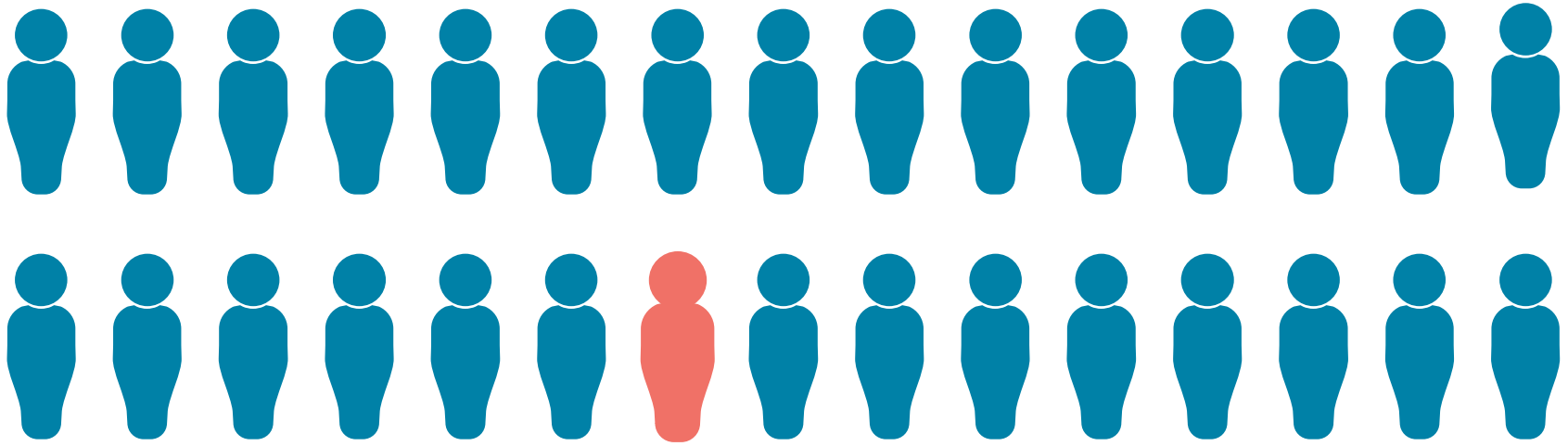
Mscreen: docking benchmarking made easy

Eduardo Mayo – Agrarian University of Havana
Estael Ochoa – University of Havana
Jacob Spiegel – Workflow Informatics Inc.
Jacob Durrant – University of Pittsburgh

@eddMayo

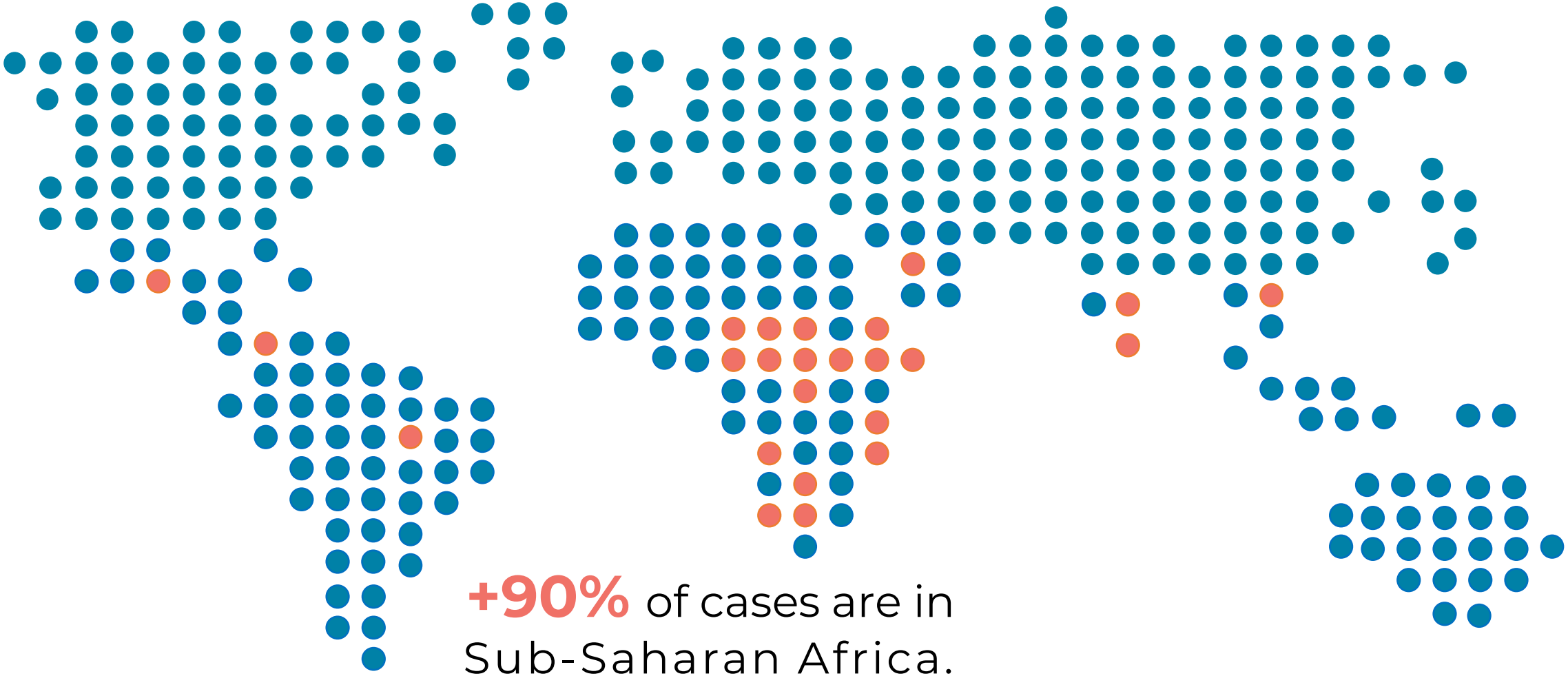
Malaria disease in 2019

- **1 out of 30** people worldwide [215 million cases]
- 6th cause of death in low-income countries [~409,000 deaths]



(2019). World malaria report. Geneva, Switzerland: World Health Organization.

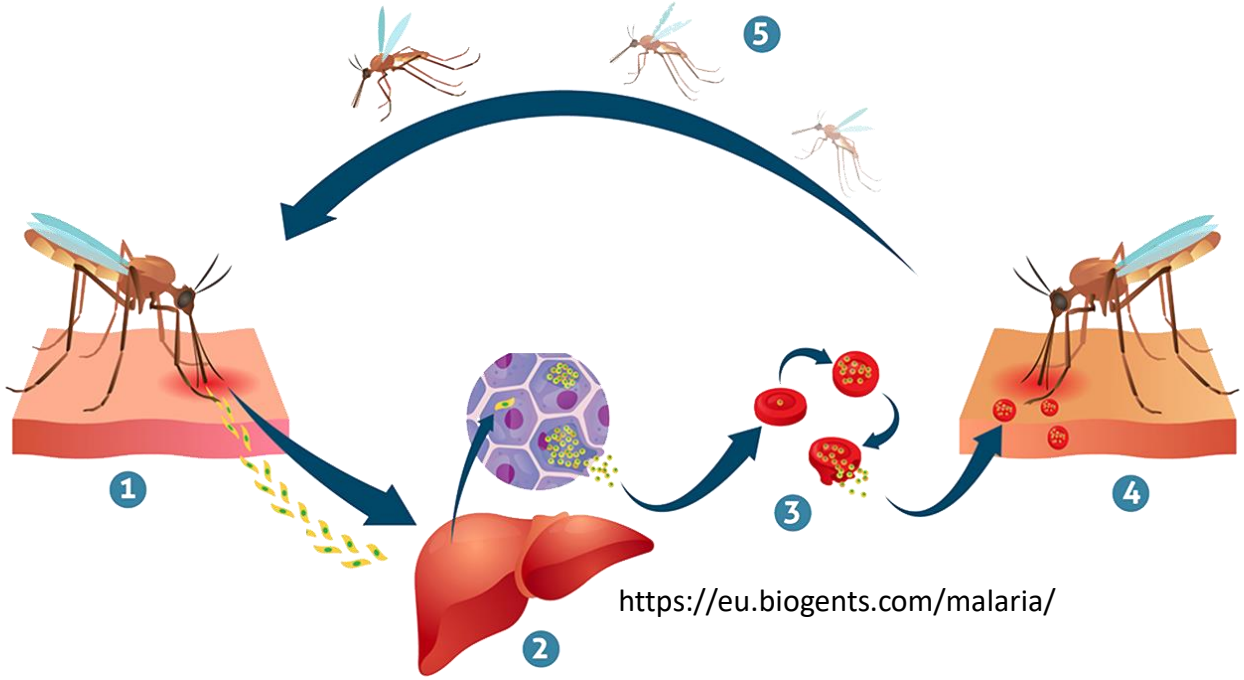
Malaria disease distribution



Malaria candidate targets

- P. falciparum
  - most lethal human parasite species
  - resistant to front-line antimalarials

- P. falciparum M1 alanyl aminopeptidase (PfA-M1) may be a potential antimalarial target:
  - 40+ structures deposited in RCSB
  - Proven to function as a hemoglobinase
  - PfA-M1 inhibitors are lethal to malaria *in situ* and reduce parasite growth in murine models



https://eu.biogents.com/malaria/

PfA-M1

We retrieved all proteins of the Organism "*Plasmodium falciprum*" and the Polymer Entity "**M1**"

We then selected 36 with **resolution ≤2Å**

The metal active site is buried in a deep pocket

Although the location of the residues is well conserved, the conformation of **R489, V459 and M1034** varies considerably.

GIF caption

Docking in drug discovery process

Drug Discovery and development stages

- Hit Discovery Process
- Hit to Lead
- Lead Optimization

Molecular docking

- Binding affinity prediction

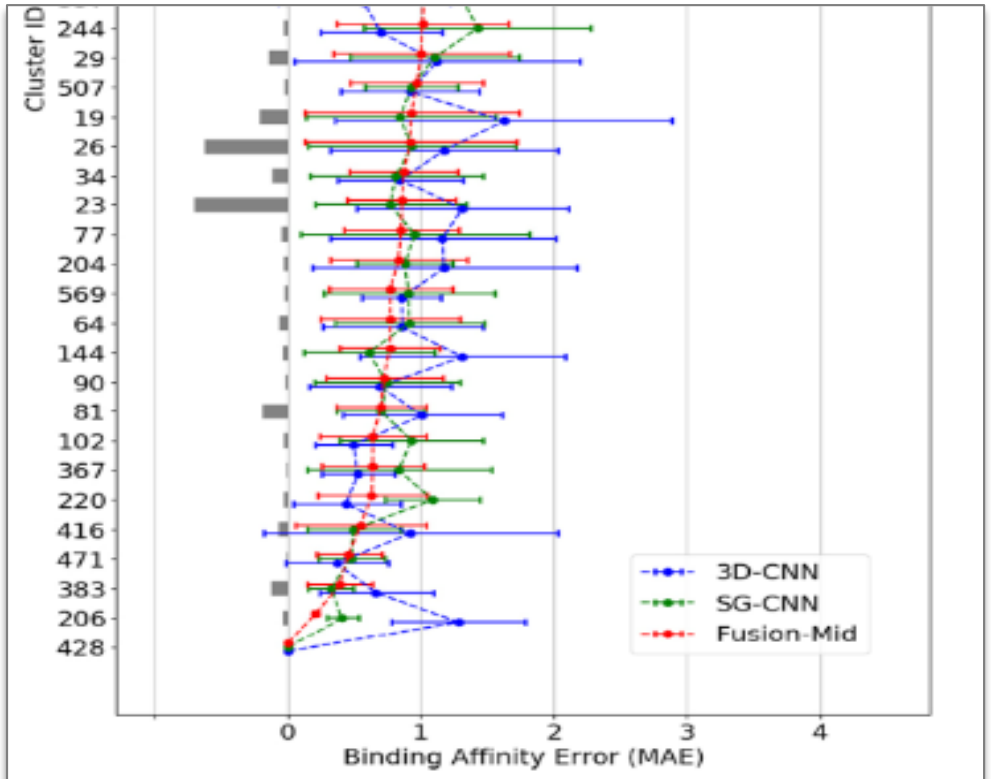


Figure 3. MAE (x-axis) with standard deviation for groups (y-axis) based on the pocket and the ligand positioning. MAE is shown for the machine learning models. The number of complexes in the refined training set is shown for each cluster (gray bars).

Jones, D., et al (2021). Improved Protein-Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference. *Journal of Chemical Information and Modeling*, *61*(4), 1583–1592. https://doi.org/10.1021/acs.jcim.0c01306

Molecular docking

- Binding affinity prediction
- Pose prediction accuracy

Table 1. RMSD values (in Å) and scoring energies [a] for the self-docking of each ligand in its binding site.

		AD Vina		AD 4.2		AD 3.0		DOCK		Glide	
		RMSD	Score	RMSD	Score	RMSD	Score	RMSD	Score	RMSD	Score
1	1ESV	**0.25**	−10.8	0.60	−8.52	0.62	−10.40	0.33	−67.42	0.37	−7.84
2	1FKD	**0.67**	−11.3	**0.71**	−11.10	0.66	−12.38	1.11	−61.33	1.14	−6.95
3	1NM6	**0.26**	−12.8	0.42	−12.79	1.22	−12.24	**0.23**	−88.46	1.07	−11.32
4	1NT1	0.65	−13.3	0.49	−13.50	1.99	−6.89	**0.20**	−85.14	0.79	−9.68
5	1PKF	**0.42**	−13.7	1.19	−12.73	1.12	−13.67	0.58	−77.96	1.97	−8.32
6	1R8Q	0.31	−12.6	0.83	−10.66	0.62	−11.69	**0.27**	−65.97	**0.28**	−11.73
7	1UU3	0.48	−11.9	0.97	−11.69	1.00	−12.88	0.84	−84.54	0.61	−10.27
8	1W96	0.48	−11.7	0.67	−12.03	0.73	−13.92	0.31	−95.40	0.60	−8.86
9	2C6H	**0.30**	−11.1	0.59	−9.99	0.65	−12.36	0.70	−79.95	1.27	−6.61
10	2E9U	**0.38**	−10.2	0.47	−9.10	0.67	−10.20	0.22	−67.52	**0.39**	−8.68
11	2IYA	0.49	−13.2	0.50	−11.60	**0.40**	−14.73	0.73	−84.69	1.39	−9.37
12	2VWC	0.31	−9.4	1.45	−9.15	0.80	−10.43	0.51	−78.13	1.04	−6.94
13	2XBK	0.62	−18.5	1.50	−15.23	1.05	−18.33	0.66	−115.34	1.05	−11.64
14	2XX5	**0.52**	−11.3	0.75	−10.58	0.83	−8.92	1.04	−73.64	0.70	−8.84
15	3DV1	0.74	−10.9	1.17	−11.57	0.94	−14.29	0.60	−94.39	1.94	−9.25
16	3DV5	1.02	−11.6	0.50	−15.36	**0.37**	−17.22	0.49	−113.99	2.37 [b]	−8.20
17	3EKS	0.91	−11.5	0.58	−11.77	0.73	−12.06	**0.35**	−75.36	**0.35**	−10.00
18	3QTF	0.51	−13.3	0.95	−10.88	1.13	−12.57	**0.44**	−78.45	0.82	−11.53
19	3UYK	0.97	−10.7	1.26	−10.03	1.39	−10.76	0.79	−55.11	**0.63**	−7.92
20	4DRU	**0.68**	−13.1	1.16	−12.99	1.16	−12.92	0.93	−71.69	1.03	−8.78
mean		0.55 ± 0.05		0.84 ± 0.08		0.90 ± 0.08		0.57 ± 0.06		0.99 ± 0.13	
SD		0.23		0.34		0.37		0.28		0.57	
median		0.50		0.73		0.82		0.55		0.93	

[a] Energies in kcal/mol, except for DOCK (in kJ·mol^{-1}), as given by each method; [b] RMSD values higher than 2.00 are written in red throughout the Tables.

Castro-Alvarez, et al. (2017). The Performance of several docking programs at reproducing protein-macrolide-like crystal structures. *Molecules*, *22*(1). https://doi.org/10.3390/molecules22010136

Molecular docking

- Binding affinity prediction
- Pose prediction accuracy
- Ligand/decoy discrimination

Table 1. Virtual screen AUC ROC values corresponding to nine protein targets and five distinct docking programs/scoring functions. [a]

	NNScore 1.0	NNScore 2.0	$AutoDock_{Fast}$	$AutoDock_{Rigorous}$	Vina
AChE	0.55	0.57	*0.48*	0.53	**0.67**
COX-2	**0.74**	*0.49*	*0.38*	*0.43*	*0.31*
DHFR	0.72	0.83	0.83	**0.95**	0.76
EGFr	*0.47*	0.51	0.51	*0.49*	**0.61**
FGFr1	**0.58**	0.55	*0.41*	*0.35*	*0.46*
FXa	**0.76**	0.52	*0.44*	*0.47*	0.63
P38	**0.75**	0.58	*0.40*	*0.37*	0.54
PDGFrb	0.60	**0.62**	0.50	*0.36*	0.53
SRC	0.58	0.63	0.65	0.58	**0.69**
Average	**0.64**	0.59	0.51	0.50	0.58

[a] Italic indicates ROC AUC less than 0.5. Bold indicates that the given docking program/scoring function is best suited to the corresponding receptor.

Durrant, J. D., & McCammon, J. A. (2011). NNScore 2.0: A neural-network receptor-ligand scoring function. *Journal of Chemical Information and Modeling*, *51*(11), 2897–2903. https://doi.org/10.1021/ci2003889

Is benchmarking worth it?

Structural Insights and Binding Analysis for Determining the Molecular Bases for Programmed Cell Death Protein Ligand-1 Inhibition

Rita C. Acurcio,[a] Carlota Leonardo-Sousa,[a] Alfonso T. García-Sosa,[b] Helena F. Florindo,[a*] Rita C. Guedes[a*]

Benchmarking the Ability of Common Docking Programs to Correctly Reproduce and Score Binding Modes in SARS-CoV-2 Protease Mpro

Shani Zev, Keren Raz, Renana Schwartz, Reem Tarabeh, Prashant Kumar Gupta, and Dan T. Major*

Cite This: *J. Chem. Inf. Model.* 2021, 61, 2957–

Aims

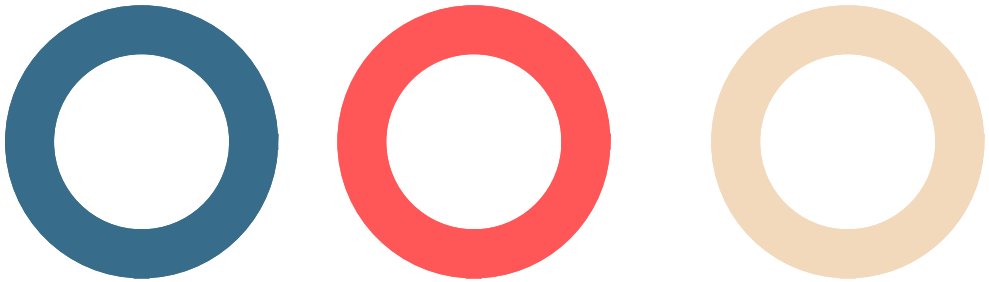
- Identify the best structural model for molecular docking studies
- Identify the docking program that performs the best regarding pose reproduction

4ZQT	4X2U	4J3B
4K5N	4K5O	4K5P
4K5L	4K5M	4R5T
4R5V	4R5X	3Q43
3Q44	4ZX3	4ZX5
4ZX4	4ZX6	4ZW3
4ZW5	4ZW6	4ZW7
4ZW8	6EA1	6EAB
6EAA	6EA2	6EE3
6EE4	6EED	6EE6
5XM7	5Y1H	5Y1K
5Y19	5Y1V	5Y1X
5Y1W	5Y1R	5Y1Q
5Y1T	5Y1S	5Y3I
3EBI	3EBG	3T8V
3EBH	6SBR	6SBQ

Methods

- **Self-docking:** Compares docking program pose reproduction and binding affinity estimation.

- **Cross-docking:** Assesses which protein structure is most suitable for docking studies.



MScreen

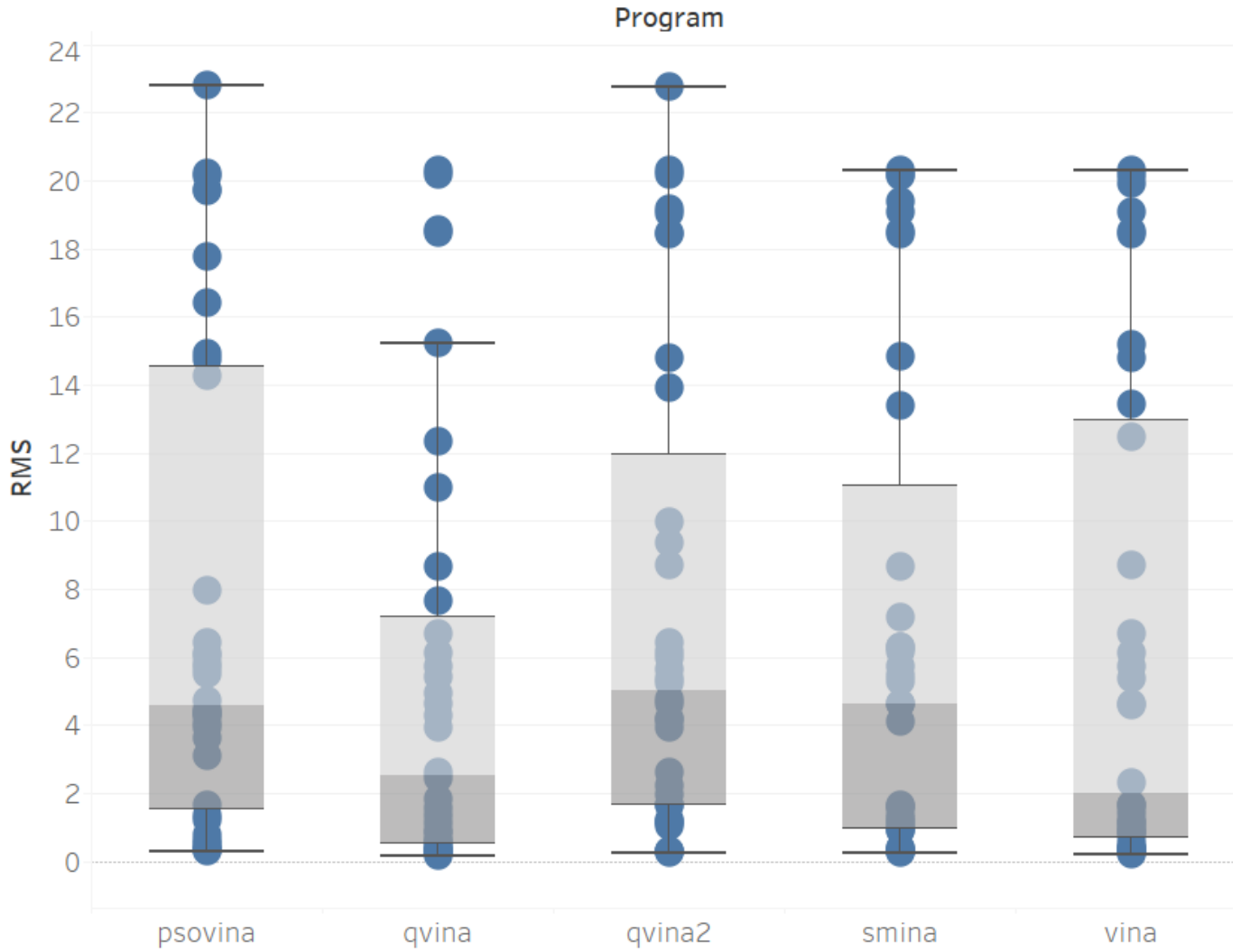
- **Carefully designed to aid** docking program benchmarking
- Over 10 different docking programs
- **Straight forward** protein and ligand preparation, docking and results

```
python mscreen.py prepare -i ligands -r receptors
                           [-d docking_program]

python mscreen.py screen -i ligands -r receptors
                          [-d docking_program]

python mscreen.py analysis -i folder -r file.sdf
                           [-d docking_program]
```

Self-Docking



The RMS distribution of the self-docking experiment shows that AutoDock Vina and Qvina reproduce the pose of the co-crystalized ligands with higher accuracy

Experimental conditions:

- 36 protein-ligand complexes
- Default pose sampling settings

Self-Docking

- AutoDock Vina 1.1.2 predicts poses with lower RMS for Pf1-M1/ligand complex
- Vina, QVina, and SMINA reproduce poses with RMS <0.5 Å for over 20% of protein-ligand complexes.
    - These three software, performed better than the others regarding the mean RMS values of their predictions.

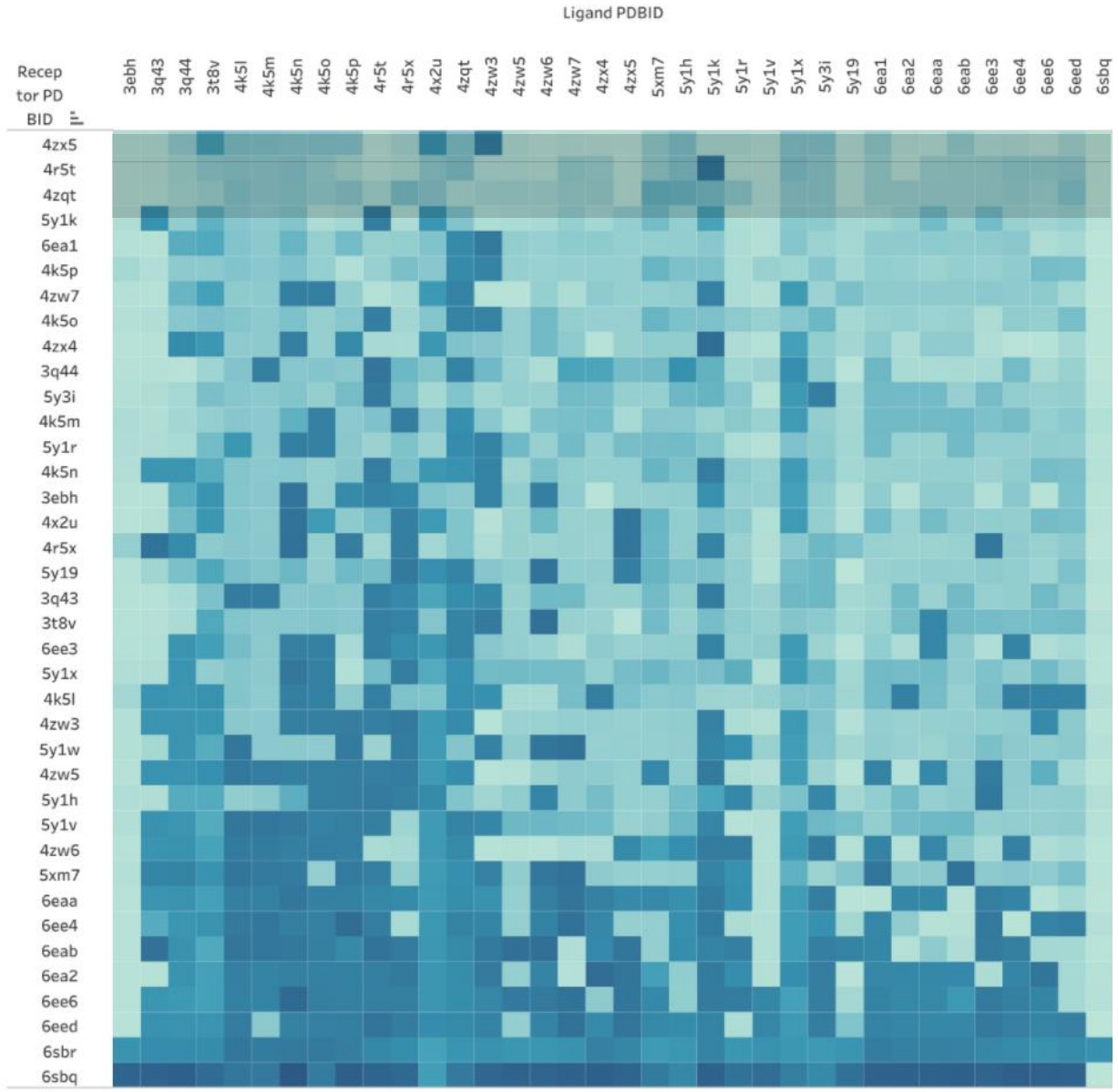
Cross-Docking:

Evaluate protein-scoring function pairings

Autodock Vina Cross-Docking

Best Docking Software: Best Structure

- Autodock Vina : 4R5T, 4ZX5, 4ZKT



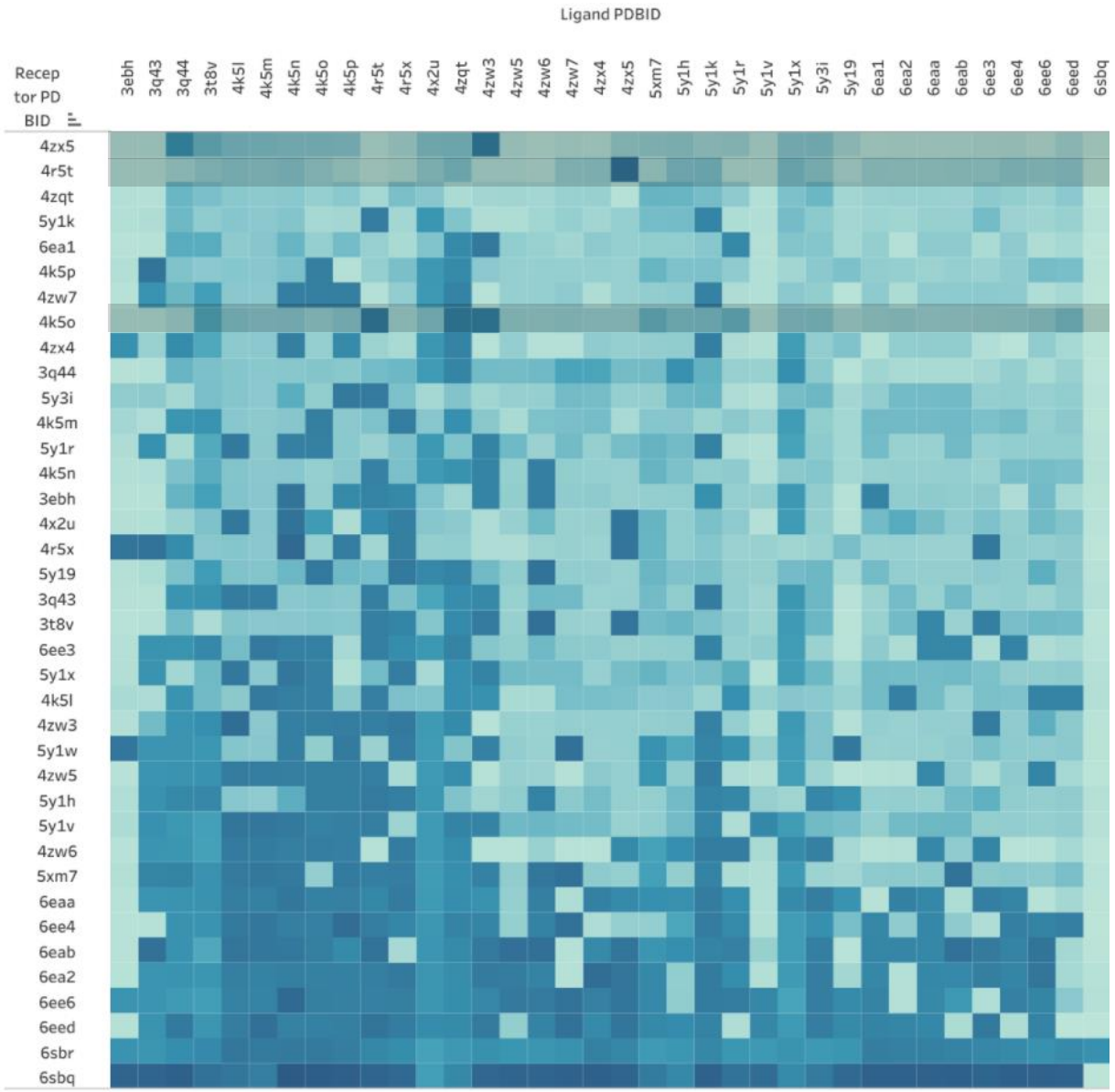
Cross-Docking:

Evaluate protein-scoring function pairings

SMINA Cross-Docking

Best Docking Software: Best Structure

- Autodock Vina : 4R5T, 4ZX5, 4ZKT
- Smina : 4ZX5, 4R5T, 4K5P



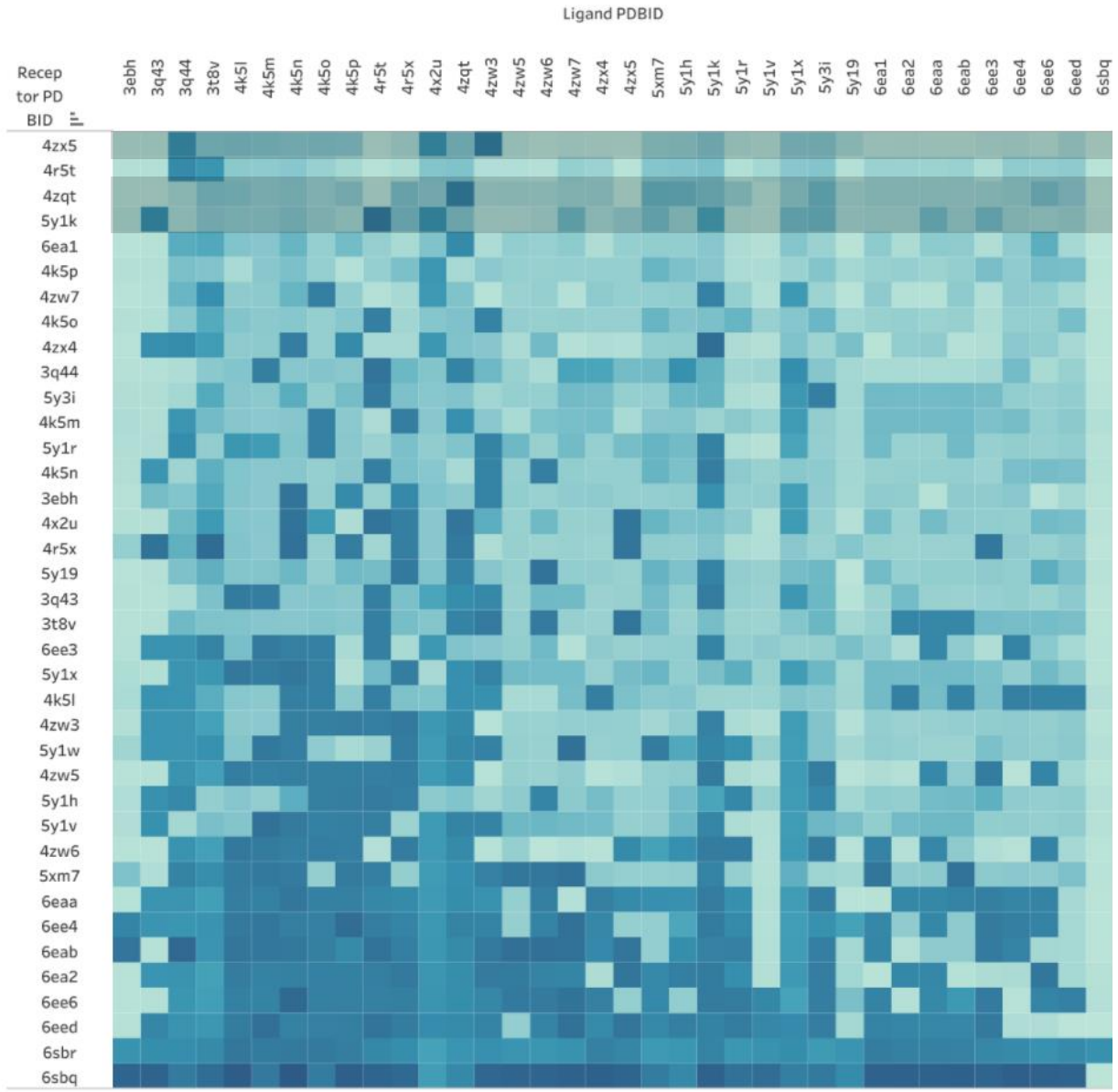
Cross-Docking:

Evaluate protein-scoring function pairings

QVina-W Cross-Docking

Best Docking Software: Best Structure

- Autodock Vina : 4R5T, 4ZX5, 4ZKT
- Smina : 4ZX5, 4R5T, 4K5P
- QVina-W : 4ZX5, 4ZQT, 5Y1K



Future directions

- Complete Mscreen development and testing
- Expand docking program support
  - DOCK6, AutoDock, AutoDock Zn, Ledock and Plants
- Manuscript preparation and release