

Lead Finder docking and scoring evaluation with a pharmaceutically relevant data set

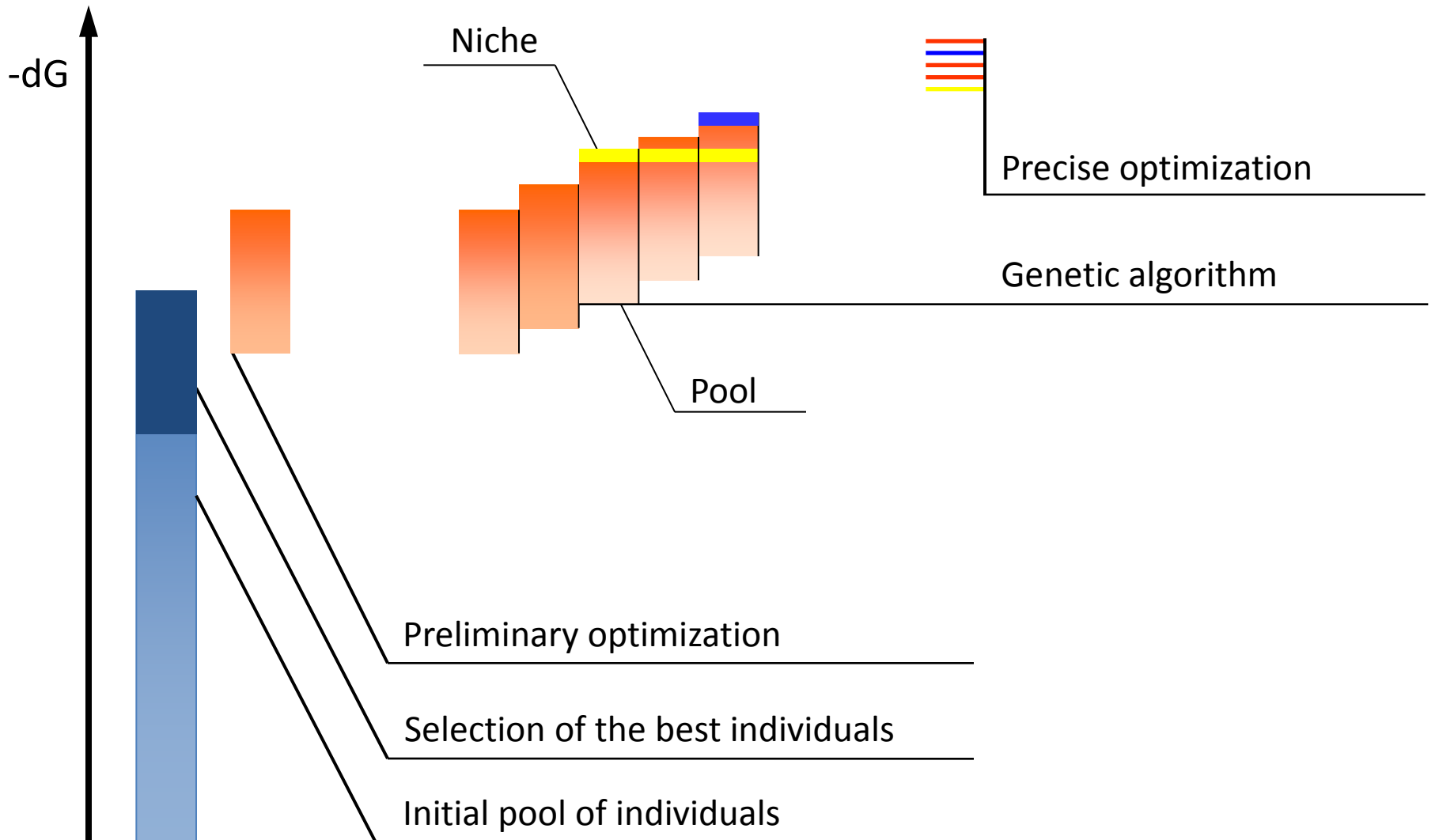
Oleg Stroganov¹, Fedor Novikov¹, Viktor Stroylov¹, Val Kulkov² and
Ghermes Chilov¹

¹ MolTech Ltd, Russian Federation, ² BioMolTech Corp, Canada

Section I

Lead Finder: Docking and Scoring Approach

Sampling: Workflow in Genetic Algorithm

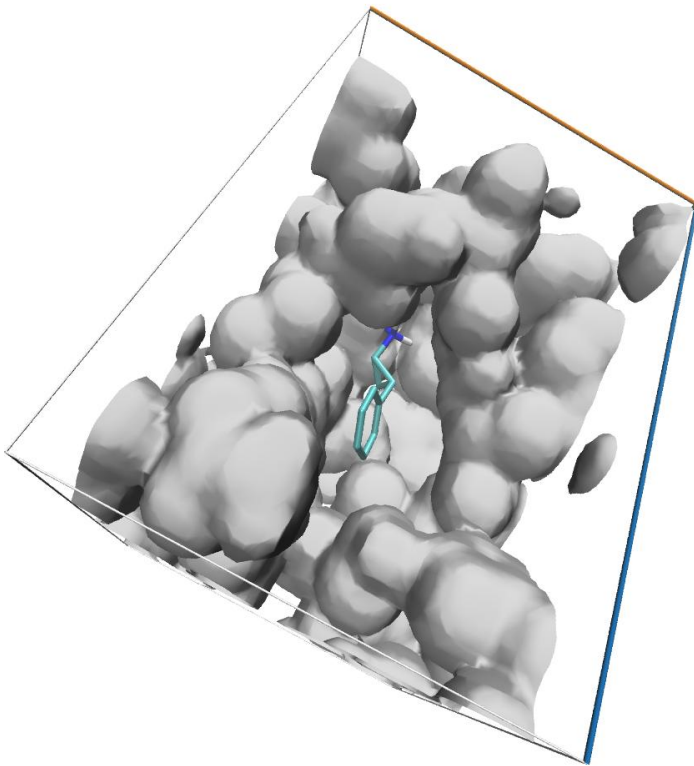


Sampling: Local Optimization

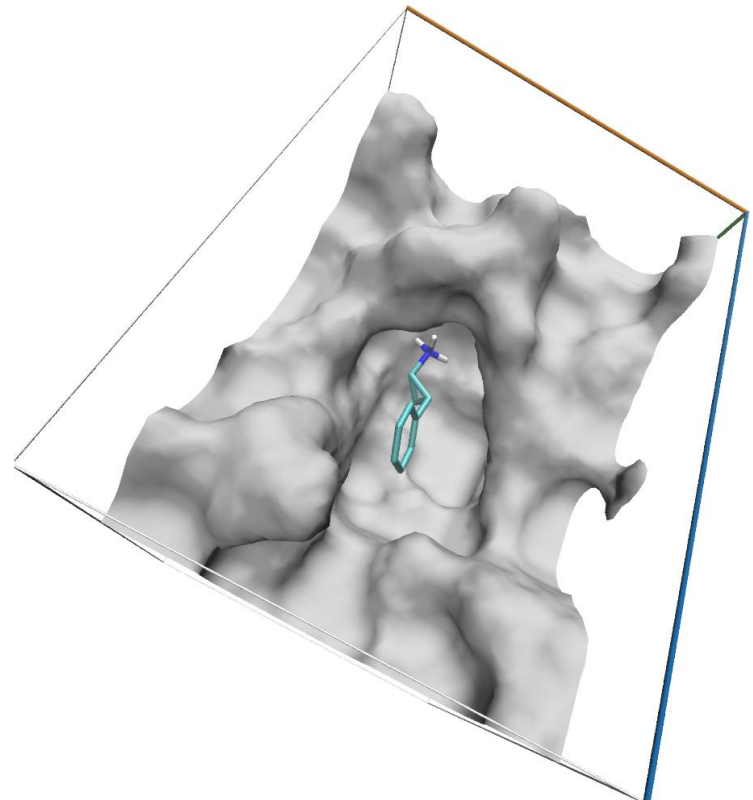
- Local optimization in GA:
 - Fast modification of Solis-Wets algorithm (90% of cases)
 - Full version of Solis-Wets algorithm (10% of cases)
- Optimization of the best niches:
 - Stochastic local minima exploration
- Post-docking optimization:
 - Free energy surface mapping, related to the internal rotation of terminal moieties
 - Exhaustive local minima exploration

Sampling: Local Optimization Tricks

Ragged VdW potential



Smoothed “immersion” potential



Significant enrichment of the evolving pool of individuals is achieved by the introduction of the “degree of immersion” parameter

Scoring: speed vs precision

dG calculations:

10-20 times per run

All available energy terms

PDO and niching:

10^3 - 10^4 times

Pairwise energy terms

GA – good poses:

10^5 - 10^6 times

All grid-based terms

GA – bad poses:

Over 10^7 times

LJ and electrostatics

Scoring: separating docking, dG and screening

$$\Delta G = k_{VdW} E_{VdW} + k_{Elec,i} E_{Elec,i} + k_{Hbond\varsigma i} E_{Hbond\varsigma i} + \dots$$

1 for Van der Waals energy,
4 for electrostatics ,
5 for hydrogen bonds,
1 for interaction with metals,
5 for Solvation,
4 for internal energy

20 coefficients

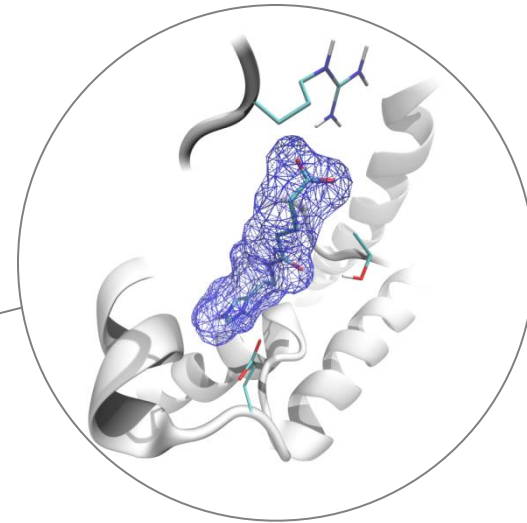
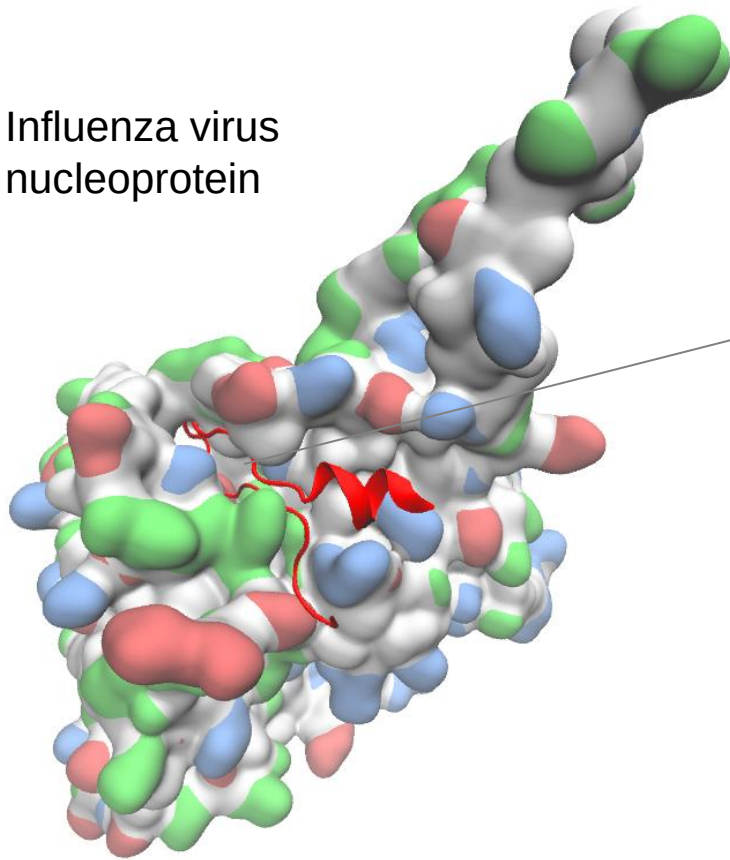
ΔG calculation

Docking

Virtual screening

Scoring: solvation

Influenza virus
nucleoprotein

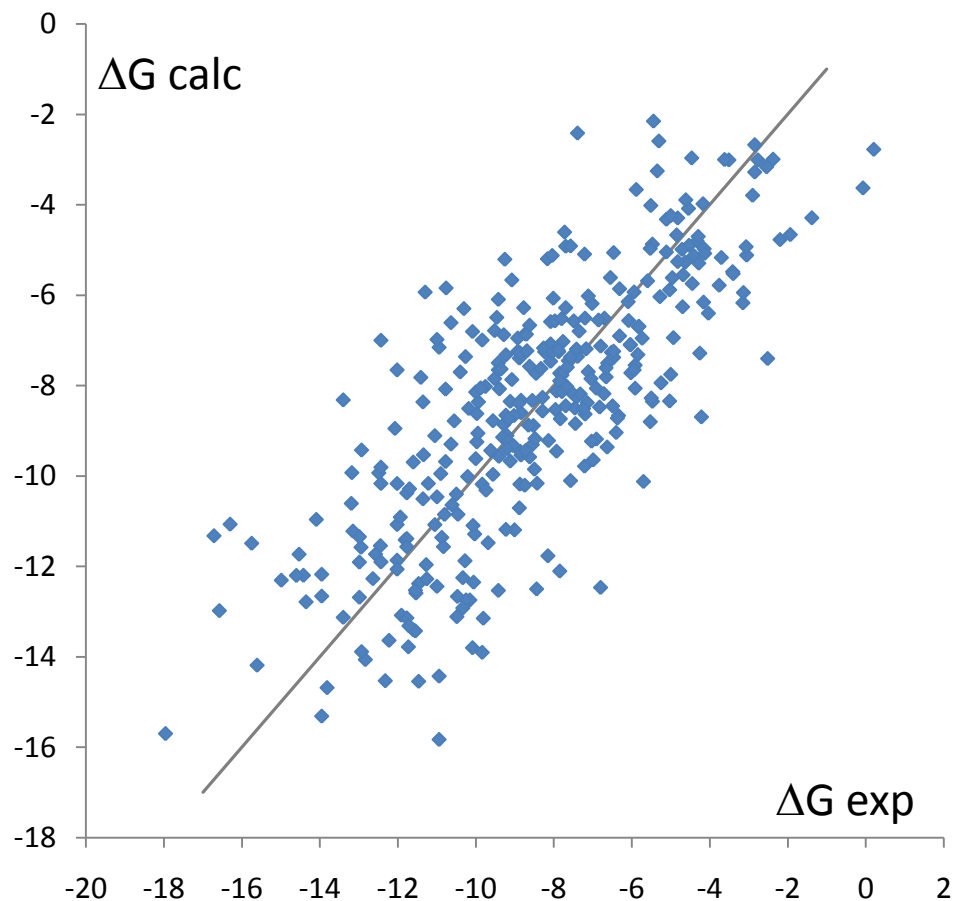


$$\Delta G_{\text{solvation}} = \sum_{\text{contact types}} k_i \cdot S_{\text{contact},i}$$

	Polar	Non-polar
Protein	-0.25	-0.40
Solvent	0.30	-0.01

Scoring: solvation

CSAR scoring challenge, ACS 240th meeting



Contribution of solvation into ΔG_{calc}

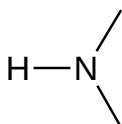
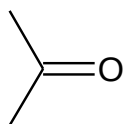
	R
Complete model	0,79
VdW + solvation	0.71

Scoring: H-bonds penalties and rewards

$$\Delta G_{HB} = \Delta G_{HB,complex} - \Delta G_{HB,solution}$$

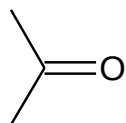
Ligand

Protein

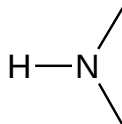
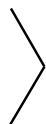


$$E_{H-bond} = E_0(r_{HA}) \cdot k_{DHA} \cdot k_{LP}$$

For the most cases $\Delta N_{hbonds} \approx 0$



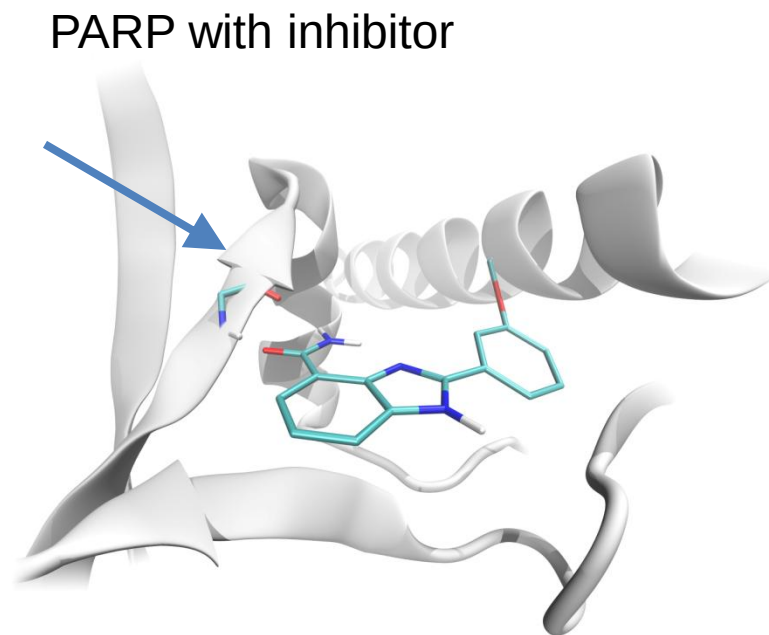
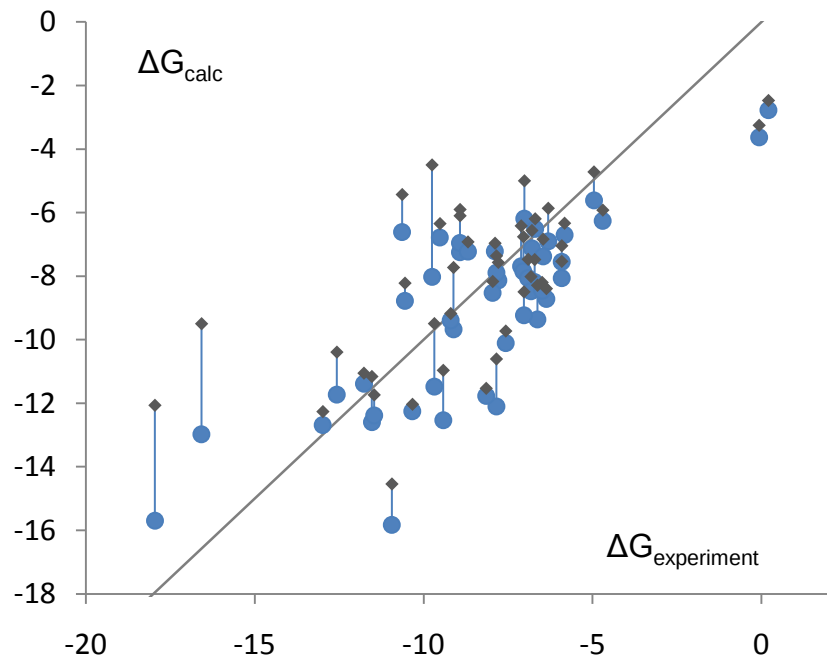
$$E_{penalty} = k \cdot N_{lost,ligand}$$



$$E_{penalty} = k \cdot N_{lost,protein}$$

H-bonds penalties serve to sieve out bad poses and poor binders

Scoring: H-bonds extra energy



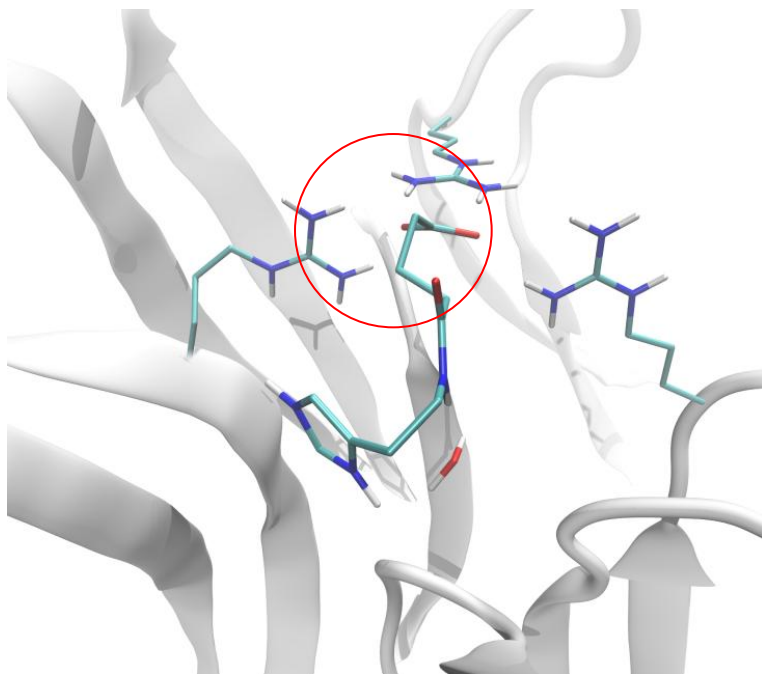
R^2 without extra H-bonds = 0.47

R^2 with extra H-bonds = 0.62

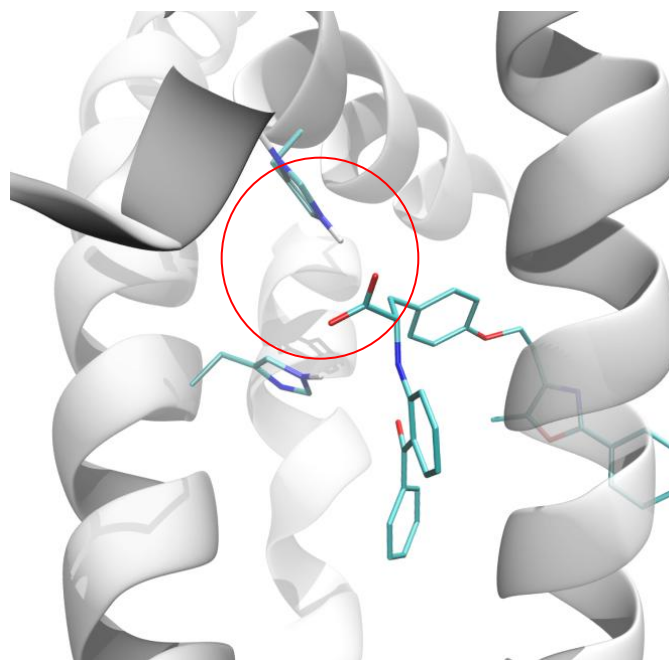
on CSAR subset of 48 structures, where systems of correlated H-bonds were found

Scoring: electrostatics

$$\Delta G_{elec} = \sum_{i \in \text{ligand}} \sum_{j \in \text{protein}} k_{elec,n}(h_i, b_i) E_{elec,n}(h_i, b_i, r_{ij}, q_i, q_j)$$



Neuraminidase



PPAR-g

Scoring: “Weak” Interactions

Statistical significance of weak interactions in the CSAR test set

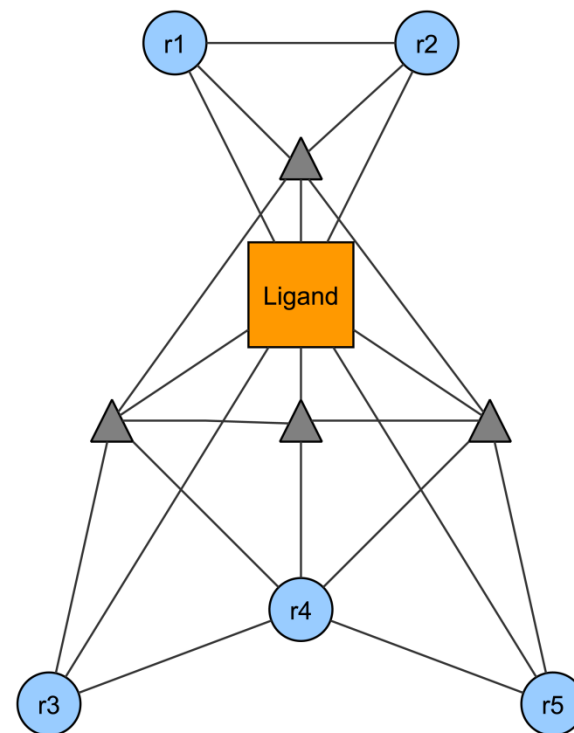
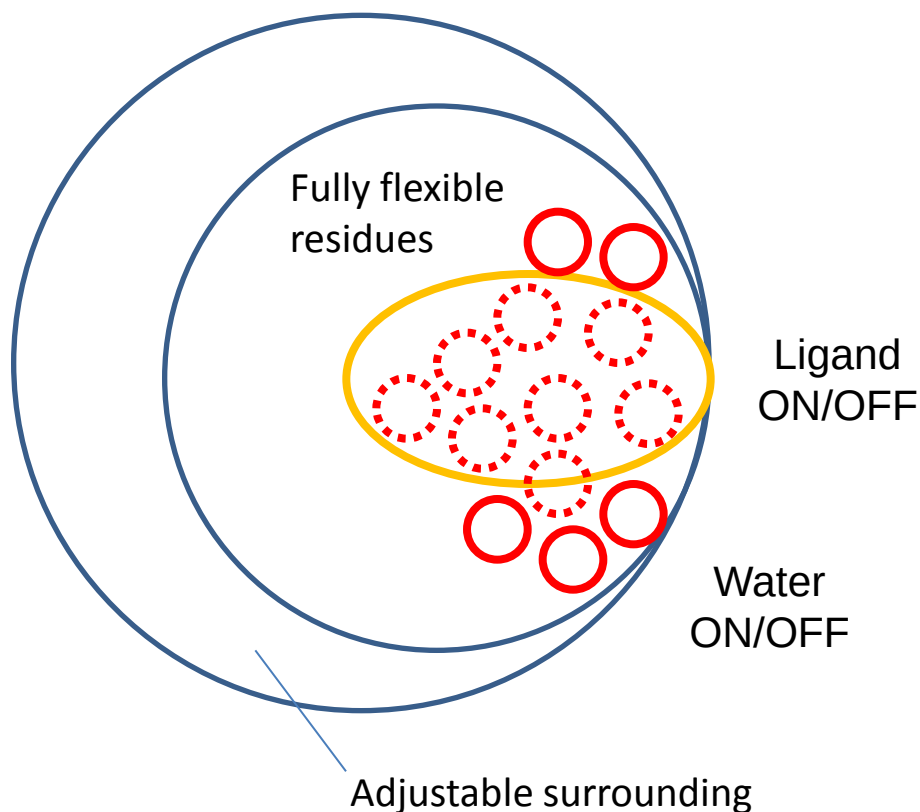
Interaction type	N	$\Delta\Delta G$, kCal/mol	RMSD. kCal/mol	p^a
X-H...Ar	3	0.1	0.98	0.62
X-H...Hal	6	1.50	2.30	0.098
C=O...Hal	7	1.70	1.99	0.0008
π -cationic interaction	18	-0.4	2.3	0.47
C-H...X	248	-0.1	1.94	0.66
Stacking	62	-0.5	1.88	0.20

a - Student's t-test probability of a null hypothesis.

TSAR: Novel Graph-Theoretical approach to account for protein flexibility and explicit water

- $\sim 10^4$ Ligand states, local/global sampling
- $\sim 10^1$ - 10^3 States per Residue, water, local/global sampling

Fixed surrounding



Section II

Binding Mode Prediction using Astex test set

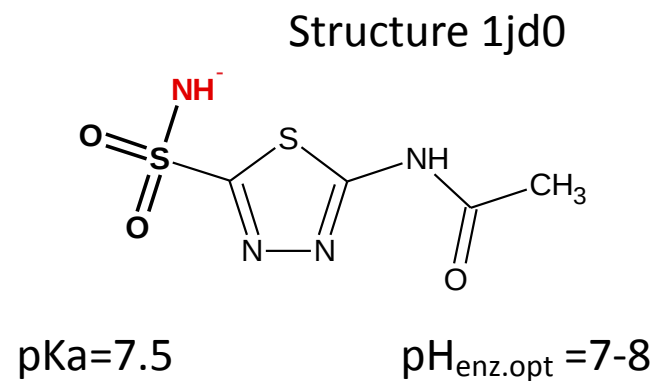
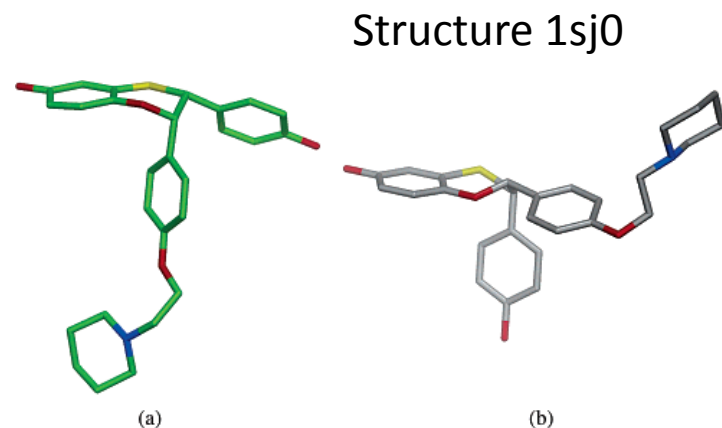
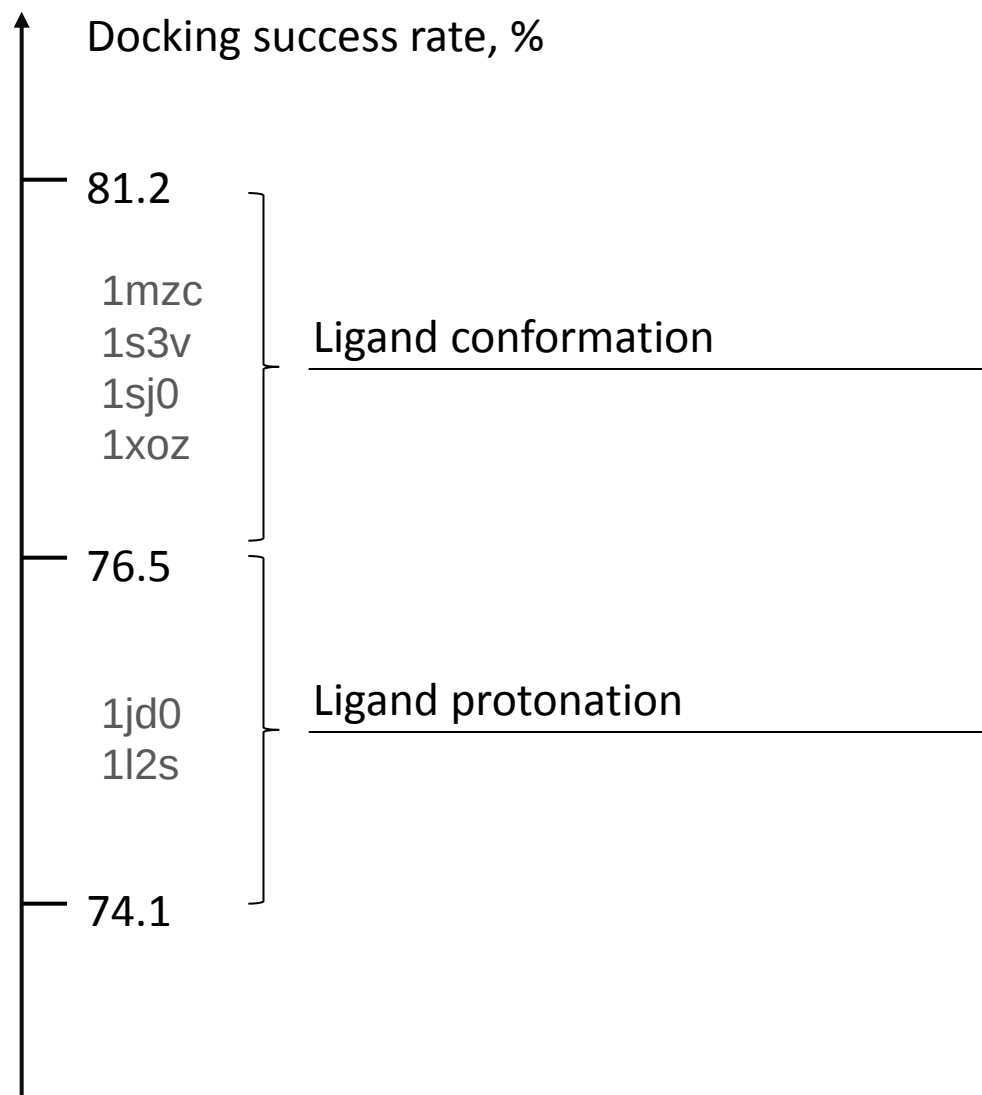
Docking protocol

- **Model preparation**
 - Astex intact structures of proteins and ligands
 - Automatic preparation of protein structures (Protein Designer) from PDB files and intact Astex ligands
 - manual correction of H-bond network (6 structures)
 - manual addition of single water molecules (4 structures)
 - Manual correction of ligand ionization states or local ring conformation (8 structures)
- **Docking**
 - Automatic docking in default regime, Lead Finder v 1.1.15
- **Processing**
 - According to the organizer's guidelines, results were averaged over 20 independent runs

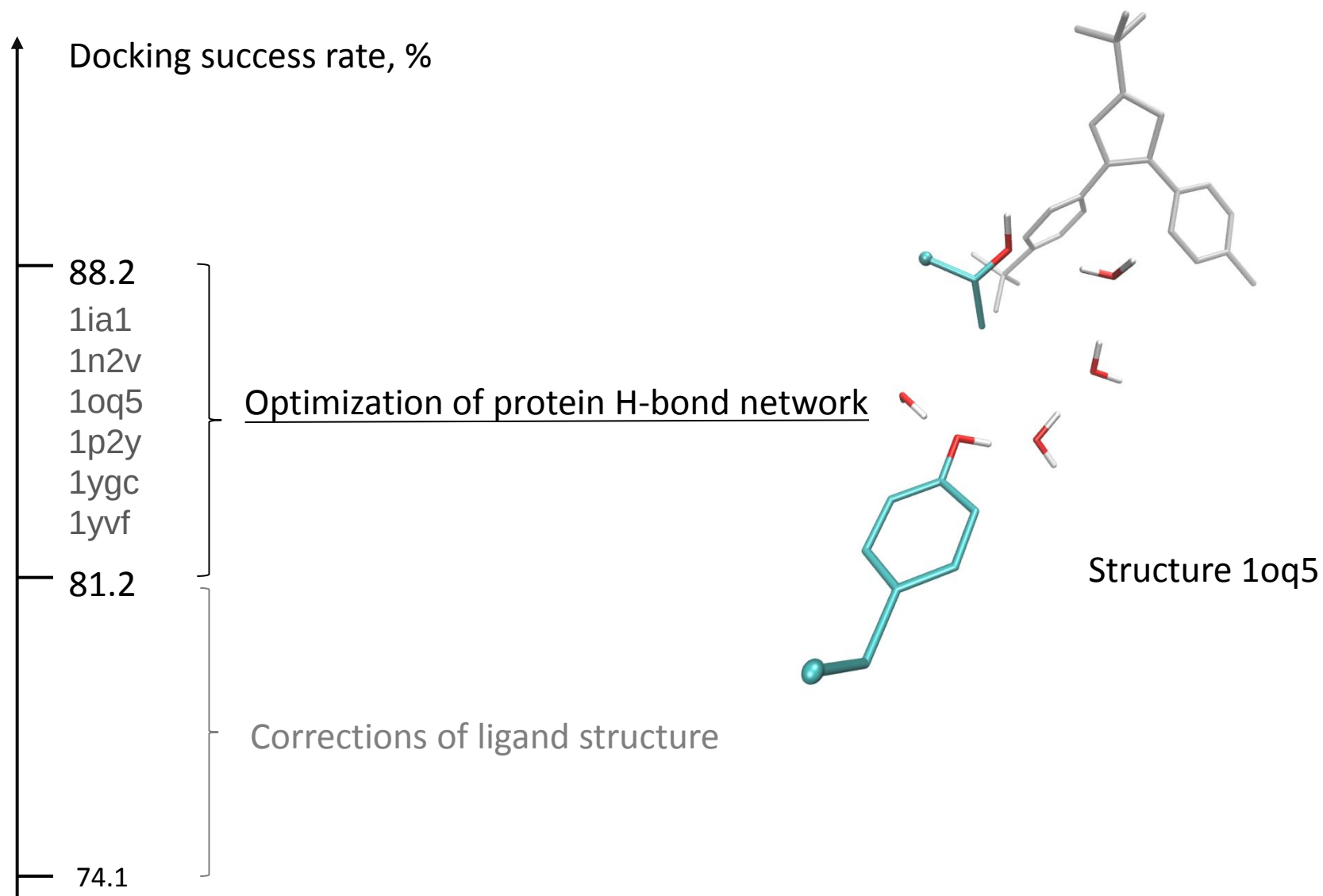
Docking success rate

RMSD	Success rate for the best pose (top 20 poses)		
	Intact set	Automatically prepared	Manually corrected
2.0 Å	72.9 (90.6)	74.1 (91.8)	90.6 (94.1)
2.5 Å	81.2 (91.8)	80.6 (91.8)	92.9 (94.1)
3.0 Å	84.8 (91.8)	82.5 (91.8)	92.9 (94.1)

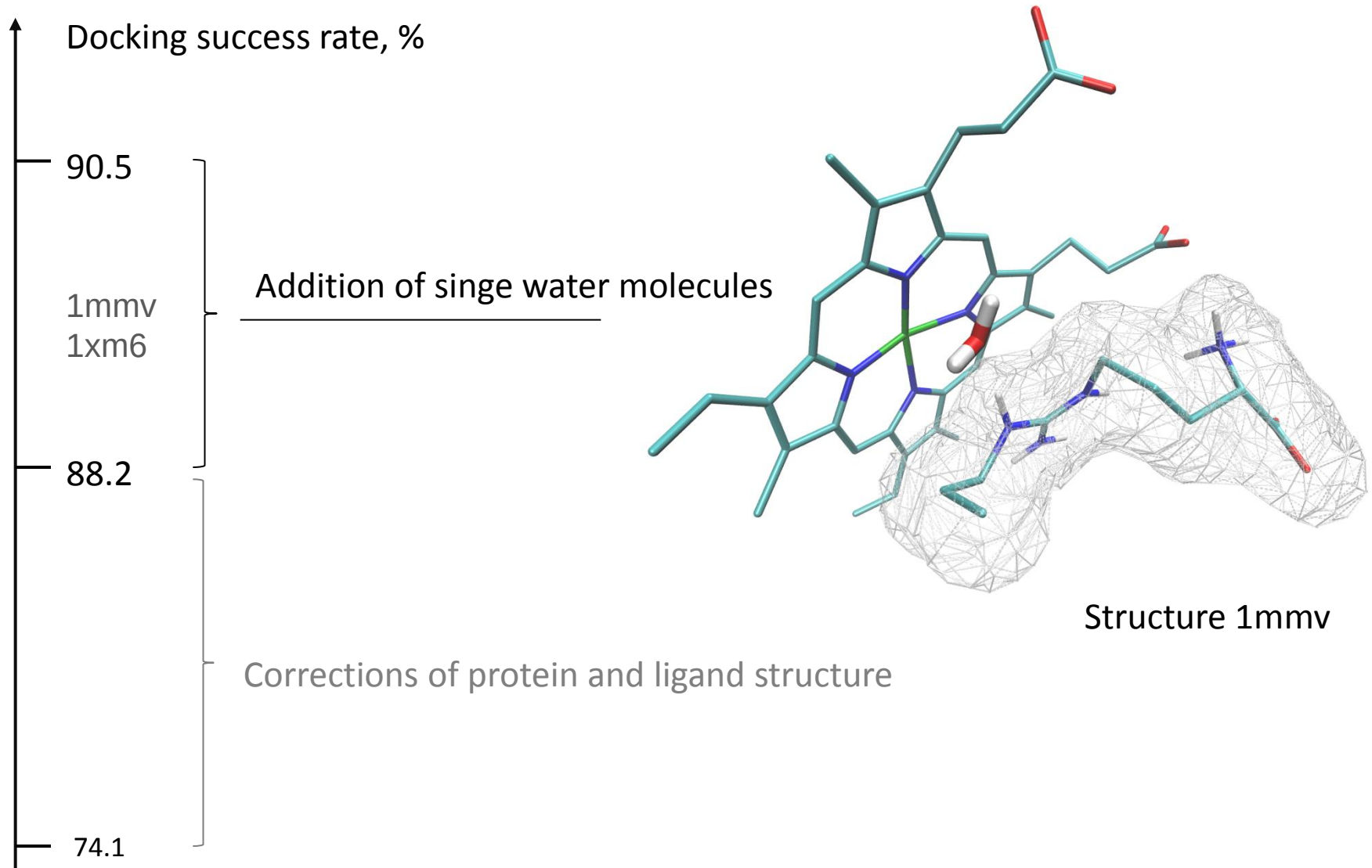
“Fixable” cases: ligand structure



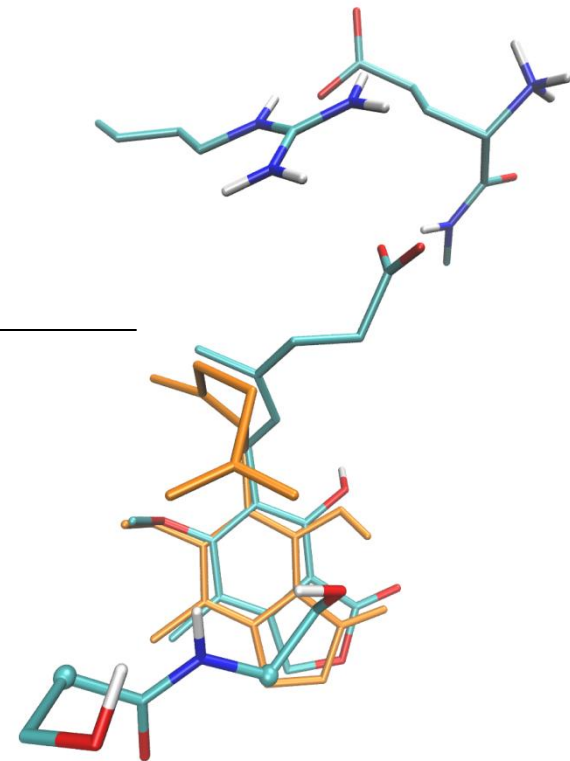
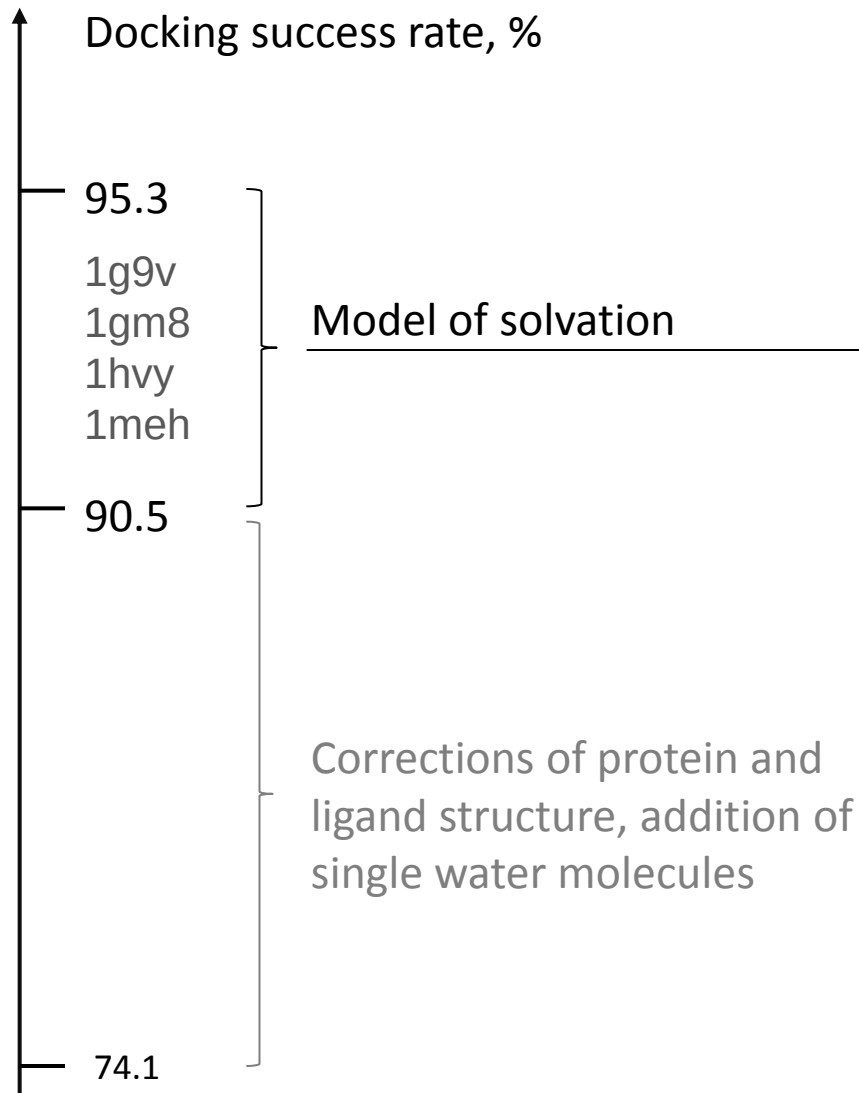
“Fixable” cases: protein structure



“Fixable” cases: water molecules

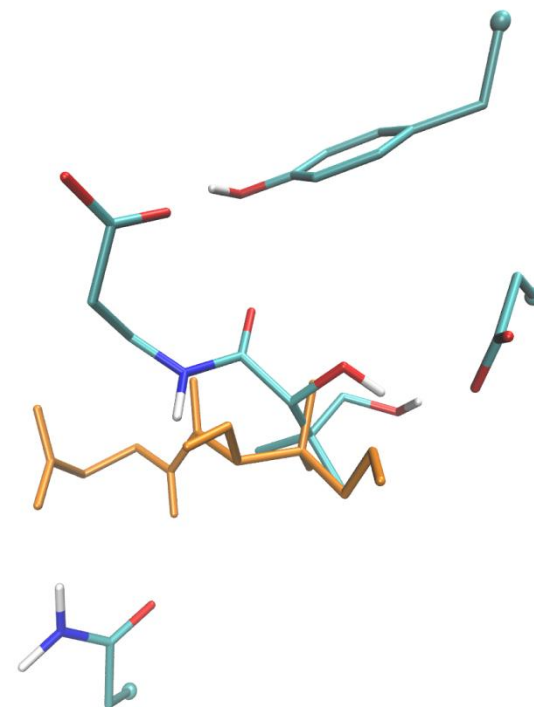
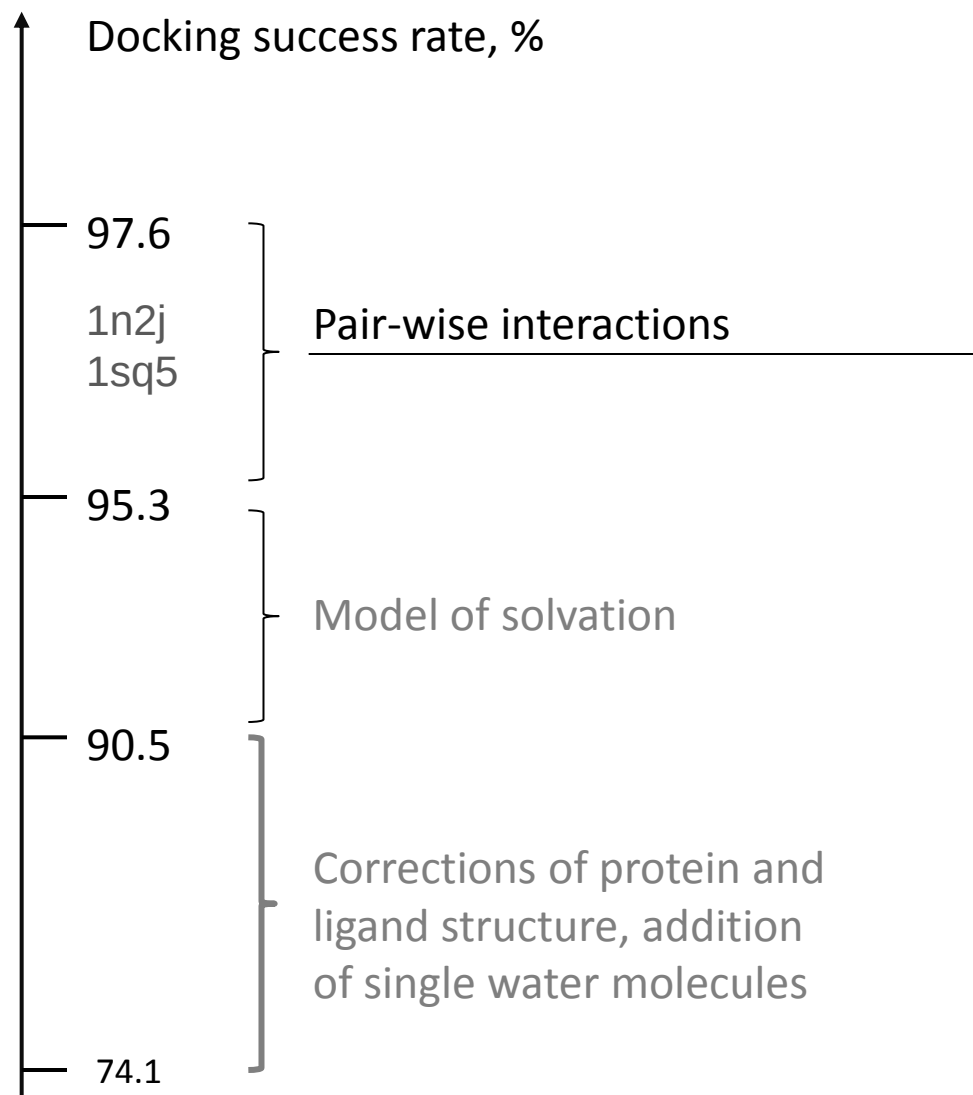


Fundamental errors



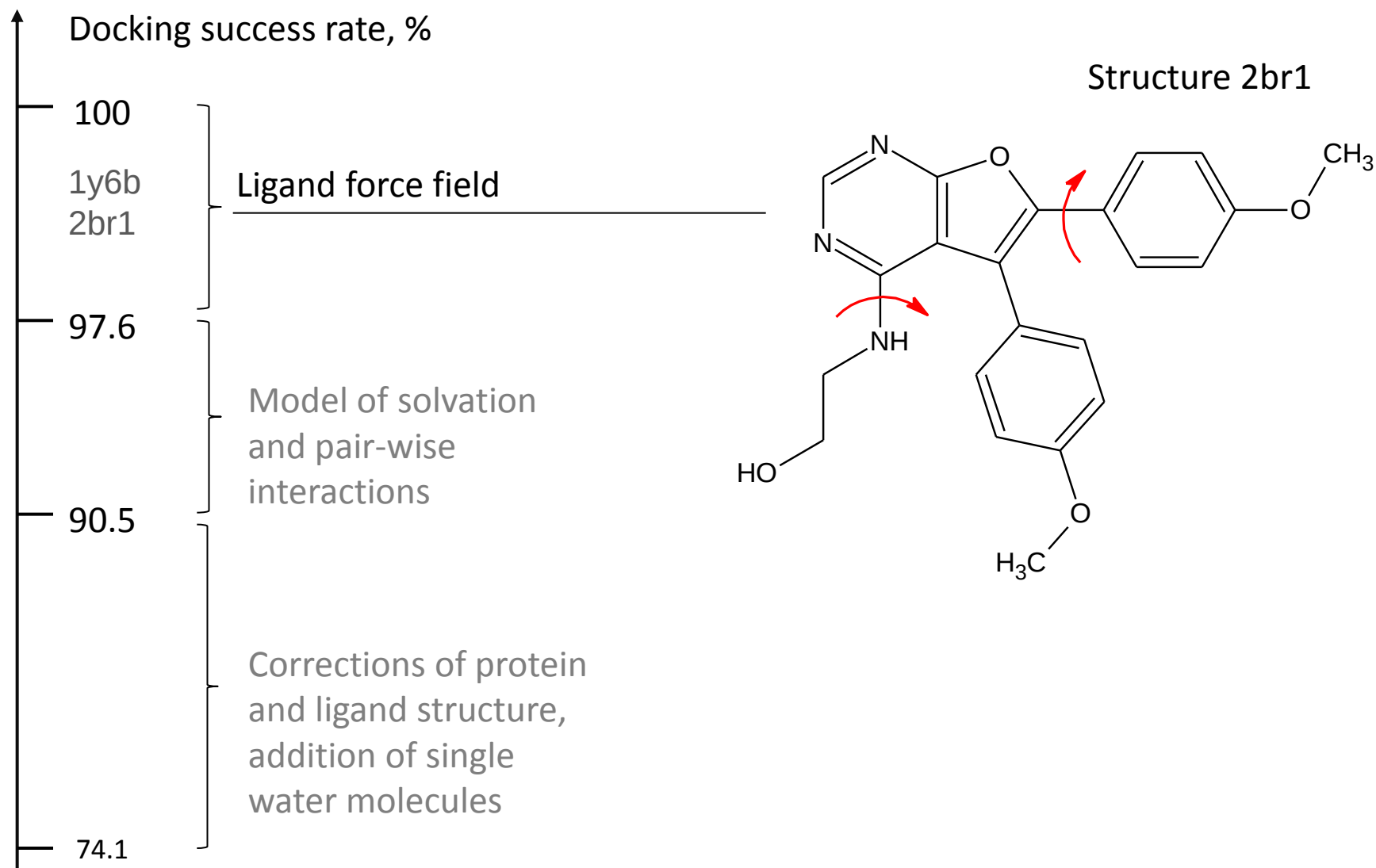
Structure 1meh

Fundamental errors



Пример: структура 1sq5

Fundamental errors

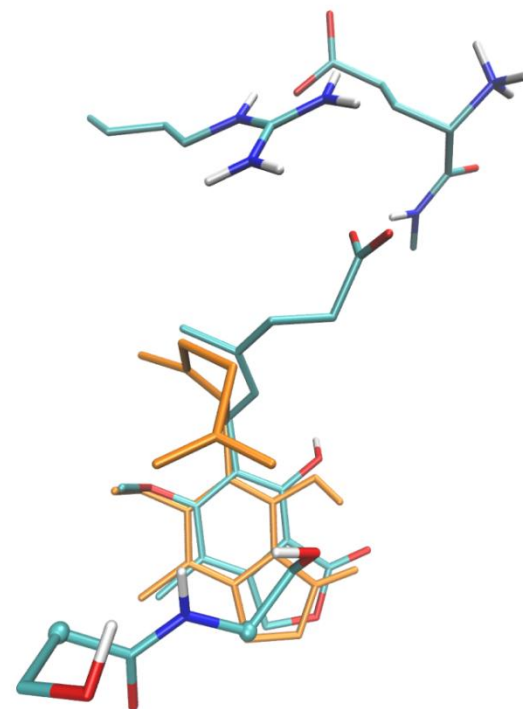
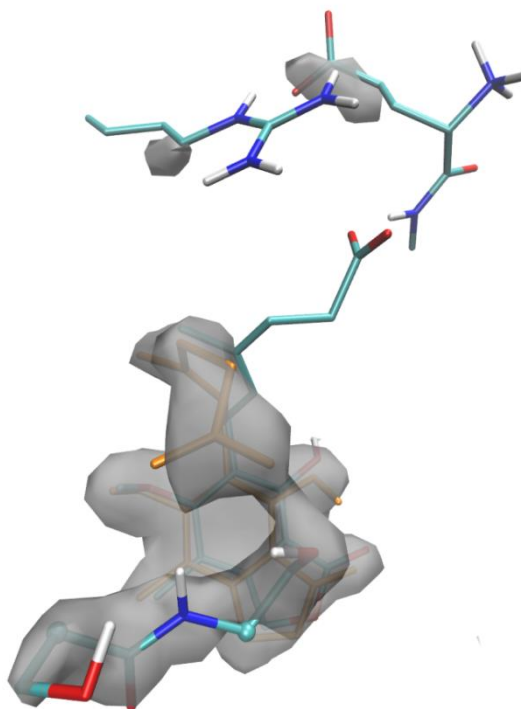


Structures with significant structural problems

- **Alternate side chain conformations:**
 - Successfully docked 'as is':
 - 1ia1 1kzk 1n46 1opk 1q4g 1r1h 1t46 1vcj 1z95
 - Docked after ligand conformation was fixed
 - 1s3v
 - Was not docked:
 - 1y6b (better ligand force field required)
- **Crystal packing interactions:**
 - Successfully docked 'as is':
 - 1gkc 1jje 1r55 1t40 1v48 2bsm
 - Was not docked (even after symmetric protein unit was included):
 - 1hvy (better solvation model required)

Structures with significant structural problems

- **Poor density:**
 - Successfully docked 'as is':
 - 1j3j 1u1c 1v4s
 - Was not docked:
 - 1meh



1meh: poorly resolved arginine guides wrong ligand pose

Structures with significant structural problems

- **Alternate ligand conformations**
 - Successfully docked 'as is':
 - 1tz8
 - 1ig3 (sulphates were automatically deleted and protein H-bond network better optimized)
 - 1sg0 (other reference molecule was chosen automatically)

Section III

Virtual screening benchmark using DUD

Virtual screening protocol

- **Model preparation**
 - Intact DUD protein models
 - Automatically prepared from raw PDB structures
- **Virtual screening**
 - Docking and scoring with Lead Finder v1.1.15
 - Rank-ordering according to Lead Finder VS-score
 - Re-ordering according to the procedure of structural filtration (correctness of ligand docking was also judged by the use of structural filtration)

Summary of Results

	Intact DUD set					<i>de novo</i> prepared DUD set				
	ES0.1%*, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
Average	3.4 (1.6)	10.6 (5.3)	15.2 (9.1)	0.735 (0.701)	0.148 (0.115)	4.3 (1.6)	12.2 (6.4)	16.9 (10.4)	0.728 (0.683)	0.133 (0.091)
SD	4.2	10.6	13.4	0.147	0.170	6.7	12.7	15.6	0.152	0.192
Median	1.7	8.2	11.2	0.760	0.106	2.1	8.8	13.3	0.717	0.123
Min	0.0	0.4	0.7	0.393	-0.211	0.1	0.5	1.1	0.420	-0.245
Max	16.0	46.2	56.5	0.961	0.467	34.0	56.3	64.9	0.954	0.525

* For ideal selection: ESX% = 100%

For random selection ES0.1% = 0,05% ES1% = 0,5% ES2% = 1,0%

** In brackets performance of virtual screening without structural filtration is provided

Δ ROC AUC = ROC AUC – ROC AUC (null model)

Nuclear Hormone Receptors

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
ER_agonist	50.5	16.0	36.1	43.9	0.961	0.052
RXR	93.7	1.8	7.5	9.8	0.924	0.467
MR	25.0	1.4	12.8	19.3	0.872	-0.015
ER_antagonist	54.2	8.9	16.2	27.3	0.870	0.370
PR	41.7	4.0	6.8	11.2	0.799	0.007
PPAR	55.3	0.3	5.0	9.9	0.791	0.267
AR	8.0	1.9	7.7	10.9	0.756	0.261
GR	21.1	1.6	8.3	10.8	0.716	-0.028
Average	43.7	4.5	12.6	17.9	0.836	0.173
SD	26.5	5.3	10.2	12.2	0.085	0.193
Median	46.1	1.8	8.0	11.0	0.835	0.156
Min	8.0	0.3	5.0	9.8	0.716	-0.028
Max	93.7	16.0	36.1	43.9	0.961	0.467

Kinases

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
SRC	73.1	3.5	10.7	18.6	0.837	0.310
EGFR	52.1	0.8	9.1	16.8	0.789	0.241
TK	29.0	0.9	4.8	6.4	0.762	0.044
CDK2	35.3	7.0	17.1	22.4	0.743	0.191
VEGFR2	16.9	0.1	1.0	1.7	0.669	0.249
HSP90*	28.0	0.0	0.4	0.8	0.595	0.017
P38	51.6	0.6	5.3	6.8	0.589	0.313
FGFR1	1.7	1.7	2.1	2.4	0.497	-0.211
PDGFRb	13.7	6.7	10.2	11.2	0.460	-0.035
Average	33.5	2.4	6.7	9.7	0.660	0.124
SD	22.3	2.7	5.5	8.0	0.133	0.180
Median	29.0	0.9	5.3	6.8	0.669	0.191
Min	1.7	0.0	0.4	0.8	0.460	-0.211
Max	73.1	7.0	17.1	22.4	0.837	0.313
*HSP90	45.3	0.1	1.1	3.0	0.704	0.127

Serine proteases

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
Trypsin	66.7	8.7	16.2	29.4	0.940	0.336
Thrombin	55.6	1.2	11.5	23.3	0.875	0.042
FXa	65.7	1.4	10.8	17.4	0.846	-0.051
Average	62.7	3.8	12.8	23.4	0.887	0.109
SD	6.2	4.3	2.9	6.0	0.048	0.202
Median	65.7	1.4	11.5	23.3	0.875	0.042
Min	55.6	1.2	10.8	17.4	0.846	-0.051
Max	66.7	8.7	16.2	29.4	0.940	0.336

Metalloenzymes

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
COMT	58.3	4.3	18.0	22.7	0.873	0.409
PDE5*	19.2	2.3	9.8	12.1	0.652	0.009
ADA	29.2	0.1	1.1	2.2	0.622	0.149
ACE	39.3	0.5	2.2	4.3	0.533	0.141
Average	36.5	1.8	7.8	10.3	0.670	0.177
SD	16.7	1.9	7.8	9.3	0.145	0.168
Median	34.3	1.4	6.0	8.2	0.637	0.145
Min	19.2	0.1	1.1	2.2	0.533	0.009
Max	58.3	4.3	18.0	22.7	0.873	0.409
*PDE5	28.8	8.1	13.0	14.6	0.719	0.083

Folate enzymes

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
DHFR	55.8	13.1	43.9	52.0	0.877	0.298
GART	53.0	7.4	19.9	30.2	0.859	0.005
Average	54.4	10.3	31.9	41.1	0.868	0.151
SD	1.9	4.0	17.0	15.4	0.013	0.207

Other enzymes

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
SAHH	43.1	1.4	8.1	19.7	0.909	0.335
COX2	69.7	2.0	13.5	22.4	0.889	0.347
NA	66.0	14.9	46.2	56.5	0.834	0.451
HMGA	54.6	3.8	12.5	21.8	0.776	-0.055
PARP*	24.5	2.9	7.4	9.8	0.762	0.277
AChE*	0.0	0.1	0.8	1.1	0.490	0.031
COX1	25.6	0.1	1.2	2.0	0.485	-0.025
InhA	3.1	0.1	0.5	0.7	0.393	0.019
Average	35.8	3.2	11.3	16.7	0.692	0.172
SD	26.9	4.9	15.0	18.6	0.204	0.200
Median	34.4	1.7	7.7	14.8	0.769	0.154
Min	0.0	0.1	0.5	0.7	0.393	-0.055
Max	69.7	14.9	46.2	56.5	0.909	0.451
AChE*	65.4	0.7	3.1	5.7	0.688	0.134
PARP*	37.3	3.4	13.5	18.1	0.879	0.354

Other enzymes

Target	Ligands docked, %	ES0.1%, %	ES1%, %	ES2%, %	ROC AUC	Δ ROC AUC
HIVRT	32.5	0.3	3.7	8.5	0.758	0.334
GPB	45.9	0.5	2.0	3.1	0.732	0.239
ALR2	12.3	3.5	6.4	7.1	0.718	0.024
AmpC*	30.3	0.1	1.0	2.0	0.656	0.023
PNP	23.1	1.5	13.9	19.2	0.654	0.072
HIVPR	14.8	10.3	11.8	12.1	0.622	0.012
Average	26.5	2.7	6.5	8.7	0.690	0.117
SD	12.5	3.9	5.3	6.3	0.054	0.136
Median	26.7	1.0	5.0	7.8	0.687	0.048
Min	12.3	0.1	1.0	2.0	0.622	0.012
Max	45.9	10.3	13.9	19.2	0.758	0.334
AmpC*	60.6	0.3	2.8	4.9	0.817	0.209