Function Assessment

Megan Egbert[1]
Dzmitry Padhorny[2]
Israel Desta[1]
Usman Ghani[1]

Dima Kozakov[2]
Sandor Vajda[1]

[1]Boston University
[2]Stony Brook University

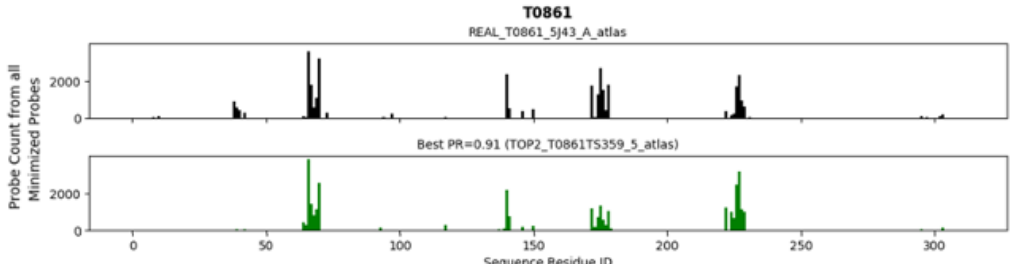
Focus on Binding

How Well the Binding Properties of the Targets Are Conserved in the Models?

1. Binding of small molecules (fragments)

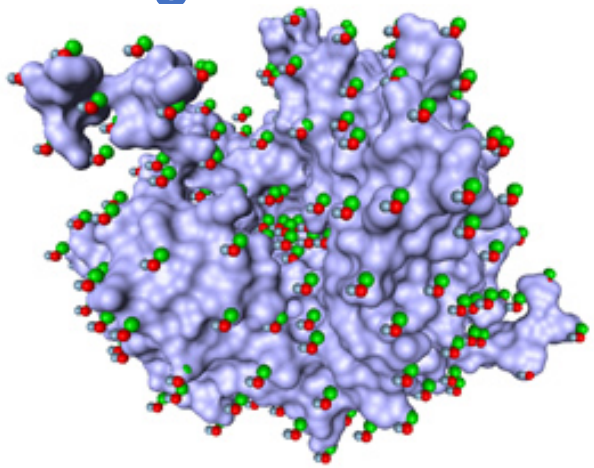
Method: Mapping the protein surface using small molecules as probes (Based on soaking protein crystals in aqueous solutions of organic solvents)

A. Global analysis: Compare binding fingerprints



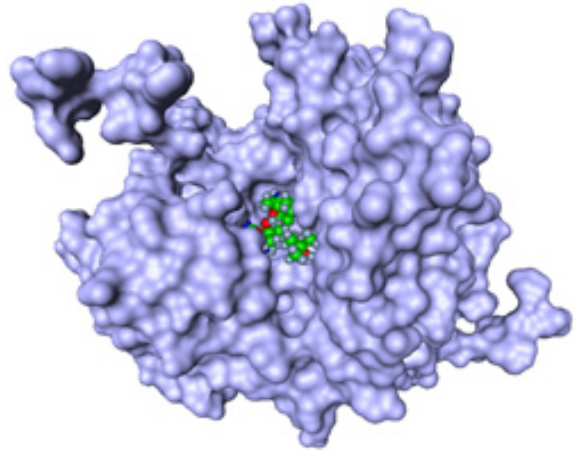
Target

Model



Advantage: Applicable to all targets

B. Local Analysis: Probe clusters identify binding hot spots

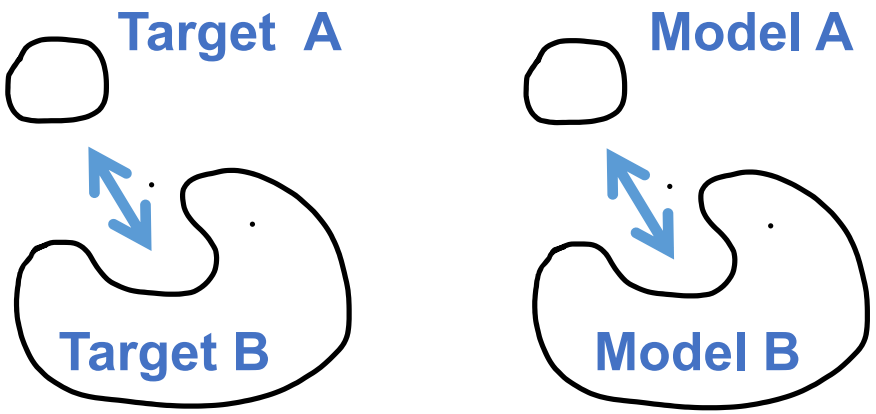


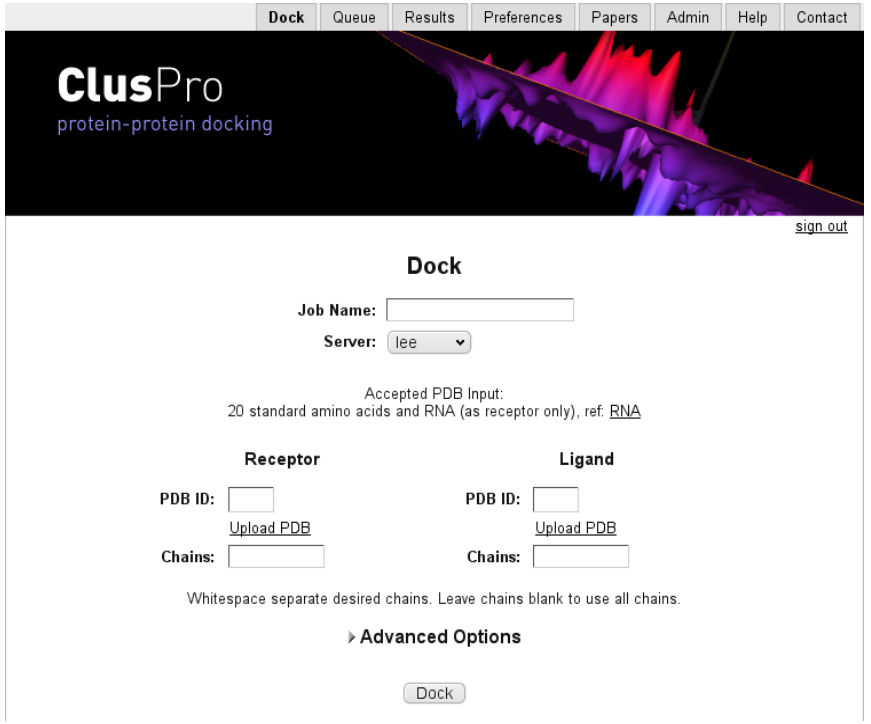
Locating binding sites

Determining binding site druggability

Fragment based drug discovery

2. Conservation of protein-protein Interactions:
Docking target A to target B versus
docking models of target A to models of target B





13 targets are parts of larger assemblies
Evaluation is based on CAPRI criteria

Direct docking using ClusPro

3. Analysis of specific protein – ligand interactions:
Docking known small ligands to targets and to their models

FTMap
computational solvent mapping

BOSTON UNIVERSITY

CASP14 Functional Assessment

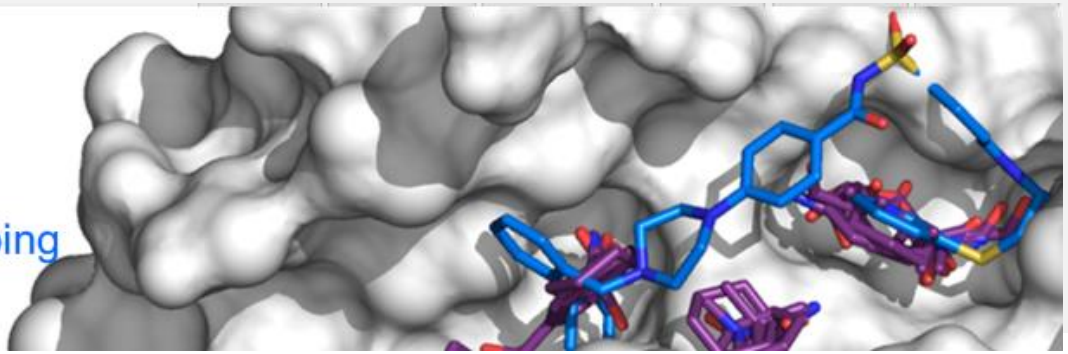
Conservation of Binding Properties in Modeled Proteins

Megan Egbert

Dr. Sandor Vajda

FTMap

computational solvent mapping

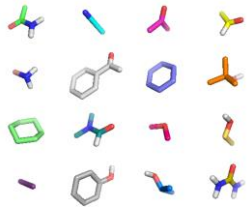


Computational alternative to MSCS (X-ray) and SAR by NMR

FFT Sampling of Probes on Grid → Energy Minimization of 2000 Best-Scoring Probe Orientations → Energy-Based & Proximity-Based Clustering → 

Crossclusters represent binding hot spots

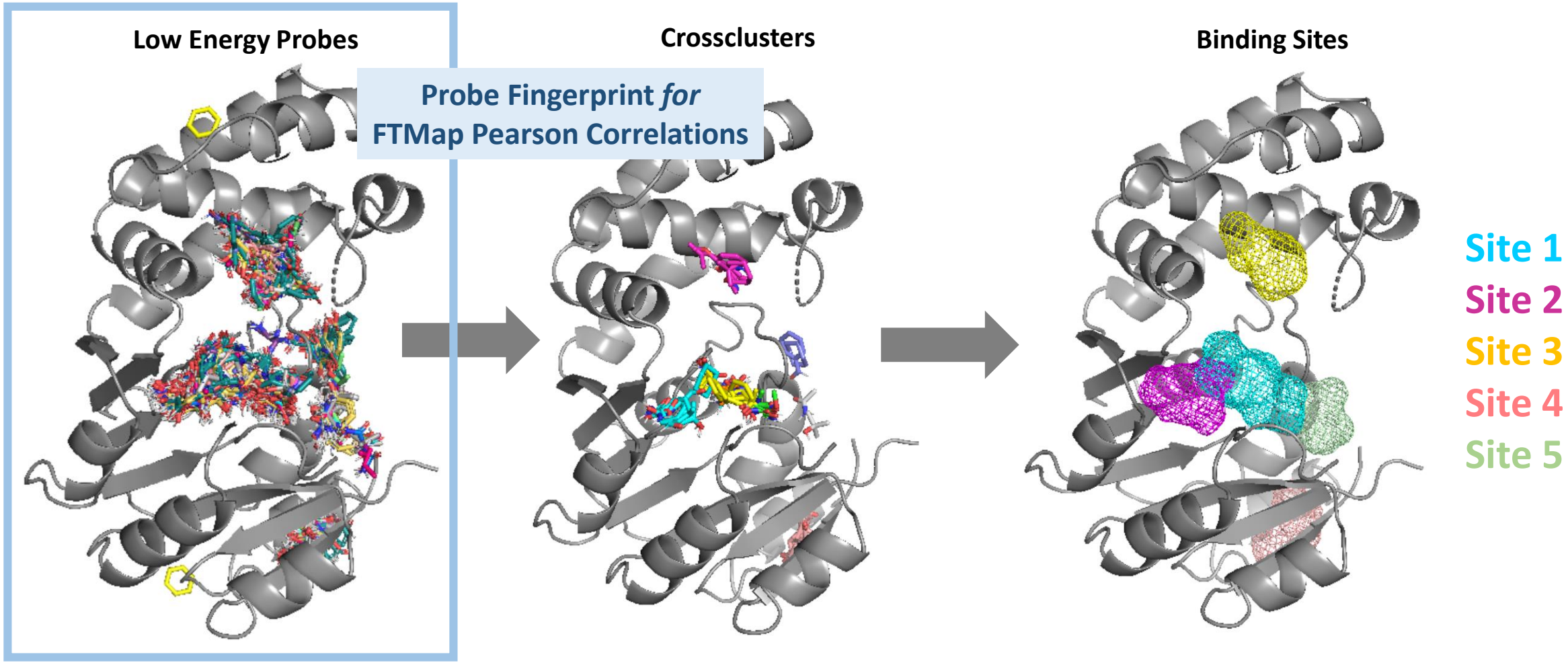
Repeat for 16 probe types



Brenke, R. et. al., (2009). *Bioinformatics*, *25*(5), 621-627. https://ftmap.bu.edu

FTMap Binding Sites

for evaluation of structure models

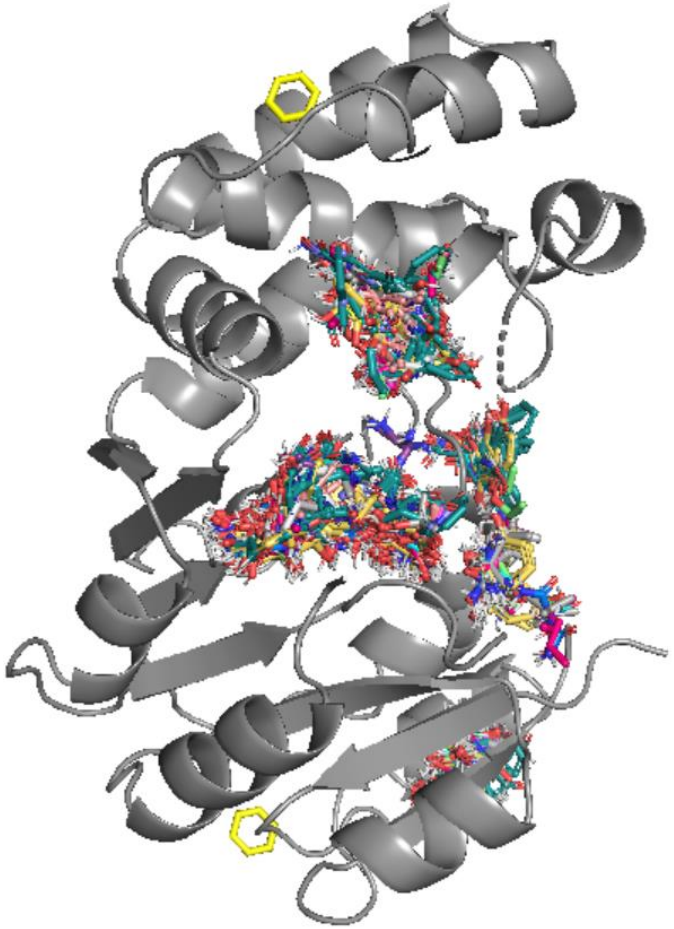


Brenke, R. et. al., (2009). *Bioinformatics*, *25*(5), 621-627.
Ngan, C. H., et. al., (2012). *Bioinformatics*, *28*(2), 286-287.

Probe Fingerprint

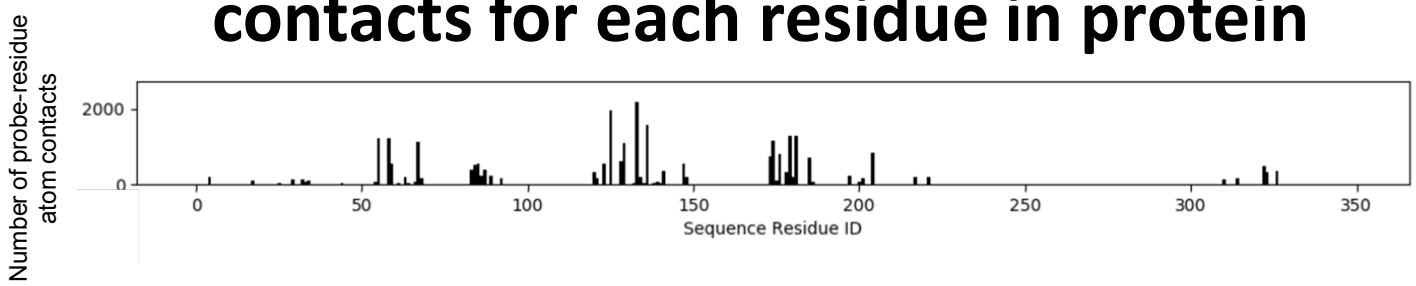
for evaluation of structure models

16 X 2000 lowest energy probes



- Consider all residues in the protein
- Consider all minimized probes, ~32,000
- Any probe-atom within 3 Å of a residue-atom == probe-residue contact

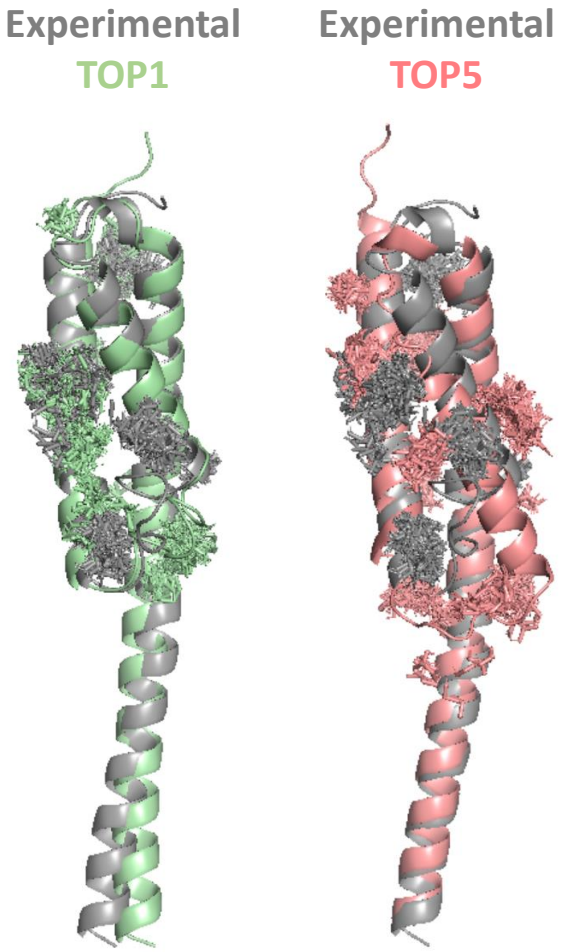
Fingerprint: Vector of # of probe-residue contacts for each residue in protein



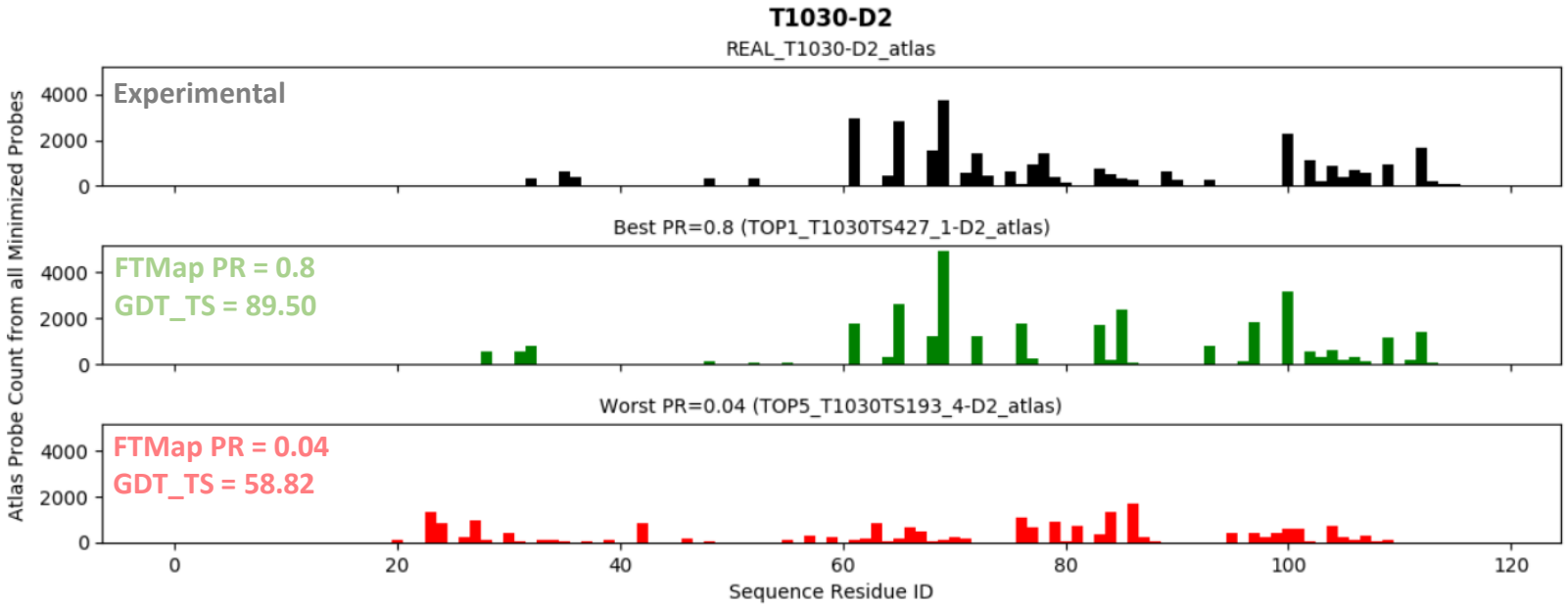
Brenke, R. et. al., (2009). *Bioinformatics*, *25*(5), 621-627.

FTMap Pearson Correlation

for evaluation of structure models

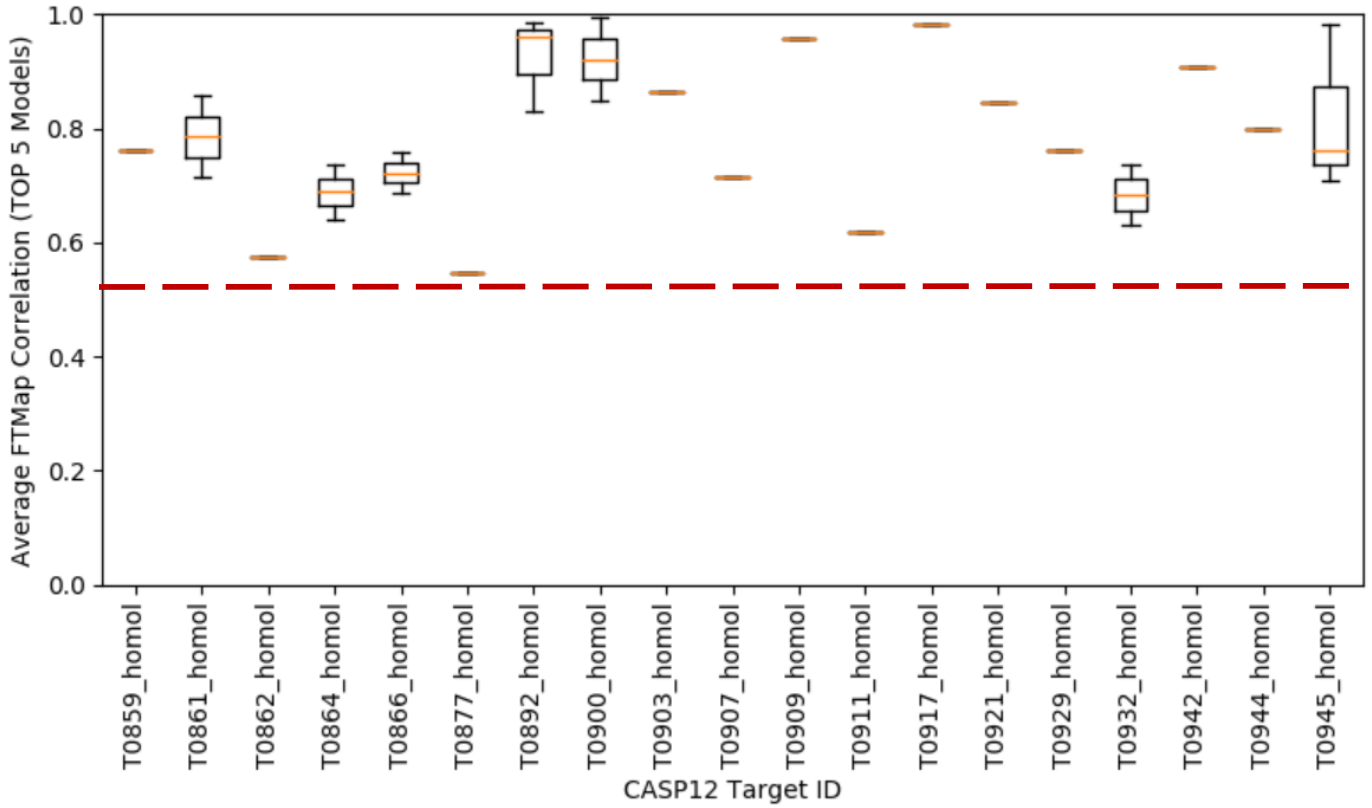


- Take Top 5 models from CASP14 Ranking (based on GDT_TS)
- Compare Probe fingerprint of model to experimental structure



FTMap Correlation of X-Ray Homologs

CASP12 Targets with homologs published in the PDB (> 95% sequence identity)

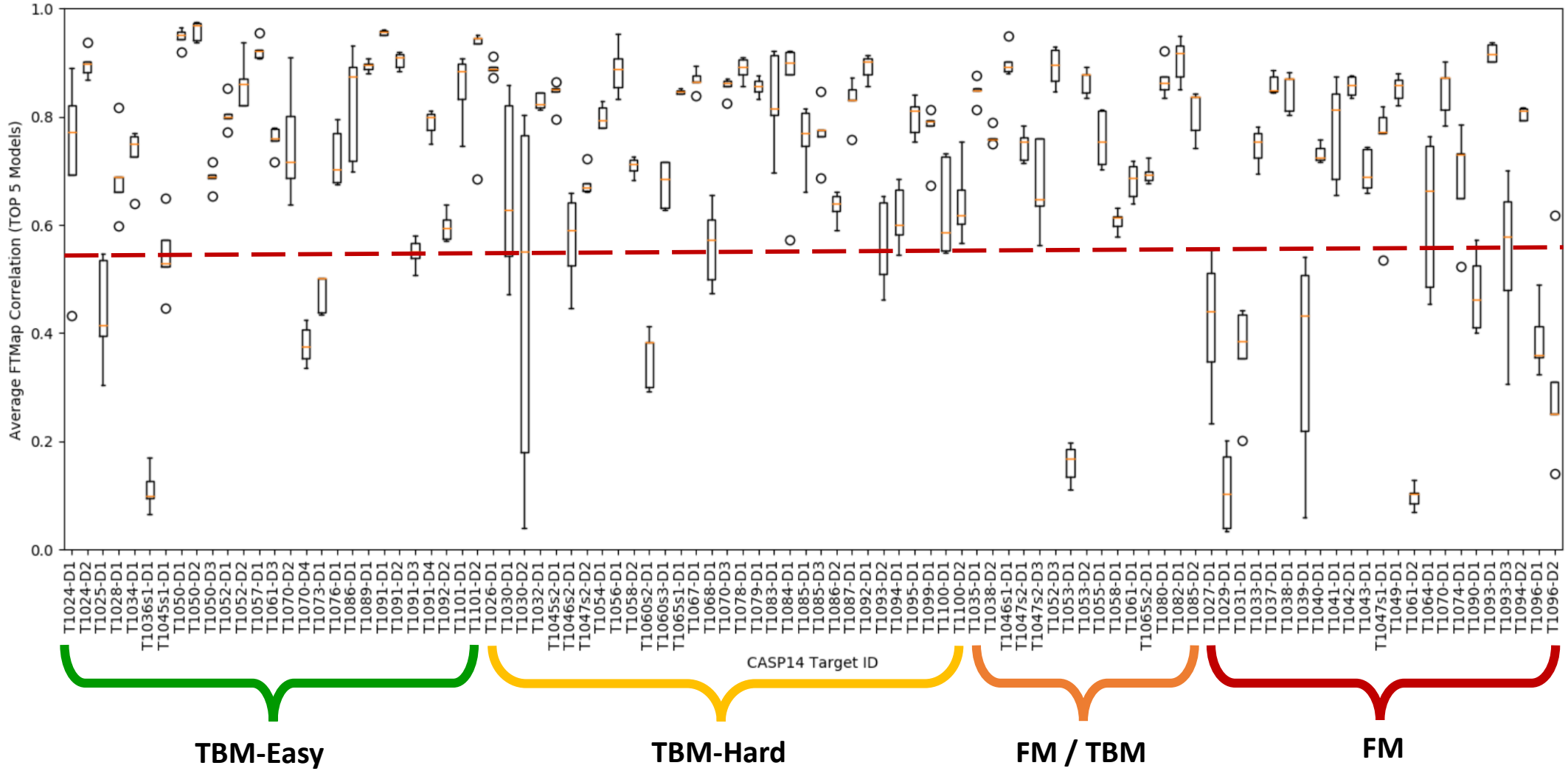


CASP12 Homologs identified by BLAST search
Sequence Similarity & Coverage ~ > 95%

Target	PDB ID	Alternate PDB IDs
T0859	5JZR_A	5FS4_A
T0861	5J43_A	1FCJ_A, 1D6S_A
T0862	5J43_B	5J5V_B
T0864	5D9G_A	5W0W_B, 5W0X_A
T0866	5UW2_A	7CGE_J, 6XBD_A
T0877	5NSJ_A	6HJ4_A
T0892	5NV4_A	5MU1_A, 6TS2_A, 6TRT_A
T0900	5AOT_A	5AOS_A, 5FU4_A
T0903	5A7D_B	4A1S_A
T0907	7CMG_A	6EY5_A
T0909	5G5N_A	5G5O_A
T0911	6E9N_A	6E9O_A
T0917	5YVR_A	5YVM_A
T0921	5AOZ_A	5M2O_A
T0942	6AIT_A	6SAR_A
T0944	5KO9_A	6OE7_A
T0945	6BW6_A	6FM9_A, 6BW5_A, 5LEV_A

CASP 14 Top5 Average FTMap Correlation

by target ID (96 domain-split targets)



78 Targets have FTMap PR >= 0.55

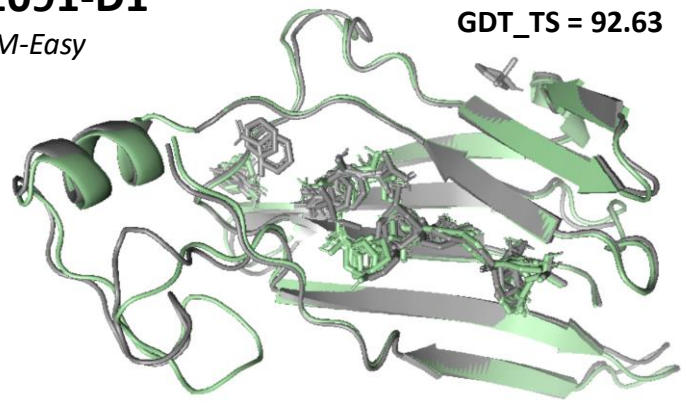
18 targets have FTMap PR < 0.55

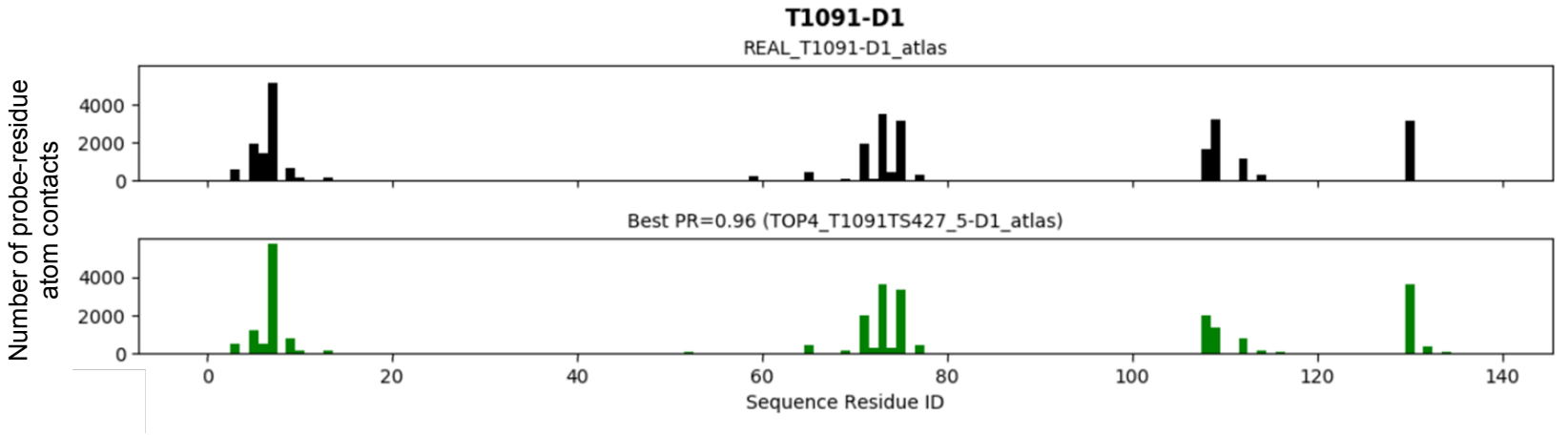
High FTMap Correlation Targets

T1091-D1

TBM-Easy

GDT_TS = 92.63

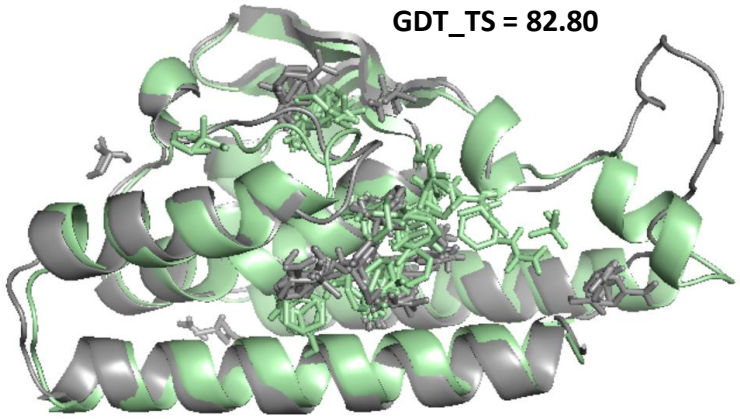


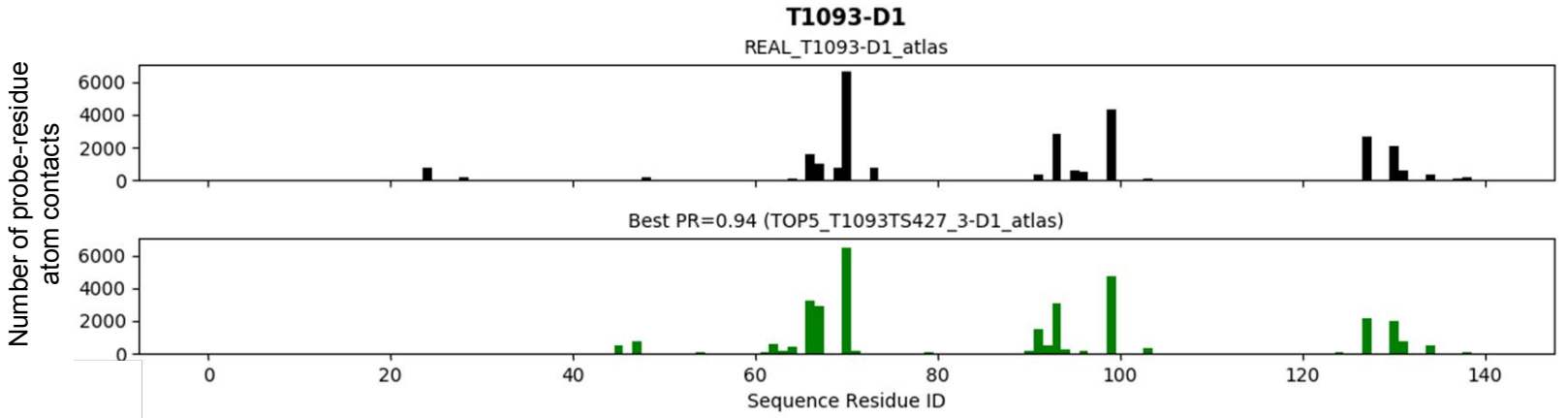


T1093-D1

FM

GDT_TS = 82.80

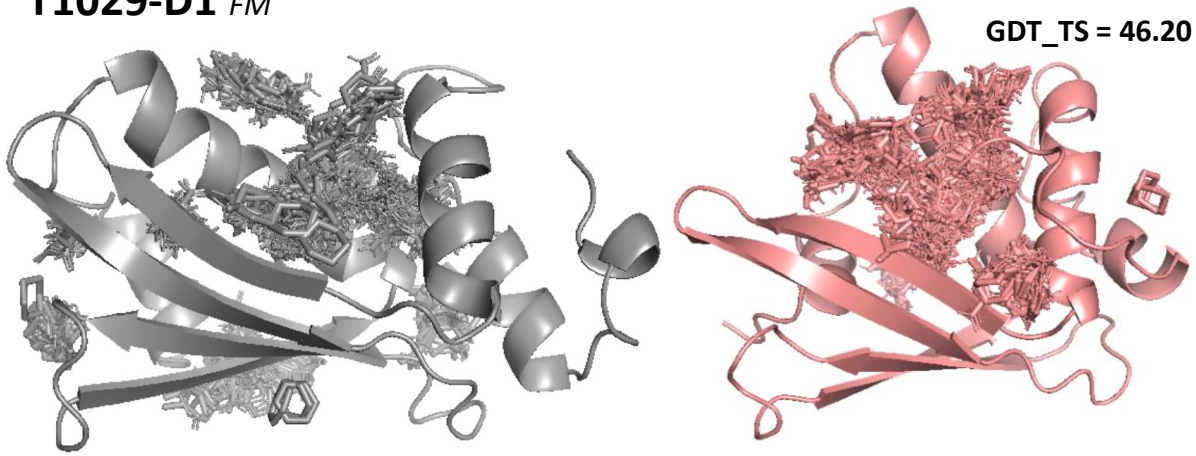




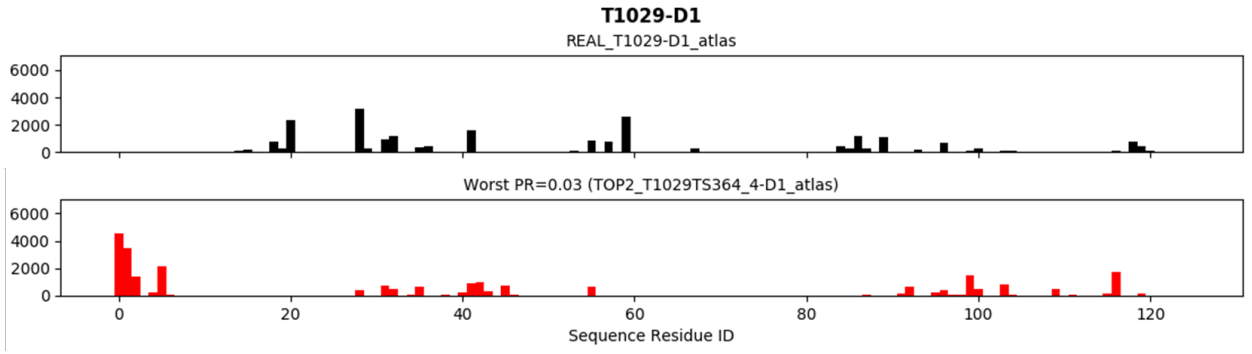
*probe cluster centers are shown in images for simplicity, rather than all ~32,000 low energy probes

Poor FTMap Correlation

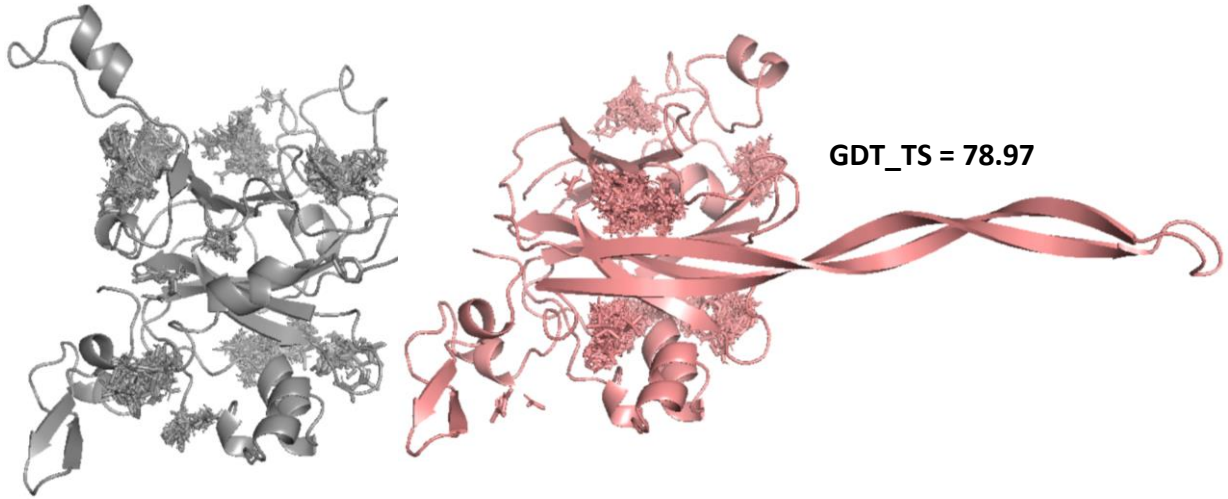
T1029-D1 *FM*



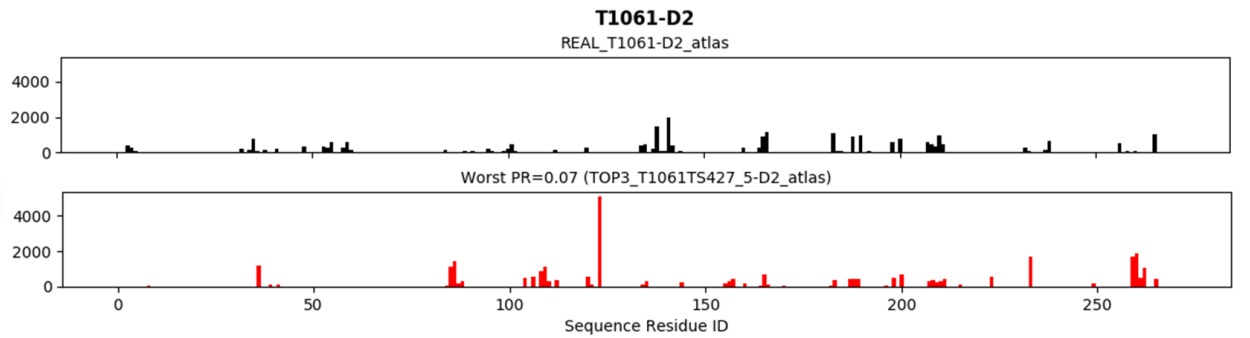
GDT_TS = 46.20



T1061-D2 *FM*

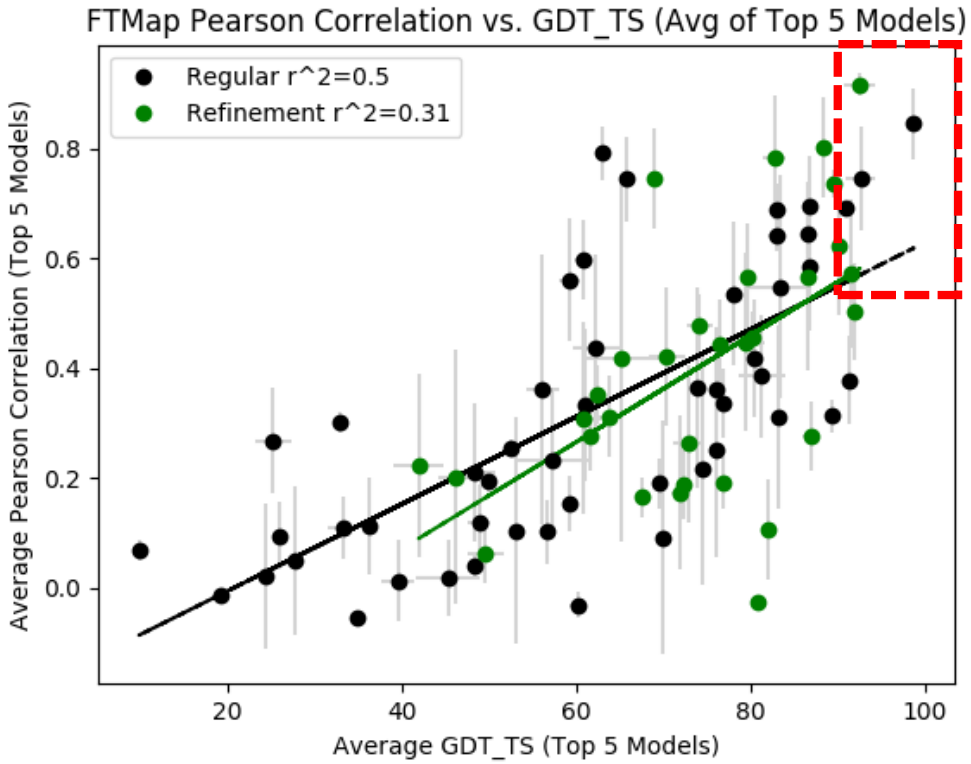


GDT_TS = 78.97

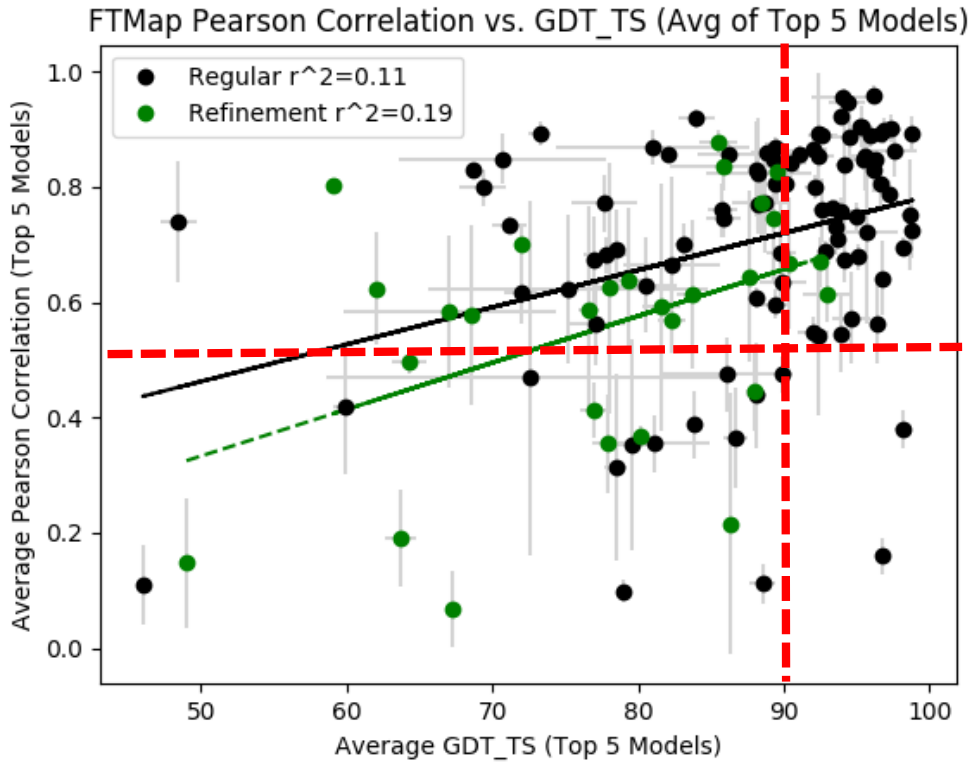


FTMap PR versus GDT_TS

CASP12



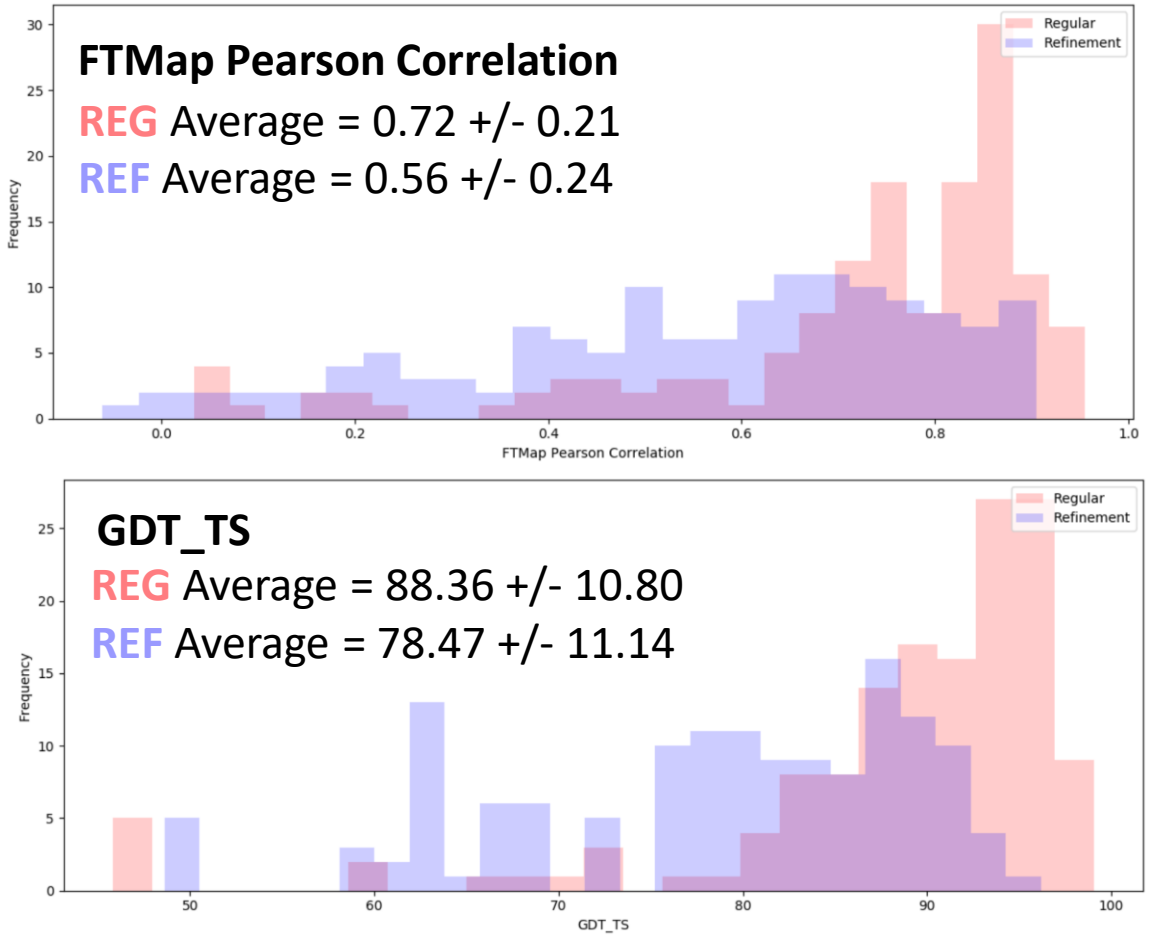
CASP14



CASP14 Regular vs. Refinement

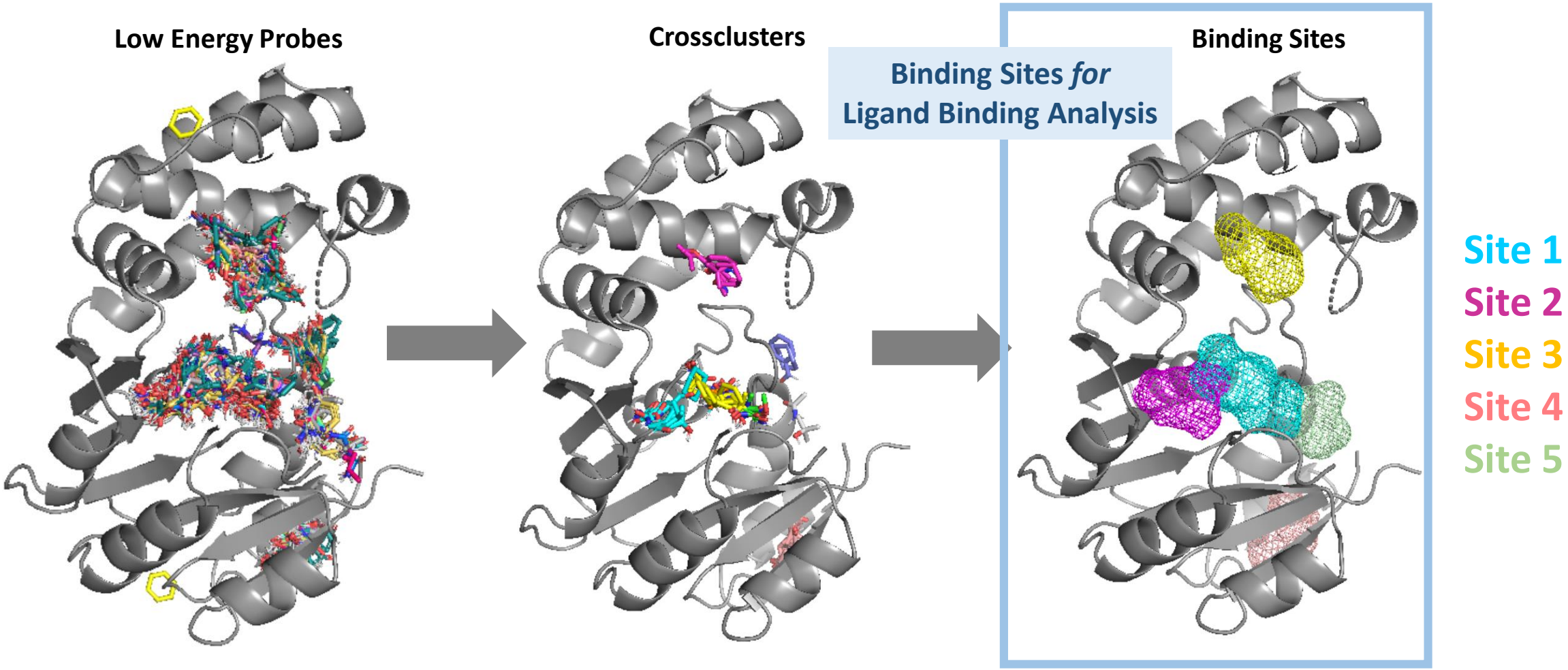
FTMap Pearson Correlation worsened with refinement

- AlphaFold structures dominated high GDT_TS & high FTMap PR correlated regular targets
- Refinement models were selected from a variety of different groups (not AlphaFold)
- AlphaFold did not submit structures for refinement cases



FTMap Binding Sites

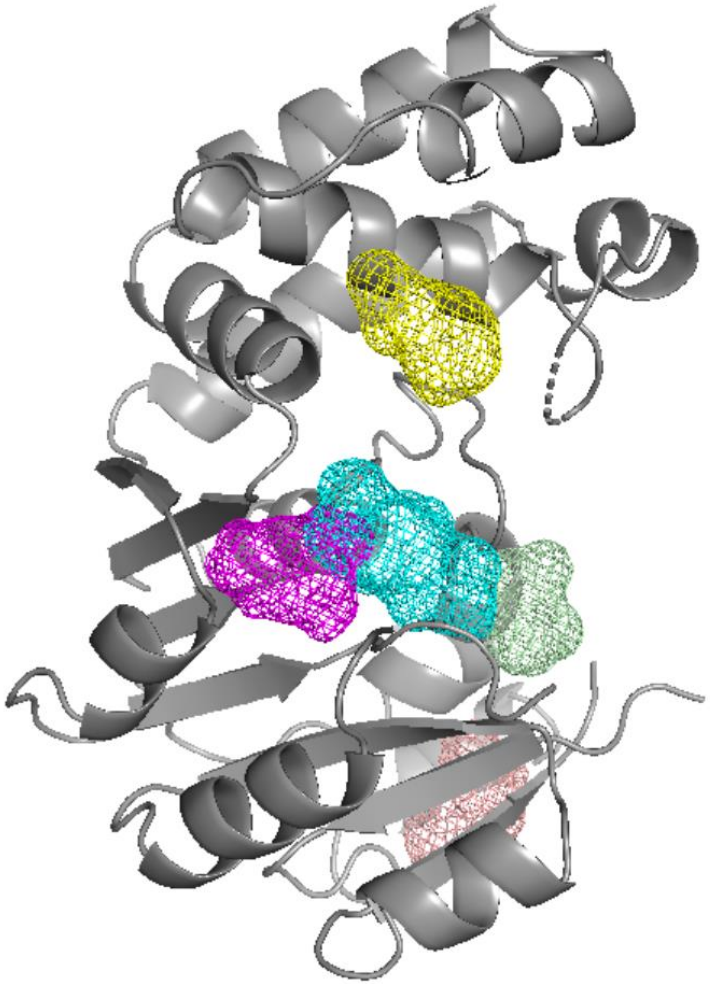
for evaluation of structure models



Brenke, R. et. al., (2009). *Bioinformatics*, *25*(5), 621-627.
Ngan, C. H., et. al., (2012). *Bioinformatics*, *28*(2), 286-287.

Binding Sites

for evaluation of ligand-binding potential



- Group together overlapping crossclusters from FTMap (using FTSite protocol)
- Ranked by number of probes clusters in site

Binding Sites: Ranked according to strength, by number of probes clusters in each site

Site 1 > Site 2 > Site 3 > Site 4 > Site 5

Brenke, R. et. al., (2009). *Bioinformatics*, *25*(5), 621-627.
Ngan, C. H., et. al., (2012). *Bioinformatics*, *28*(2), 286-287.

CASP14 Targets with co-crystalized ligands

for evaluation of structure models

			Real Binding Site		TOP Model Binding Site Strength				
Regular Targets	**PDB ID**	**Ligand**	**Site # | Probes**	**Strength**	**TOP1**	**TOP2**	**TOP3**	**TOP4**	**TOP5**
T1024	6T1Z_A	HT1	1 | 29	29	31	53	3	36	69
T1024	6T1Z_A	XP4	3|16, 4|14, 5|10	40	42	72	0	85	89
T1049-D1	6Y4F_A	GLU	3 | 12	12	15	14	17	16	23
T1053-D1	-	ADP	2 | 18	18	41	44	37	34	29
T1057-D1	-	SAM	1|41, 2|18	59	64	57	62	65	61
T1076-D1	-	ADP	1|33, 6|2	35	42	47	34	37	33

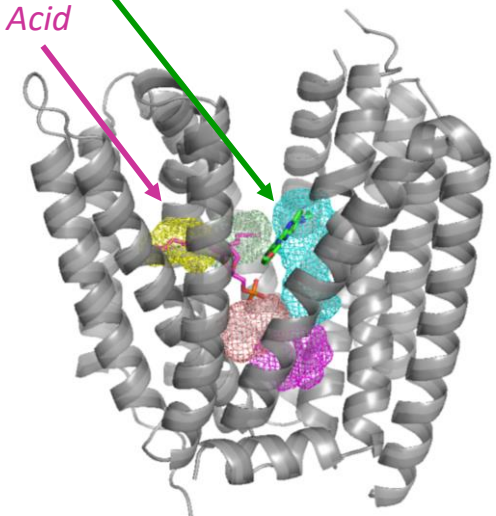
T1024 – LmrP Integral Membrane Protein

Co-Crystalized with XP4 (Phosphatidic Acid) and HT1 (Hoechst 33342) – Multi Drug Transporter,

Experimental Structure

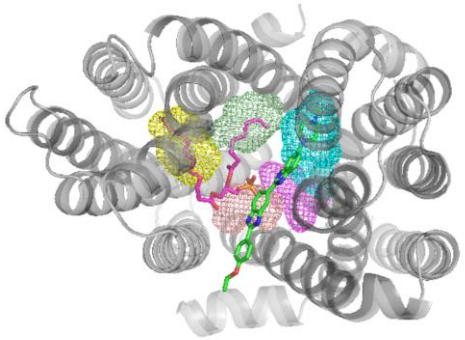
HT1 *Hoechst 33342*

XP4 *Phosphatidic Acid*



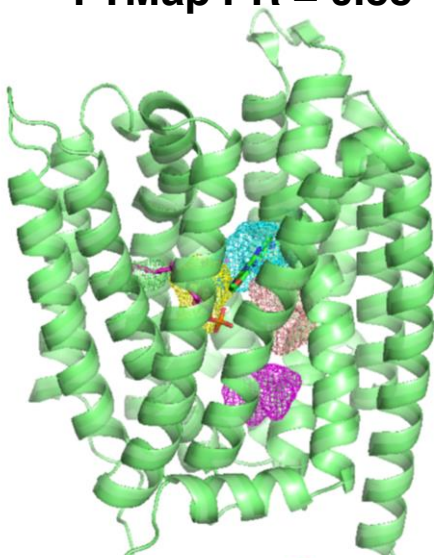
Side View

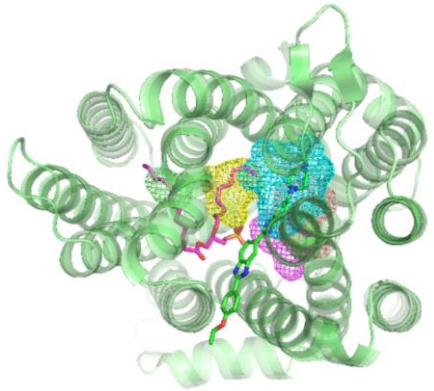
Site 1 > Site 2 > Site 3 > Site 4 > Site 5



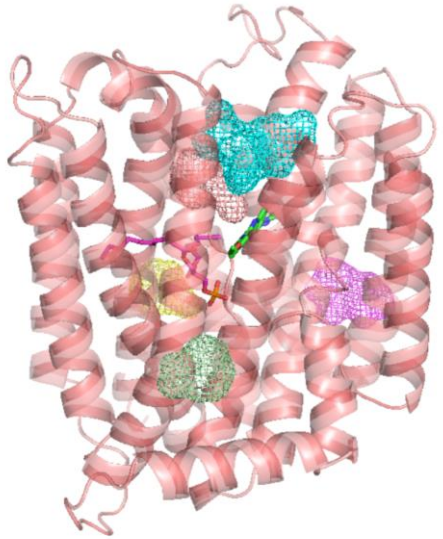
Aerial View

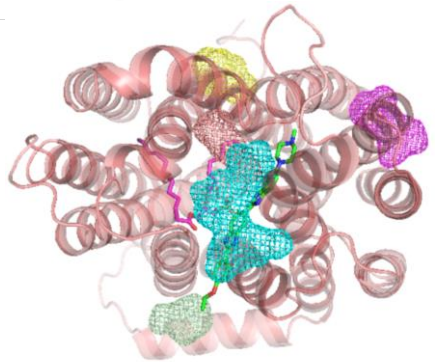
TOP1_T1024TS427_3
FTMap PR = 0.55





TOP3_T1024TS226_5
FTMap PR = 0.06

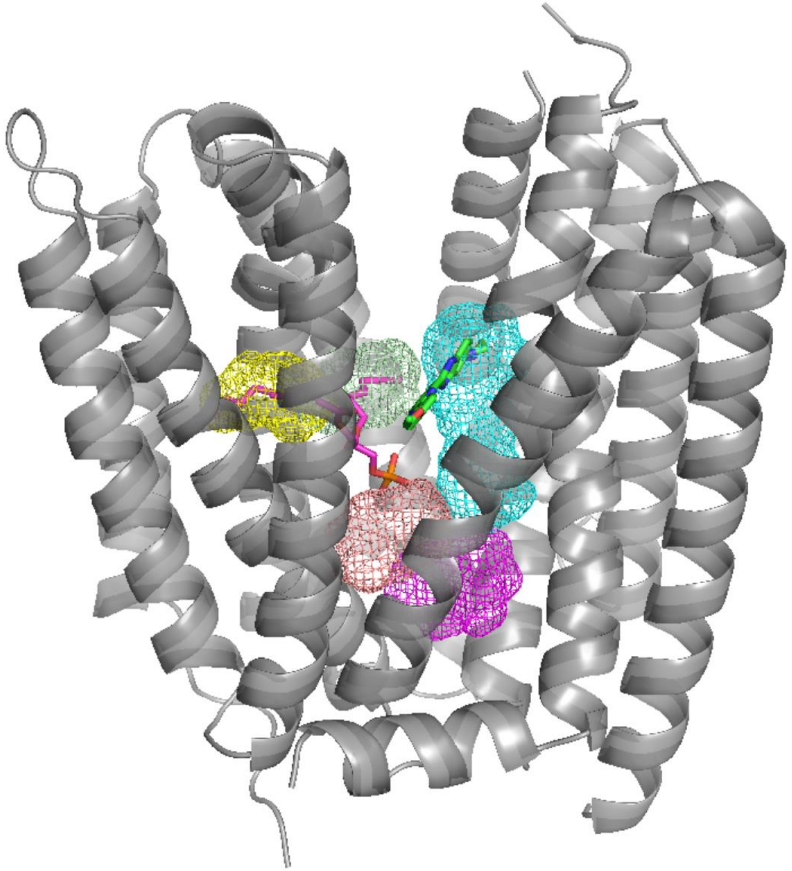


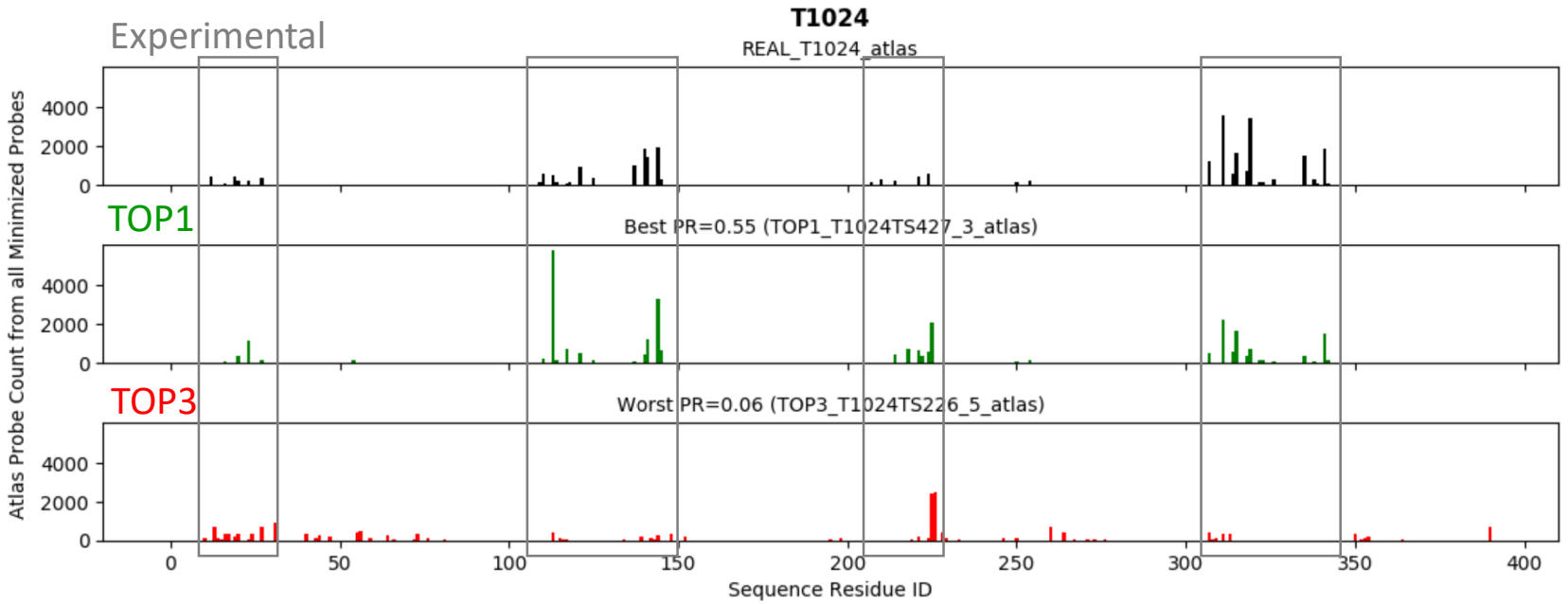


Debruycker, V., Hutchin, A., Masureel, M., Ficici, E., Martens, C., Legrand, P., ... & Govaerts, C. (2020). An embedded lipid in the multidrug transporter LmrP suggests a mechanism for polyspecificity. *Nature structural & molecular biology*, *27*(9), 829-835.

T1024 – LmrP Integral Membrane Protein

Co-Crystalized with XP4 (Phosphatidic Acid) and HT1 (Hoechst 33342)





HT1 Binding Site Residues: 50, 53, 54, 143, 147, 231, 232, 234, 235, 266, 269, 270, 273, 323, 327, 358, 359, 362, 363

XP4 Binding Site Residues: 21, 22, 24, 25, 29, 52, 53, 56, 83, 111, 112, 115, 116, 119, 123, 146, 147, 150, 172, 175, 358

Debruycker, V., Hutchin, A., Masureel, M., Ficici, E., Martens, C., Legrand, P., ... & Govaerts, C. (2020). An embedded lipid in the multidrug transporter LmrP suggests a mechanism for polyspecificity. *Nature structural & molecular biology*, *27*(9), 829-835.

CASP14 Targets with co-crystalized ligands

for evaluation of structure models

BOSTON UNIVERSITY

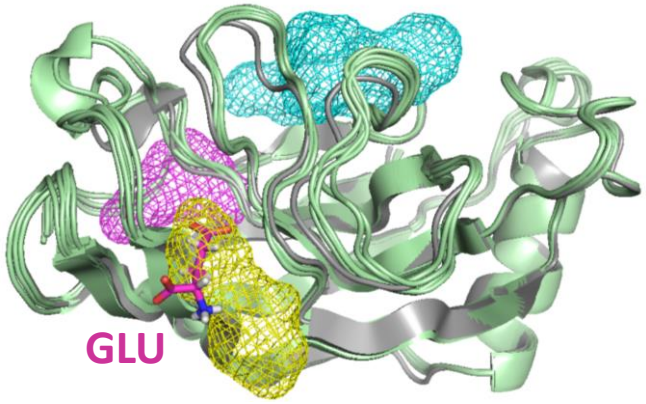
			Real Binding Site		TOP Model Binding Site Strength				
Regular Targets	**PDB ID**	**Ligand**	**Site # | Probes**	**Strength**	**TOP1**	**TOP2**	**TOP3**	**TOP4**	**TOP5**
T1024	6T1Z_A	HT1	1 | 29	29	31	53	3	36	69
T1024	6T1Z_A	XP4	3|16, 4|14, 5|10	40	42	72	0	85	89
T1049-D1	6Y4F_A	GLU	3 | 12	12	15	14	17	16	23
T1053-D1	-	ADP	2 | 18	18	41	44	37	34	29
T1057-D1	-	SAM	1|41, 2|18	59	64	57	62	65	61
T1076-D1	-	ADP	1|33, 6|2	35	42	47	34	37	33

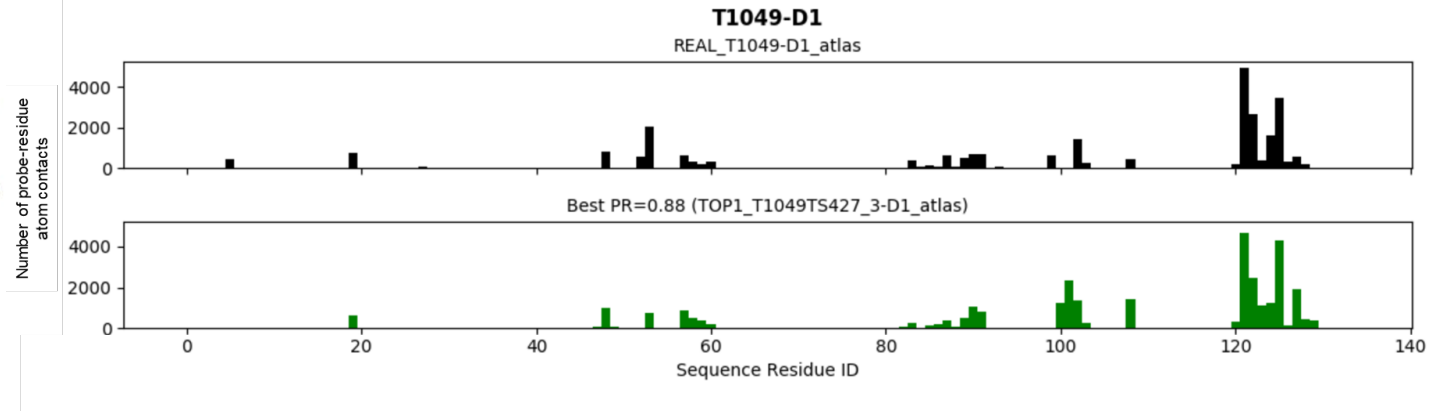
Strong coverage across all Top 5 models

(all models submitted by Group 427 - AlphaFold)

T1049 & T1057 – All models are good

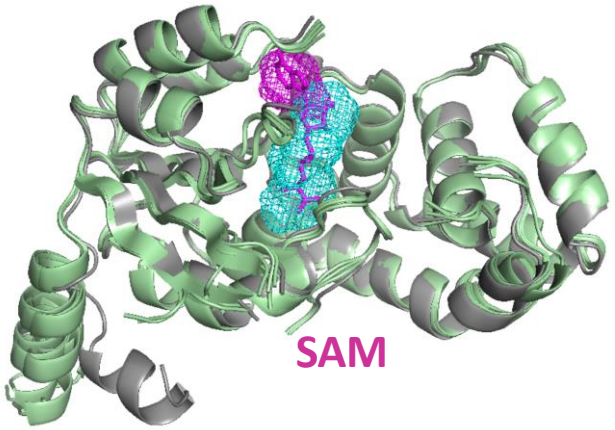
T1049-D1: Fimbrial adhesion protein bound to glutamic acid

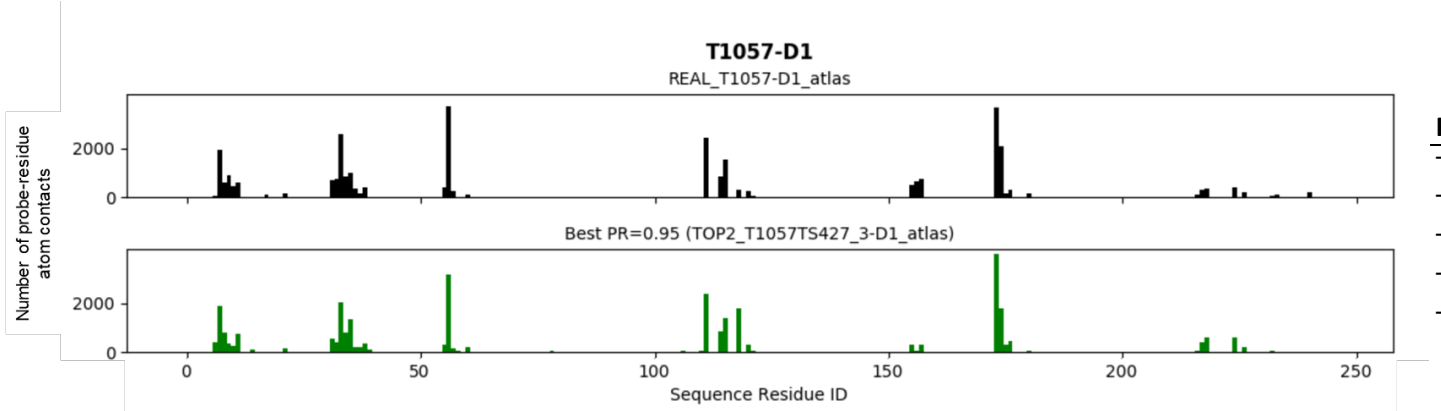




RANK	GDT_TS	FTMap PR
TOP1	93.28	0.88
TOP2	93.1	0.87
TOP3	93.1	0.82
TOP4	91.79	0.86
TOP5	90.3	0.84

T1057-D1: N4-Cytosine Methyltransferase bound to S-adenosylmethionine



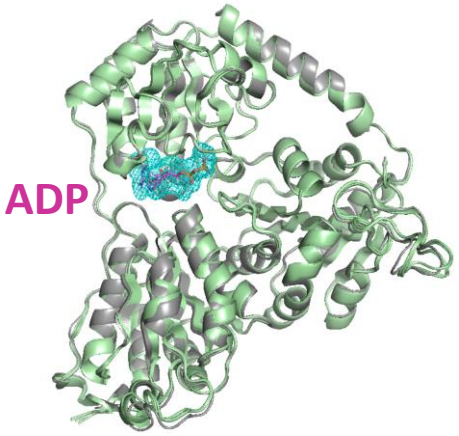


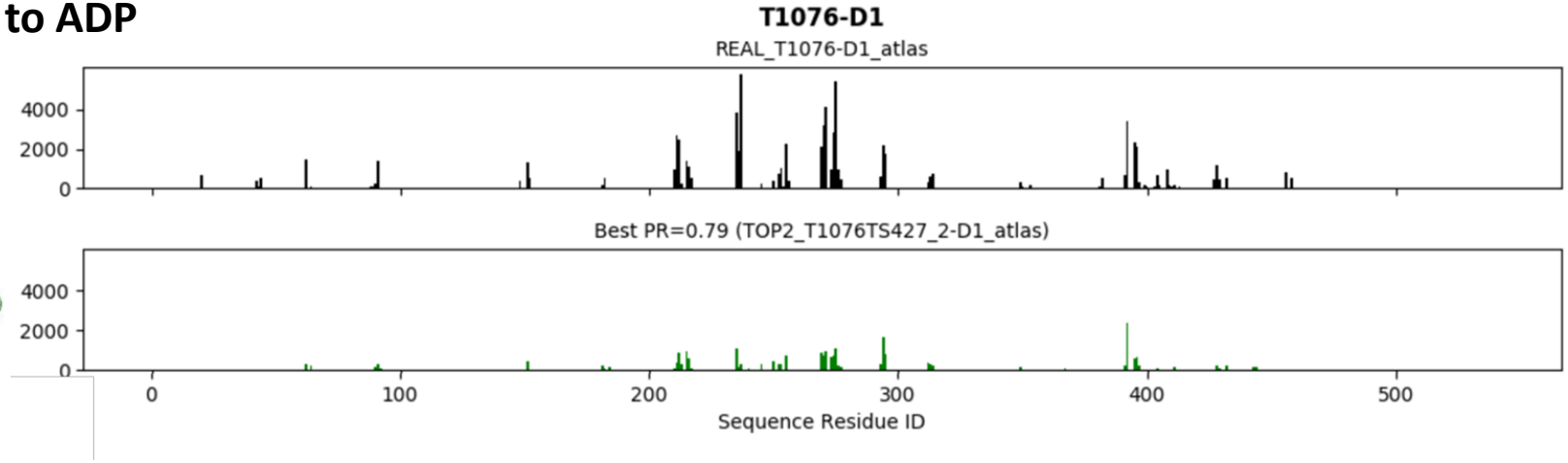
RANK	GDT_TS	FTMap PR
TOP1	94.41	0.91
TOP2	94	0.95
TOP3	93.9	0.92
TOP4	93.8	0.92
TOP5	93.29	0.91

T1076 Binding Site Conserved

Good overall FTMap correlation

T1076-D1: HACL bound to ADP



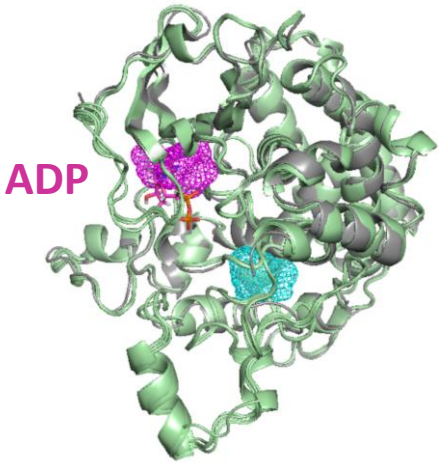


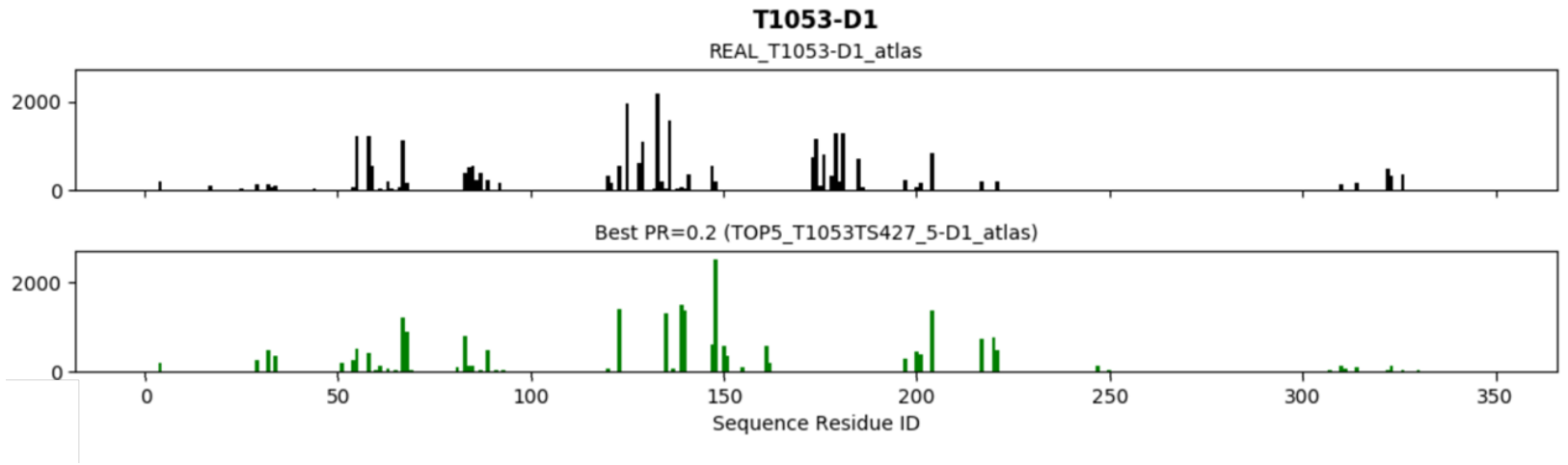
RANK	GDT_TS	FTMap PR
TOP1	99.07	0.77
TOP2	99.07	0.79
TOP3	98.75	0.68
TOP4	98.61	0.68
TOP5	98.56	0.70

T1053-D1 Binding Site Conserved

Overall correlation poor

T1053-D1: T4SS effector bound to ADP





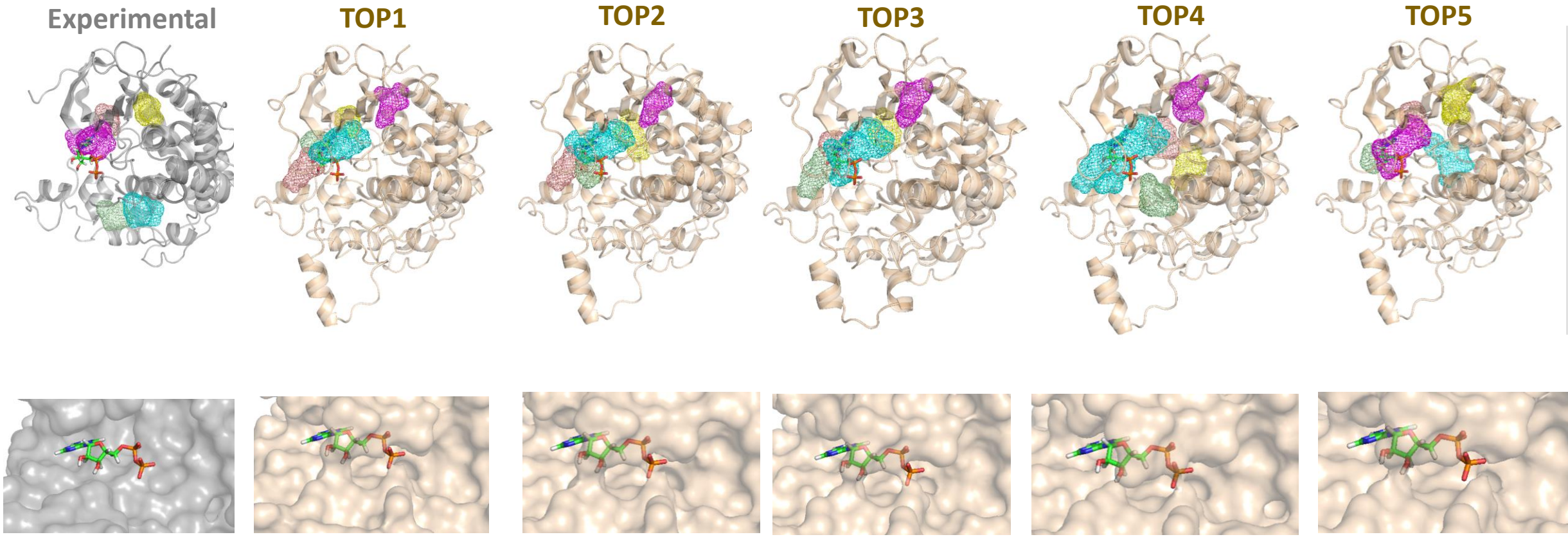
RANK	GDT_TS	FTMap PR
TOP1	97.14	0.17
TOP2	97.06	0.11
TOP3	96.78	0.14
TOP4	96.49	0.19
TOP5	96.49	0.20

T1053 –D1 Binding Sites Conserved

T4SS effector bound to ADP

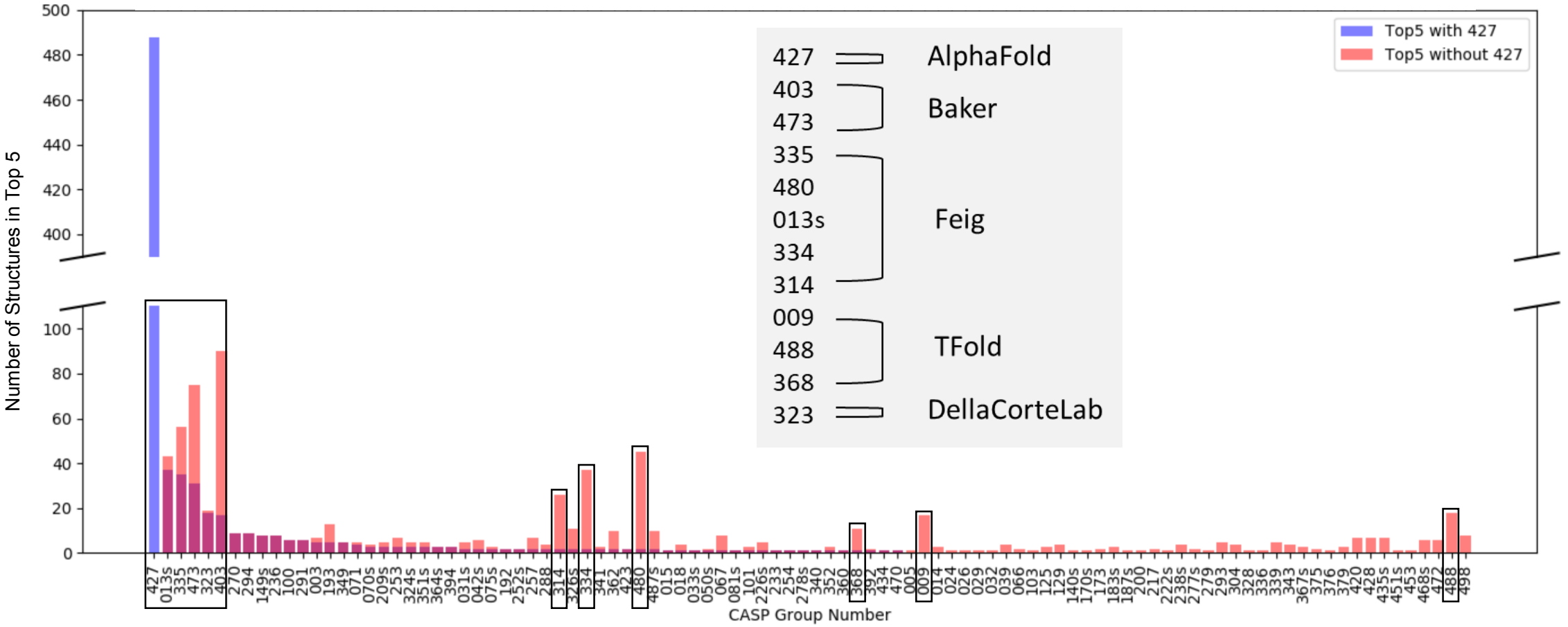
RANK	GDT_TS	FTMap PR
TOP1	97.14	0.17
TOP2	97.06	0.11
TOP3	96.78	0.14
TOP4	96.49	0.19
TOP5	96.49	0.20

BOSTON UNIVERSITY



Group Performance in TOP5 Models

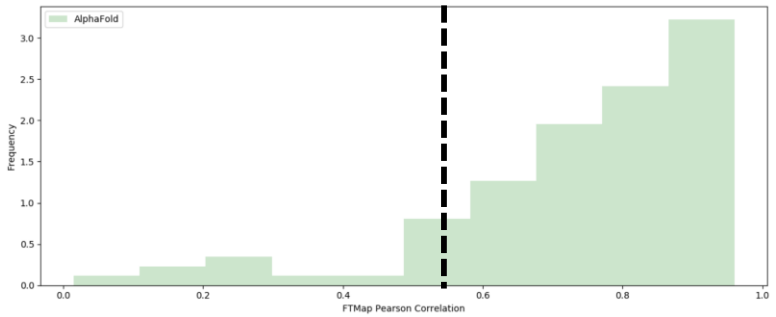
Mostly all group 427…



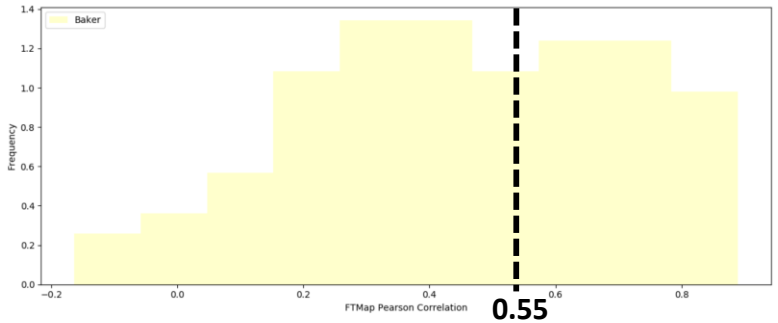
FTMap PR Performance by Group

Evaluation of highest GDT_TS model by each group for all regular targets

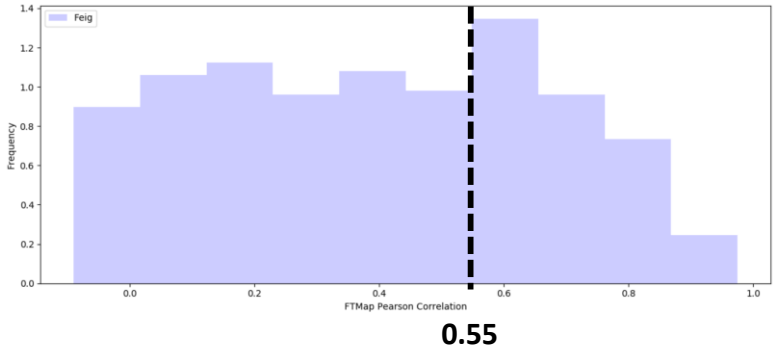
AlphaFold



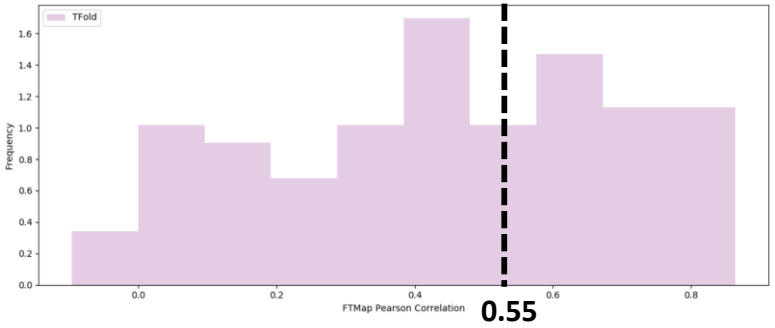
Baker



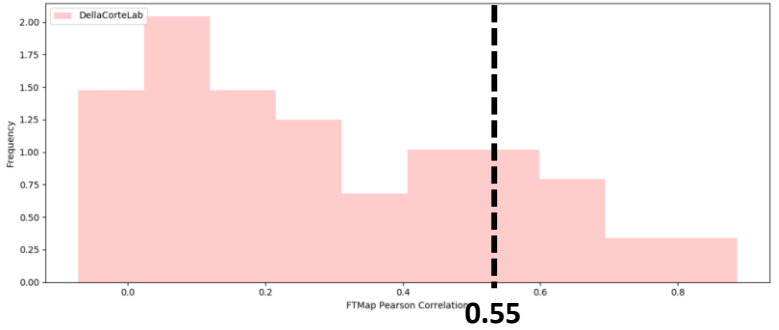
Feig



TFold



DellaCorteLab



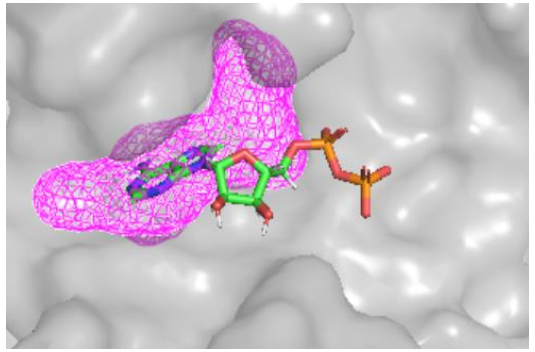
427	AlphaFold
403	Baker
473	Baker
335	Feig
480	Feig
013s	Feig
334	Feig
314	Feig
009	TFold
488	TFold
368	TFold
323	DellaCorteLab

Group Conservation of Ligand-Binding Sites

			Experimental Structure Binding Site		Top Model Total Binding Site Strength											
					AlphaFold	Baker		Feig					Tfold			DellaCorteLab
Regular Targets	PDB ID	Ligand	Site # | Probes	Strength	427	403	473	335	480	013s	334	314	9	488	368	323
T1049-D1	6Y4F_A	GLU	3 | 12	12	15	19	3	16	12	35	0	0	31	16	36	0
T1053-D1	-	ADP	2 | 18	18	41	54	16	11	20	17	31	0	38	26	39	4
T1057-D1	-	SAM	1|41, 2|18	59	64	50	50	49	41	31	44	42	58	34	44	72
T1076-D1	-	ADP	1|33, 6|2	35	44	19	21	0	39	14	24	17	16	0	16	78

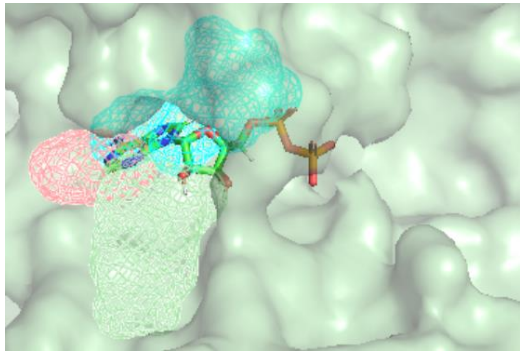
T1053-D1 : *T4SS effector bound to ADP*

REAL



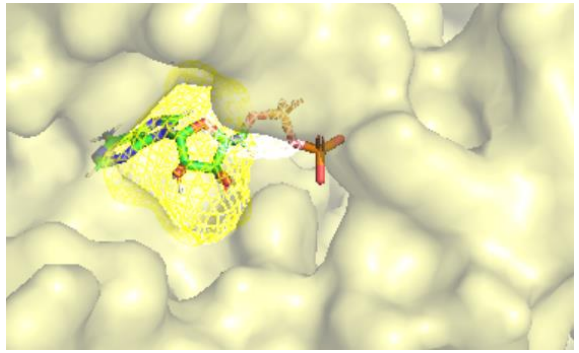
18

AlphaFold (427) TOP 1



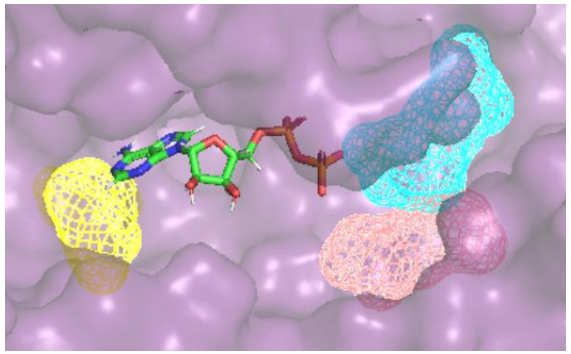
41

Baker (473) TOP 12



16

Tfold (488) TOP 6



26

Conclusions

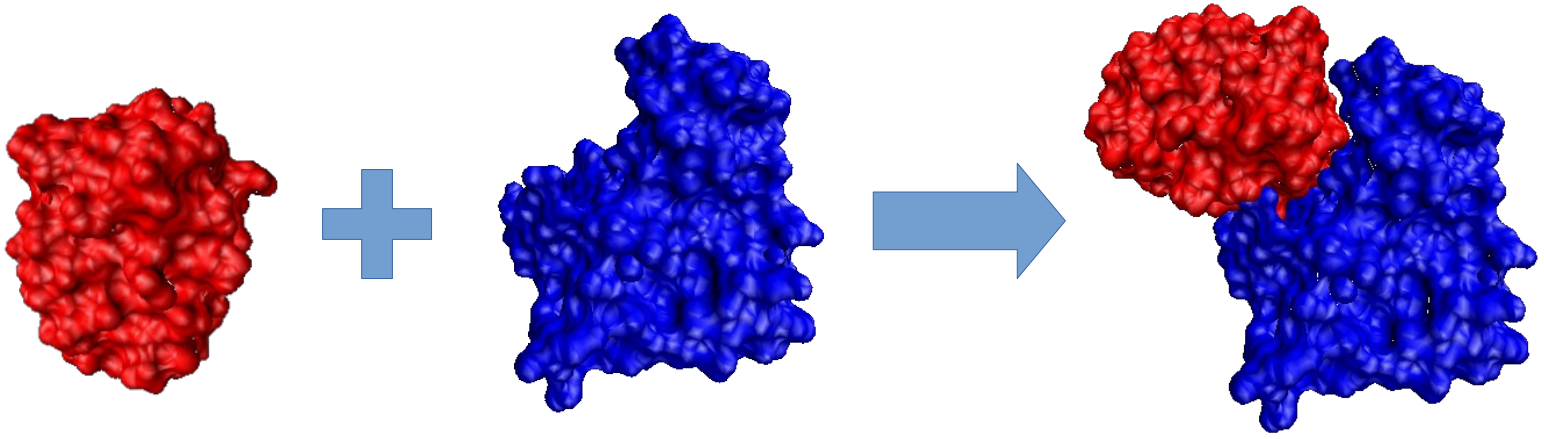
- **78/96** CASP14 domain-split targets have x-ray homolog-level surface conservation
- Positive correlation between FTMap PR & GDT_TS
- GDT_TS > 90 almost guarantees FTMap PR > 0.55 (in x-ray homolog range)
- Good identification of ligand binding sites among high FTMap PR, high GDT_TS models
- Small changes in side chains near the binding site can have major impacts on the binding site strength and structure

Conservation of Binding Properties in Modeled Proteins: part 2

Protein-protein docking

Protein-protein docking

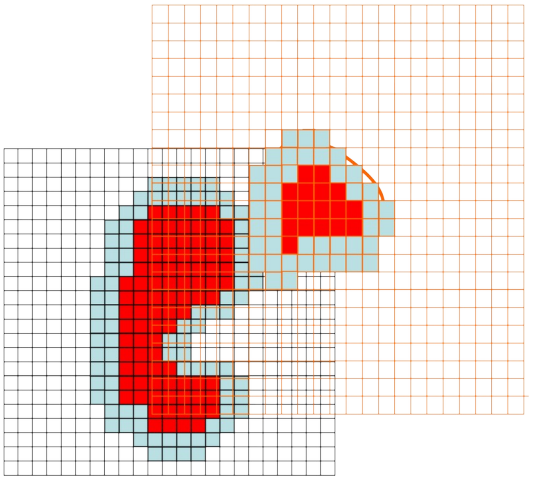
- Docking: sampling the 6D configurational space of 2 rigid bodies.
- What can we learn from the statistical analysis of the sampled energy landscape?



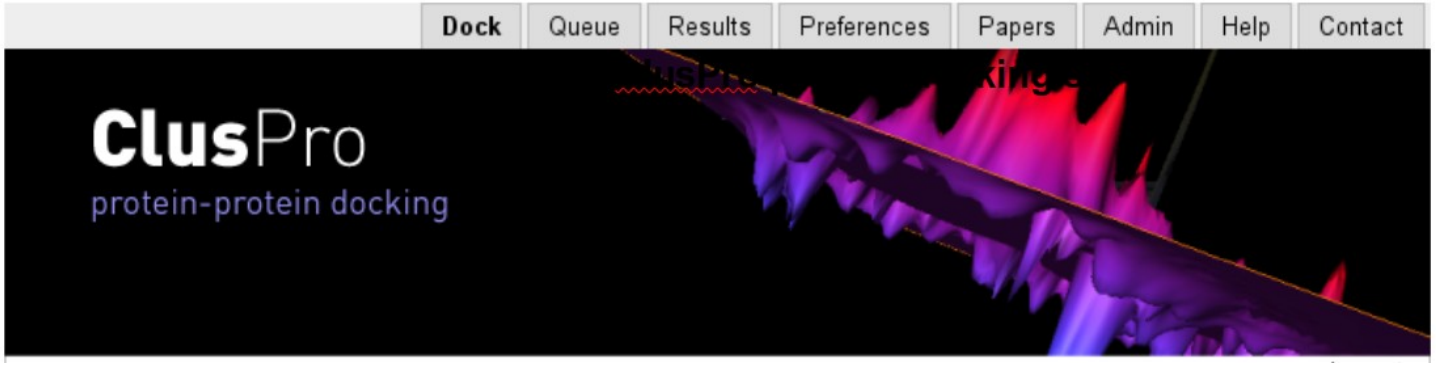
Rigid body sampling by Fast Fourier Transform

- FFT-based global sampling on a grid using “smooth” potentials (PIPER)
- Top-1000 docking poses are clustered, cluster centers are refined and reported as final models.

Ligand representation
(e.g. charge density)

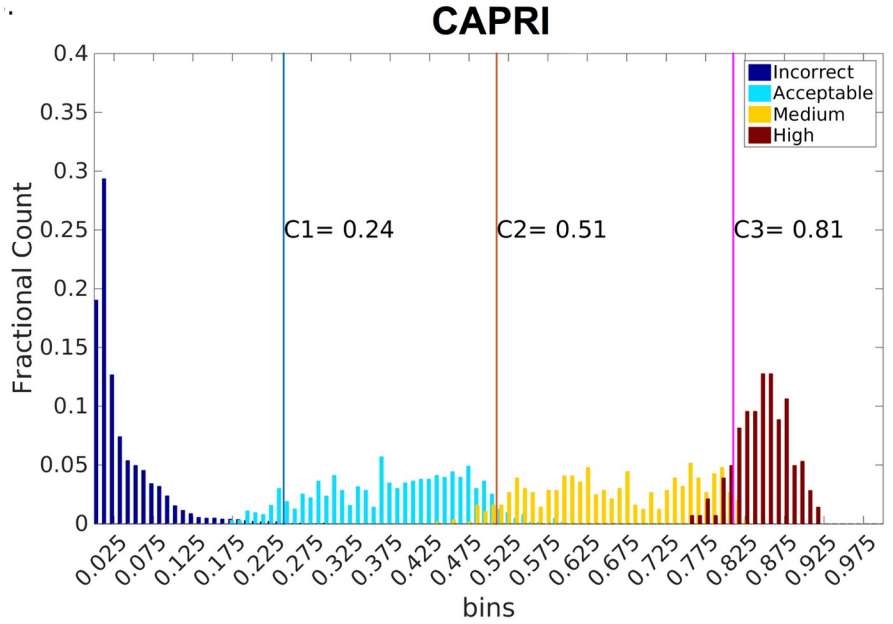


Receptor representation
(e.g. Electrostatic potential)



Docking quality assessment

- DockQ: a single continuous quality measure for protein docked models based on CAPRI evaluation protocol



$$DockQ = f(F_{nat}, LRMSD, iRMSD)$$

Basu S, Wallner B (2016) DockQ: A Quality Measure for Protein-Protein Docking Models. PLOS ONE 11(8): e0161879

Summary:
CASP12 vs CASP14

- CASP12 – 14 targets (Refinement)
- CASP14 – 13 targets
- Two docking scenarios:
  - [X-ray to X-ray] – baseline
  - [X-ray to model]
- Larger structure selected as receptor.
- For each target, docked top-5 CASP models
- For each docking run, selected the best DockQ result in top-10 ClusPro models.

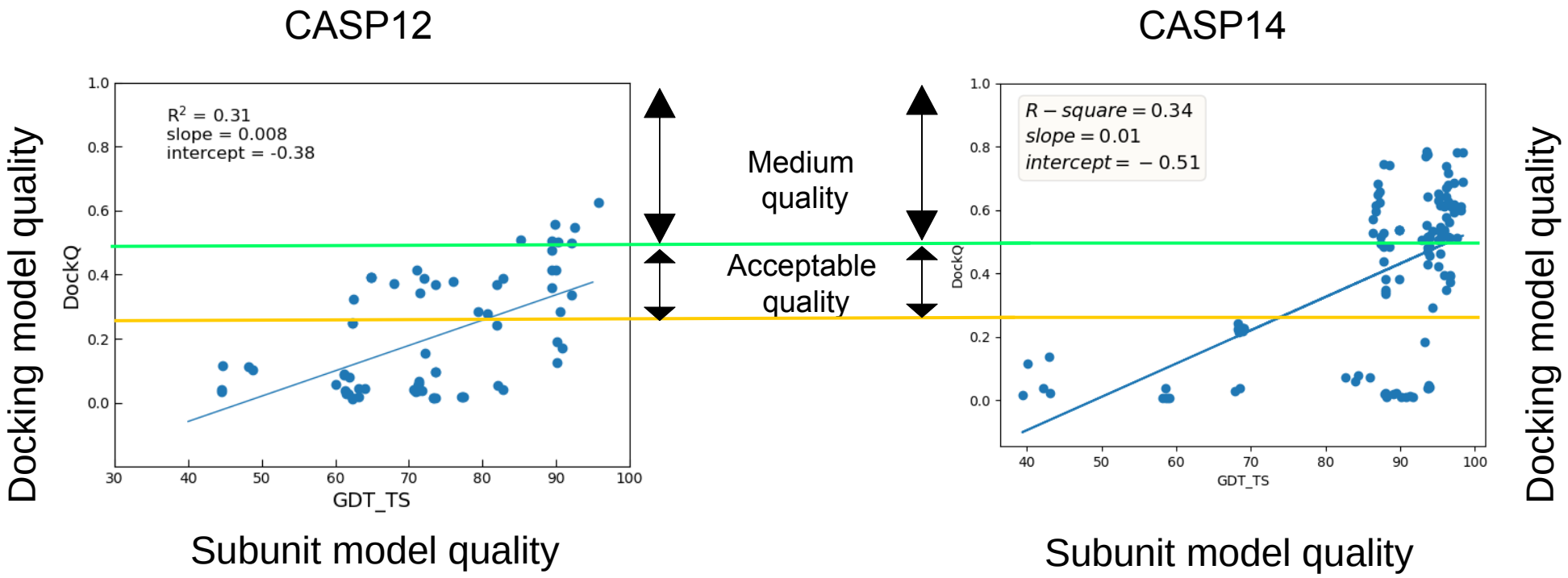
Docking Results: CASP12 Refinement Targets

CASP ID	Native PDB ID[a]	Docking of X-ray structures					Docking of models to X-ray structure				
		Protein 1 Ligand	Protein 2 Receptor	Rank	DockQ	Quality	Protein 1 Ligand	Protein 2 Receptor	Rank	DockQ	Quality
TR594[b]	5HKQ_AI	---	---								
TR862	5J5V_AB	**5J5V_B**	5J5V_A	3	0.876	high	TOP5	5J5V_A	7	0.045	incorrect
TR866[c]	5UW2_A6	**5UW2_A**	5UW2_B	7	0.685	medium	TOP1	5UW2_B	1	0.388	acceptable
TR868	5J4A_AB	5J4A_B	**5J4A_A**	3	0.833	high	5J4A_B	TOP5_	2	0.524	medium
TR870[d]	5J5V_BC	5J5V_B	**5J5V_C**	1	0.715	medium	5J5V_B	TOP4_	5	0.270	acceptable
TR881	3EEV_ABC	**3EEV_A**	3EEV_BC	3	0.574	medium	TOP2	3EEV_BC	8	0.05	incorrect
TR884	5T87_AE	**5T87_E**	5T87_A	1	0.595	medium	TOP1_	5T87_A	3	0.539	medium
TR885	5T87_AE	5T87_E	**5T87_A**	1	0.595	medium	5T87_E	TOP2_	9	0.448	acceptable
TR887	6F03_AB	**6F03_A**	6F03_B	7	0.769	medium	TOP1	6F03_B	5	0.496	medium
TR893[e]	5IDJ_A2	---	---								
TR894	5HKQ_AI	**5HKQ_A**	5HKQ_I	5	0.634	medium	TOP1	5HKQ_I	2	0.687	medium
TR895	5HKQ_AI	**5HKQ_I**	5HKQ_A	1	0.655	medium	TOP1	5HKQ_A	9	0.392	acceptable
TR909	5G5N_ABC	**5G5N_A**	5G5N_BC	1	0.749	medium	TOP1	5G5N_BC	2	0.393	acceptable
TR917	5YVR_AB	**5YVR_A**	5YVR_B	3	0.768	medium	TOP2	5YVR_B	2	0.578	medium
TR921/ TR922	5M2O_AB	**5M2O_B**	5M2O_A	2	0.047	incorrect	5M2O_B	TOP3	9	0.125	incorrect
TR945[f]	5LEV_AB	6BW6_B	6BW6_A	2	0.671	medium	6BW6_B	TOP1	7	0.425	acceptable

Docking results: CASP14 targets

CASP ID	Docking of X-ray structures					Docking of model to X-ray structure				
	Receptor	Ligand	Rank	Dockq	Quality	Receptor	Ligand	Rank	Dockq	Quality
T1032	T1032.orig_B	T1032.orig_A	2	0.781	medium	T1032.orig_A	TOP4_B	5	0.235	acceptable
T1034	T1034.orig_BCD	T1034.orig_A	0	0.879	high	T1034.orig_BCD	TOP3_A	0	0.786	medium
T1038	T1038.orig_A	T1038.orig_B	0	0.885	high	T1038.orig_A	TOP3_B	0	0.65	medium
H1045	H1045.orig_B	H1045.orig_A	5	0.838	high	H1045.orig_B	TOP1_A	1	0.293	acceptable
H1046	H1046.orig_BRT	H1046.orig_A	1	0.681	medium	H1046.orig_BRT	TOP2_A	0	0.783	medium
H1046	H1046.orig_ART	H1046.orig_B	0	0.547	medium	H1046.orig_ART	TOP4_B	0	0.554	medium
H1065	T1065.orig_C	H1065.orig_A	2	0.822	high	H1065.orig_C	TOP2_A	1	0.783	medium
T1070	T1070.orig_XY	T1070.orig_Z	9	0.301	acceptable	T1070.orig_XY	TOP2_Z	6	0.139	incorrect
T0173	T1073.orig_BCD	T1073.orig_A	4	0.891	high	T1073.orig_BCD	TOP4_A	8	0.537	medium
T1078	T1078.orig_B	T1078.orig_A	6	0.56	medium	T1078.orig_B	TOP2_A	5	0.681	medium
T1083	T1083.orig_B	T1083.orig_A	2	0.713	medium	T1083.orig_B	TOP4_A	0	0.528	medium
T1084	T1084.orig_B	T1084.orig_A	5	0.424	acceptable	T1084.orig_B	TOP1_A	7	0.507	medium
T1087	T1087.orig_B	T1087.orig_A	7	0.511	medium	T1087.orig_B	TOP3_A	4	0.738	medium

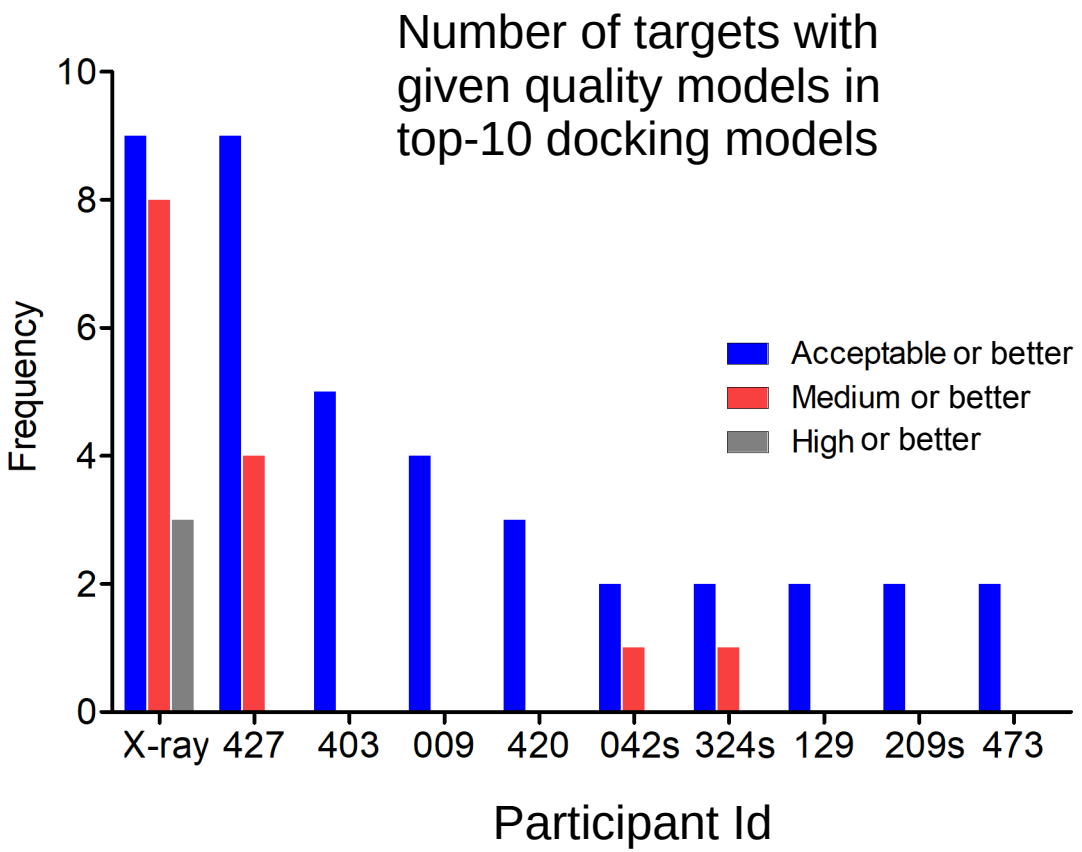
CASP12/14:
DockQ vs GDT_TS



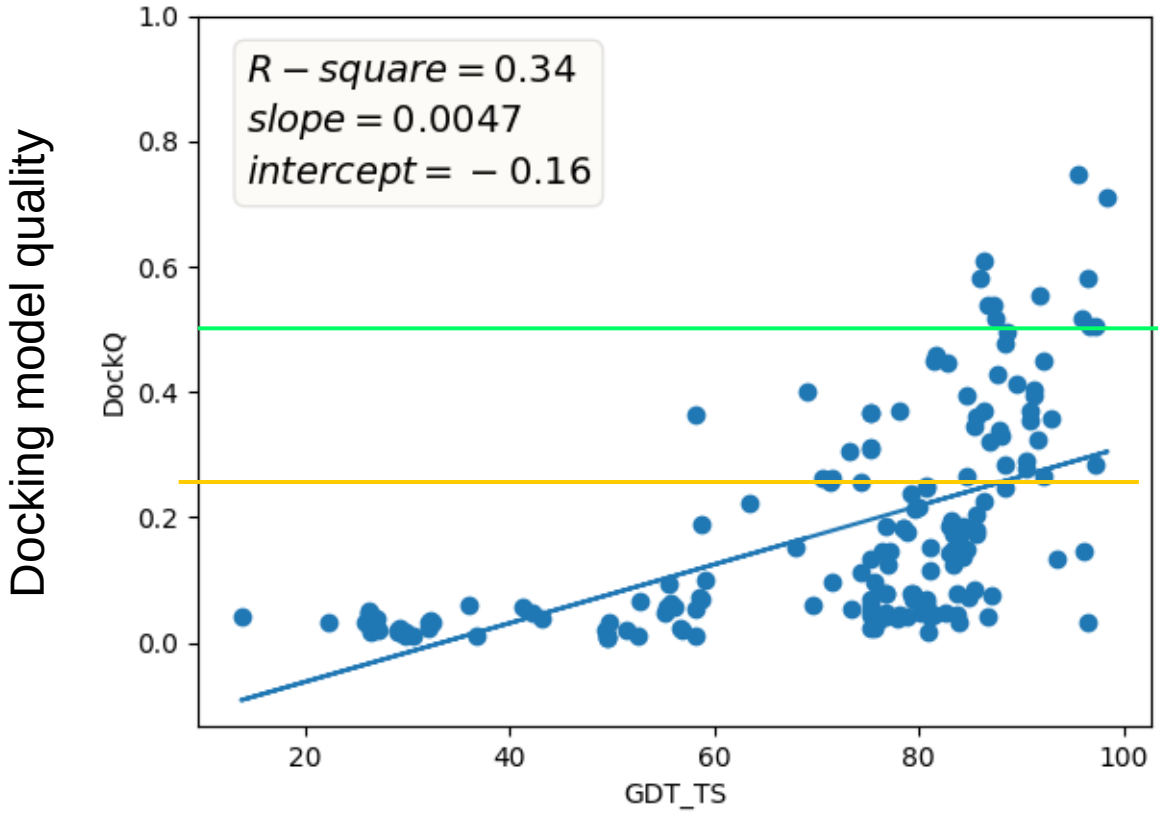
Summary:
CASP14 model to model

- CASP14 – 11 targets:
  - Dimers and Cn simmetry cases.
  - Subunit models available from top-3 groups.
- Two docking scenarios:
  - [X-ray to X-ray] – baseline
  - [model to model]
- For each target, selected the best DockQ result in top-10 ClusPro models.

Docking results: CASP14 – model to model



CASP14: DockQ vs GDT_TS

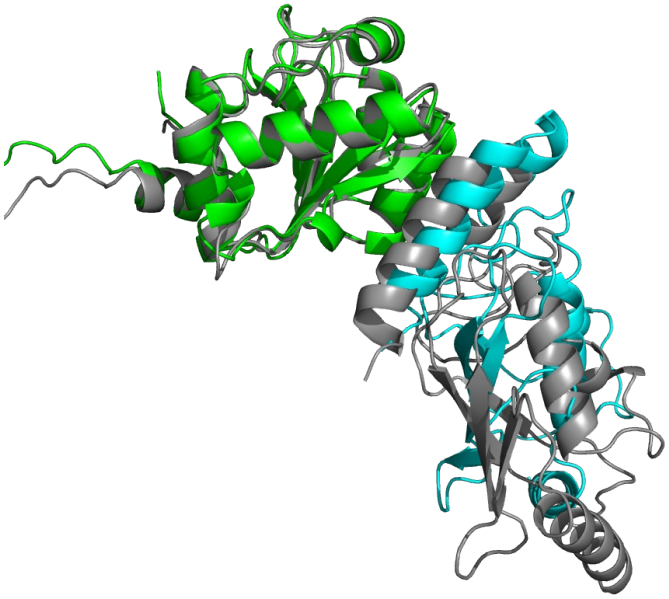


Medium quality models

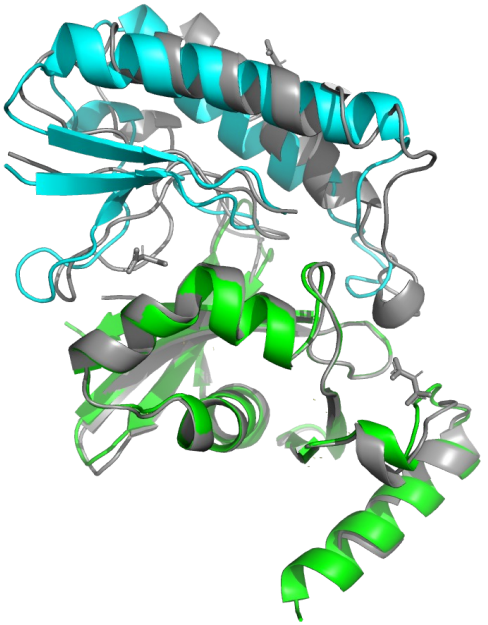
Acceptable quality models

CASP14: docking models to models

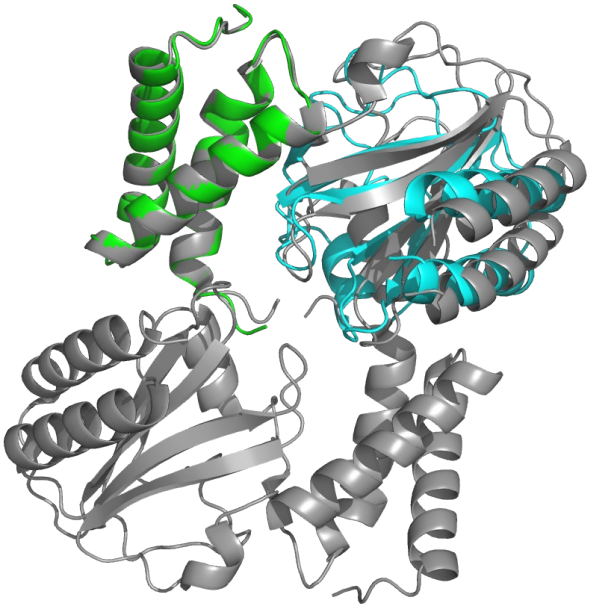
- Sample heterodimer docking results:



T1045, model 00
DockQ: 0.305
Acceptable quality



T1065, model 02
DockQ: 0.749
Medium quality

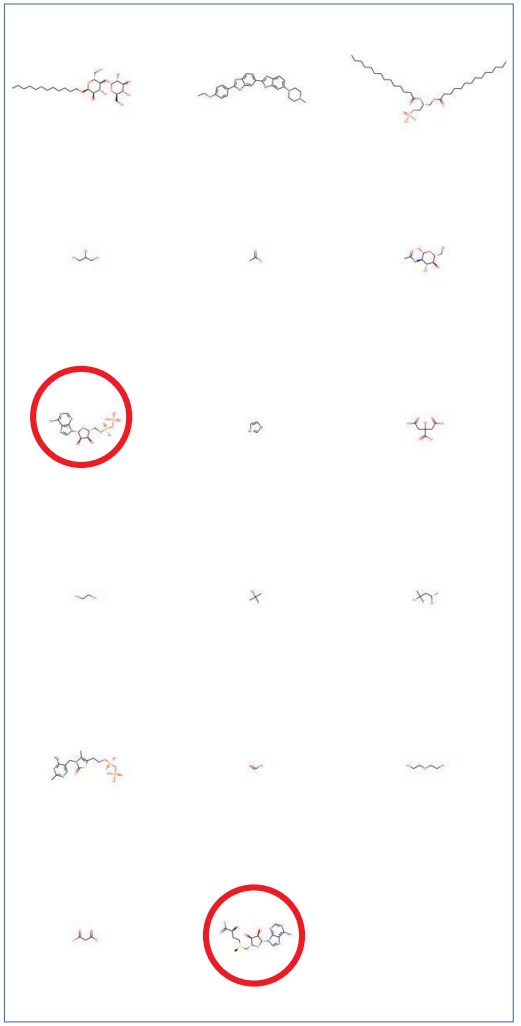


T1046, model 01
DockQ: 0.629
Medium quality

Protein-ligand docking

CASP14 targets co-crystallized with ligands

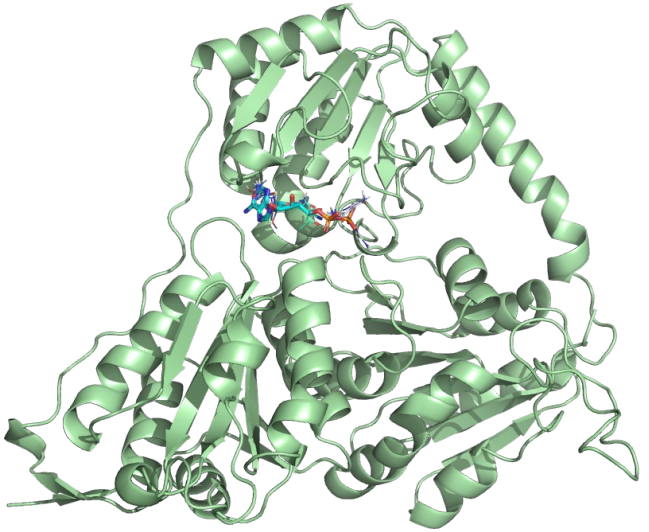
Target	Ligands	Reason for exclusion
T1024	LMU, HT1, XP4, GOL	size, low GDT_TS
T1025	ACT	size
T1028	GOL	size
T1034	FMTB, FMTA, FMTD, PEGA, FMTC	size
T1038	NAG	size
T1053	**ADP**	
T1052	FMT, GOL	size
T1057	**SAM**, EDO	EDO-size
T1058	HEM	size
H1065	MLI	size
T1074	ACT, IMD, TBU, MPD	size
T1076	**ADP**, TZD	TZD - interchain binding site
T1079	CIT, EDO, GOL	size
T1101	IMD	size



Summary:
CASP14 ligand docking

- Autodock Vina used for ligand placement.
- Docking restricted to 15 Angstrom box centered at binding site.

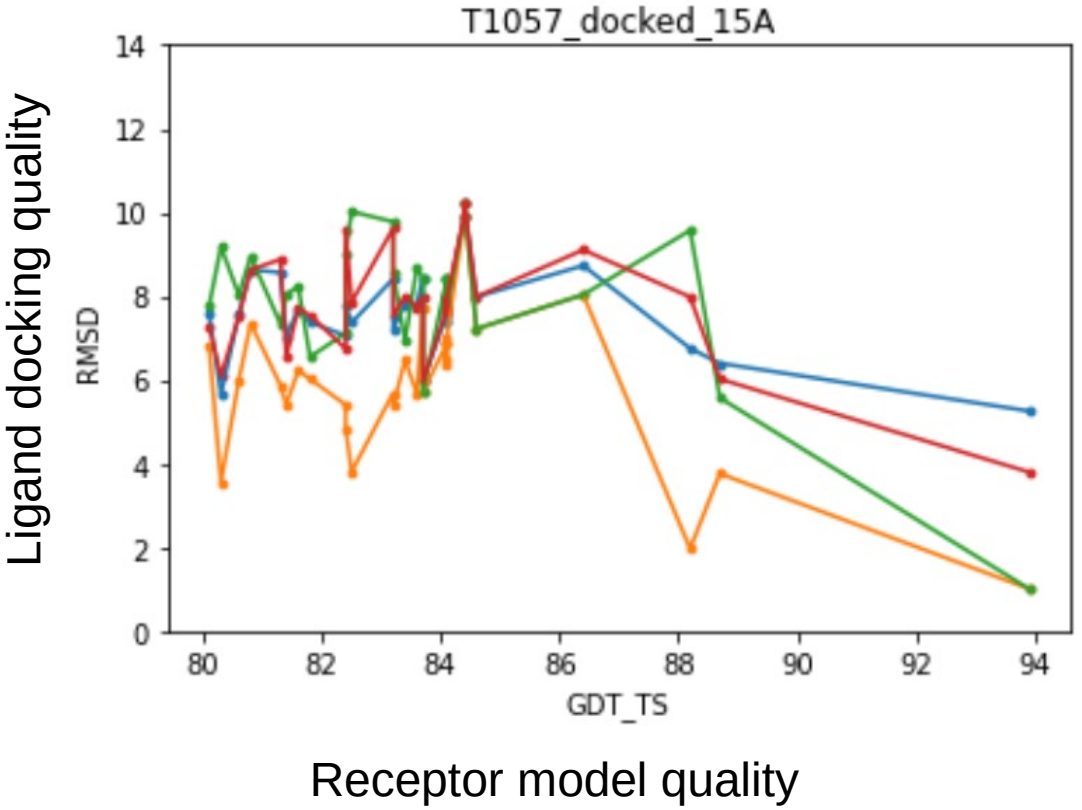
Top ftmap cluster

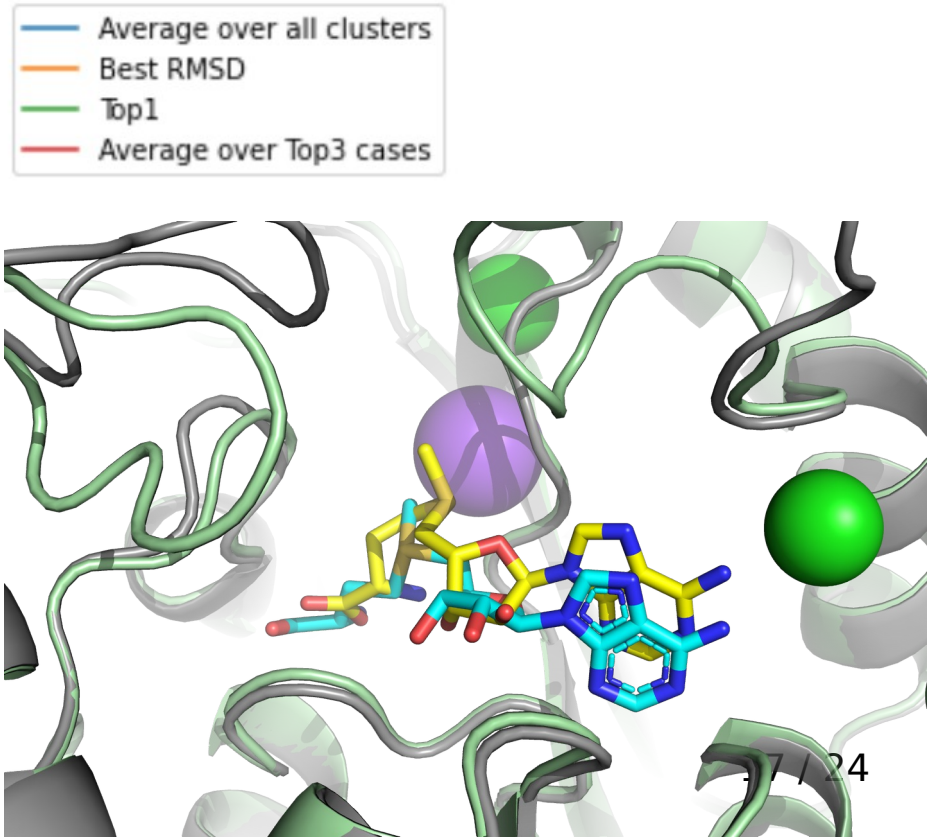


X-ray ligand position

Ligand docking: T1057

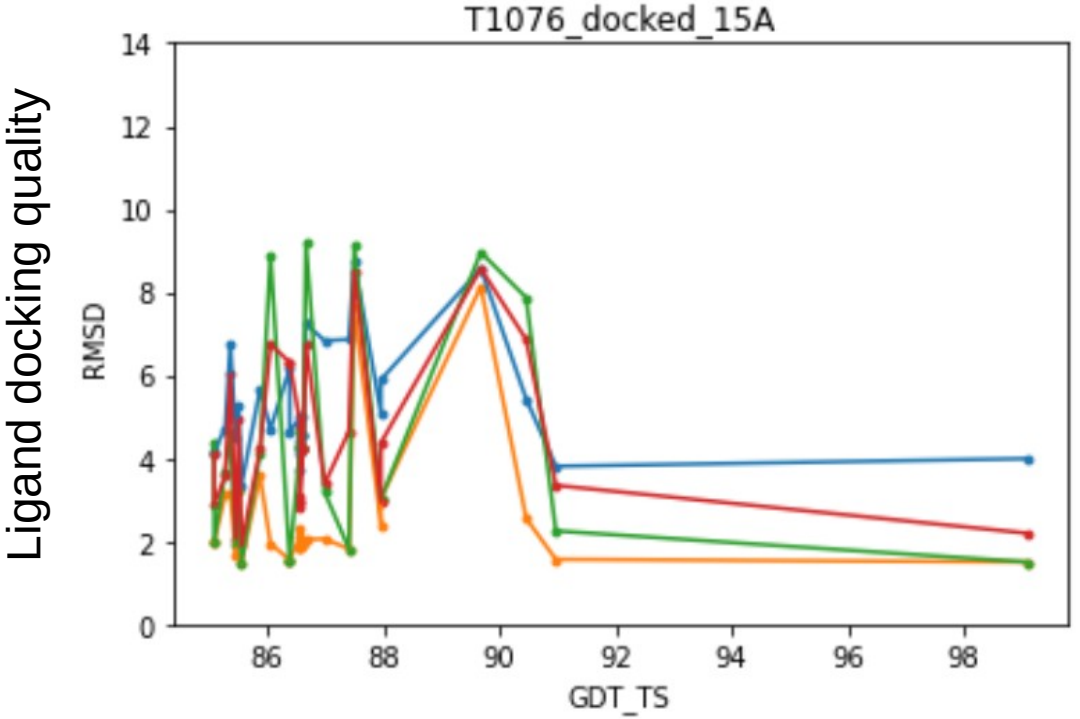
- Receptor model quality vs ligand docking results





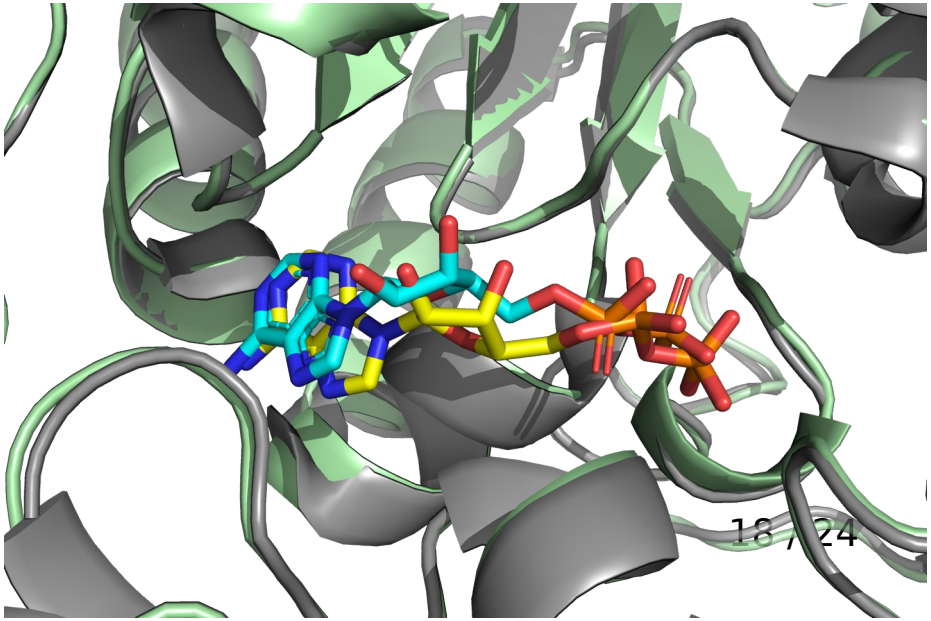
Ligand docking: T1076

- Receptor model quality vs ligand docking results



Ligand docking quality

Receptor model quality



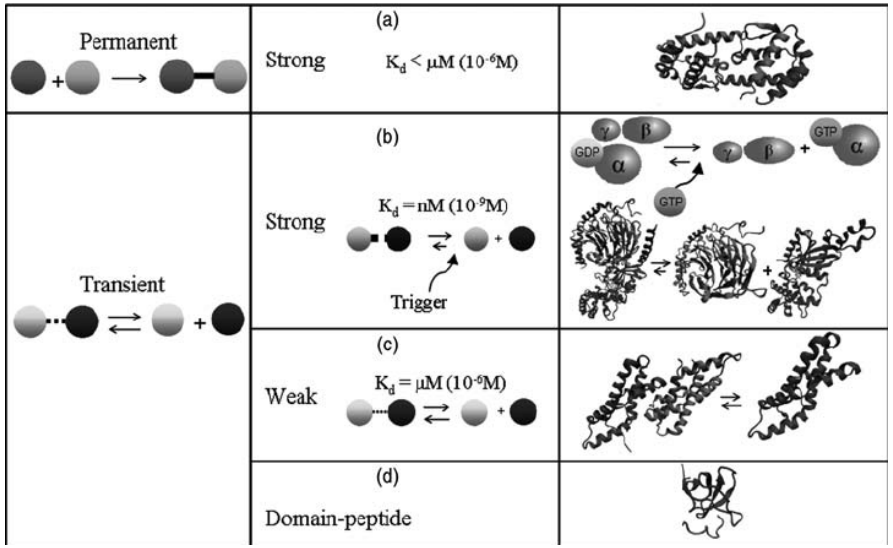
Summary

High quality models (GDT_TS>90) behave similarly to crystal structures in docking programs

Locally accurate models in the binding sites can be useful as well, in spite of overall lower GDT_TS (conservation of binding sites)

It appears there is (will be) big progress for modeling of obligatory (permanent) protein complexes (both home and hetero)

Ensemble of crystal like models and/or dynamic simulations is needed to capture monomer conformational change upon transient protein-protein and protein ligand interactions



Ozbacan et.al. Transient Protein Interactions; 2011