

CASP 15 Ligand Evaluation

Xavier Robin and Pat Walters

Acknowledgements



Jerome Eberhardt



Andriy Kryshatafovych



Aaron Sweeney



Jay Durairaj

Outline

Problem statement

Evaluation criteria – RMSD and IDDT-PLI

Challenges associated with submissions

Overall summary

Specific targets

Some thoughts for next time

The Challenge

Given:

- Protein/RNA sequence(s)
- Ligands as SMILES
- Stoichiometry

```
ID Name SMILES Relevant
001 ATP C1=NC(=C2C(=N1)N(C=N2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)OP(=O)(O)OP(=O)(O)O)O)N yes
002 ATP C1=NC(=C2C(=N1)N(C=N2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)OP(=O)(O)OP(=O)(O)O)O)N yes
003 2MG [Mg+2] yes
004 2MG [Mg+2] yes
```

Predict

- The protein/RNA (complex) structure
 - Up to 5 models / group / target
- The pose of the given ligands, returned in SDF format
 - Up to 5 poses / model

Scores

Requirements:

- Compare prediction with the reference target
- Diverse scores representing different aspects of the prediction accuracy

No well-established scores exist for this challenge

We implemented two new scores – used in CASP for the 1st time

- Symmetry-Corrected Pose RMSD – similar to D3R – assesses the accuracy of the pose
- IDDT-PLI – assesses protein-ligand contacts

Chain mapping (for oligomers)

Ligand assignment (when > 1 ligand in target)

Scores

Requirements:

- Compare prediction with the reference target
- Diverse scores representing different aspects of the prediction accuracy

No well-established scores exist for this challenge

We implemented two new scores – used in CASP for the 1st time

- **Symmetry-Corrected Pose RMSD** – similar to D3R – assesses the accuracy of the pose
- **IDDT-PLI** – assesses protein-ligand contacts

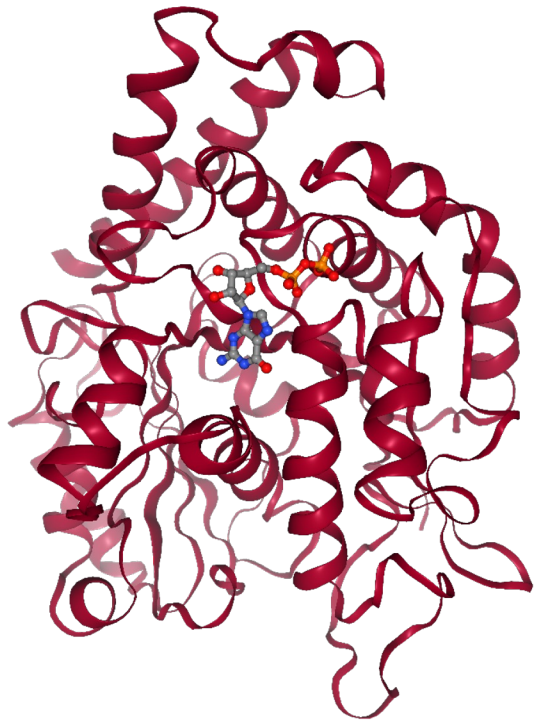
Chain mapping (for oligomers)

Ligand assignment (when > 1 ligand in target)

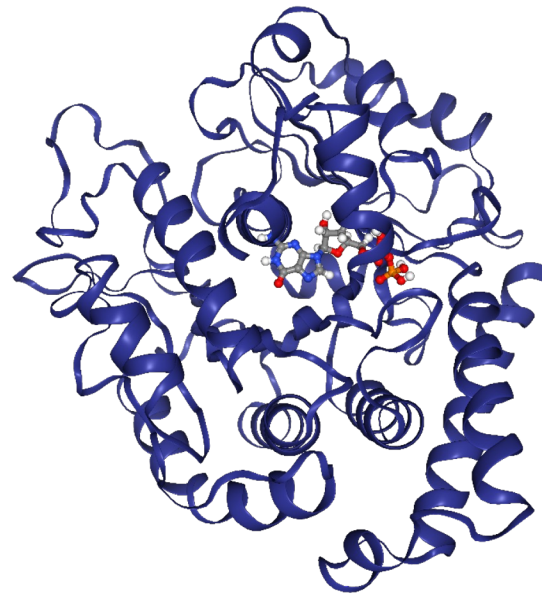
Symmetry-Corrected Pose RMSD

7EVE - ADP

Reference

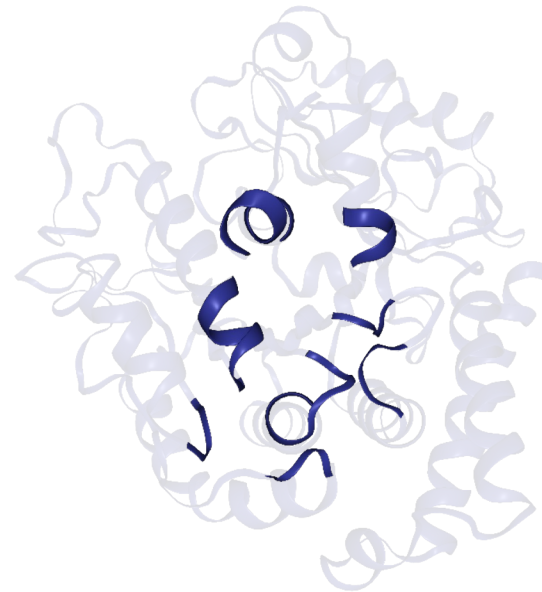
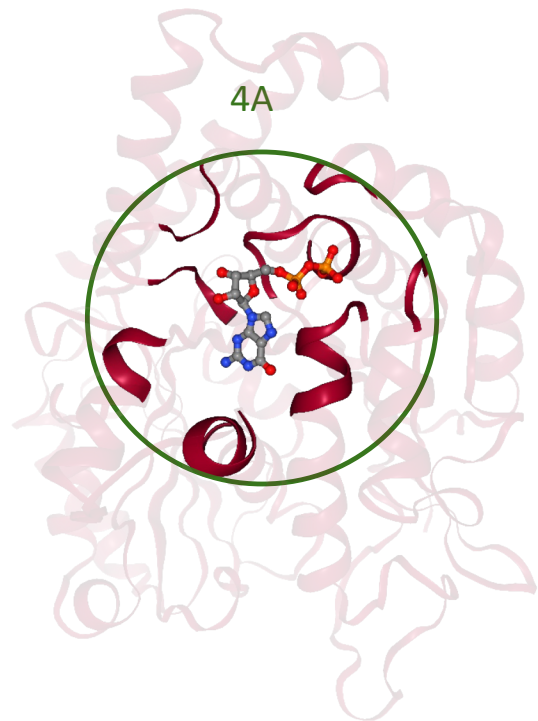


Model



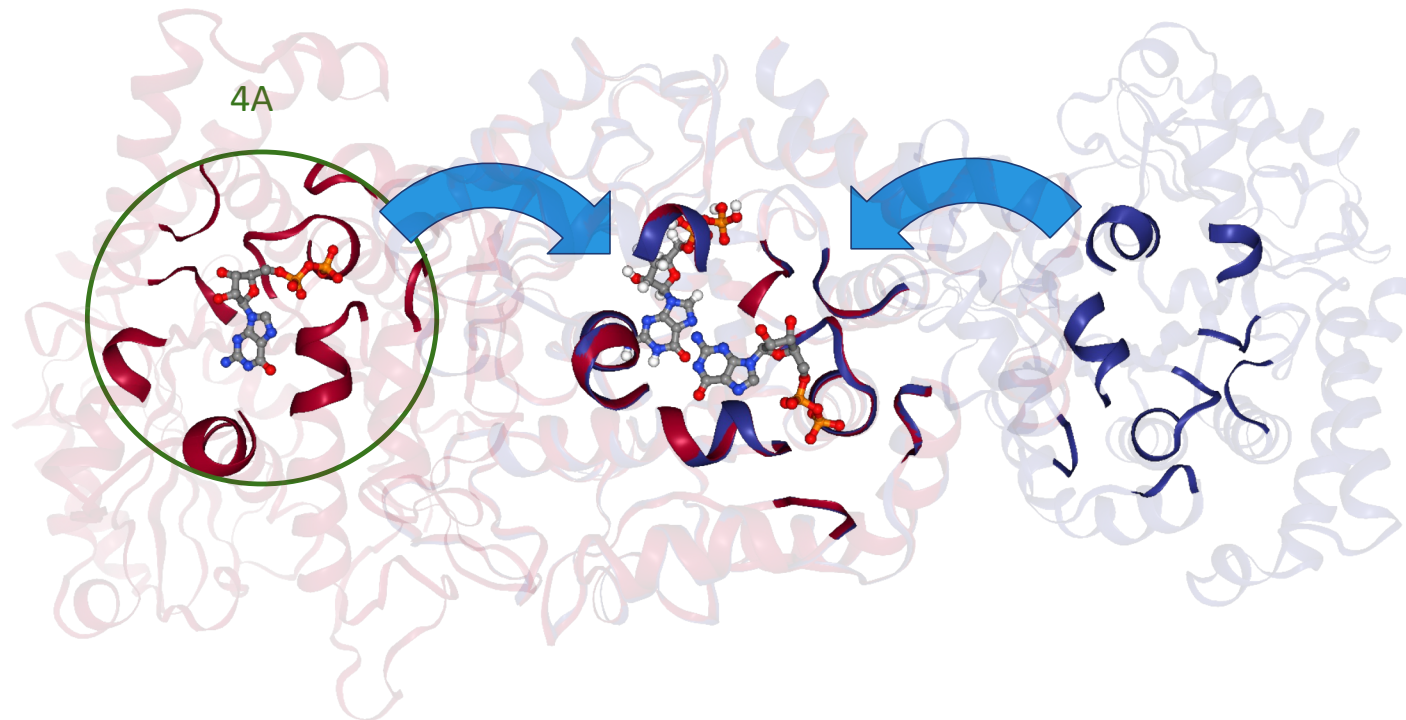
Symmetry-Corrected Pose RMSD – Binding site extraction

7EVE - ADP



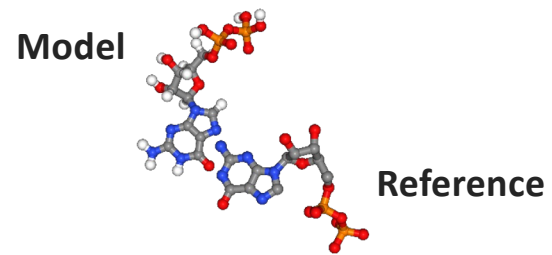
Symmetry-Corrected Pose RMSD – binding site superposition

7EVE - ADP



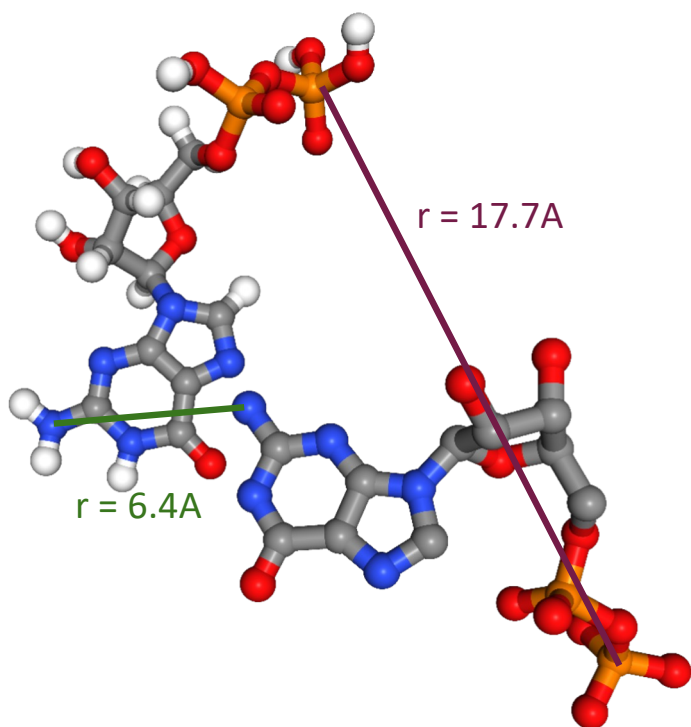
Symmetry-Corrected Pose RMSD

7EVE - ADP



Symmetry-Corrected Pose RMSD

7EVE - ADP



$$\text{RMSD}_{\text{standard}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{j=0}^2 (A_{ij} - B_{ij})^2}$$

$$\text{RMSD}_{\text{standard}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \mathbf{r}_i^2}$$

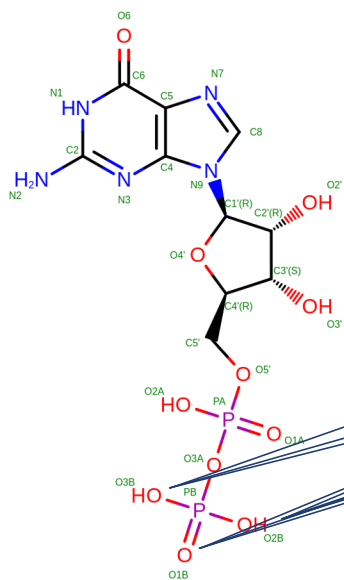
=> RMSD: 12.54

Symmetry-Correction

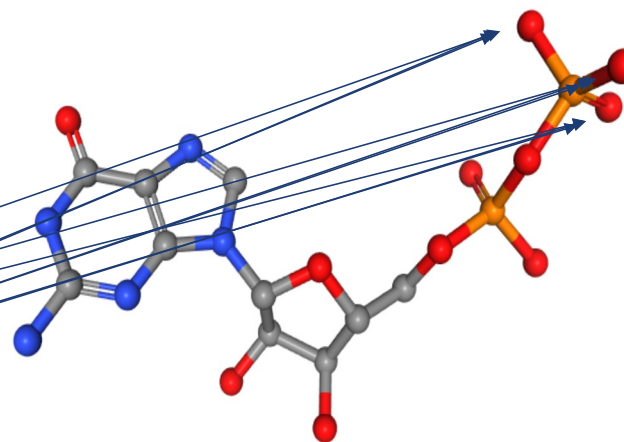
7EVE - ADP

Reference

Model



Symmetry



Solution:
Enumerate graph isomorphisms
Requires connectivity

Scores

Requirements:

- Compare prediction with the reference target
- Diverse scores representing different aspects of the prediction accuracy

No well-established scores exist for this challenge

We implemented two new scores – used in CASP for the 1st time

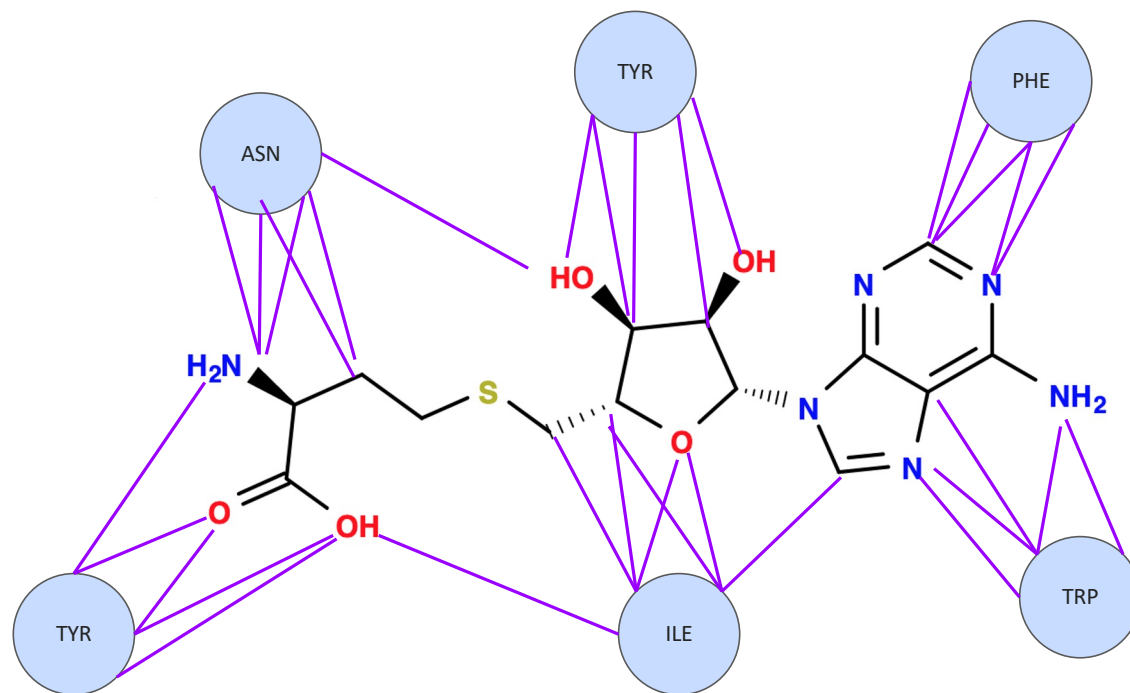
- Symmetry-Corrected Pose RMSD – similar to D3R – assesses the accuracy of the pose
- **IDDT-PLI – assesses protein-ligand contacts**

Chain mapping (for oligomers)

Ligand assignment (when > 1 ligand in target)

IDDT-PLI

Only considers contacts between the ligand and the protein/RNA chain
Symmetry correction



Parameters:

- Binding Site radius: 4Å
- IDDT inclusion radius: 6Å
- Thresholds: 0.5, 1, 2, 4 (default)

Caveats:

- Sequences must match (see T1146 and T1158v4)
- No penalty for added contacts

Scores

Requirements:

- Compare prediction with the reference target
- Diverse scores representing different aspects of the prediction accuracy

No well-established scores exist for this challenge

We implemented two new scores – used in CASP for the 1st time

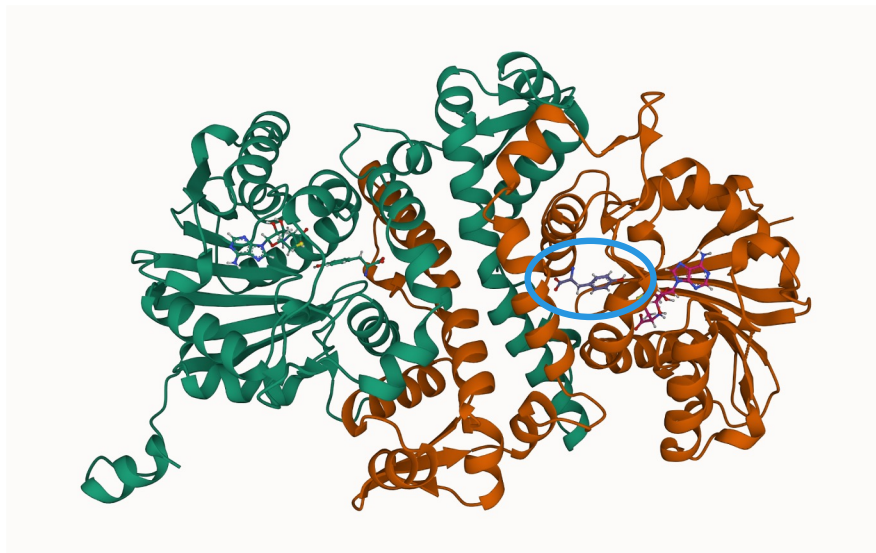
- Symmetry-Corrected Pose RMSD – similar to D3R – assesses the accuracy of the pose
- IDDT-PLI – assesses protein-ligand contacts

Chain mapping (for oligomers)

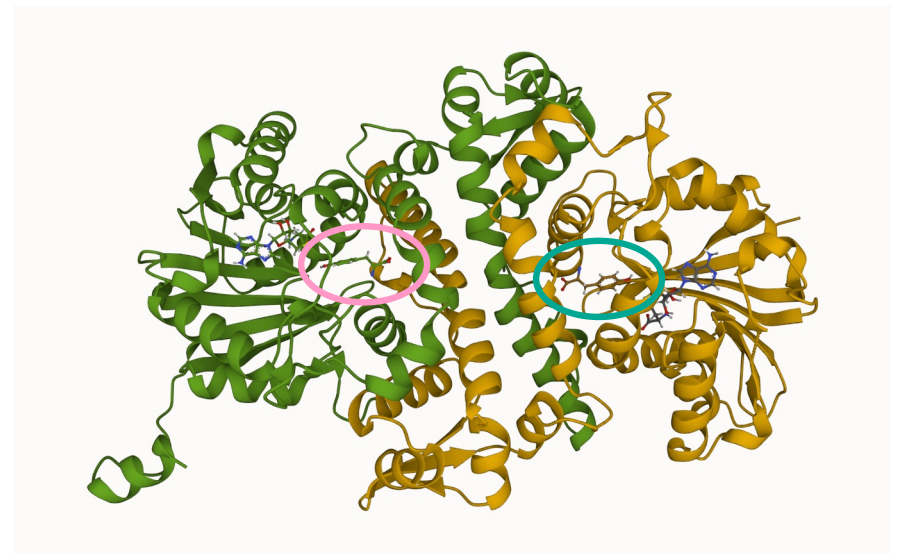
Ligand assignment (when > 1 ligand in target)

Chain mapping

T1124



Model



A → A, B → B

A → B, B → A

Ligand Assignment – T1124 example

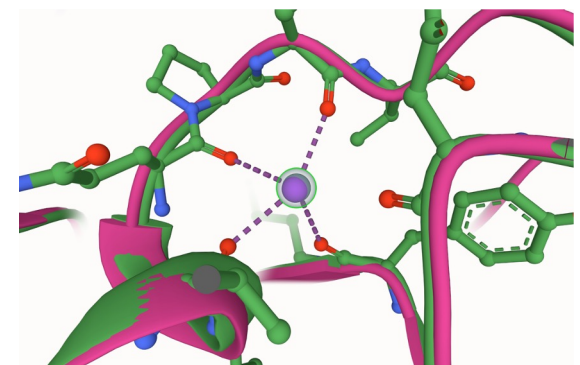
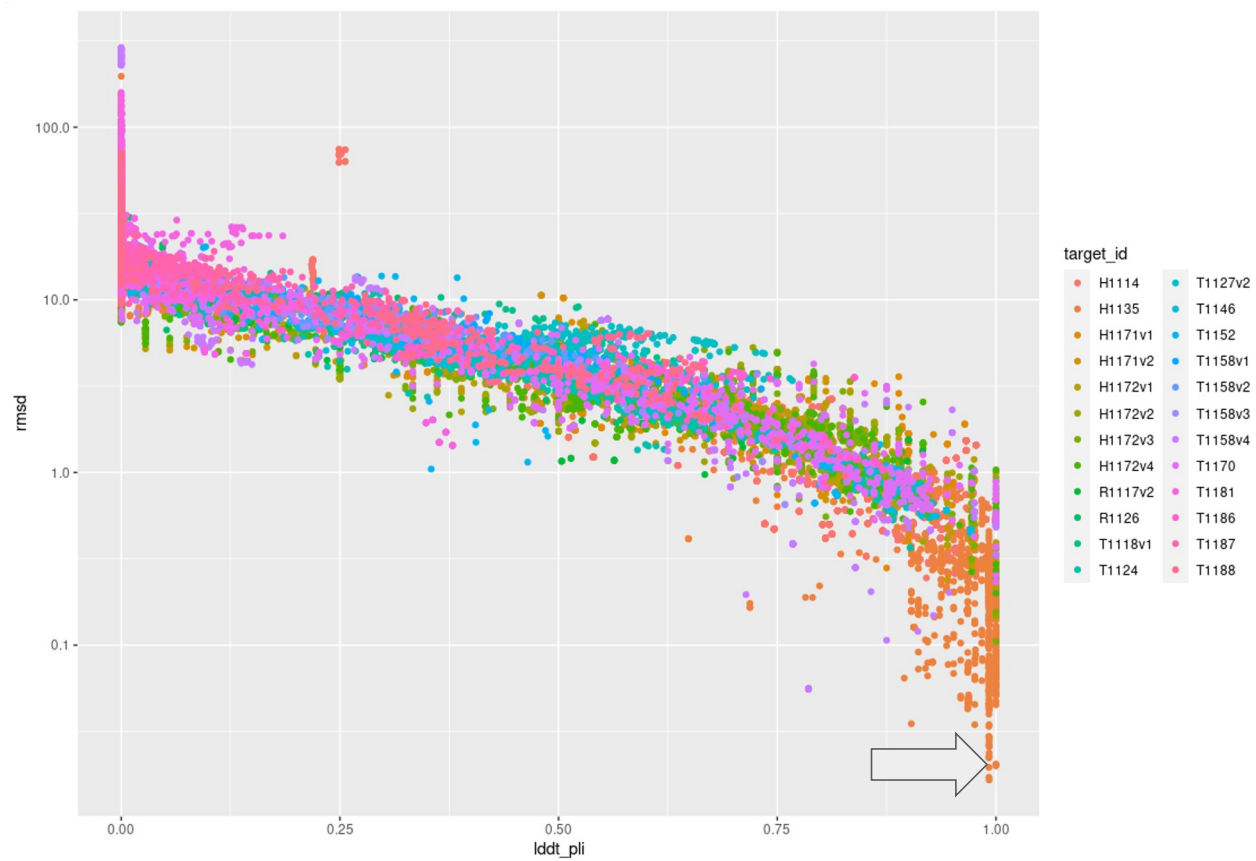
IDDT-PLI

		Model			
		001 SAH	002 SAH	003 TYR	004 TYR
Reference	SAH:A:401	0.8598	0.8598		
	SAH:B:402	0.8786	0.8786		
	TYR:A:402			0.7898	0.7898
	TYR:B:403			0.8017	0.8017

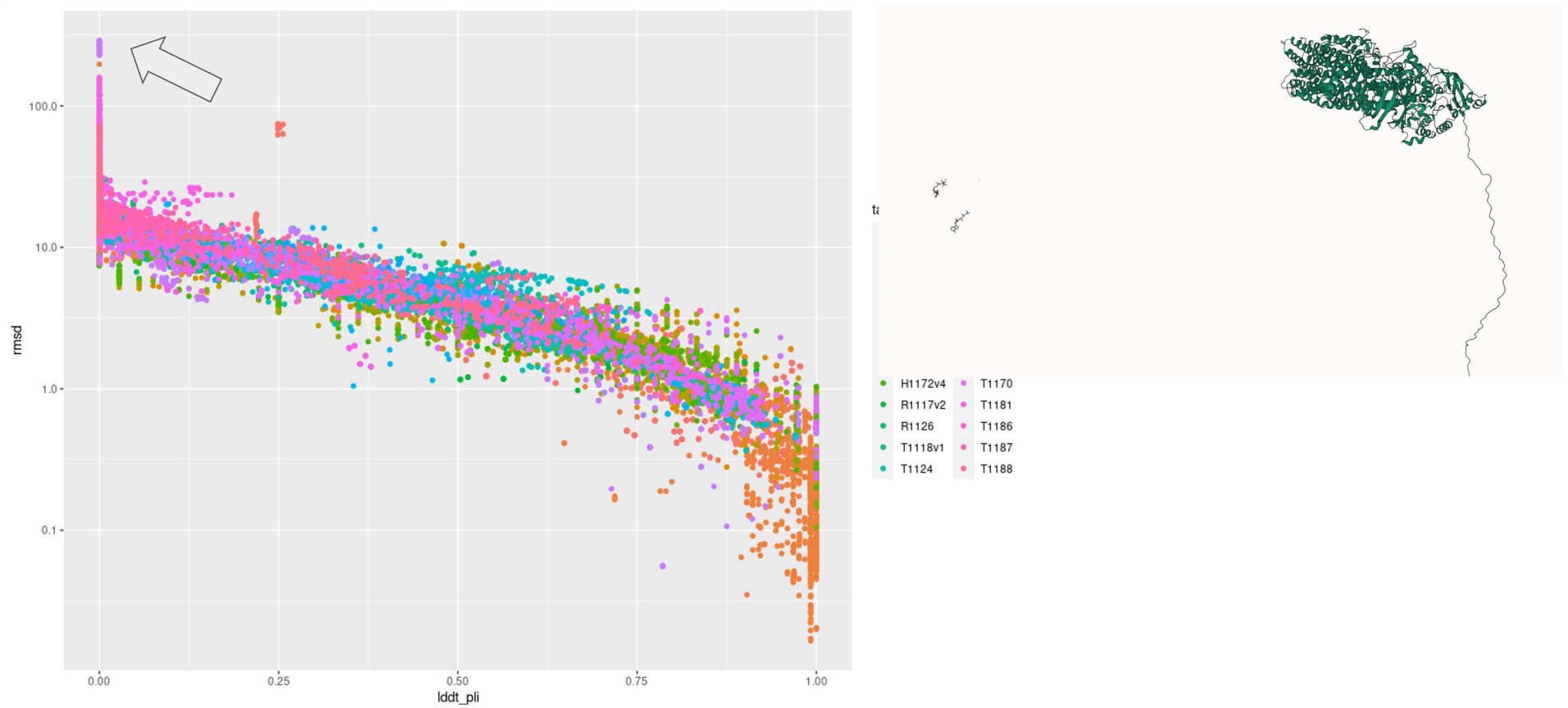
Scores comparison



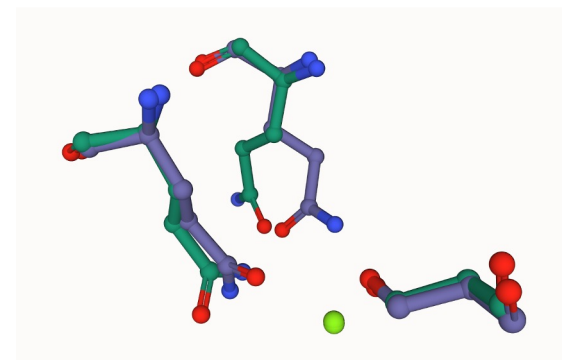
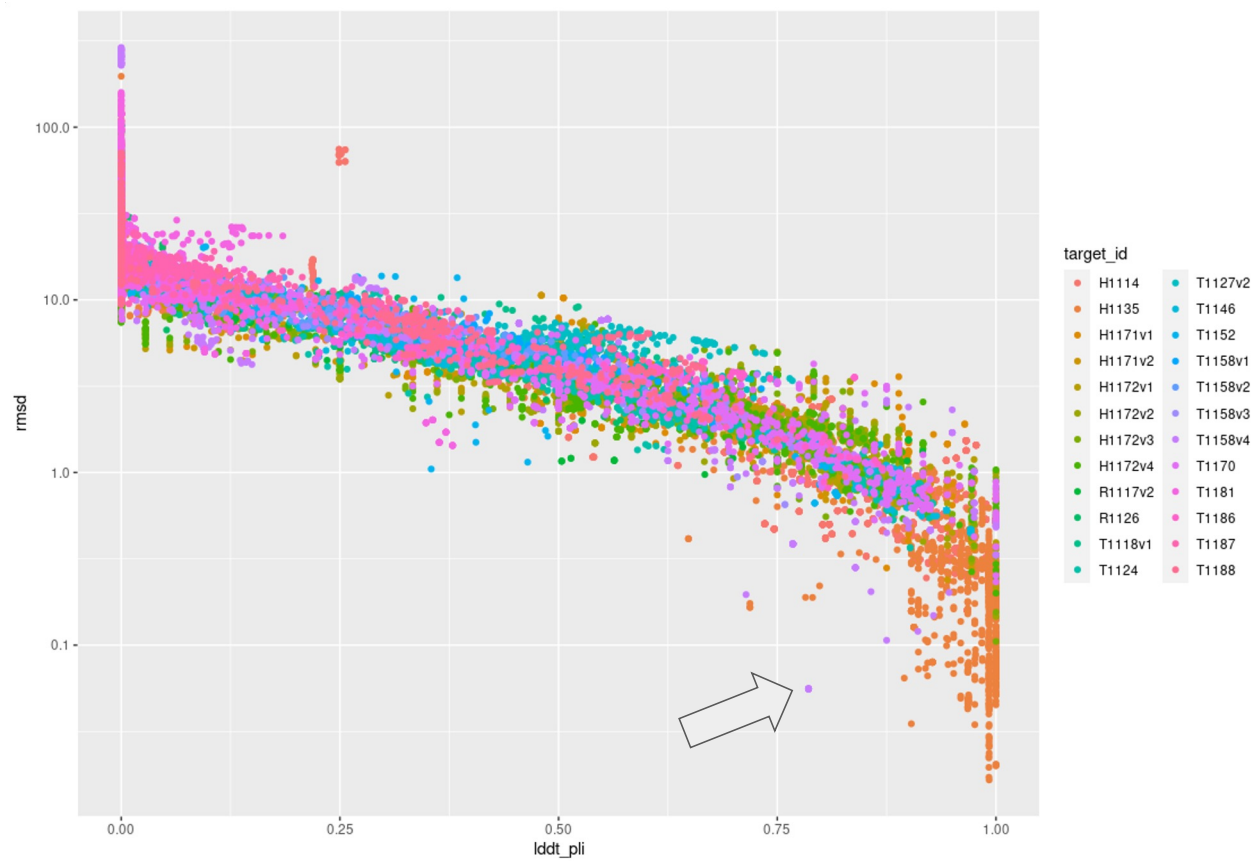
Scores comparison



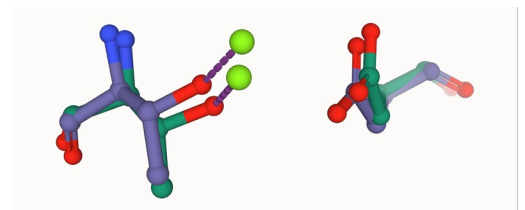
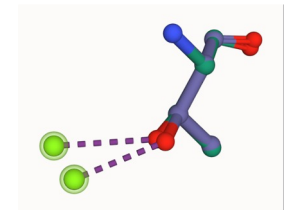
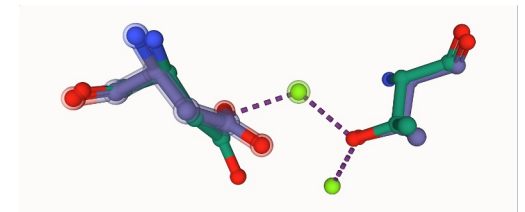
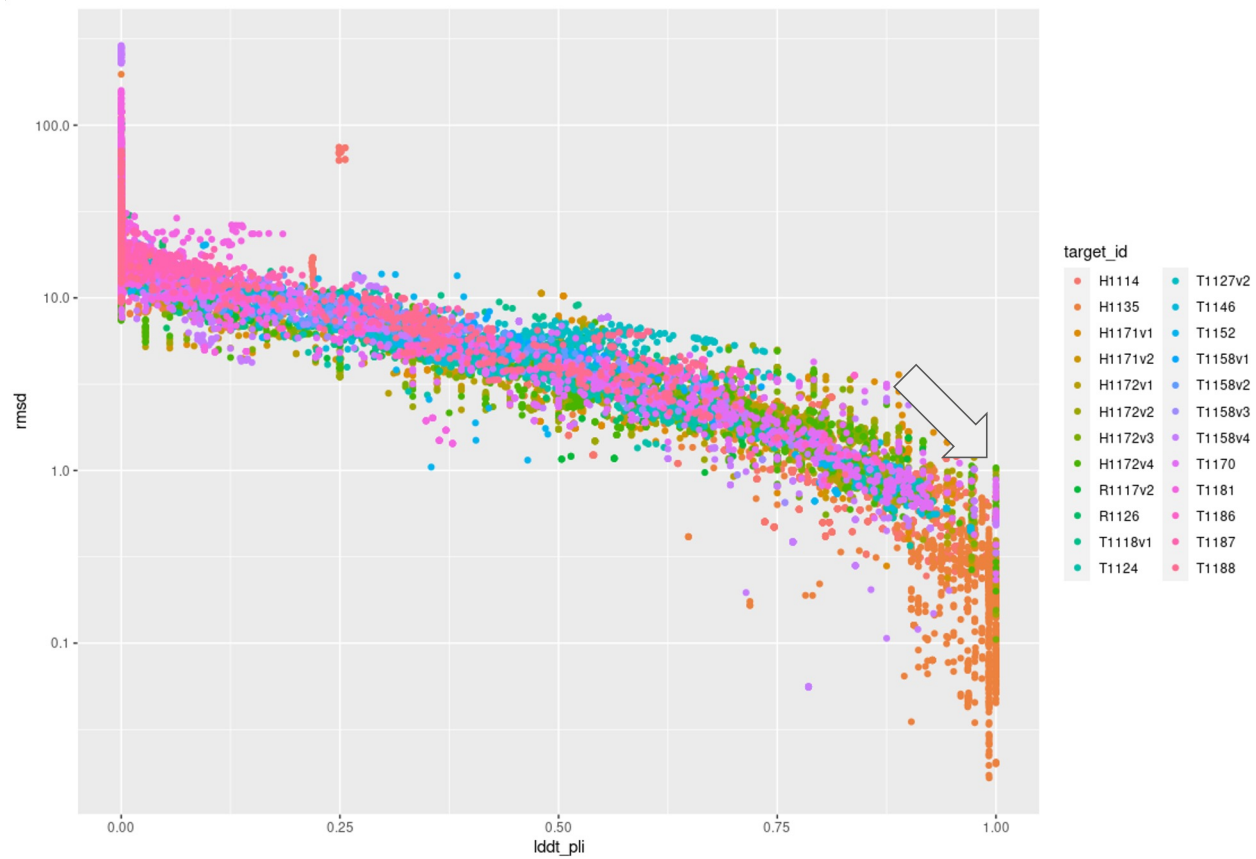
Scores comparison



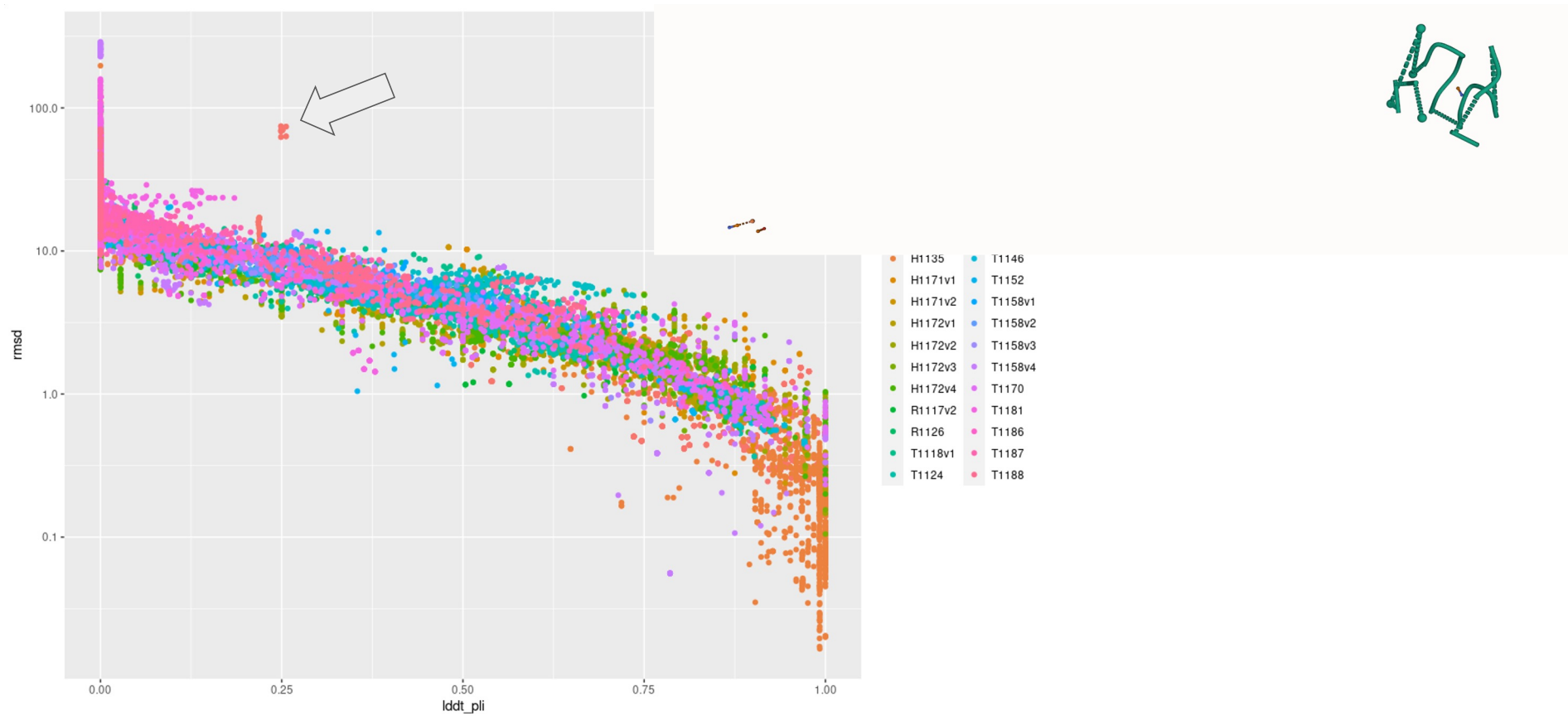
Scores comparison



Scores comparison



Scores comparison



CASP 15 is a Different Kind of Ligand Prediction Challenge

	Protein Coordinates	Binding Site Known	Ligands	Number of Ligands
D3R	Yes	Yes	Drug-like	1
CASP 15	No	No	Mixed	1-56

Published: 30 September 2016

D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions

Symon Gathiaka, Shuai Liu, Michael Chiu, Huanwang Yang, Jeanne A. Stuckey, You Na Kang, Jim Delproposto, Ginger Kubish, James B. Dunbar Jr., Heather A. Carlson, Stephen K. Burley, W. Patrick Walters, Rommie E. Amaro [✉](#), Victoria A. Feher [✉](#) & Michael K. Gilson [✉](#)

Journal of Computer-Aided Molecular Design **30**, 651–668 (2016) | [Cite this article](#)

2237 Accesses | 143 Citations | 12 Altmetric | [Metrics](#)

Published: 04 December 2017

D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies

Zied Gaieb, Shuai Liu, Symon Gathiaka, Michael Chiu, Huanwang Yang, Chenghua Shao, Victoria A. Feher, W. Patrick Walters, Bernd Kuhn, Markus G. Rudolph, Stephen K. Burley, Michael K. Gilson [✉](#) & Rommie E. Amaro

Journal of Computer-Aided Molecular Design **32**, 1–20 (2018) | [Cite this article](#)

3549 Accesses | 119 Citations | 11 Altmetric | [Metrics](#)

Published: 10 January 2019

D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings

Zied Gaieb, Conor D. Parks, Michael Chiu, Huanwang Yang, Chenghua Shao, W. Patrick Walters, Millard H. Lambert, Neysa Nevins, Scott D. Bembenek, Michael K. Ameriks, Tara Mirzadegan, Stephen K. Burley, Rommie E. Amaro [✉](#) & Michael K. Gilson [✉](#)

Journal of Computer-Aided Molecular Design **33**, 1–18 (2019) | [Cite this article](#)

1638 Accesses | 71 Citations | [Metrics](#)

Published: 23 January 2020

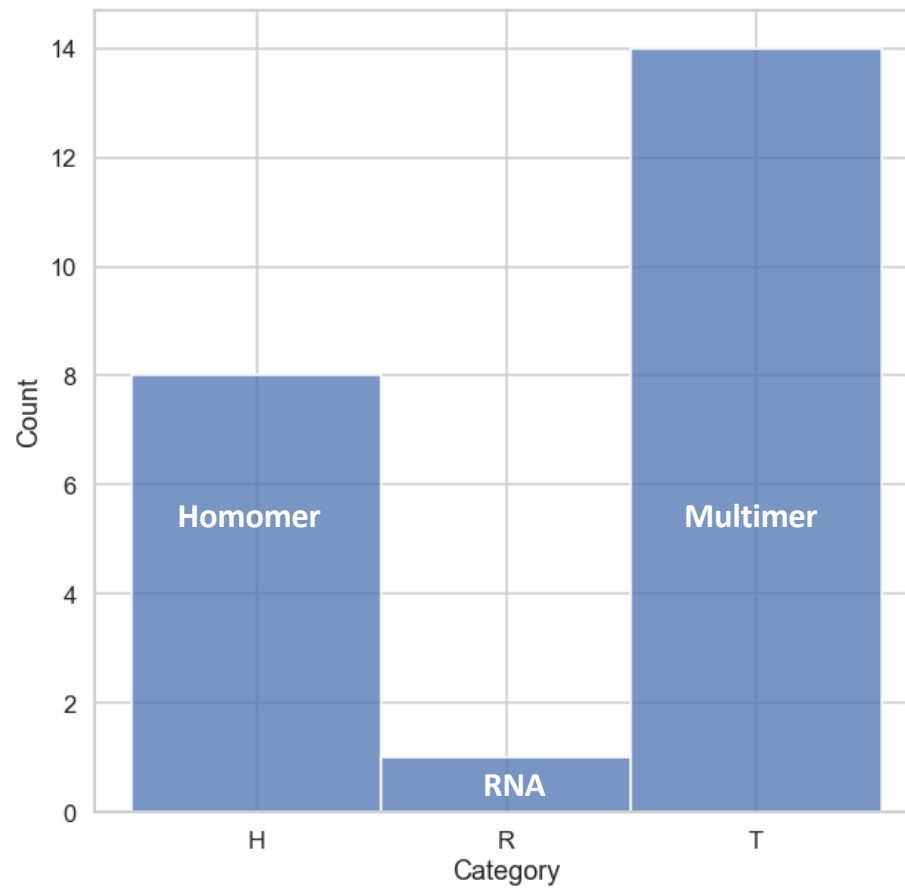
D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies

Conor D. Parks, Zied Gaieb, Michael Chiu, Huanwang Yang, Chenghua Shao, W. Patrick Walters, Johanna M. Jansen, Georgia McGaughey, Richard A. Lewis, Scott D. Bembenek, Michael K. Ameriks, Tara Mirzadegan, Stephen K. Burley, Rommie E. Amaro [✉](#) & Michael K. Gilson [✉](#)

Journal of Computer-Aided Molecular Design **34**, 99–119 (2020) | [Cite this article](#)

1485 Accesses | 47 Citations | 1 Altmetric | [Metrics](#)

CASP 15 Ligand Target Overview



R1126 was removed

Data Processing Challenges

Submissions without protein structures

Ligands incorrectly named

Ligands as PDB files

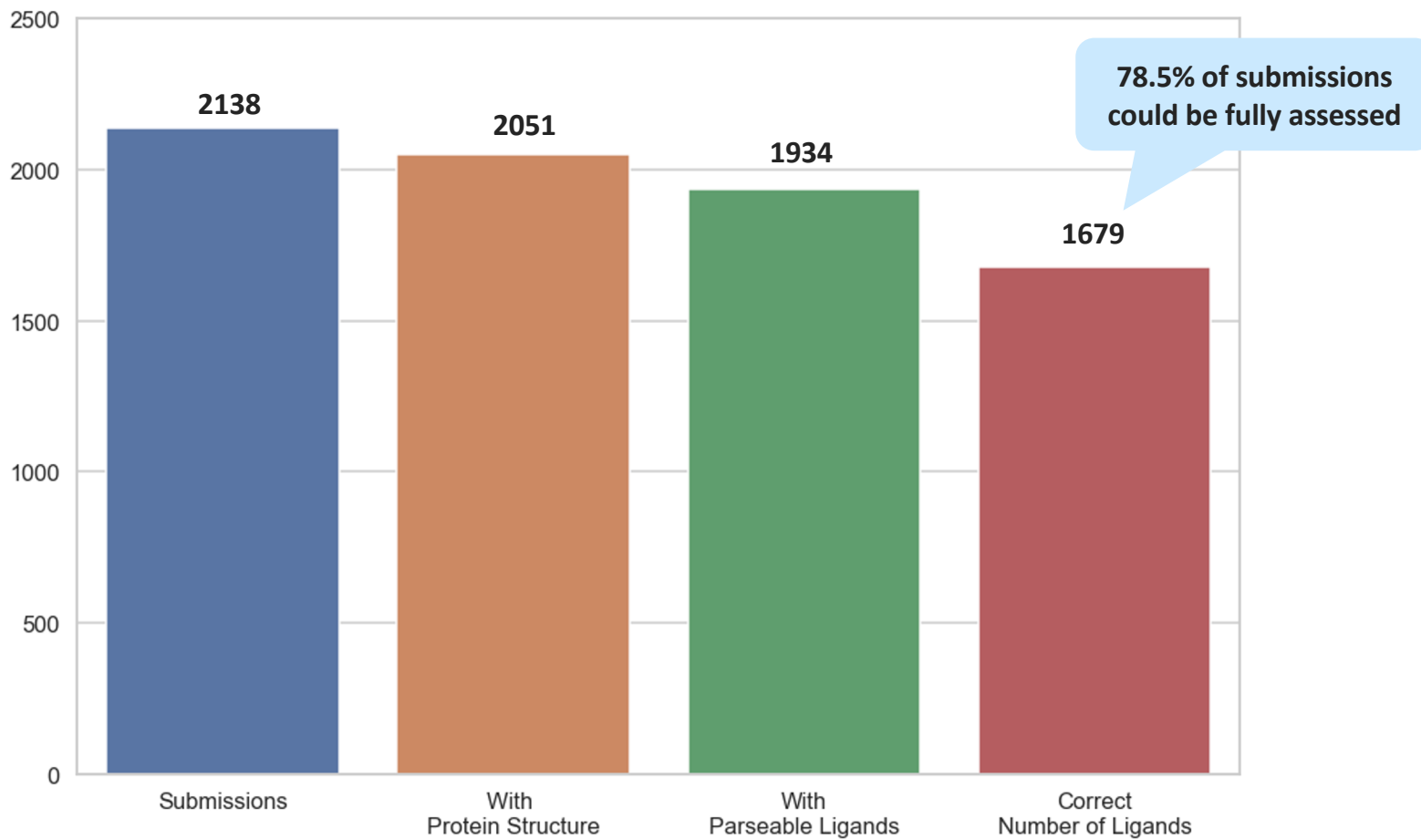
Proteins as V3000 molfiles

Ligands without bond orders

Ligands with and without charges

Submissions with an incorrect number of ligands

9.5% of Submissions Could Not Be Processed



Structure Corrections

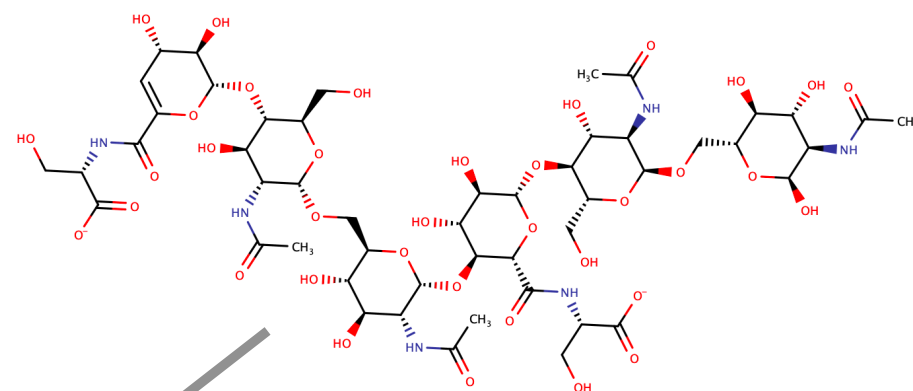
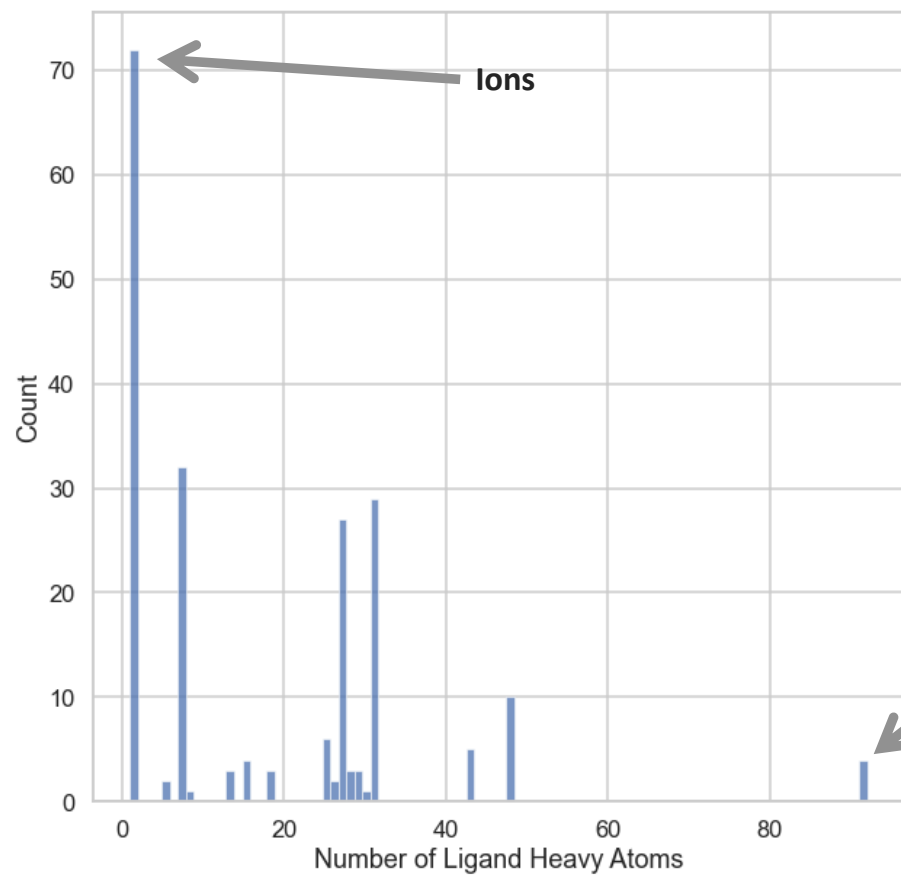
T1187 – specified 4 NAG ligands but the crystal structure has 2

T1146 – specified 2 NAG ligands but the crystal structure has 1

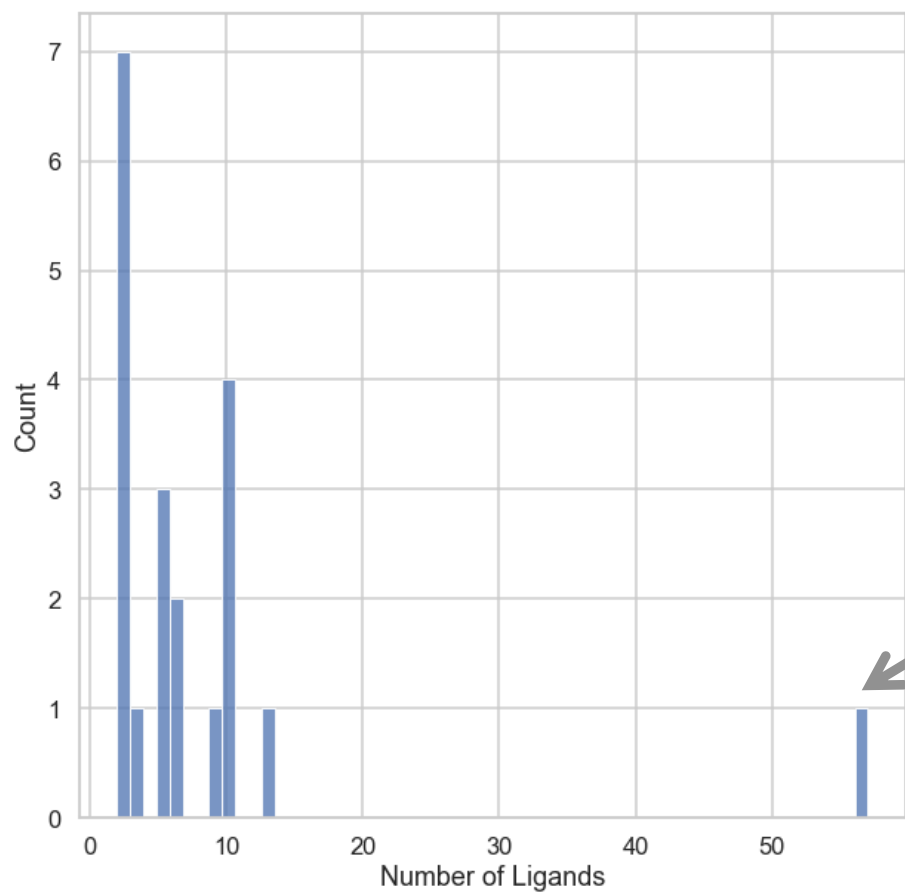
T1181 – crystal ligand OAA is a subset of the provided ligand structure

H1114 – crystal ligand MQ7 is a subset of the provided ligand structure

Number of Atoms in Ligands



Number of Ligands in Structures



Assessment is complicated when averaging
IDDT PLI or RMS

H1114 has 56 ligands

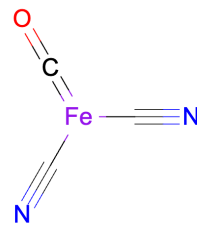
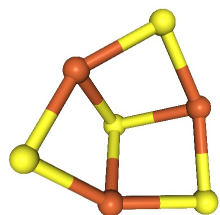
H11114

0.86

**EM
Multimer**

H1114 Has 56 Ligands

Ni^{+3}



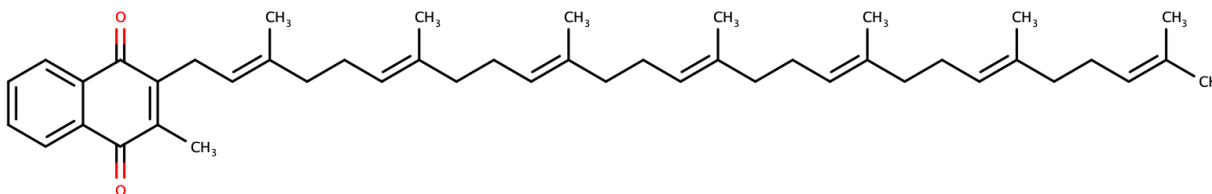
Mg^{+2}

$3\text{Ni} * 8$
NICKEL (III) ION

$\text{F3S} * 24$
FE3-S4 CLUSTER

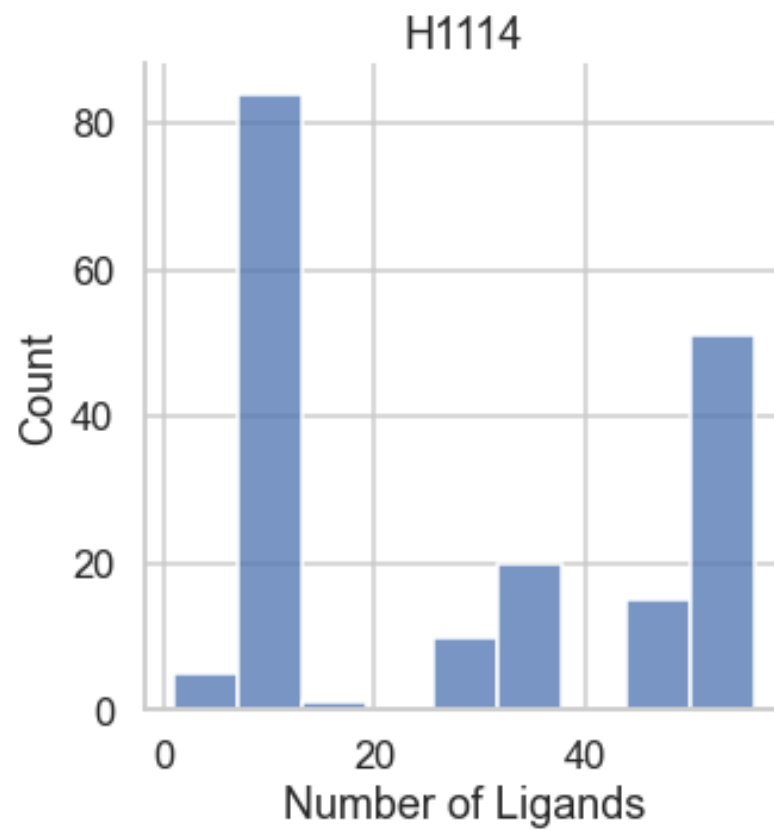
$\text{FCO} * 8$
CARBONMONOXIDE-(DICYANO) IRON

$\text{MG} * 8$
MAGNESIUM ION

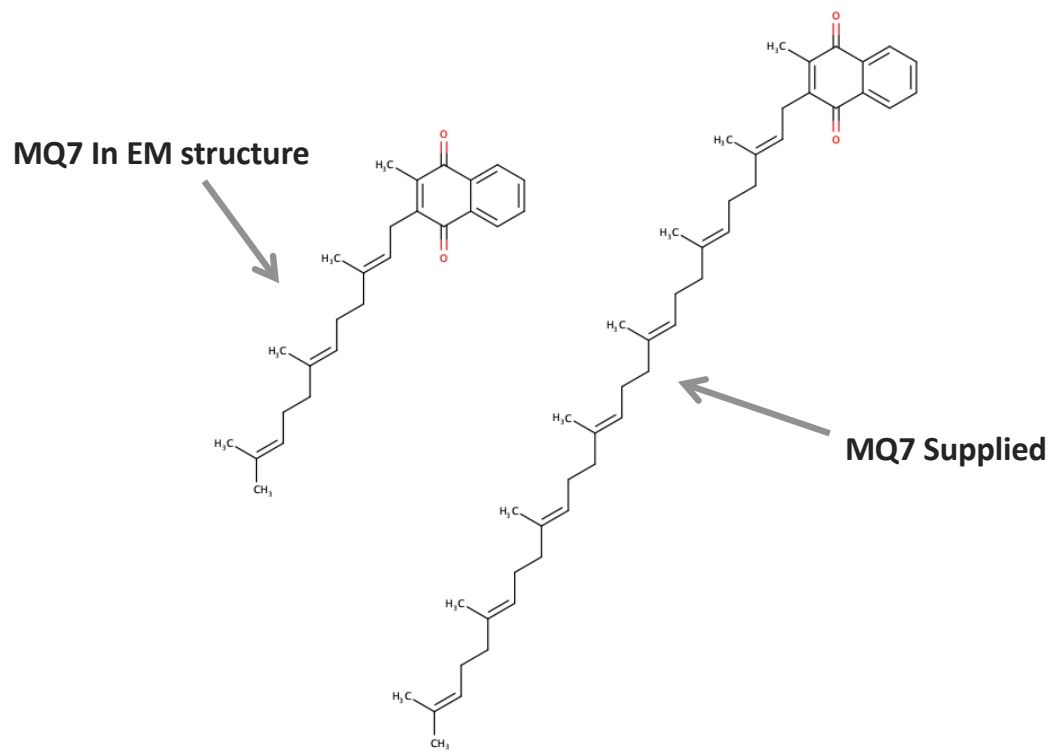
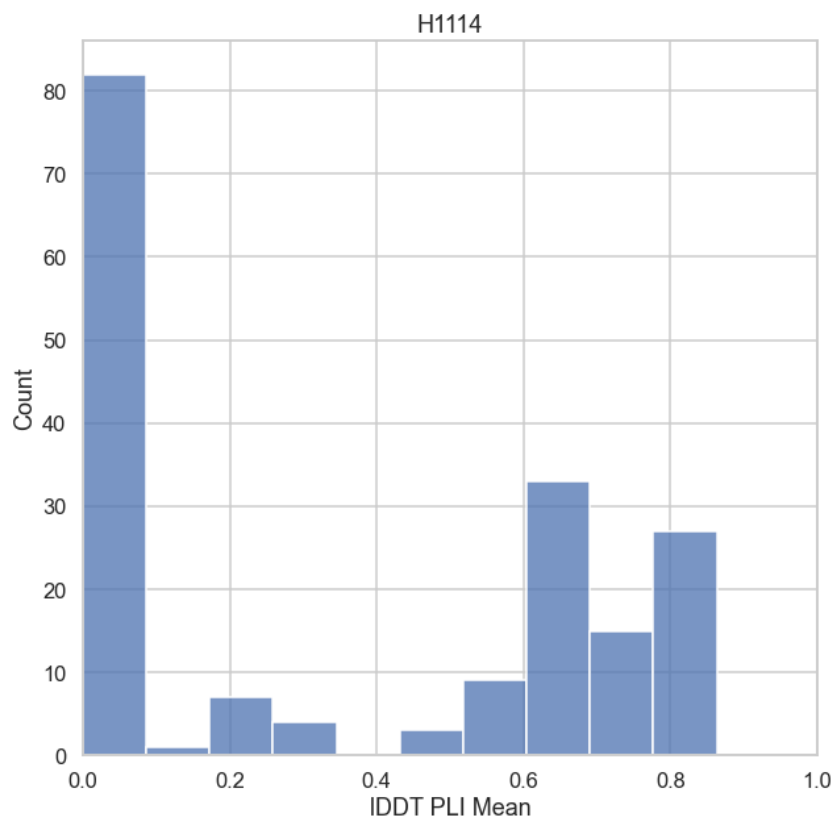


$\text{MQ7} * 8$
MENAQUINONE-7

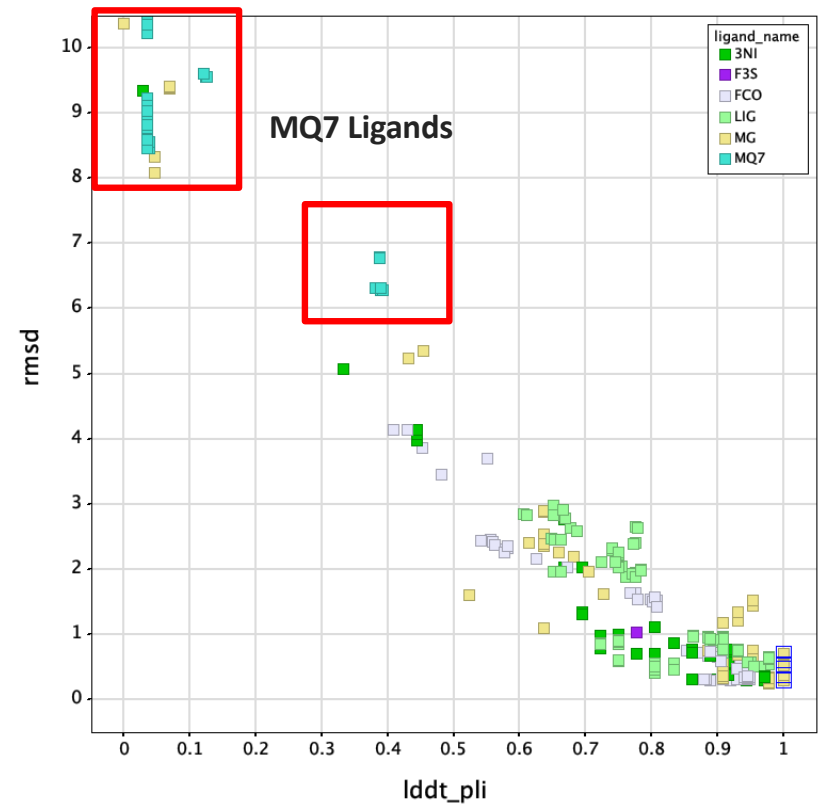
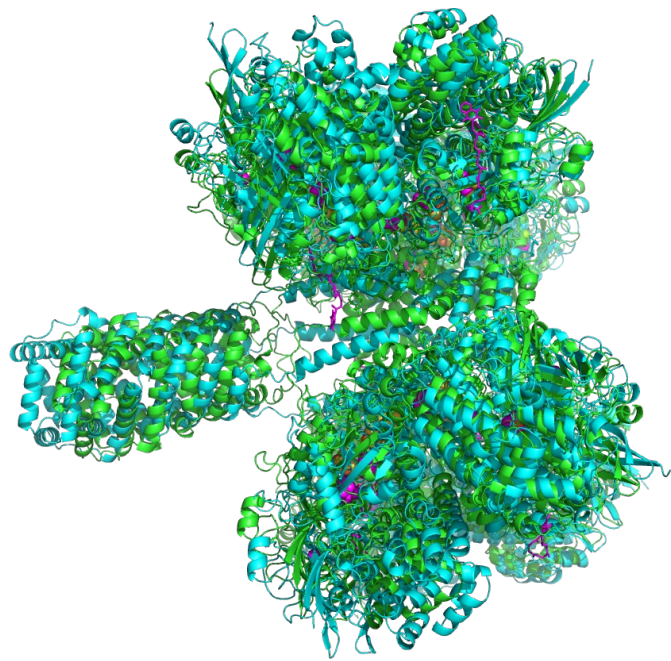
H1114 – Number of Submitted Ligands



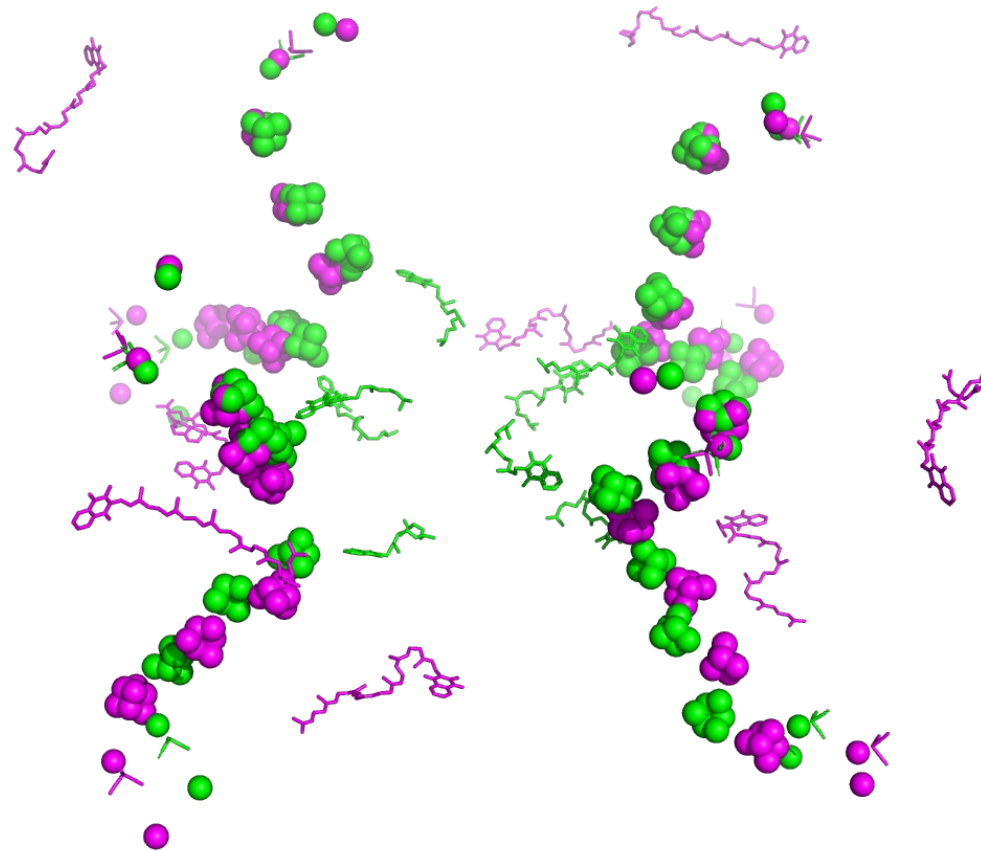
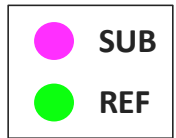
H1114



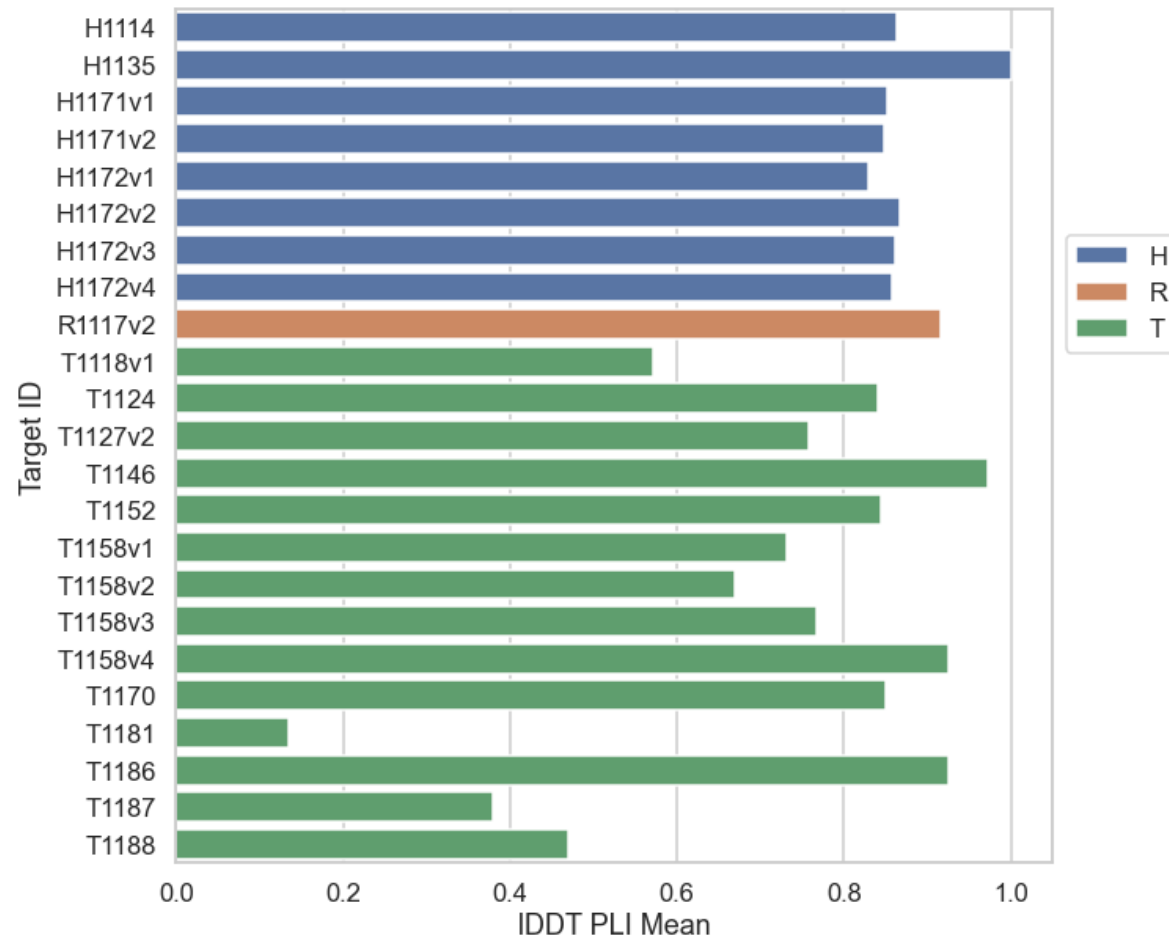
H1114



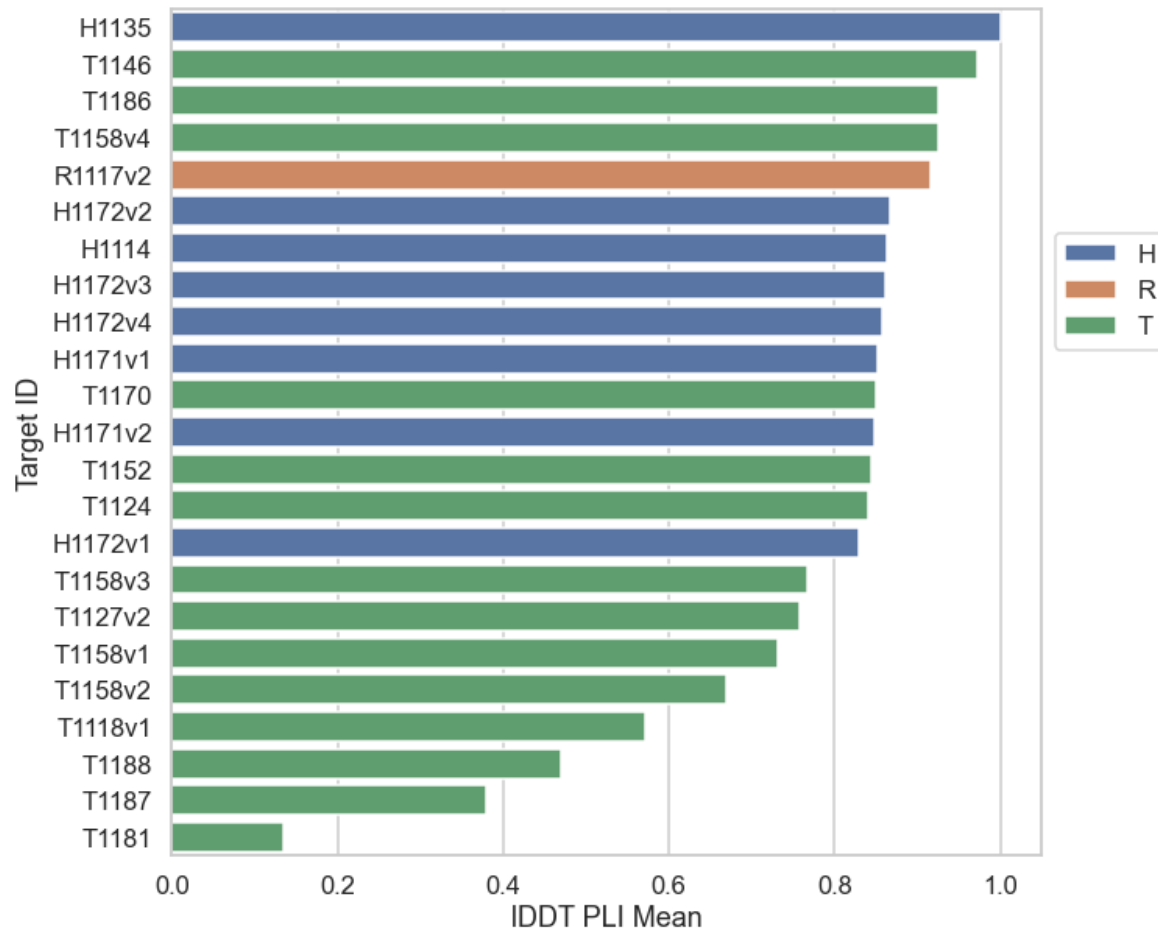
H1114



Best Submissions Ordered By Target



Best Submissions Ordered by Mean IDDT PLI



Best Performance By Group

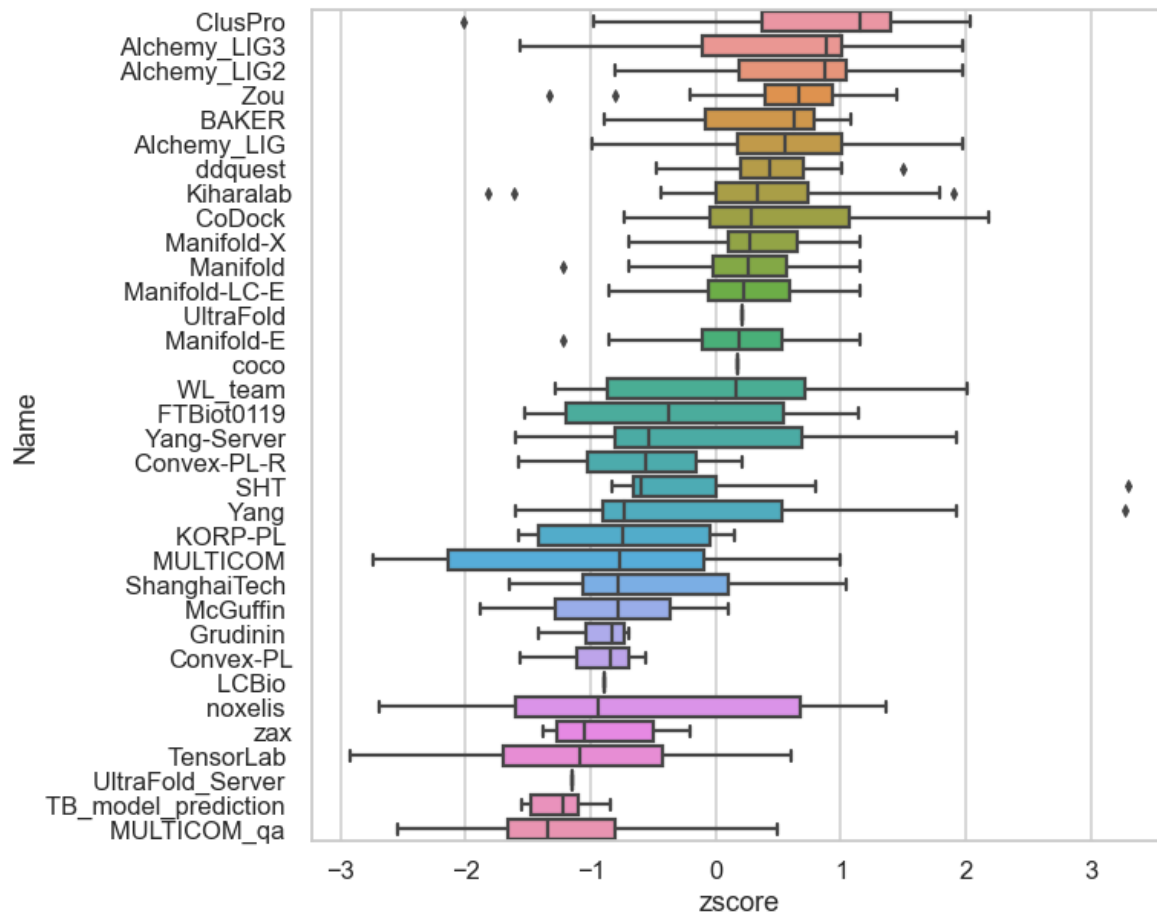
		Type	H1114	H1135	H1171v1	H1171v2	H1172v1	H1172v2	H1172v3	H1172v4	R1117v2	T1118v1	T1124	T1127v2	T1146	T1152	T1158v1	T1158v2	T1158v3	T1158v4	T1170	T1181	T1186	T1187	T1188	sum
Zou	Alchemy_LIG2	Human	0	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0	0	1	1	0	0	1	14
	ClusPro	Human	0	1	0	0	1	1	1	1	1	0	1	0	1	1	0	0	1	0	1	0	1	0	0	12
	Alchemy_LIG3	Human	0	0	1	1	1	1	1	1	0	0	0	1	1	0	1	1	0	0	0	0	1	0	0	11
	Alchemy_LIG	Human	0	1	0	0	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	1	0	0	0	10
	CoDock	Human	0	1	0	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0	0	1	0	1	0	10
		Human	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	0	0	9
		Human	0	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	7
		Human	0	0	0	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	7
		Server	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	6
		Human	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	5
	Human	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	5	
LG248	Manifold	Human	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
LG035	Manifold-E	Server	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
LG046	Manifold-LC-E	Server	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	
LG119	Kiharalab	Human	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	
LG236	noxelis	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
LG472	ddquest	Human	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
LG367	MULTICOM	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
LG165	FTBiot0119	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
LG225	ShanghaiTech	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
LG147	SHT	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
LG460	Convex-PL-R	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG054	UltraFold	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG086	MULTICOM_qa	Server	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG088	coco	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG392	LCBio	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG180	McGuffin	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG352	KORP-PL	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG122	zax	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG199	TB_model_prediction	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG125	UltraFold_Server	Server	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG132	TensorLab	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG150	Grudinin	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LG338	Convex-PL	Human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Time Resolved croEM, 7 structures
All inputs specify 3*ADP, 3*ATPGS, 3*Mg+2

	ADP	AGS	Mg ²⁺	
T1170	2	4	4	10
H1171v1	1	5	5	11
H1171v2	3	3	3	9
H1172v1	3	3	3	9
H1172v2	4	2	3	9
H1172v3	4	2	2	8
H1172v4	3	3	3	9

kea indicates a prediction within 10% of the best

Best Performance By Z-Score



Conclusions

Good predictions on ions and endogenous ligands

Less so for larger, more flexible ligands

IDDT PLI provides a good assessment of ligand interactions

- Do we need additional metrics?

Averaging scores over multiple ligands doesn't reflect overall performance

Does CASP need different types of ligand prediction challenges?

Questions?

