Introduction

In the COMBINE method interaction energies between receptors and ligands are calculated for predicting the interaction energy of ligands with unknown bioactivity. In the most cases no receptor-ligand-complex structures are available, too. For this reason, the new ligands have to dock into the active site in the correct binding mode, and subsequently based on these docking solutions, interaction energies can be calculated.

Method

SD file

Creating and correcting an SD file [CACTVS browser and editor] containing information of 293 ligands with kinetic data for several trypsin-like serine proteases. Charges were set manually for a neutral pH (e.g. charged amidino group, uncharged phenol oxygen). From the 2D coordinates 3D coordinates were calculated by CORINA [all_300805_3D]. The ligands with available X-ray structures in complex with urokinase were selected by SDF toolkit and were used in the following docking experiments [PDB_uPA_300805.sdf].

residue	docking#	ID	CRA#	uPA	PDB_uPA
A04	1	21	8696	0.072	1o5a
A12	2	10	1144	8.9	1gi8
A13	3	16	6669	6	1gi9
A14	4	20	7806	0.45	1gjb_1gjc
A15	5	3	10655	0.22	1o3p
A16	6	5	10950	0.11	1gj8
A17	7	8	11092	0.033	1o5c
A18	8	2	10302	0.013	1gj7
A19	9	288	6860	0.21	1c5x_1c5w_1o5b
A20	10	289		5	1c5z_1f5k
A21	11	290	7538	63	1c5y
A22	12	291	10273	31	1gi7
A23	13	292	11421	6	1gjd
A24	14	293	7136	3.8	1gja
A25	15	294		2.4	1ejn
A26	16	295		5.3	1f5l
A28	17	297		0.64	1f92
A29	18	298		0.00062	1sqa
A30	19	299		0.035	1sqo
A31	20	300		0.63	1sqt
A32	21	301		0.145	1owd
A33	22	302		0.631	1owe
A34	23	303		0.04	1owh
A35	24	304		0.104	1owi
A36	25	305		0.048	1owj
A37	26	306		0.0235	1owk
A38	27	307		1.31	1u6q
A39	28	308		0.02	1vja
A40	29	309		0.028	1vj9
A41	30	310		10	1sc8

Receptor

The available X-ray structures of urokinase in complex with the small molecular ligands were downloaded from PDB and were superimposed by Pymol to structure 1o5a. Based on the structure 1o5a six urokinase models were built. The Ala190 (1o5a) were replaced by Ser190 (1o3p) as well as the residues His99 (1sqt, 1o3p), Gln192 (1c5z, 1gi8, 1o3p) and Arg217 (1owd) for building models with alternative conformations.
The model structures were protonated by WHATIF and were minimized by a short molecular mechanics calculation in AMBER8 (50 cycles, dielec=1.0, cut=20 Å, constrain: backbone=500/ side chains without replaced ones=50/ replaced side chains=1).

1o5a	His99	Ser190	Gln192	Arg217
model1	1sqt	1o3p	1c5z	1owd
model2	1o3p	1o3p	1c5z	1owd
model3	1sqt	1o3p	1gi8	1owd
model4	1o3p	1o3p	1gi8	1owd
model5	1sqt	1o3p	1o3p	1owd
model6	1o3p	1o3p	1o3p	1owd

Conserved water sites

Conserved water sites in urokinase structures were detected by a cluster analysis with WatCH. This program analysis the continuum of overlaps between water sites into the set of maximally dense microclusters of overlapping water molecules. 36 X-structures of urokinase complexes together with their water molecules were aligned by Pymol and were prepared for WatCH. The *B*-factor column of the resulting PDB file with positions of the calculated conserved water sites represented the conservation in percent. A threshold of 2.4 Å for separating two clusters gave in the ranges of 80-100 %, 70-100 %, 50-100 %, and 30-100 % gave 11, 23, 53 and 129 conserved water sites, respectively.

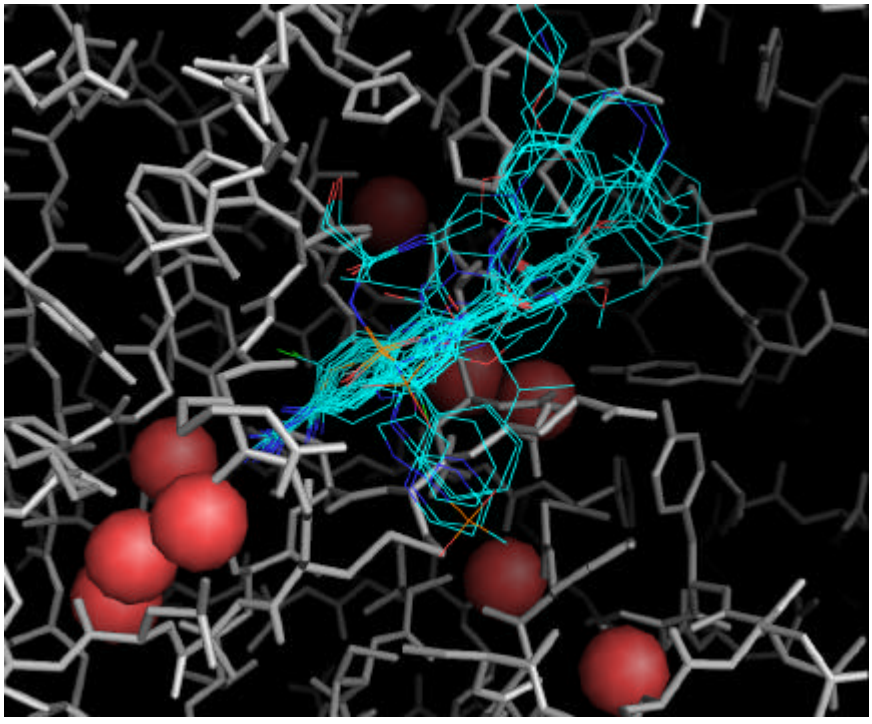

Figure: Urokinase model5 (white sticks) together with superimposed ligands of 30 X-ray structures (blue lines) used for docking. Water sites with a conservation of more than 80 % are drawn as red spheres.

Docking

The docking calculations were done with GOLD on the computers priamos, cronus and prometheus. As a receptor the minimized, protonated urokinase models with/without explicit water molecules were used in PDB format. The ligands were given in one SD file. The active site was defined within a radius of 10 Å around Gly228-Ca.

Substructure-based distance constraints

used distances:
C: max 4.0 Å, min, 3.6 Å, spring constant 25.0
N: max 3.4 Å, min 2.3 Å, spring constant 25.0

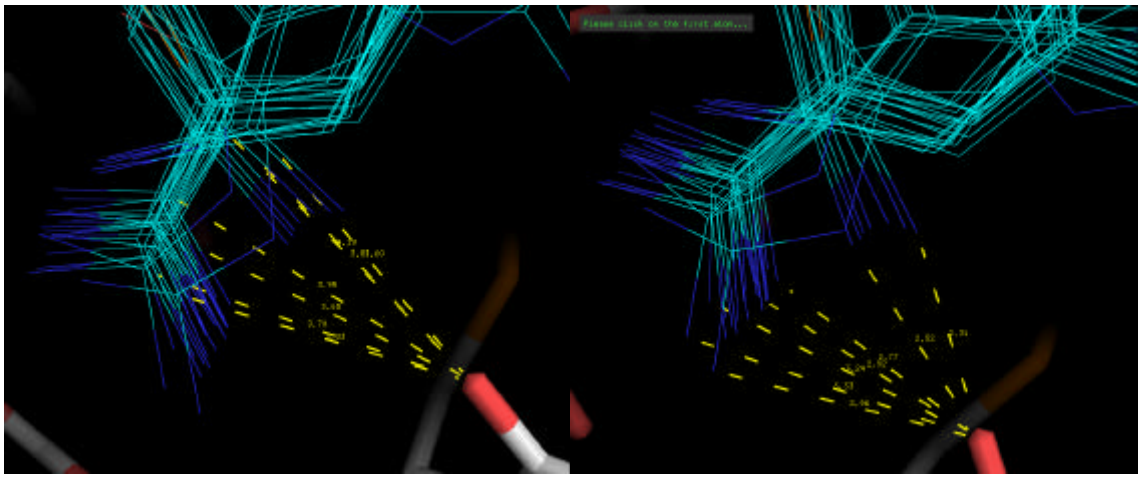

Figure: Distances between Gly230-O and the C atom of the functional group of amidino/guanidine group (left) and the N atom (right). The distances Gly230-O – C atom and Gly230-O – N atom were around 3.3 to 4.3 Å and 2.3 to 3.3 Å, respectively.

Docking with explicit water molecules

GOLD runs:

1	-
2	model1
3	model1, water80-100
4	model2, water80-100
5	model2
6	model2, amidino-N distance constraints
7	model2, amidino-C distance constraints
8	model5, amidino-C distance constraints
9	model5, amidino-N distance constraints
10	model2, water80-100, amidino-N distance constraints
11	model5
12	model5, water80-100
13	model5, water80-100, amidino-N distance constraints
14	model5, water80-100, amidino-C distance constraints

RMSD calculation

The PDB coordinates of the X-ray ligand conformations were converted by Babel into an SD file, which was compared afterwards to each SD file containing one docking solution. The RMSD of the heavy atoms was calculated by the utility program smart_rms of the GOLD program. This program calculates the RMSD between two conformations of the same structure, while taking account of symmetry effects. With a script [gold_analysis.sh] these values were tabled for each ligand. The best RMSD, their position among the GOLD score ranked docking solutions as well as the average and standard deviation over all ten docking solutions were calculated. For each run the summary values of the ligands were averaged (table).

Results

Three receptor models of urokinase (model1, model2, model5) were used for docking of 30 ligands. Run 2, 5 and 11 (see table) were done without using explicit water molecules or substructure-based distance constraints. In runs 3, 4 and 12 explicit water molecules with a conservation of more than 80 % were added to the model before docking the ligands. During the procedure of run 6, 9 and 10 as well as 7 and 9 substructure-based distance constraints between Gly230-O and the N atom or the C atom of the amidino group, respectively, were used.

receptor	water sites	con-straints	run	ranking # of best rmsd dock. sol.	best rmsd	best GOLD ranked rmsd	aver-age rmsd	sd of rmsd	# best rmsd <=2Å	# best GOLD <=2Å
model1			**2**	5.2	3.87	4.62	4.81	0.81	8	7
model2			**5**	5.2	3.91	4.93	4.81	0.70	7	4
model5			**11**	5.6	3.69	4.24	4.59	0.78	10	9
model1	**80 %**		**3**	5.9	3.73	4.32	4.55	0.55	10	7
model2	**80 %**		**4**	6.4	3.90	4.34	**4.52**	0.50	8	8
model5	**80 %**		**12**	6.4	**3.65**	4.30	4.55	0.68	**11**	**10**
model2	**80 %**	**N**	**10**	5.5	3.83	4.45	4.56	**0.48**	8	6
model2		**N**	**6**	**4.2**	3.80	4.97	4.64	0.71	9	6
model5		**N**	**9**	4.9	3.80	**4.13**	4.61	0.78	9	9
model2		**C**	**7**	4.3	3.78	4.93	4.77	0.85	9	6
model5		**C**	**8**	4.7	3.74	4.24	4.61	0.72	10	8

Table: Summary of 11 docking experiments. The values of the columns 5 to 9 are the average over 30 ligands of each run with 10 docking solutions for every ligand. The last two columns are listing the numbers of best RMSD values and best GOLD score ranked RMSD values of each run with an RMSD below 2 Å.

Models 1 and 2 (run 2, 5, 3, 4) showed (with and without water molecules) similar RMSD values, but model 5 (run 11, 12) led to better results, whereas model 5 in combination with using water molecules (run 12) gave the lowest RMSD of 3.65 Å. This run showed also the most best- or best-GOLD score ranked docking solutions of all the runs with an RMSD below 2 Å. The corresponding numbers of 11 and 10, respectively, are equivalent to a hit rate of 37 and 33 %. The lost RMSD of best GOLD score ranked docking solution could be found in the docking experiment of model 5 with distance constraints between Gly230-O and the N atom of the amidino group (run 9).

Distance constraints of 2.3 – 3.4 Å and 3.6 – 4.0 Å (spring constant 25.0) between Gly230-O and the N or C atom did not show much difference (for model 2 run 6, 7 and for model 5 run 9, 8).

For one ligand (10/A20) no correct and for some others (6/A16, 21/A32, 24/A35, 25/A36, 26/A37) very bad binding modes could predicted in any of the runs. These docking solutions with the best RMSD out of 10 showed a completely wrong even inverted binding modes.