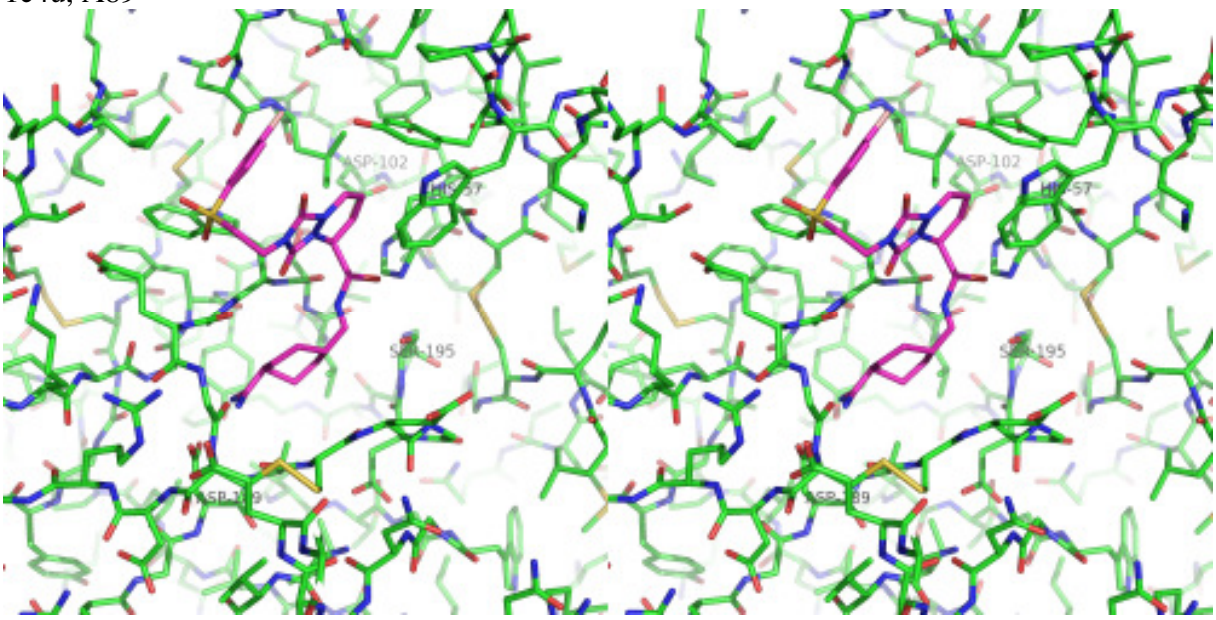
1 a-c) image of active site of the three taken crystal structures with the ligand, in standard orientation

colours, orientation, labels, stereo?
labels: S1, S2, S3

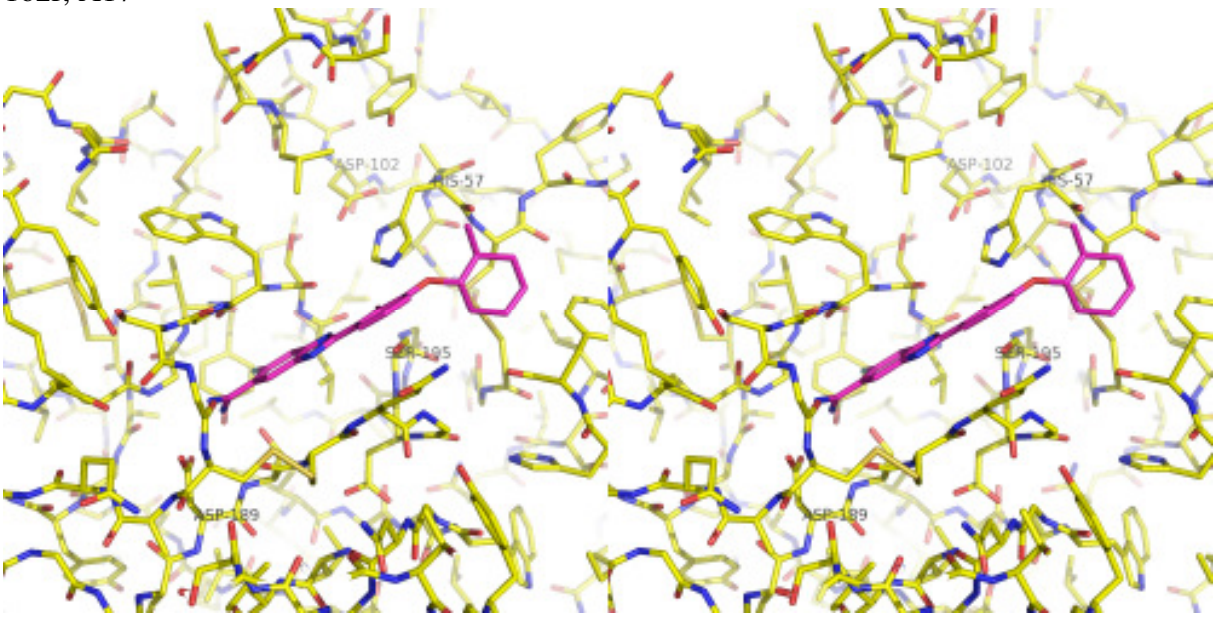
thrombin

thrombin model 3
1c4u, A89



trypsin

trypsin-model-del-ions
1o2r, A17



urokinase

model7-1owd
1owd, A32

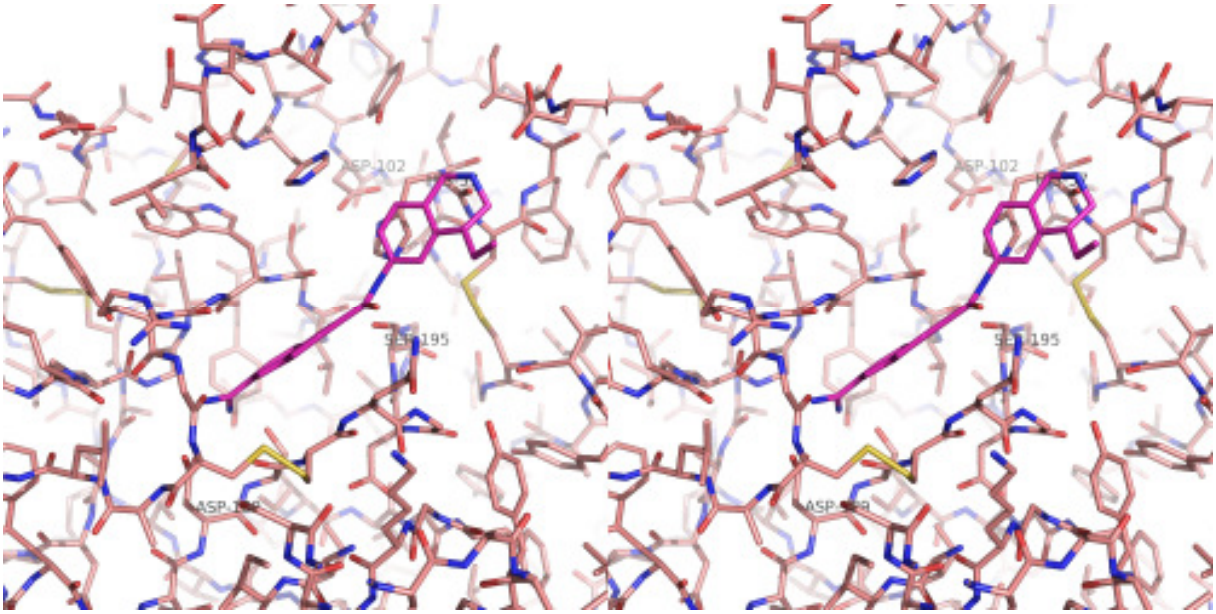
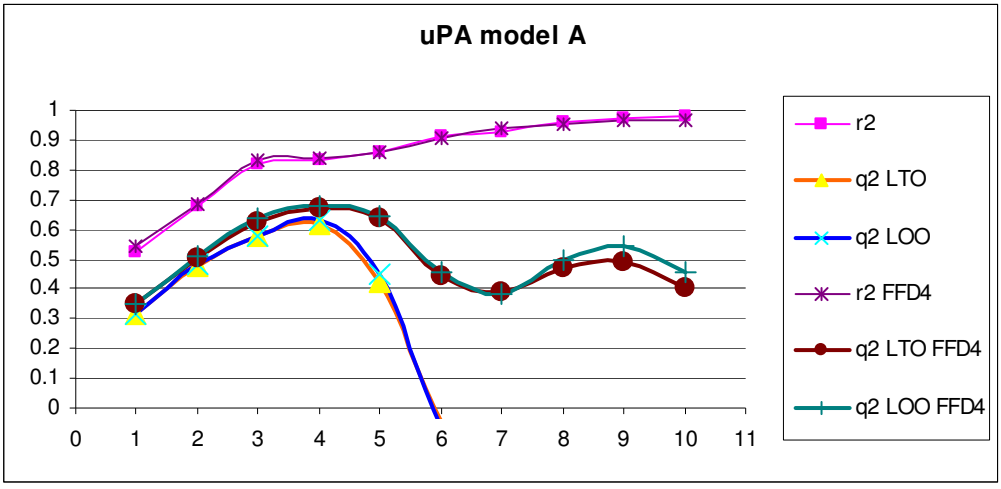


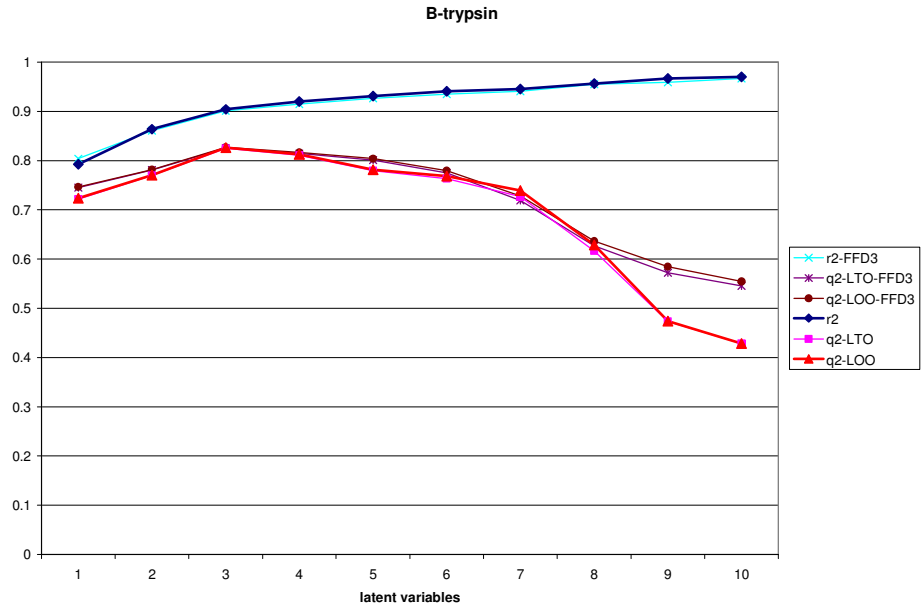
Figure 1:
View into the active site of the used receptor models of thrombin, trypsin and urokinase. The residues Ser195, His57 and Asp102 of the catalytic triads as well as Asp189 at the bottom of the S1 are labelled. The ligands (A89, A17, A32), coloured in pink, are representing orientation taken from the original crystal structure conformation where the models based on (thrombin: 1c4u; trypsin: 1o2r; urokinase: 1owd).

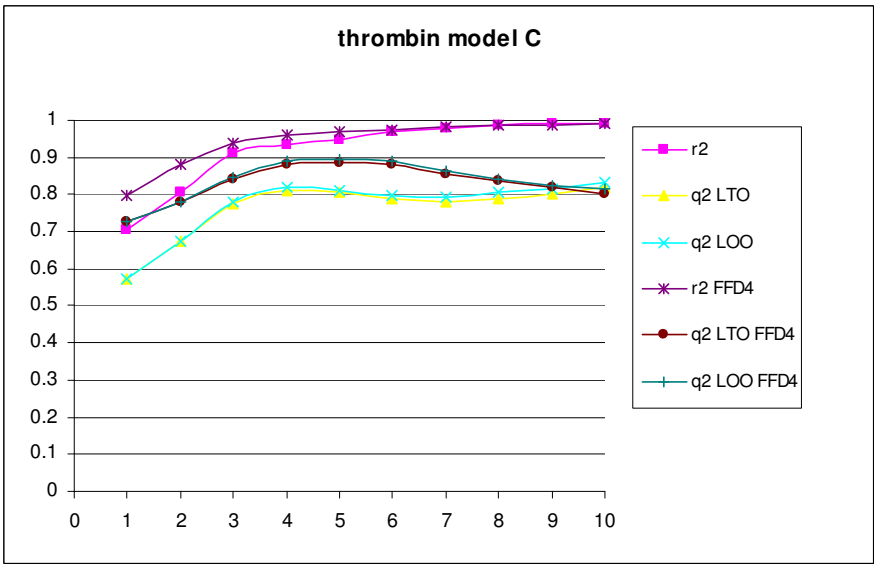
3. r2, q2

Figure 3:
The coefficients of determination R2 and the predictive correlation coefficient Q2 for the three COMBINE models were plotted in dependency of the number of latent variables (LV). The best Q2-LOO values for thrombin, trypsin and urokinase were 0.89 (LV5), 0.83 (LV3), and 0.68 (LV4). In trypsin, variable selection did not improve the model, but in thrombin and urokinase the models could be improved by using D-optimal pre-selection (D-opt) and fractional factorial design (FFD) variable selection at LV4. For thrombin LV4 and LV5 resulted in nearly the same values so the lower one was chosen (see table).

model	variable selection	LV	R2	SDEC	Q2 LTO	SDEP LTO	Q2 LOO	SDEP LOO
thrombin	-	4	0.93	0.59	0.81	1.00	0.82	0.98
	D-opt LV4 FFD LV4	4 5	0.96 0.97	0.46 0.40	0.88 0.89	0.79 0.77	0.89 0.89	0.77 0.76
trypsin	-	3	0.90	0.72	0.82	0.97	0.83	0.97
	D-opt LV3 FFD LV3	3	0.90	0.73	0.83	0.97	0.83	0.97
urokinase	-	4	0.83	0.62	0.62	0.93	0.63	0.91
	D-opt LV4 FFD LV4	4	0.84	0.60	0.67	0.87	0.68	0.86




B-trypsin
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
latent variables
r2-FFD3
q2-LTO-FFD3
q2-LOO-FFD3
r2
q2-LTO
q2-LOO


thrombin model C
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
8
9
10
r2
q2 LTO
q2 LOO
r2 FFD4
q2 LTO FFD4
q2 LOO FFD4

4. a-c) REC curves
training xray, training docked, testset active docked

Pseudo training sets: docking of ligands, which were already used in training sets for model building.

Figure 4:
In the Regression Error Characteristic Curves (Bi and Bennett, 2003) the error tolerance, defined as the absolute difference between the experimental and predicted ΔG, is plotted versus the points predicted within the tolerance. The 'training-xray' curves base on the internal prediction of the crystal structure ligand conformations used as a training set. The error tolerance of the null model is calculated of the absolute difference between experimental ΔG and the average of it. The rest of the curves are retrieved from the best ranked predicted ΔG values. The ranking was performed according to highest GoldScoreFitness (best GoldS), smallest error (best abs), lowest ΔG bind elec of UHBD (best dG bind elec), lowest RMSD (best RMSD) of the ten docking solutions. The curves 'x. ranked pred dG' were calculated after ranking the ten predicted ΔG values for each ligand (lowest first) and selecting the x. ranked value. The median is taken from all ten values.

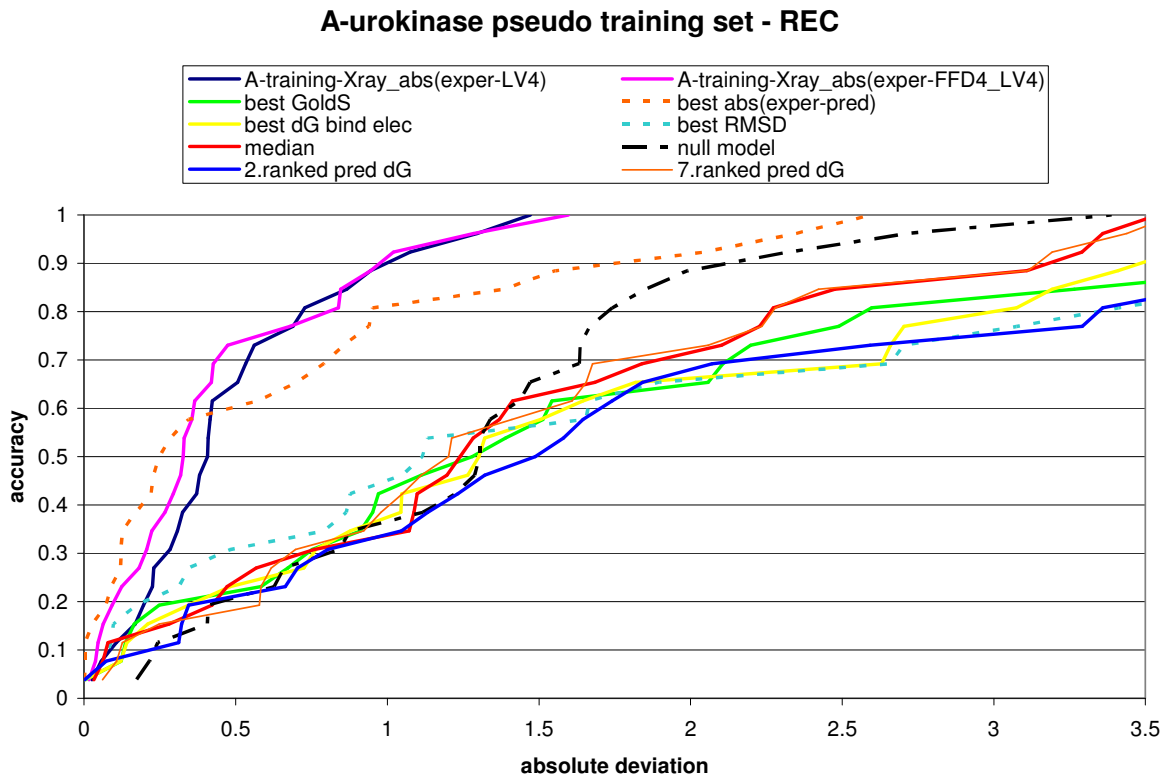
absolute($\Delta G_{experimental}$-$\Delta G_{predicted}$)
pseudo test set

after variable selection (ffd4)

urokinase	**r2**	**r2spearman**	**SDEP**
xray before var sel	0.83	0.90	0.61
xray	0.84	0.90	0.60
GoldS	0.14	0.33	2.12
best abs	0.65	0.81	0.98
best dG bind elec	0.18	0.35	2.13
best RMSD	0.20	0.41	2.23
1.	0.25	0.47	2.51
2.	0.25	0.53	2.27
7.	0.15	0.28	1.79
10.	0.15	0.29	1.80
mean	0.21	0.40	1.84
median	0.19	0.35	1.80
null model			1.49

mean of experimental=-8.91

standard deviation of the experimental ΔG =1.54

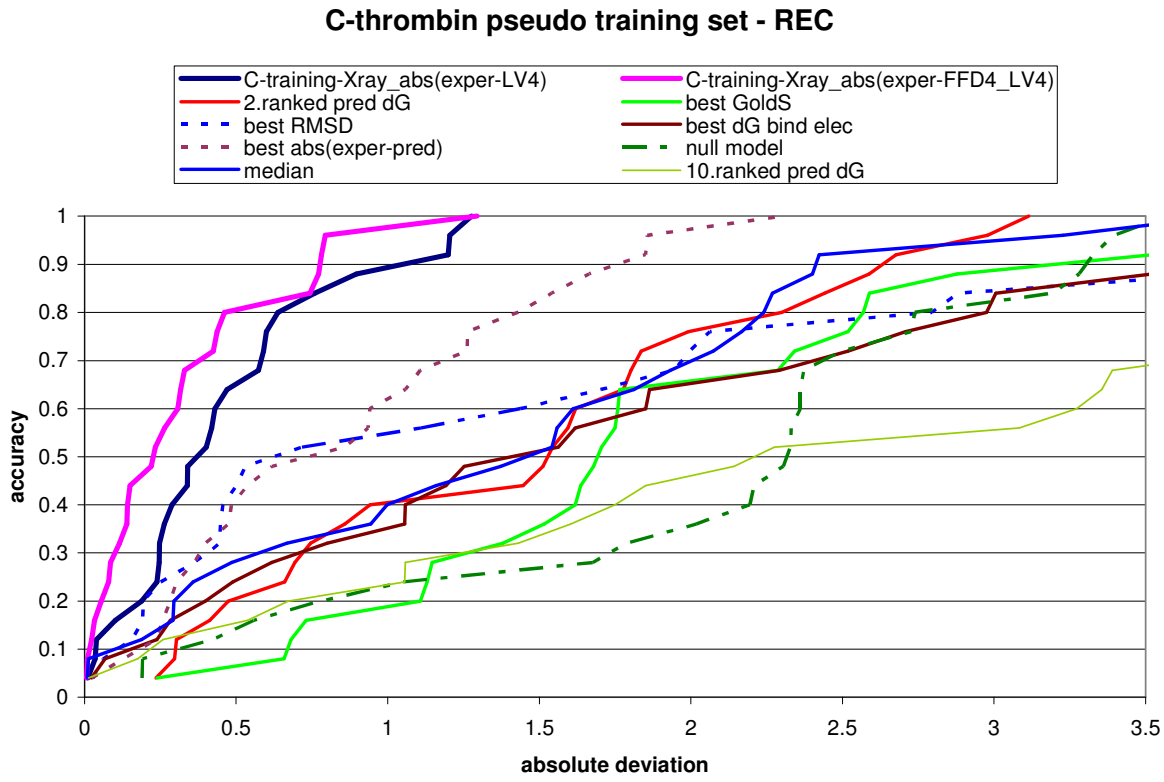


after variable selection (ffd4)

thrombin	**r2**	**r2spearman**	**SDEP**
before variable selection	0.93	0.96	0.59
xray	0.96	0.98	0.46
GoldS	0.31	0.59	2.09
best abs	0.78	0.81	1.08
best dG bind elec	0.42	0.64	2.12
best RMSD	0.24	0.40	2.47
1.	0.48	0.69	2.17
2.	**0.53**	0.66	**1.71**
6.	0.47	**0.72**	1.75
10.	0.12	0.23	3.37
mean	0.46	0.73	1.75
median	0.46	0.69	1.74
null model			2.30

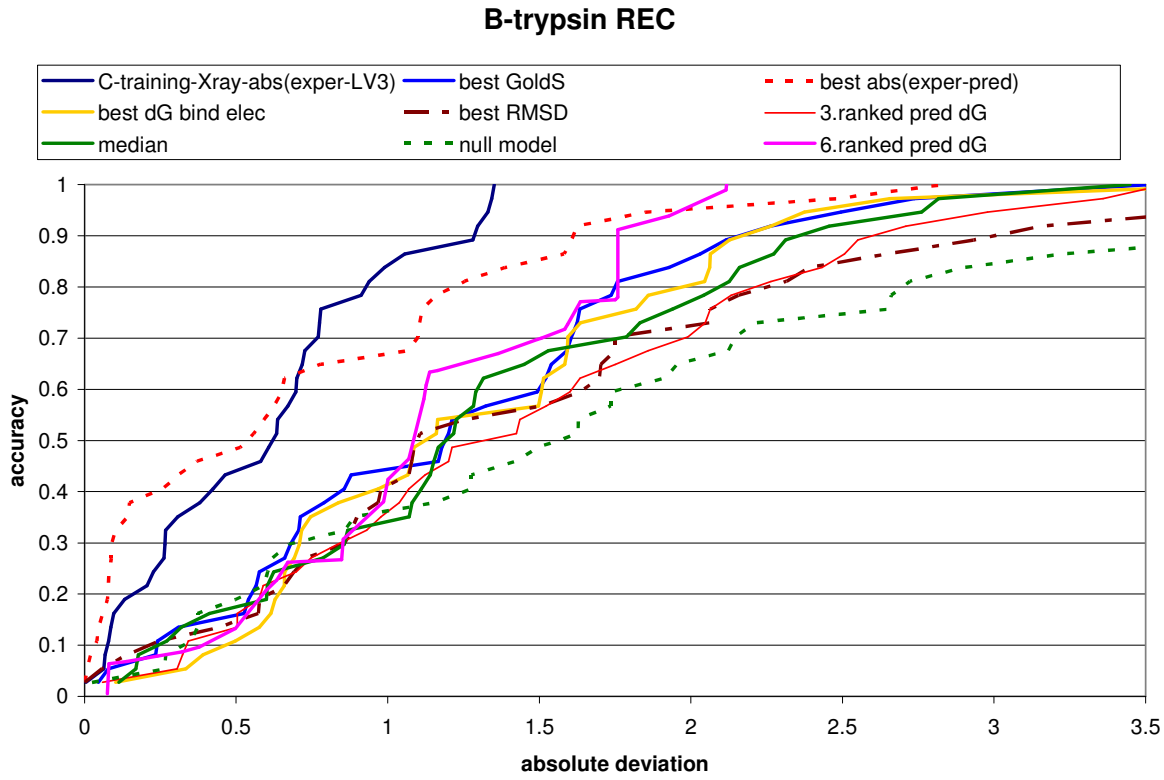
mean of experimental ΔG = -9.12

standard deviation of the experimental ΔG = 2.35



trypsin	r2	r2spearman	SDEP
xray	0.90	0.93	0.55
GoldS	0.77	0.78	1.18
best abs	0.87	0.87	0.80
best dG bind elec	0.71	0.78	1.21
best RMSD	0.54 (not all RMSD calculated so far)	0.66	1.68
1.	0.72	0.72	1.77
3.	0.74	0.78	1.39
6.	0.71	0.75	1.26
10.	0.72	0.71	1.94
mean	0.69	0.75	1.27
median	0.71	0.76	1.29
null model			2.13

mean of experimental=-7.81
standard deviation of the experimental ΔG =2.36



selectivity between trypsin and thrombin