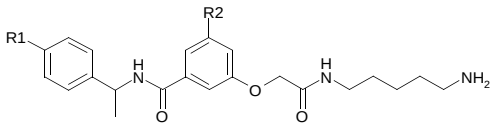
Supporting Information

In silico discovery of β-secretase inhibitors

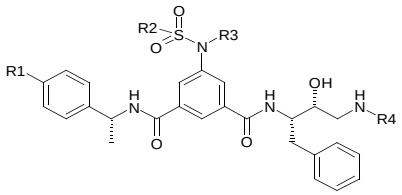
Danzhi Huang[a], Urs Lüthi[b], Peter Kolb[a], Marco Cecchini[a],
Alcide Barberis[b], and Amedeo Caflisch[a*]

Contents

1 Evaluation of binding free energy with LIECE



Compounds	Stereochem	R1	R2	IC_{50} (μM)	ΔG (kcal/mol)	LIECE Ki (kcal/mol)
1	*R*	H	C(O)N(allyl)2	25	-6.3	-6.3
2	*R*	H	2-CN-Ph	21	-6.4	-5.3
3	*R*	H	OSO2CH2Ph	3.4	-7.5	-6.8
4	*S*	H	OSO2CH2Ph	>100	>-5.5	-4.8
5	*R*	F	OSO2CH2Ph	1.4	-8.0	-7.3



Compounds	R1	R2	R3	R4	IC_{50} (μM)	ΔG (kcal/mol)	LIECE Ki (kcal/mol)
6	H	Me	Me	cyclopropyl	15	-10.9	-9.9
7	F	Me	H	cyclopropyl	63	-9.8	-8.6
8	F	Me	Me	cyclopropyl	10	-10.9	-9.8
9	F	Me	Me	H	23	-10.4	-10.2
10	F	Me	Me	ethyl	15	-10.7	-10.4
11	F	Me	Me	*tert-butyl*	15	-10.7	-10.0
12	H	*iPr*	Me	cyclopropyl	41	-10.1	-10.0
13	H	N(Me)2	Me	cyclopropyl	14	-10.7	-9.8

Table 1: Structures, experimental binding energies and LIECE energies of a set of 13 inhibitors [1, 2]

2 Preparation of the BACE-1 structure

The crystal structure of BACE-1 was solved in the complex with two peptidic inhibitors, OM99-2 [3] and OM00-3 [4]. Coordinates of BACE-1 in complex with the inhibitor OM00-3 (Glu-Leu-Asp-Leu-ψ{CHOH-CH2}-Ala-Val-Glu-Phe where ψ{CHOH-CH2} is a hydroxyethylene isostere of the peptide bond) were downloaded from the PDB database [5] (PDB entry 1M4H [4]). The inhibitor and all

water molecules were removed. Particular attention was addressed to the ionization state of the cleavage site, which contains the aspartyl dyad (Asp32/Asp228). At optimal pH for enzymatic activity ($\sim 3.5-4.5$ [6]), the aspartyl dyad is most probably monoprotonated in the uncomplexed enzyme as well as in the complex with peptidomimetic inhibitors with a hydroxyethylene isostere of the peptide bond. The choice of which of the two catalytic aspartates to protonate should have little effect on the relative binding affinity because all of the inhibitors have the same binding motif at the catalytic site [7]. Asp228 was protonated in this study [8]. Two other protein conformations, 1W51 [9] and 1TQF [1], were prepared by the same procedure.

3 Computational approach

The *in silico* screening approach consists of four steps which are presented in the following four subsections. In subsection **3.1** the first step (DAIM) is explained in detail whereas for the remaining three steps the main differences with respect to the original methods are highlighted (subsections **3.2-3.4**).

3.1 DAIM (Decomposition and Identification of Molecules)

The decomposition of a ligand into fragments and the choice of the anchor fragments for FFLD (see **3.3**) have been automatized recently (P. Kolb and A. Caflisch, manuscript in preparation). The major rules are listed here. The decomposition is guided by the fact that SEED (see **3.2**) treats all molecules as rigid. Hence, preference is given to aromatic rings but also other small rings and molecules that contain several amidic, double or triple bonds. The fact that non-aromatic ring systems might have several distinct conformations can be accounted for by the ability of SEED to dock multiple (predefined) conformations at the same time. If one of these conformations can be docked with a lower binding energy than the others, it will automatically be chosen in the subsequent steps, since it will receive higher ranks. FFLD requires three not-necessarily different fragments to place a

flexible ligand unambiguously in the binding site. The fragment determination and selection of the three most suitable fragments for flexible ligand docking by SEED-FFLD follows a few simple rules:

1. All atoms in a fragment must be connected by rigid or terminal bonds.

2. Large fragments are preferred since there are more steric constraints for large entities, as a consequence these should be positioned first.

3. Cyclic fragments are preferred because they usually are more rigid than acyclic moieties.

4. Since the fragments should be involved in the most significant interactions, those that contain hydrogen bond donors and acceptors are selected. Charged groups usually do not make such good anchors, since they tend to be positioned at the borders of the binding site, which are more exposed to the solvent. (However, there are exceptions as in the case of thrombin, where a very favorable electrostatic interaction is provided by a charged aspartic acid in the specificity pocket).

5. Fragments that are close to the center of the molecule are omitted, especially if they have a high number of substituent groups. Such "central" or "scaffold" fragments will hardly ever form significant interactions.

6. Finally, fragments should not overlap (i.e. one atom should not be part of two fragments), since this would mean that there are no rotatable bonds in between, so their relative position can not be changed.

The DAIM rules for fragment identification and selection of the three most suitable fragments for flexible docking by SEED-FFLD can be exemplified with the molecule XK263 (Dupont Merck, Fig. 1). In principle, there are three fragment types that could be chosen: naphthalene, benzene and the cyclic urea in the center. The largest fragment would be the cyclic urea. According to rule 5, this is not a good choice however, as it is the core fragment and has 4 substituents. Furthermore, it

is the most flexible of the three types, which is another point against its choice. The remaining two types are aromatic and thus a recommended choice (rule 1). Finally, DAIM selects two naphthalenes and one benzene and not vice versa (rule 2).

A more difficult choice is presented by acetyl-pepstatin (Fig. 2), since it has no rings and almost no rigid bonds. All the fragments obtained by the application of rule 1 are therefore very small. All the larger fragments with a rigid bond (the amide groups) are located in the backbone and will not make good anchors (rule 5). One of the few choices remaining is to select three i-butanes (the "side chains") which are preferable with respect to the terminal carboxylic group, which is charged (rule 4).

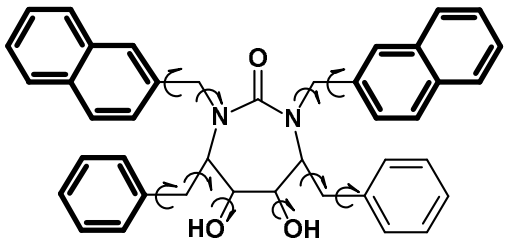


Figure 1: *XK263 (Dupont Merck) is a nanomolar inhibitor of HIV-1 aspartic protease (PDB accession code of the complex: 1HVR). Fragments selected by DAIM for SEED-FFLD docking are bold. Curly arrows denote rotatable bonds.*

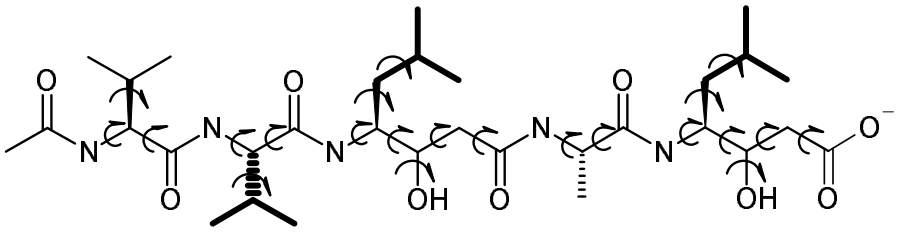


Figure 2: *Acetyl-pepstatin is a micromolar inhibitor of HIV-1 aspartic proteinase (PDB accession code of the complex: 5HVP). Fragments selected by DAIM for SEED-FFLD docking are bold. Curly arrows denote rotatable bonds.*

3.2 SEED (Solvation Energy for Exhaustive Docking)

The docking approach implemented in the program SEED [10] determines optimal positions and orientations of small to medium-size molecular fragments in the bind-

ing site of a protein. Apolar fragments are docked into hydrophobic regions of the receptor while polar fragments are positioned such that at least one intermolecular hydrogen bond is formed. Each fragment is placed at several thousand different positions with multiple orientations (for a total of in the order of 10^6 conformations) and the binding energy is estimated whenever severe clashes are not present (usually about 10^5 conformations). The binding energy is the sum of the van der Waals interaction and the electrostatic energy. The latter consists of screened receptor-fragment interaction, as well as receptor and fragment desolvations.

As an improvement with respect to previous versions of SEED [10, 11, 12] the screened electrostatic interaction as well as fragment desolvation energy were evaluated using an empirical correction of the Coulomb field approximation, i.e., equation 8 of Ref. [13]. The SEED input parameters used for this application to BACE-1 are identical to those used in Ref. [14]. SEED version 3.0 of March 2003 was used in this study.

3.3 FFLD (Fragment-based Flexible Ligand Docking)

The flexible-ligand docking approach FFLD uses a genetic algorithm and a very efficient scoring function [15]. The genetic algorithm perturbations affect only the conformation of the ligand; its placement in the binding site is determined by the SEED anchors and a least square fitting method [16]. In this way the position and orientation of the ligand in the binding site are determined by the best binding modes of its fragments previously docked using an accurate energy function with electrostatic solvation [17]. The scoring function used in FFLD is based on van der Waals and hydrogen bond terms and does not explicitly include solvation for efficiency reasons. Solvation effects are implicitly accounted for as the binding mode of the fragments are determined with electrostatic solvation.

The FFLD version 2.1 of July 2003, which contains the improvements presented in Ref. [12] as well as the ligand dihedral energy was used in the present application to BACE-1.

The FFLD poses are postprocessed by minimization with CHARMM [18].

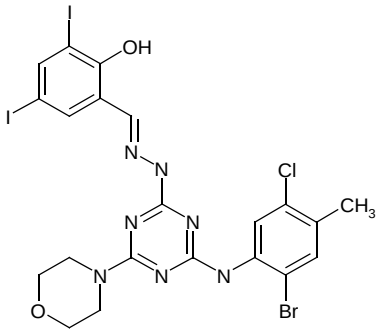
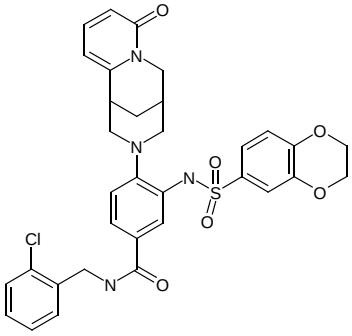
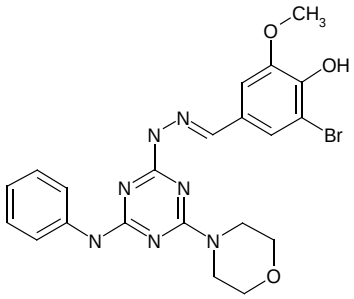
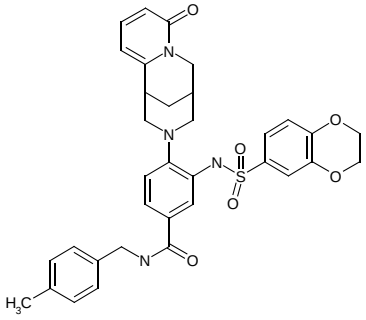
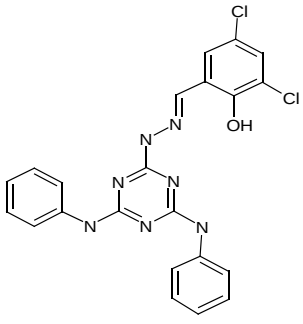
4 Structures, enzymatic activities and LIECE values of 88 tested compounds

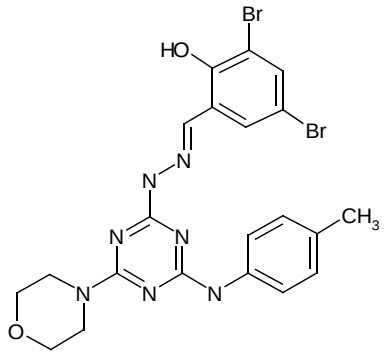
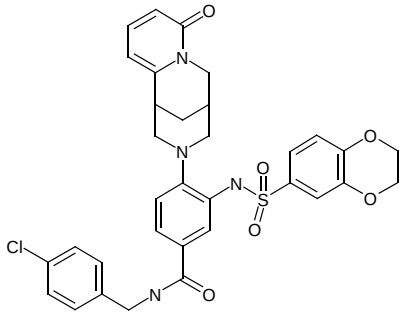
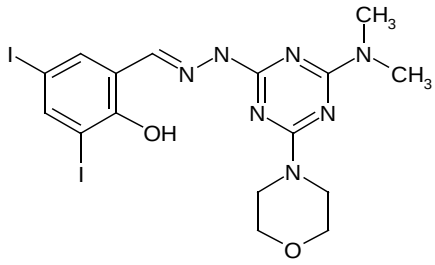
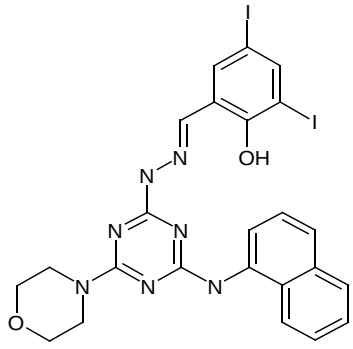
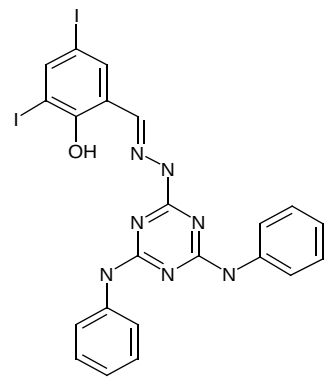
Sixty-four and 24 tested compounds from two high throughput dockings are listed in Table 2 and Table 3, respectively, and the rankings are according to their LIECE energies.

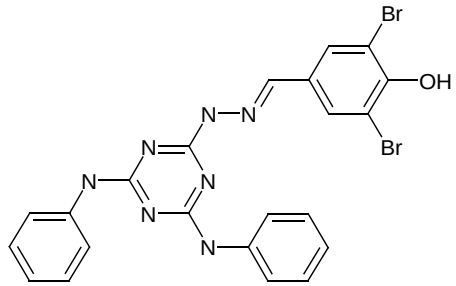
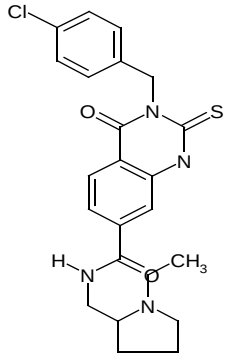
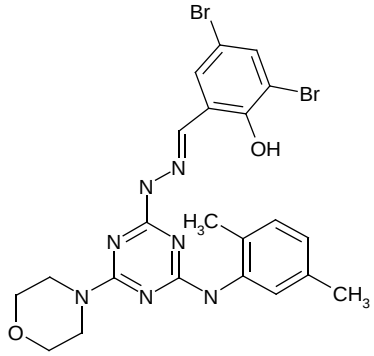
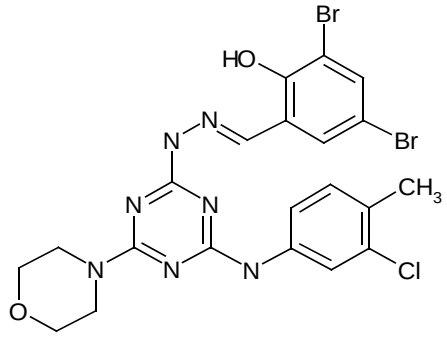
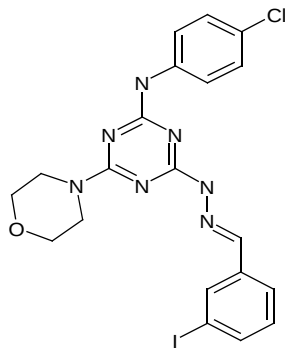
5 LIECE energy values of Figure 3

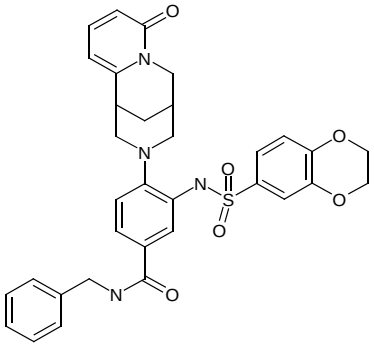
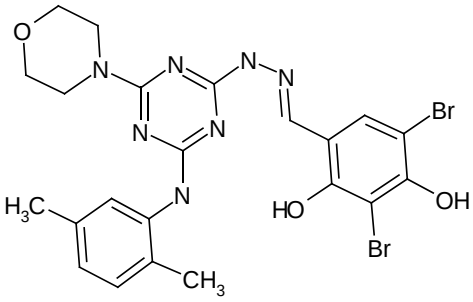
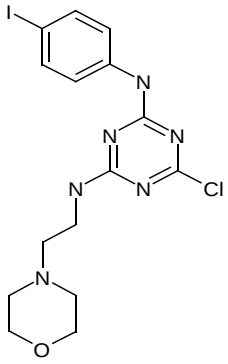
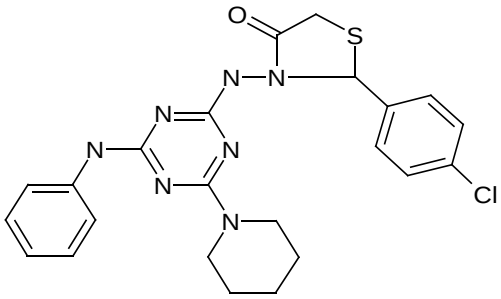
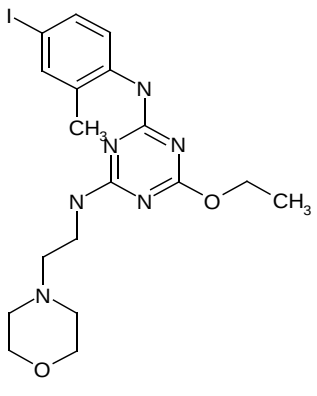
The LIECE energy values of 37 known inhibitors and the top100 library compounds (111 poses) of Figure 3 are listed in Table 4 and 5, respectively.

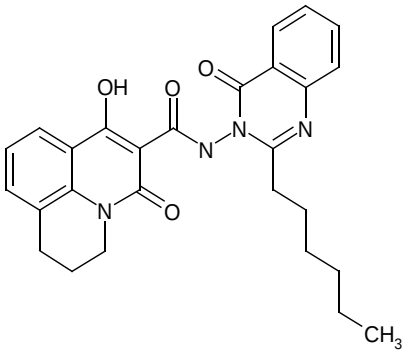
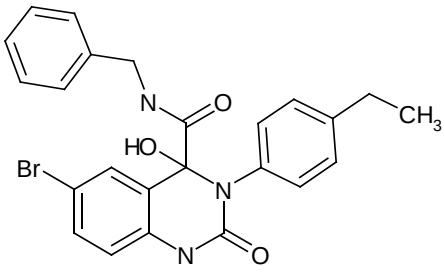
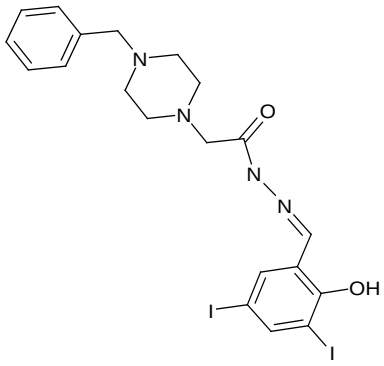
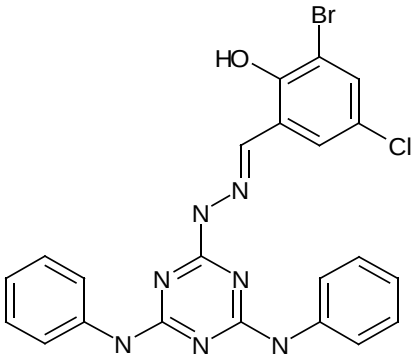
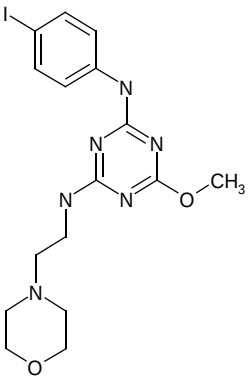
Table 2: Sixty-four compounds from the first *in silico* screening

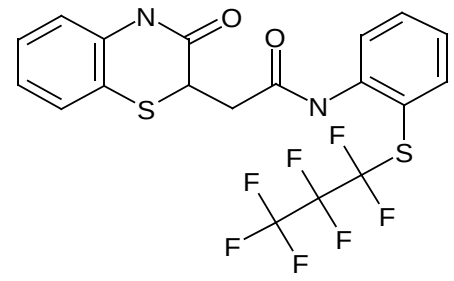
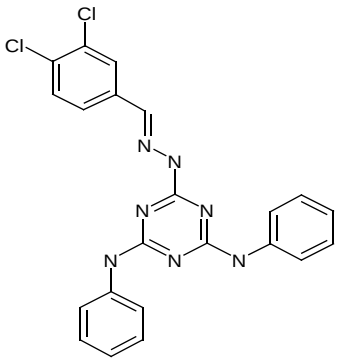
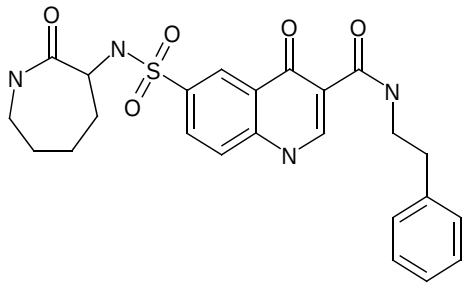
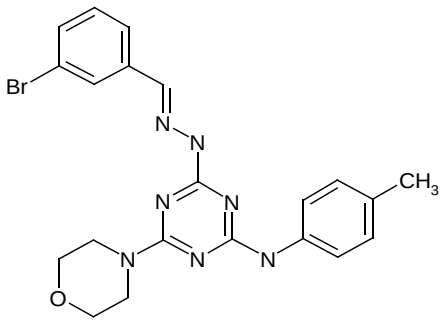
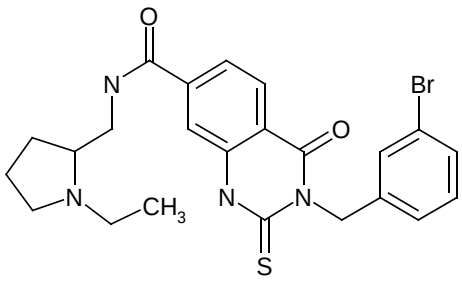
COMPOUND	STRUCTURE	MW (g mol^{-1})	BACE-1[a] IC_{50} (μM)	LIECE K_i (μM)
1		771	>100	9.1
2		647	>100	10.5
3		500	>100	10.9
4		627	>100	18.7
5		466	11.2 ± 0.2	20.2

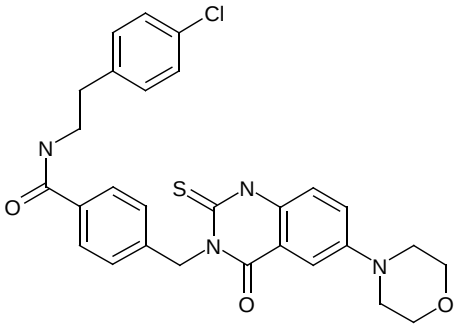
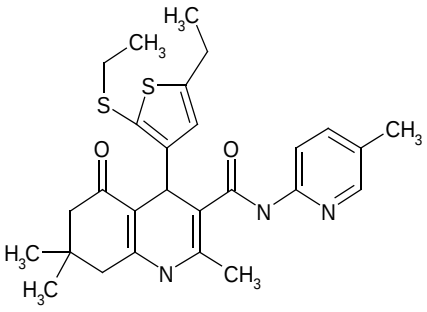
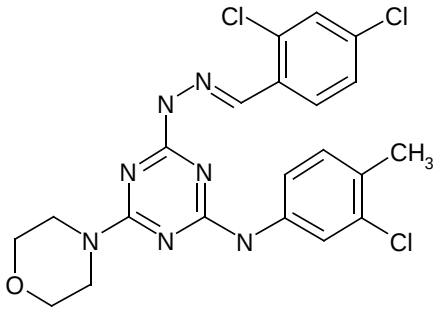
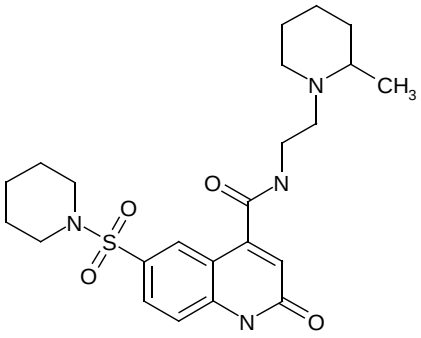
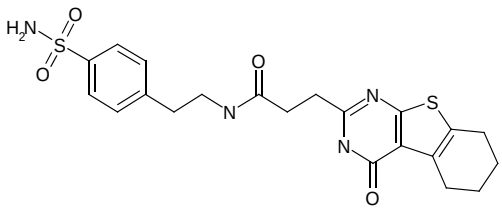
6		563	48.1 ± 15.8	30.9
7		647	>100	31.8
8		595	>100	37.5
9		693	>100	38.0
10		649	25.5 ± 8.4	38.8

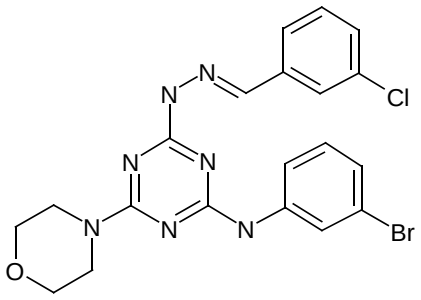
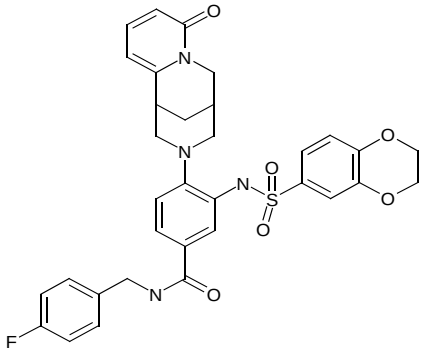
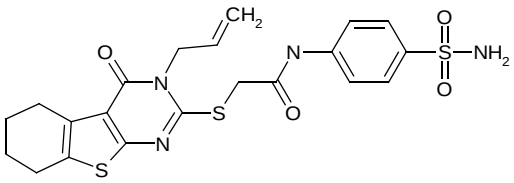
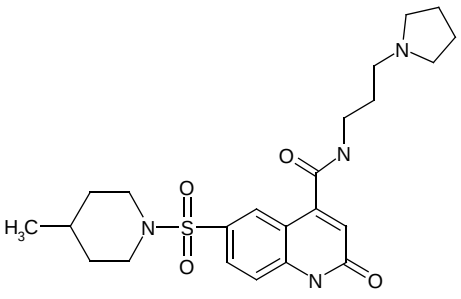
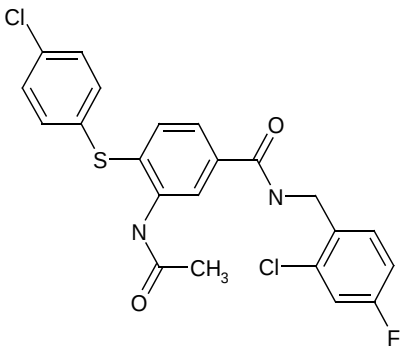
11		555	53.9 ± 16.4	41.4
12		457	>100	44.9
13		577	>100	48.1
14		598	20.6 ± 1.8	49.4
15		536	>100	56.3

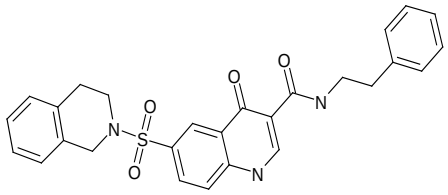
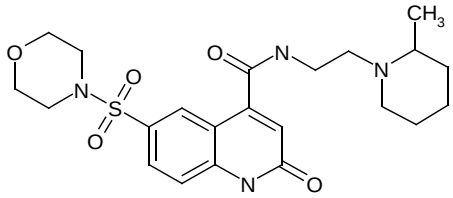
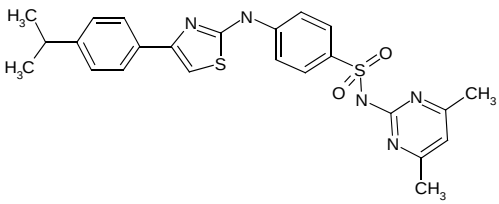
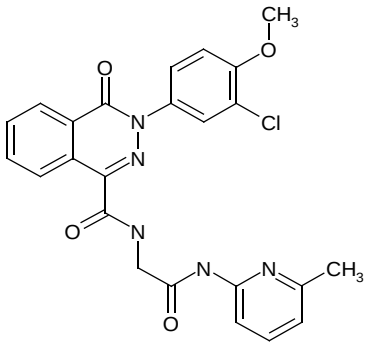
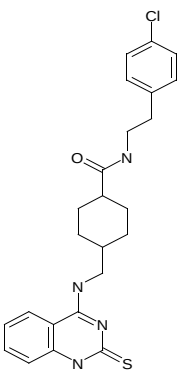
16		613	>100	56.4
17		593	81.8 ± 12.9	56.5
18		461	>100	60.0
19		482	>100	60.9
20		484	>100	61.0

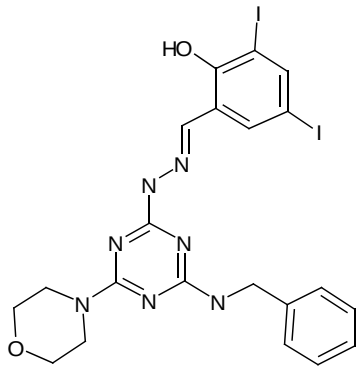
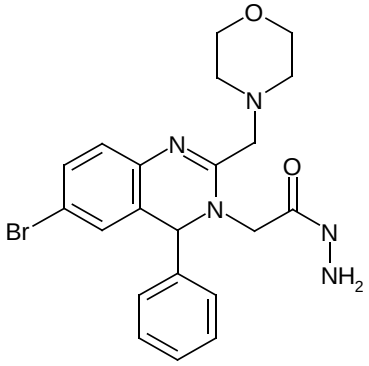
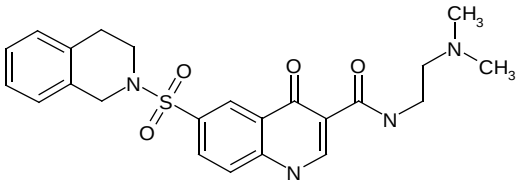
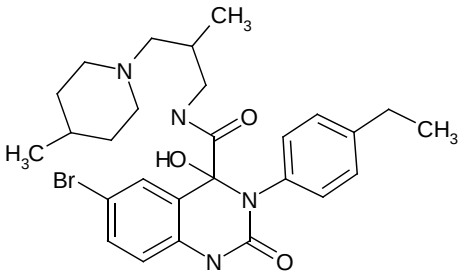
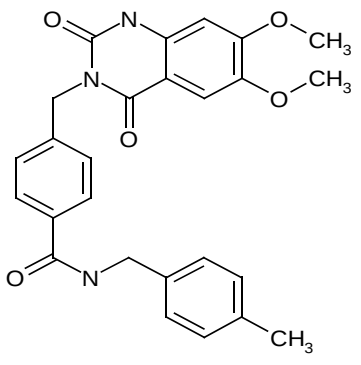
21		473	>100	62.4
22		480	>100	65.4
23		604	>100	65.7
24		511	11.9 ± 4.9	74.2
25		456	>100	75.1

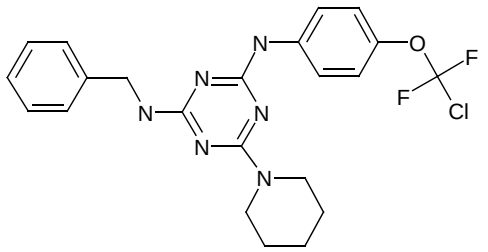
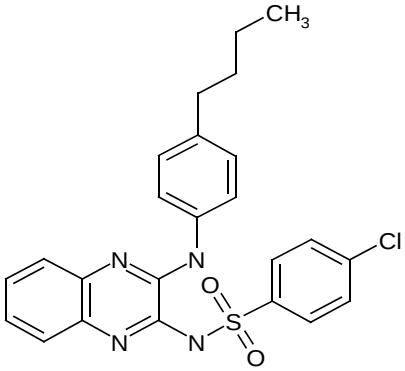
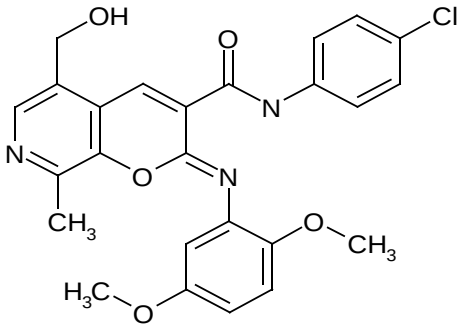
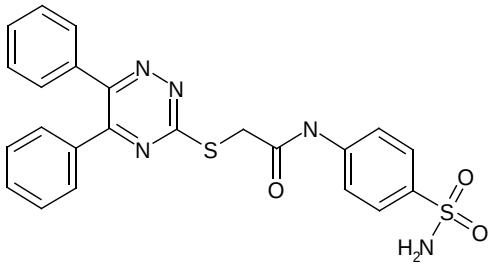
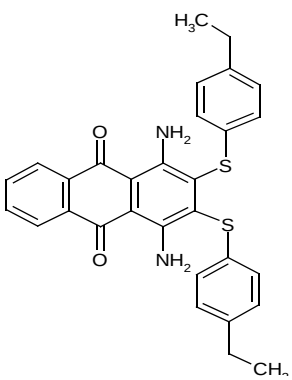
26		498	>100	76.0
27		450	>100	87.1
28		483	>100	88.0
29		468	>100	93.4
30		501	>100	94.4

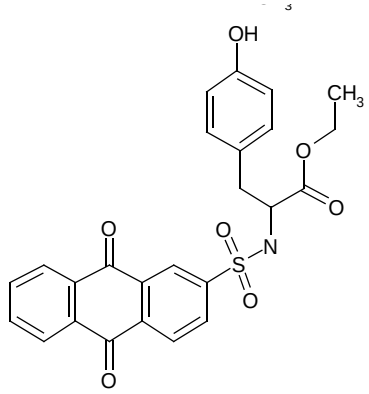
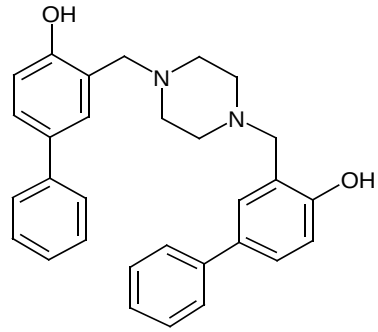
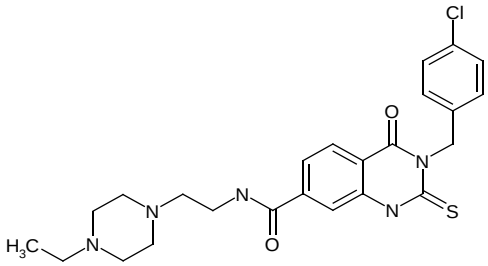
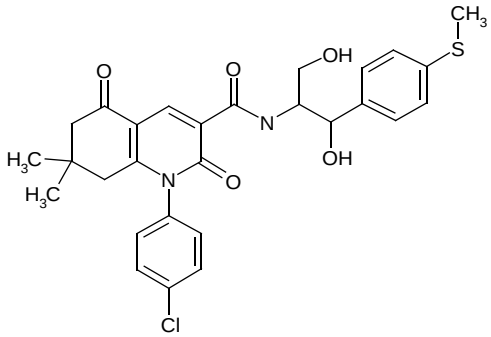
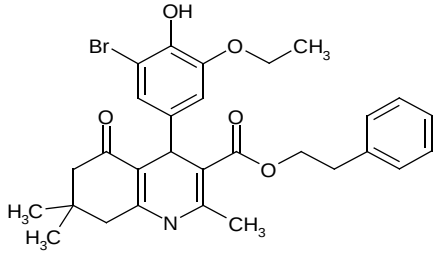
31		535	>100	96.7
32		496	>100	97.7
33		493	>100	105.4
34		461	>100	111.0
35		461	>100	115.8

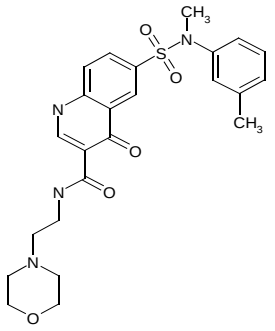
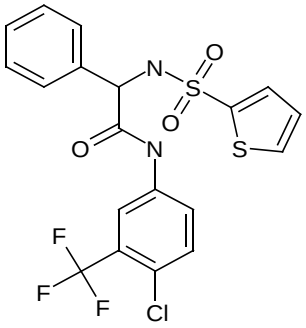
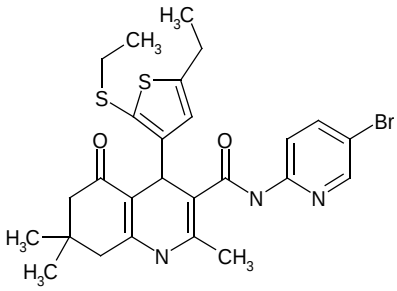
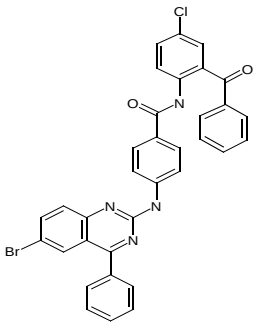
36		489	>100	118.5
37		631	>100	118.8
38		491	>100	119.5
39		461	>100	122.1
40		463	>100	123.0

41		488	>100	131.7
42		463	>100	134.2
43		480	No	135.1
44		478	>100	139.2
45		455	>100	176.3

46		657	>100	180.4
47		458	>100	181.6
48		455	>100	187.2
49		544	>100	189.0
50		460	>100	196.9

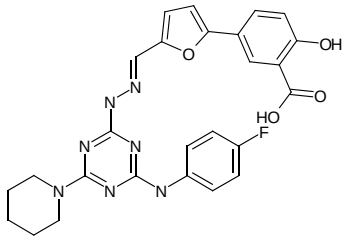
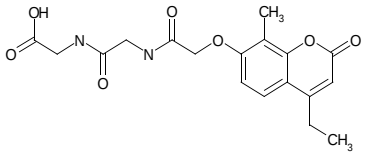
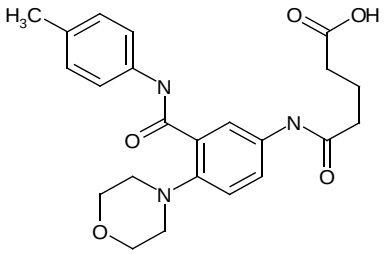
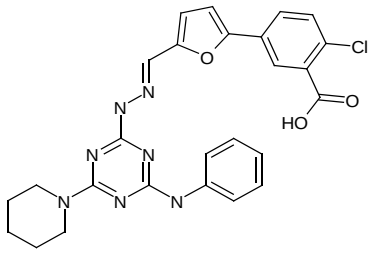
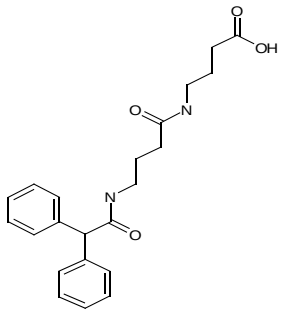
51		461	>100	199.6
52		467	>100	210.1
53		480	>100	227.0
54		478	>100	249.6
55		511	253	258.7

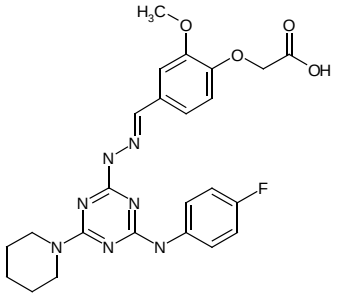
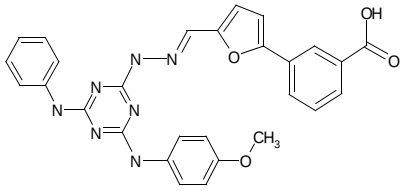
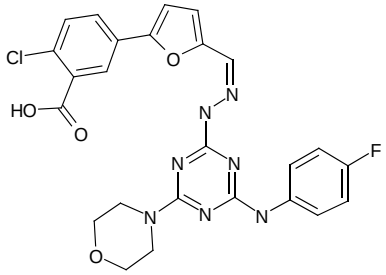
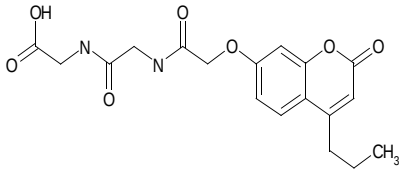
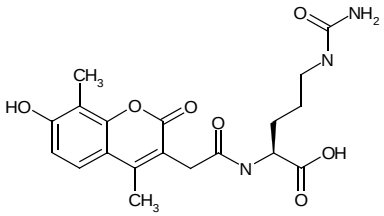
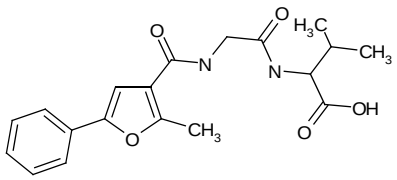
56		480	>100	276.7
57		451	>100	280.2
58		486	>100	294.0
59		541	>100	317.2
60		554	>100	357.5

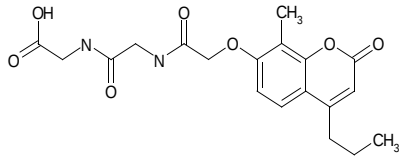
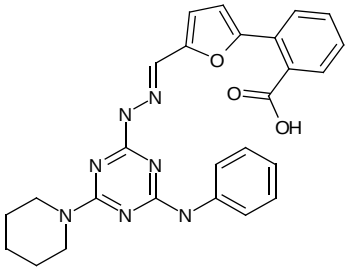
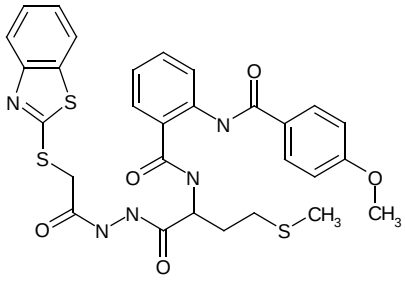
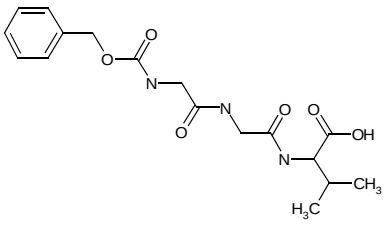
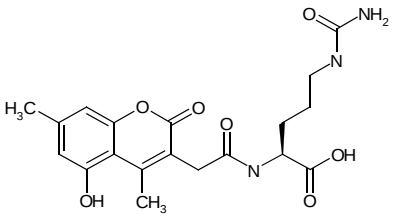
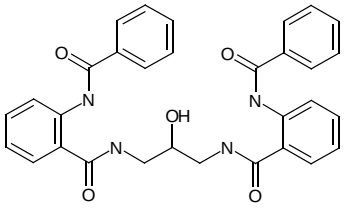
61		485	>100	370.7
62		475	>100	378.3
63		561	>100	388.1
64		634	>100	416.5

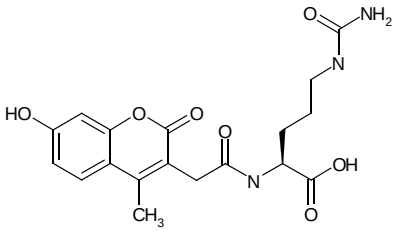
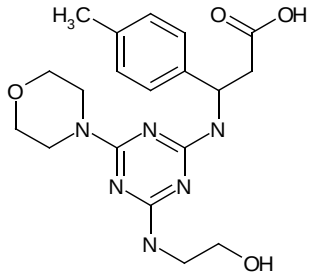
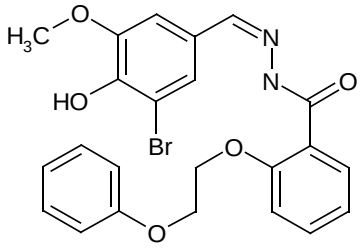
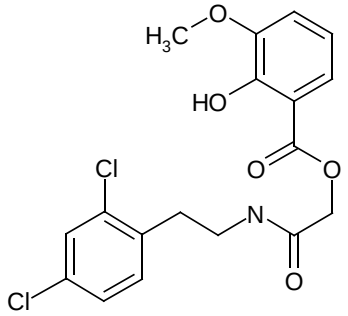
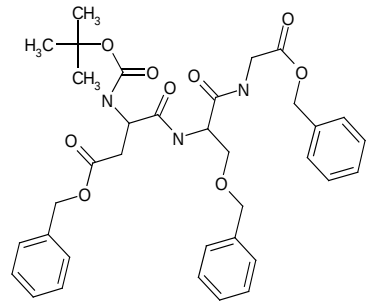
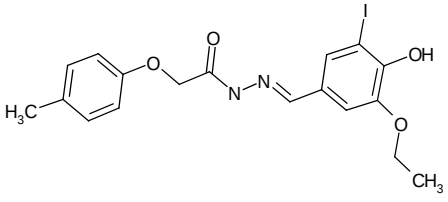
a The BACE-1 fluorescence resonance energy transfer assay kit was purchased from PanVera (Madison, WI; cat. P2985).

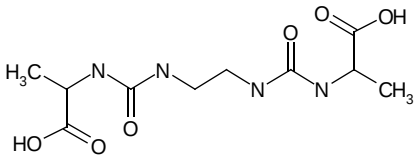
Table 3: Twenty-four compounds from the second *in silico* screening

COMPOUND	STRUCTURE	MW (g mol^{-1}	BACE-1[a] IC_{50} (μM)	LIECE K_i (μM)
1		518	27.9 ± 4.2	2.9
2		376	>500	5.2
3		425	>500	5.6
4		518	151.8 ± 14.5	6.6
5		382	>500	6.8

6		496	>100	11.4
7		522	66.6 ± 11.0	12.6
8		538	7.1 ± 1.2	12.8
9		376	>500	16.5
10		405	>500	19.5
11		358	>500	21.3

12		390	>100	22.4
13		484	>100	24.7
14		624	>50	30.8
15		365	>500	32.4
16		405	>500	35.9
17		537	>100	36.5

18		391	>500	40.4
19		402	>500	50.3
20		485	>500	52.0
21		398	>500	59.5
22		647	>100	62.6
23		454	>100	135.9

24		290	>500	135.9

[a] The BACE-1 fluorescence resonance energy transfer assay kit was purchased from PanVera (Madison, WI; cat. P2985).

Table 4: LIECE energy values of 37 known inhibitors

Compound		LIECE energy (kcal/mol)
Set1	1	-6.48
	2	-7.97
	3	-7.67
	4	-9.48
	5	-11.54
	6	-10.21
	7	-11.13
	8	-8.80
	9	-10.21
	10	-11.64
	11	-8.51
	12	-13.00
	13	-10.06
Set2	14	-6.31
	15	-5.32
	16	-6.80
	17	-7.29
	18	-9.87
	19	-8.57
	20	-9.83
	21	-10.21
	22	-10.36
	23	-9.97
	24	-9.77
	25	-9.08
Set3	26	-6.20
	27	-6.80
	28	-6.90
	29	-7.00
Set4	30	-5.60
	31	-6.00
	32	-5.86
	33	-6.40
	34	-7.57
	35	-7.07
	36	-6.69
	37	-6.68

Table 5: LIECE energy values of top 100 library compounds (111 poses)

Pose	Compound	LIECE energy (kcal/mol)	Pose	Compound	LIECE energy (kcal/mol)	Pose	Compound	LIECE energy (kcal/mol)
1	1	-8.20	38	34	-7.11	75	69	-6.75
2	2	-7.83	39	17	-7.10	76	13	-6.73
3	2	-7.79	40	35	-7.10	77	70	-6.72
4	1	-7.66	41	36	-7.08	78	71	-6.72
5	3	-7.63	42	37	-7.06	79	72	-6.7
6	4	-7.56	43	38	-7.03	80	73	-6.7
7	5	-7.48	44	39	-7.03	81	74	-6.68
8	6	-7.48	45	40	-7.03	82	75	-6.68
9	7	-7.47	46	41	-7.01	83	76	-6.66
10	8	-7.46	47	42	-7.01	84	77	-6.66
11	9	-7.46	48	43	-7.00	85	22	-6.65
12	10	-7.42	49	44	-6.98	86	78	-6.64
13	11	-7.40	50	45	-6.96	87	79	-6.63
14	12	-7.39	51	46	-6.95	88	80	-6.63
15	13	-7.35	52	47	-6.95	89	81	-6.63
16	14	-7.34	53	48	-6.94	90	82	-6.62
17	15	-7.33	54	49	-6.94	91	83	-6.62
18	16	-7.32	55	50	-6.93	92	84	-6.61
19	17	-7.30	56	51	-6.92	93	1	-6.6
20	18	-7.29	57	52	-6.92	94	85	-6.6
21	19	-7.28	58	53	-6.92	95	9	-6.6
22	10	-7.27	59	54	-6.91	96	86	-6.59
23	20	-7.26	60	2	-6.9	97	87	-6.59
24	21	-7.24	61	55	-6.89	98	88	-6.58
25	22	-7.23	62	56	-6.87	99	89	-6.58
26	23	-7.19	63	57	-6.86	100	90	-6.58
27	24	-7.19	64	58	-6.85	101	91	-6.55
28	25	-7.18	65	59	-6.84	102	92	-6.55
29	26	-7.18	66	60	-6.84	103	93	-6.55
30	27	-7.18	67	61	-6.84	104	94	-6.54
31	1	-7.16	68	62	-6.82	105	95	-6.54
32	28	-7.16	69	63	-6.82	106	65	-6.54
33	29	-7.15	70	64	-6.8	107	96	-6.54
34	30	-7.14	71	65	-6.79	108	97	-6.53
35	31	-7.12	72	66	-6.79	109	98	-6.53
36	32	-7.11	73	67	-6.76	110	99	-6.53
37	33	-7.11	74	68	-6.76	111	100	-6.52

6 Complete Refs. 3, 11 and 36

[3] Luo, Y.; Bolon, B.; Kahn, S.; Bennett, B. D.; Babu-Khan, S.; Denis, P.; Fan, W.; Kha, H.; Zhang, j.; Gong, Y.; Martin, L.; Louis, J.; Yan, Q.; Richards, W. G.; Citron, M.; Vassar, R. Nat. Neurosci., 2001, 4, 231-232.

[11] Hanessian, S.; Yun, H.; Yang, G.; Bayrakdarian, M.; Therrien, E.; Moitessier, N.; Roggo, S.; Veenstra, S.; Tintelnot-Blomley, M.; Rondeau, J.-M.; Ostermeier, C.; Strauss, A.; Ramage, P.; Paganetti, P.; Neumann, U.; Betschart, C. J. Med. Chem., 2005, 48, 5175-5190.

[36] Stachel, S. J.; Coburn, C. A.; Steele, T. G.; Jones, K. G.; Loutzenhiser, E. F.; Gregro, A. R.; Rajapakse, H. A.; Lai, M. T.; Crouthamel, M. C.; Xu, M.; Tugusheva, K.; Lineberger, J. E.; Pietrak, B. L.; Espeseth, A. S.; Shi, X. P.; Chen-Dodson, E.; Holloway, M. K.; Munshi, S.; Simon, A. J.; Kuo, L.; Vacca, J. P. J. Med. Chem., 2004, 47, 6447-6450.

References

[1] Coburn, C. A.; Stachel, S. J.; Li, Y. M.; Rush, D. M.; Steele, T. G.; et al. Identification of a small molecule nonpeptide active site β-secretase inhibitor that displays a nontraditional binding mode for aspartyl proteases. *J. Med. Chem.* **2004**. *47*, 6117–6119.

[2] Stachel, S. J.; Coburn, C. A.; Steele, T. G.; Jones, K. G.; Loutzenhiser, E. F.; et al. Structure-based design of potent and selective cell-permeable inhibitors of human β-secretase (BACE-1). *J. Med. Chem.* **2004**. *47*, 6447–6450.

[3] Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; et al. Structure of the protease domain of memapsin 2 (β-secretase) complexed with inhibitor. *Science* **2000**. *290*, 150–153.

[4] Hong, L.; Turner, R. T.; Koelsch, G.; Shin, D.; Ghosh, A. K.; et al. Crystal structure of memapsin 2 (β-secretase) in complex with an inhibitor OM00-3. *Biochemistry* **2002**. *41*, 10963–10967.

[5] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; et al. The protein data bank. *Nucl. Acids Res.* **2000**. *28*, 235–242.

[6] Grüninger-Leitch, F.; Schlatter, D.; Küng, E.; Nelböck, P.; Döbeli, H. Substrate and inhibitor profile of BACE and comparison with other mammalian aspartic proteases. *J. Biol. Chem.* **2002**. *277*, 4687–4693.

[7] Tounge, B. A.; Reynolds, C. H. Calculation of the binding affinity of β-secretase inhibitors using the linear interaction energy method. *J. Med. Chem.* **2003**. *46*, 2074–2082.

[8] Rajamani, R.; Reynolds, C. H. Modeling the protonation states of the catalytic aspartates in β-secretase. *J. Med. Chem.* **2004**. *47*, 5159–5166.

[9] Patel, S.; Vuillard, L.; Cleasby, A.; Murray, C. W.; Yon, J. Apo and inhibitor complex structures of BACE (β-secretase). *J. Mol. Biol.* **2004**. *343*, 407–416.

[10] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. *Proteins: Structure, Function, and Bioinformatics* **1999**. *37*, 88–105.

[11] Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins: Structure, Function, and Bioinformatics* **2001**. *42*, 256–268.

[12] Cecchini, M.; Kolb, P.; Majeux, N.; Caflisch, A. Automated docking of highly flexible ligands by genetic algorithms: A critical assessment. *J. Comput. Chem.* **2004**. *25*, 412–422.

[13] Lee, M. S.; Salsbury, F. R.; Brooks III, C. L. Novel generalized born methods. *J. Chem. Phys.* **2002**. *116*, 10606–10614.

[14] Huang, D.; Lüthi, U.; Kolb, P.; Edler, K.; Cecchini, M.; et al. Discovery of cell-permeable non-peptide inhibitors of β-secretase by high-throughput docking and continuum electrostatics calculations. *J. Med. Chem.* **2005**. *48*, 5108–5111.

[15] Budin, N.; Majeux, N.; Caflisch, A. Fragment-based flexible ligand docking by evolutionary opimization. *Biol. Chem.* **2001**. *382*, 1365–1372.

[16] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **1976**. *A32*, 922–923.

[17] Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A* **1997**. *101*, 8098–8106.

[18] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**. *4*, 187–217.