free energy1	7	200.0-10000
967-5	TIE2	6	0	free energy2	7	3.4- 3200
971-3	ABL1	12	0	mutagenesis	11	49.0-10000
		1176	**24**		**465**	

JAK2 subchallenge 3, activity cliff



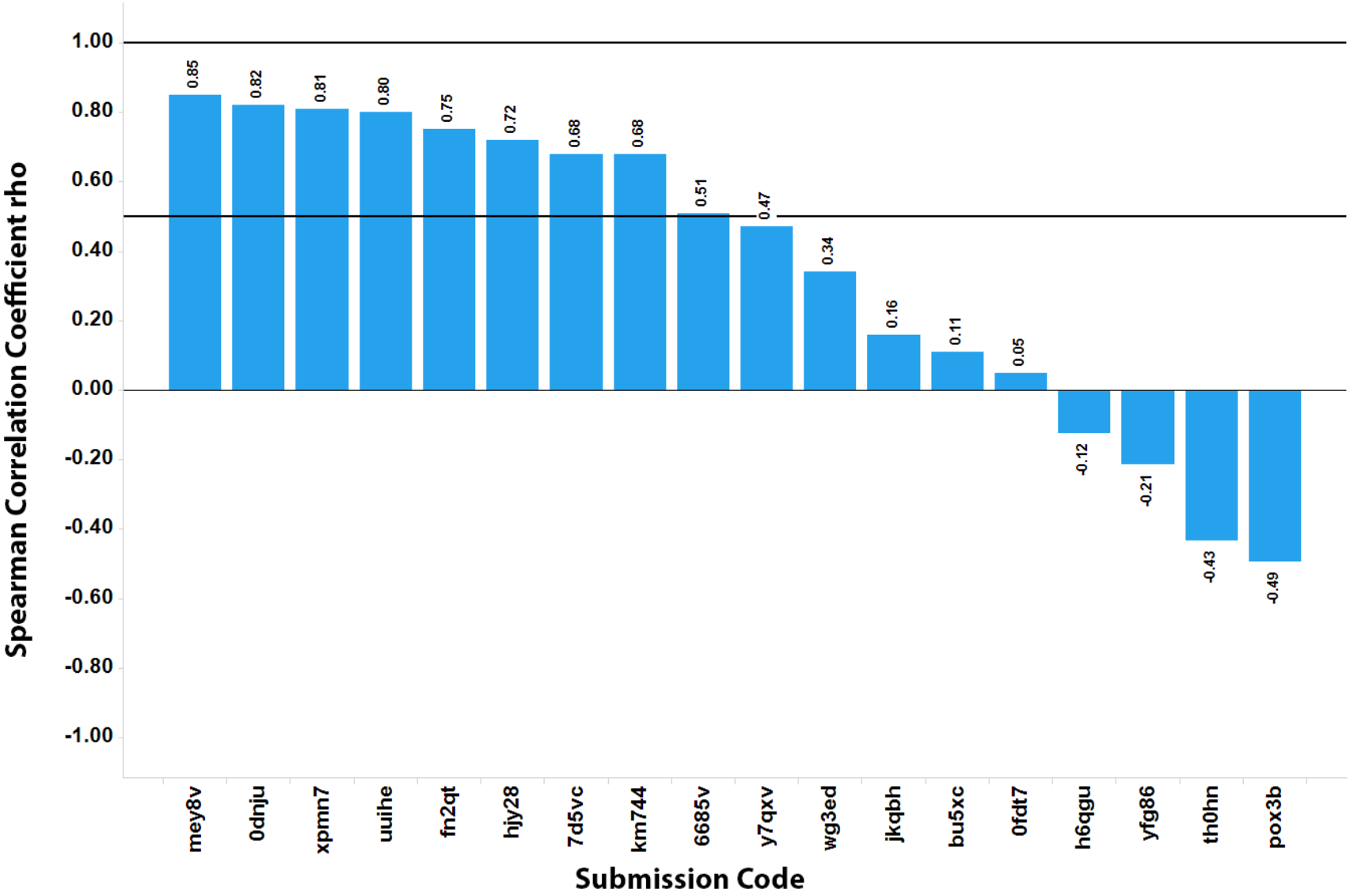
JAK2 SC3, yellow=structure-based, blue=free-energy



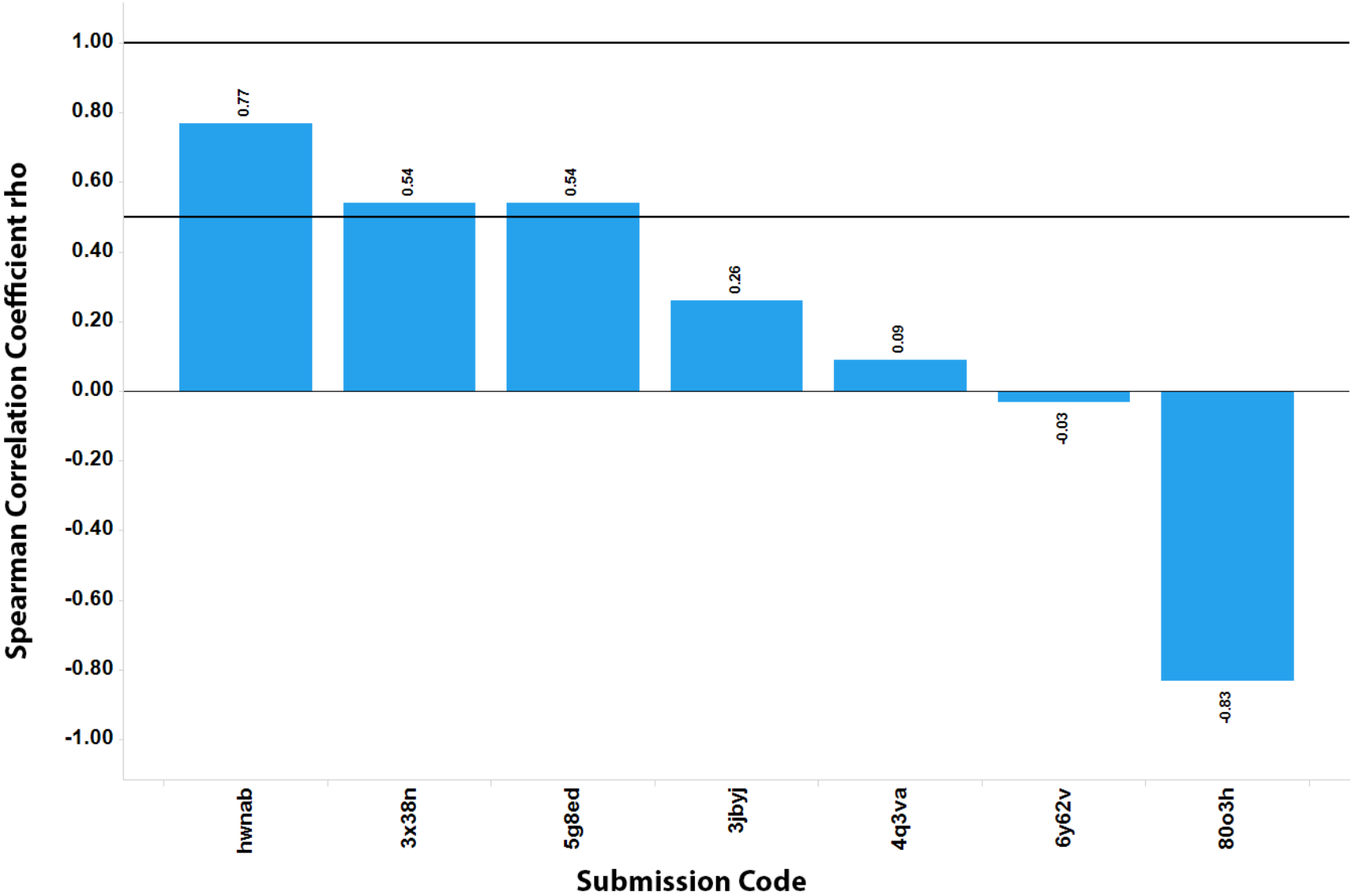
Subchallenges in Grand Challenge 3

challenge	target	ic50s	xrays	type	submissions	ic50-range nM
968-1	CatS	0	24	pose	52	
968-2	CatS	136	0	ligand-based	11	3.0-8520
968-3	CatS	136	0	structure-based	43	3.0-8520
968-4	CatS	33	0	free energy	35	
972-1	CatS	0	24	pose	47	
1009-2	CatS	136	0	ligand-based	9	3.0-8520
1009-3	CatS	136	0	structure-based	72	3.0-8520
1009-4	CatS	33	0	free energy	34	
969-2	JAK2-SC2	89	0	ligand-based	3	0.66-10000
969-3	JAK2-SC2	89	0	structure-based	28	0.66-10000
965-2	p38a	72	0	ligand-based	3	0.23-10000
965-3	p38a	72	0	structure-based	26	0.23-10000
966-2	VEGFR	85	0	ligand-based	3	0.62-10000
966-3	VEGFR	85	0	structure-based	31	0.62-10000
970-3	JAK2-SC3	17	0	structure-based	18	53.0-10000
970-4	JAK2-SC3	17	0	free energy	7	53.0-10000
967-3	**TIE2**	**18**	**0**	**structure-based**	**18**	**3.4-10000**
967-4	**TIE2**	**4**	**0**	**free energy1**	**7**	**200.0-10000**
967-5	**TIE2**	**6**	**0**	**free energy2**	**7**	**3.4- 3200**
971-3	**ABL1**	**12**	**0**	**mutagenesis**	**11**	**49.0-10000**
		1176	**24**		**465**	

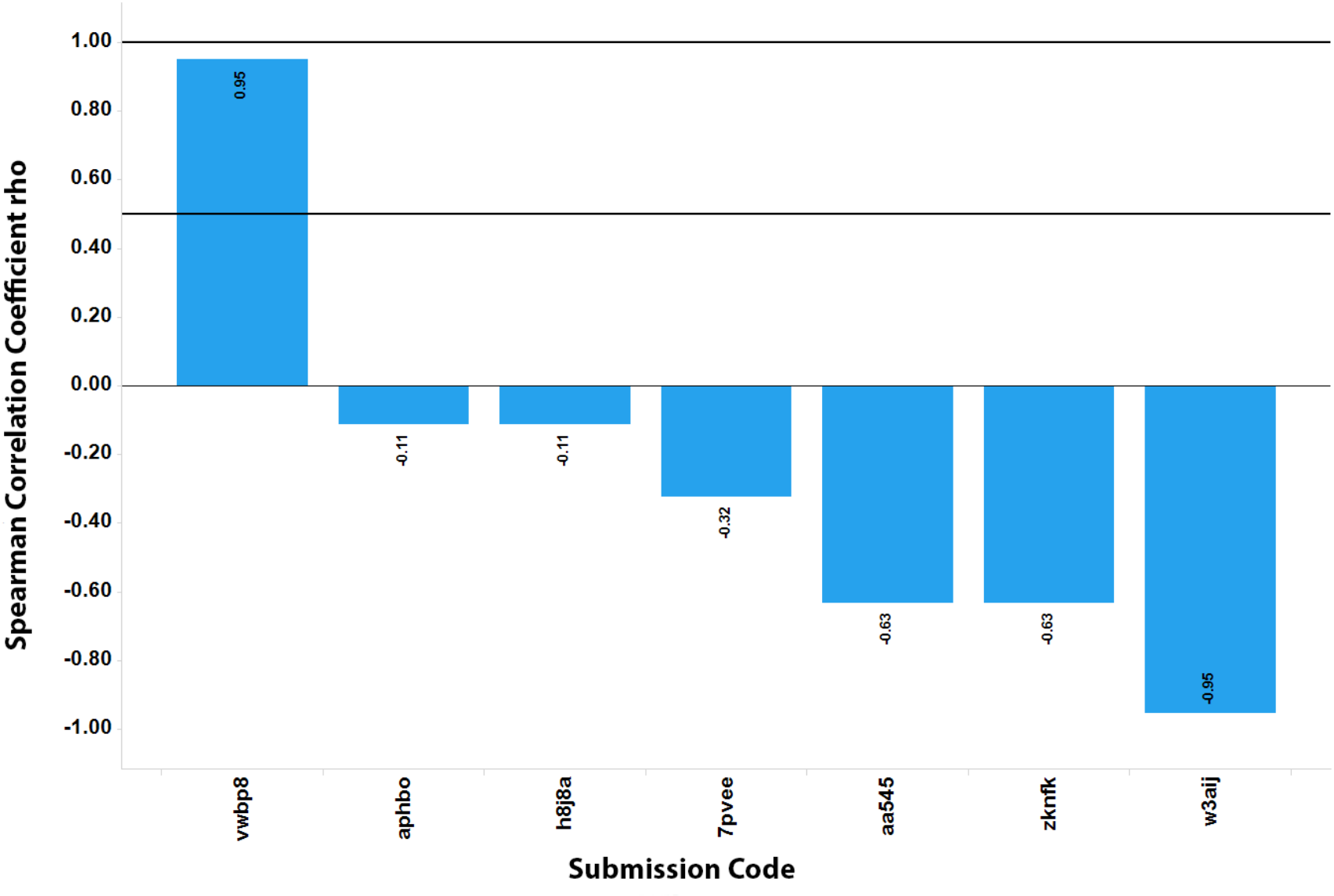
Affinity correlations for TIE2, str/lig-based set of 18 cmpds



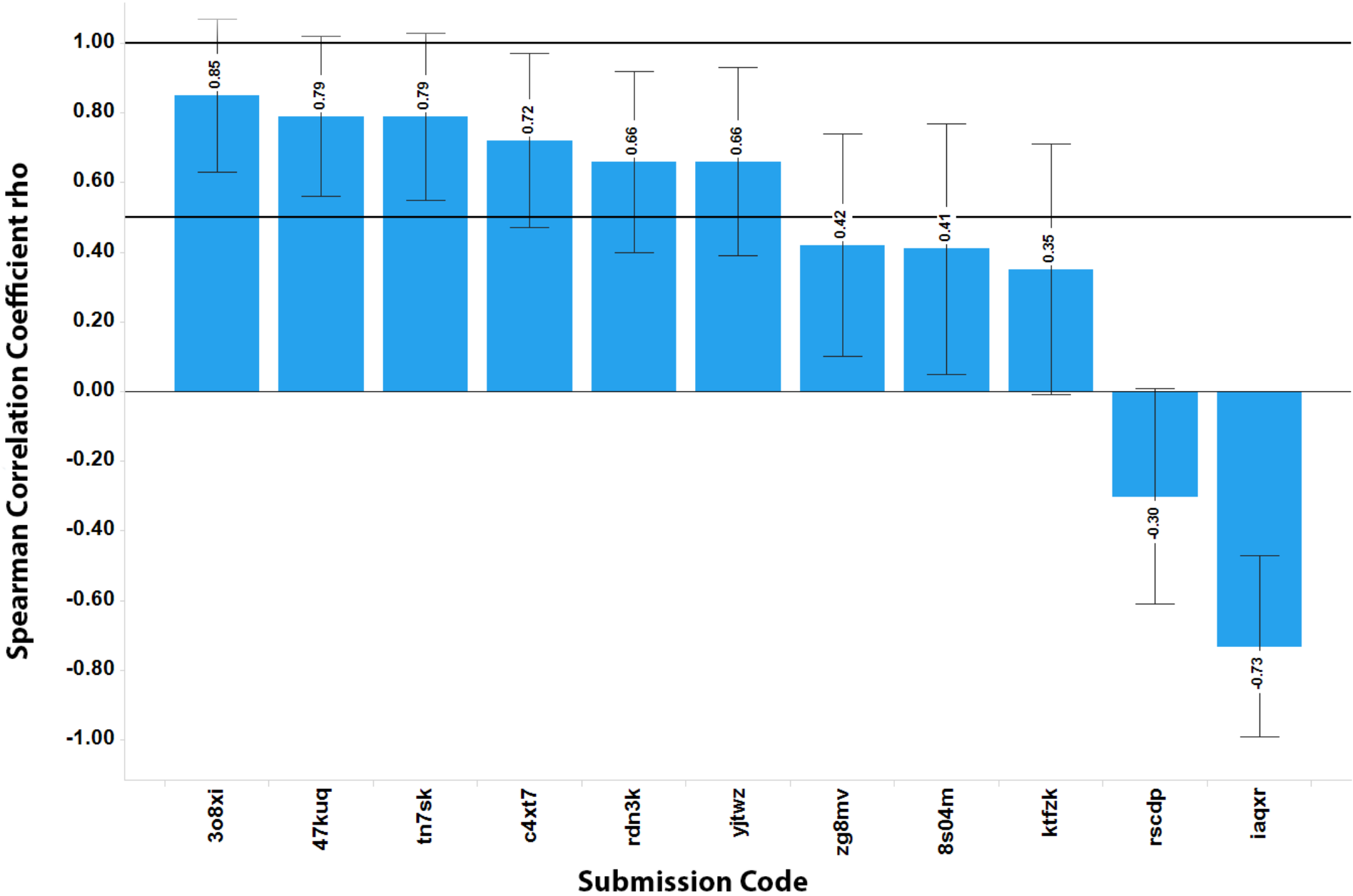
Affinity correlations for TIE2, free energy set of 6 cmpds



Affinity correlations for TIE2, free energy set of 4 cmpds



ABL1 mutagenesis challenge



Conclusions

Docking and Scoring

- To get the pose right, most (all?) participants needed to use prior knowledge
- Several participants predicted the flipped pose correctly (no prior knowledge!)
- In scoring, ligand-based and machine-learning methods are beating many structure-based methods and Free Energy Perturbation.

Suggestions for Future D3R Evaluations

- Participants with multiple predictions should identify their “best” prediction
- Metadata: rather than giving protocol in paragraph form, the submission process should break out workflow steps, have submitters cite software used in each step
- Submission process should lay out categories of methods, such as MM, MD, FEP, machine learning, QSAR, and participants should specify the category.
- Submission process should link poses and affinities more tightly

Acknowledgements

– Mike Gilson and Rommie Amaro
– Conor Parks
– Zied Gaieb
– Mike Chiu
– Pat Walters (Relay Therapeutics)

– Jannsen Pharma (for the CatS dataset)
– Bill Zuercher, David Drewry, Tim Willson (kinase dataset)