Comprehensive 3D-QSAR and binding mode of BACE-1 inhibitors using R-group search and molecular docking

Dandan Huang, Yonglan Liu, Bozhi Shi, Yueting Li, Guixue Wang, Guizhao Liang*

Key Laboratory of Biorheological Science and Technology, Ministry of Education, Bioengineering College, Chongqing University, Chongqing 400044, China

ARTICLE INFO

ABSTRACT

The β-enzyme (BACE), which takes an active part in the processing of amyloid precursor protein, thereby leads to the production of amyloid-β (Aβ) in the brain, is a major therapeutic target against Alzheimer's disease. The present study is aimed at studying 3D-QSAR of BACE-1 inhibitors and their binding mode. We build a 3D-QSAR model involving 99 training BACE-1 inhibitors based on Topomer CoMFA, and 26 molecules are employed to validate the external predictive power of the model obtained. The multiple correlation coefficients of fitting modeling, leave one out cross validation, and external validation are 0.966, 0.767 and 0.784, respectively. Topomer search is used as a tool for virtual screening in lead-like compounds of ZINC databases (2012); as a result, we successfully design 30 new molecules with higher activity than that of all training and test inhibitors. Besides, Surflex-dock is employed to explore binding mode of the inhibitors studied when acting with BACE-1 enzyme. The result shows that the inhibitors closely interact with the key sites related to ASP93, THR133, GLN134, ASP289, GLY291, THR292, THR293, ASN294, ARG296 and SER386 of BACE-1.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Alzheimer's disease (AD) is a common and serious neurological disease characterized by memory loss and cognitive decline [1], which is like an eraser in patients' brain, removing the memory of their friends and relatives and even their own, to some extent. AD plagues elderly people all over the world; meanwhile, it results in heavy economical and moral pressure to their families, leading to a sharp decline in the quality of their life and increasing social instability. At present, at least 35 million people around the world suffer from AD, and the total annual cost is up to $200 billion [2]. It also shows an exponent increase for the number of AD patients [2] over the coming several decades.

Driven by the harm of AD, Scientists are struggling to seek for effective drugs against this disease. To date, many drugs have come into clinical applications, but none of them can effectively relieve AD; therefore, it is imminent to develop effective medications against AD. The major pathological features of AD include two aspects: one is insoluble neurofibrillary tangles (NFT) [3] which are generated in cell mediated by tau protein, the other is commonly known as senile plaques (SPs) resulting from the mis-aggregation of extracellular amyloid-β (Aβ) peptide [4], which causes brain dysfunction in patients by producing toxicity to cell.

In recent years, more and more research show that excessive aggregation and tangles of Aβ are main factors that lead to AD [5], which is strongly supported by genetic, histopathological and clinical evidences [6]; hence, clinical intervention to reduce Aβ levels in brain has been an attractive and promising approach for the development of therapeutics for AD.

Amyloid precursor protein (APP), as a kind of transmembrane protein in biological tissue, is indispensable for the normal function of human body, and its expression mainly occurs in the brain. Aβ is produced through a first cleavage of a membrane APP by a protease known as β-secretase, which has been identified and designated as the membrane-anchored aspartic protease BACE-1. The matured Aβ is then generated by a second cleavage at the C-terminus called γ-secretase [7]. Thus the work of β-secretase and γ-secretase is a prerequisite for the generation of Aβ. However, γ-secretase itself has many physiological functions in the growth of cell, and whether inhibiting its function in Aβ production can interferes with other important functions remains unknown [8]. Besides, the mice with removal of β-secretase have less Aβ than that of the normal mice, and are healthy and fertile, and appear to be normal in cognitive ability [9,10]. Therefore, reducing the secretion of β-secretase or decreasing expression of β-secretase has become one of the important strategies of drug development for AD. In recent years, we have witnessed hundreds of publications and patents related to BACE-1 inhibitors [11], but the current AD treatments can only attenuate slowing of cognitive and the underlying pathological mechanism is completely unknown.

* Corresponding author. Tel.: +86 2365112677.
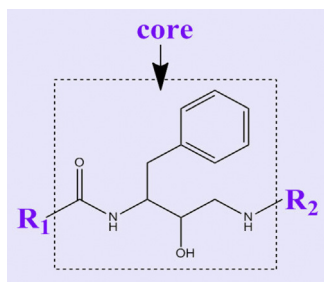 *E-mail addresses:* gzliang@cqu.edu.cn, sdqdlgz@163.com (G. Liang).

Fig. 1. The common structure of BACE-1 inhibitors (the part inside the dotted line box is the core group, and R_1 and R_2 represent substituents.)

Quantitative structure–activity relationship (QSAR) describes the linkage between structural features and biological activities of compounds, widely used as valuable tools to assist design of new inhibitors in clinical medicine [12,13]. Comparative molecular field analysis (CoMFA) [14] is one of well-established 3D-QSAR methods. Since its introduction in 1988, CoMFA has made significant contribution to ligand-based drug design. However, there are some restrictions for input ligands, i.e., the ligands should have typical 3D-structures, and they have to align with molecules in database. The whole procedure is time-consuming, highly subjective and has low repeatability [15].

To overcome the shortcomings, Cramer's group proposed the second generation of CoMFA method, i.e., Topomer CoMFA [16], which belongs to one of fragment-based 3D-QSAR methods. In this case, fragments can automatically generate according to molecular 3D-pose, and experientially complete identification and alignment of pose of the fragments; as a result, a reliable QSAR model may yield in a few minutes. Moreover, it is achievable for Topomer search using CoMFA model, as a powerful screening tool, to do automatic alignment on new compounds, thus the potency of the screened compounds with desired structural similarity can be directly predicted by the QSAR model obtained; thereby, R-groups or substituent groups which can be benefit to the activity of the designed molecules are obtained.

In this study, we built a 3D-QSAR model of BACE-1 inhibitors based on Topomer CoMFA, and analytically discussed its field-effect relationship. Then, we carried out virtual screening in ZINC molecular databases by employing Topomer search as a screening tool; as a result, a total of 30 new molecules with higher activity than that of the training molecules were designed. Finally, we used surflex-dock to explore binding mode of the newly designed molecules, by comparison with that of the training molecules. The present work supplies technical references for drug design against AD.

2. Principles and methods

2.1. Data set

The data set used in this work is constituted by 125 BACE-1 inhibitors which were synthesized and assayed by Emmanuel Demont et al. [17–23]. The potency of the molecules was identified under the same experimental conditions, as measured by $pIC_{50}(-\log IC_{50})$, which was the negative logarithm of half maximal inhibitory concentration (IC_{50}) value measuring the effectiveness of compound inhibition. The inhibitor as depicted in Fig. 1 is composed by a common structural skeleton also called as a core, together with 2R-groups which are influential on molecular biological activity.

2.2. Three-dimensional structure construction

The 3D structures of the BACE-1 inhibitors were constructed using SYBYL 8.1 molecule modeling package followed by energy minimization with gradient descent method. Then, each single optimized pose of each molecule in the data set was minimized on energy employing Tripos force field and Gasteiger–Huckel charges. The corresponding parameters were set as following: the maximum iteration coefficient was 1000, the gradient descent coefficient was 0.001, and the energy convergence limit was 0.001 kj/mol. According to their structural similarity, a training set including 99 molecules (1–99; Table 1) was chosen to build the QSAR model, and the others (100–125; Table 1) as test set which were not used in the generation of the QSAR model were applied in the external validation to test the predictive ability of the model.

2.3. Topomer CoMFA modeling

The Topomer CoMFA model was developed using SYBYL8.1. Briefly, the molecular 3D-pose was conducted and optimized by adding hydrogen atoms and charges. Then, identification and alignment of pose of the fragments were quickly complemented by empirical rules, and a 3D-grid containing all training set molecules was created; meanwhile, the steric and electrostatic field energy between the molecules and a sp^3 hybridized carbon atom with a charge of +1 was calculated. The descriptors obtained were used as independent variables, and pIC_{50} values were considered as dependent variables in partial least square (PLS) [24]. The Topomer CoMFA model was built based on the optimum number of principal components, and evaluated by a cross validation approach.

2.4. Molecular screening

Topomer search [25], as a fast 3D ligand-based virtual screening tool, is capable of searching fragments similar to the chemical structures of the known lead compounds in large libraries of compounds. Intrinsically, it is a kind of screen for whole molecules, side chains, or scaffolds using posed independence of topomer similarity. In the present work, Topomer search was employed to screen R-groups in a total of 534597 lead-like compounds of ZINC molecule database (2012) [26]. Topomer distance was set as 185 to evaluate the binding degree, and other parameters were defaulted by Sybyl 8.1.

2.5. Molecular docking

Surflex-dock was employed to explore binding mode of the inhibitors when acting with BACE-1 enzyme. The crystal structure (PDB ID: 2VIZ) of BACE-1 was prepared by a series of steps including removing water of crystallization and adding hydrogen atoms, etc. The prototype molecule (protomol) was generated on the condition that the coordinate of ligand was extracted from the crystal structure as the location center. Then, the method for the preparation for inhibitor molecules was in accordance with what was previously described in establishing Topomer CoMFA modeling section. The ligand extracted from the crystal structure of BACE-1, named N-[(1S,2R)-1-benzyl-3-{[(1S)-2-(cyclohexylamino)-1-methyl-2-oxoethyl]amino}-2-hydroxypropyl]-3-(2-oxo-2,3-dihydro-1H-pyrrol-1-yl)-5-propoxybenzamide (Fig. S1), as a reference molecule, was used to define the docking site with BACE-1. The number of the maximum output poses was set as 20, and other set parameters were defaulted by Sybyl 8.1. The output poses were evaluated by scoring functions. In the present work, we chose output poses using C-score [27], combining with a relatively large total score containing crash score and polar score.

Table 1

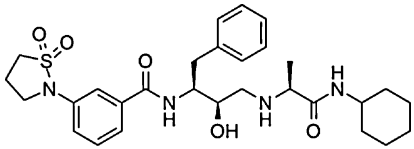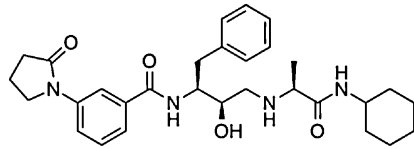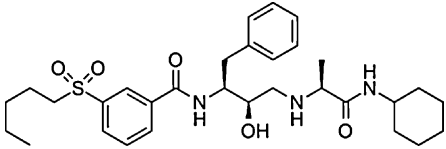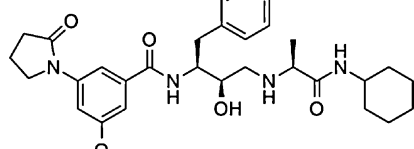Structures and pIC$_{50}$ values of inhibitors used for modeling.
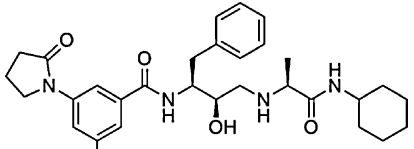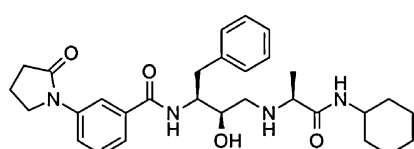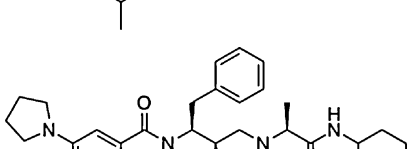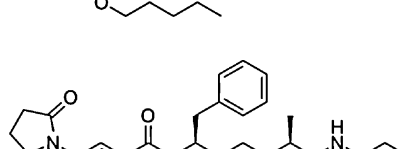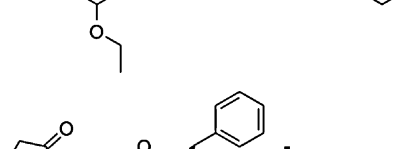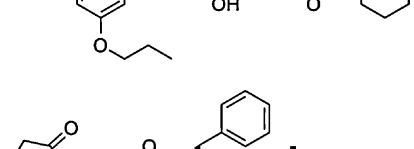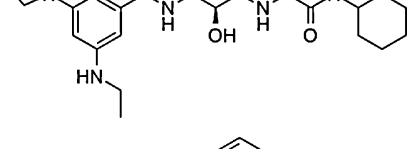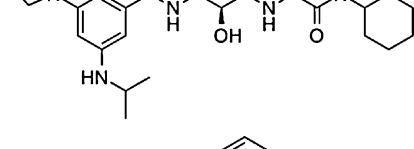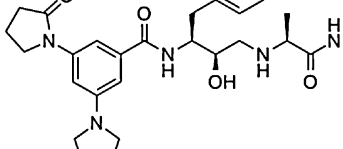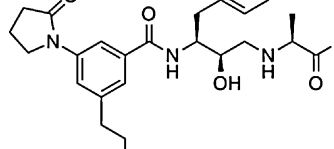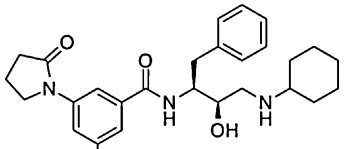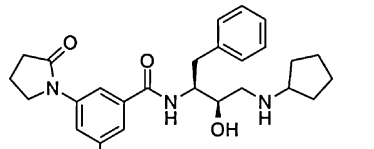
Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
1		5.72	2		5.28
3		5.74	4		6.13
5		6.92	6		7.49
7		5.84	8		6.22
9		7.89	10		7.41
11		6.07	12		7.14
13		6.74	14		5.61
15		6.19	16		5.23

Table 1 (*Continued*)
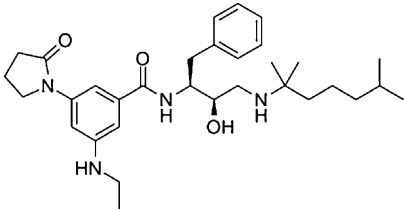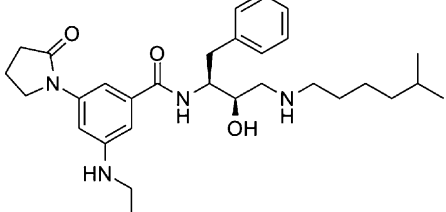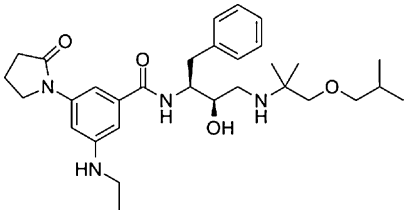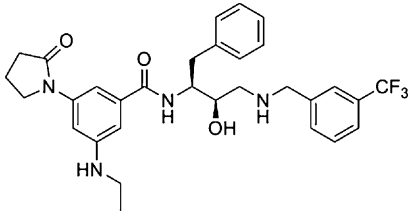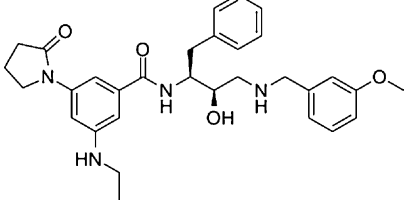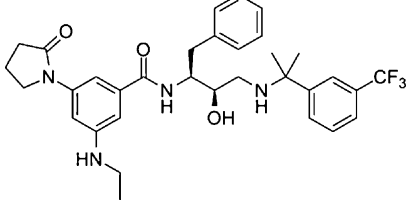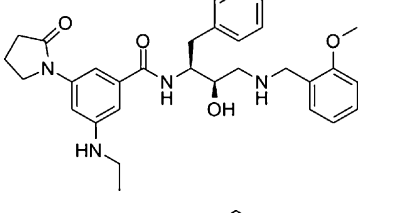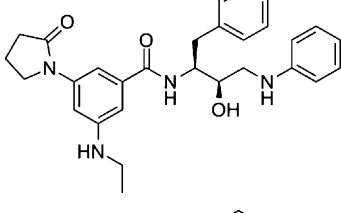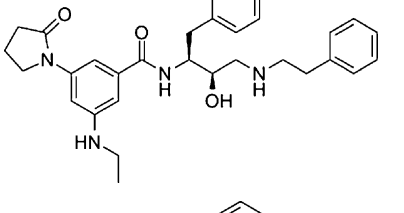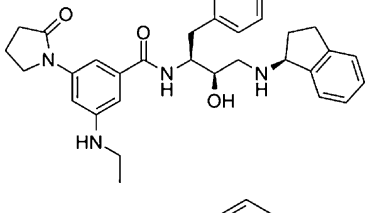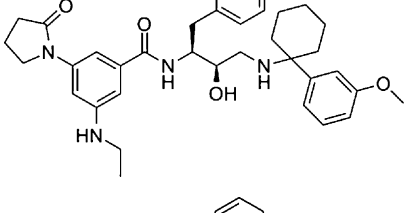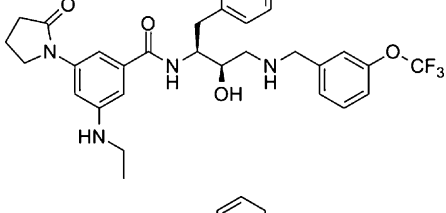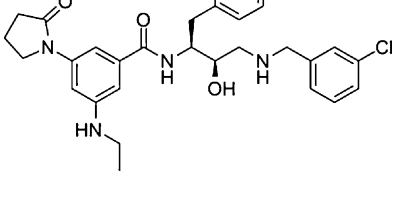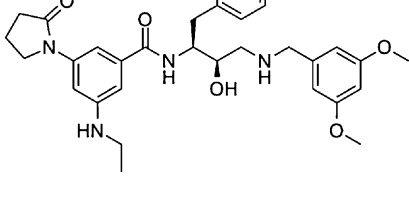
Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
17		7.48	18		6.52
19		6.10	20		7.40
21		7.37	22		7.77
23		5.87	24		5.77
25		6.22	26		6.80
27		7.85	28		7.70
29		7.08	30		7.00

Table 1 (*Continued*)
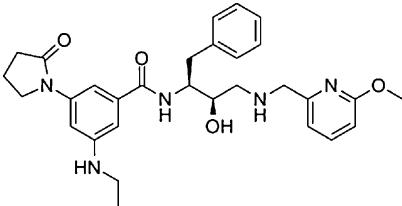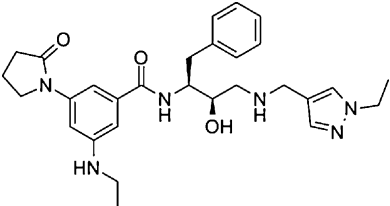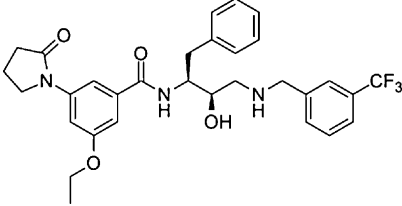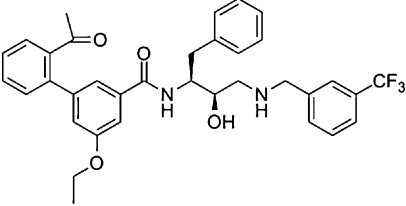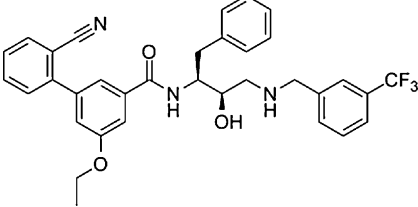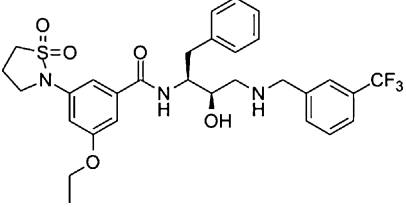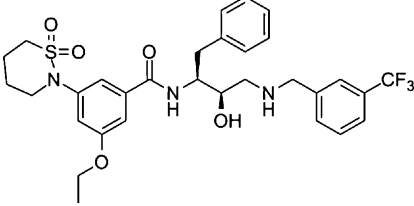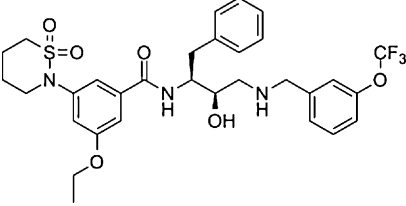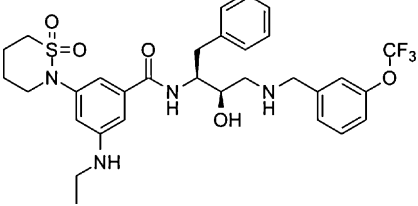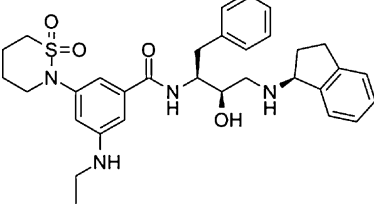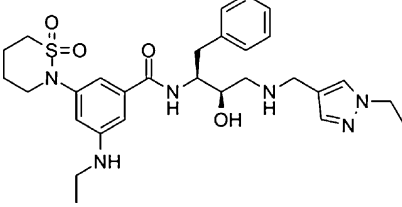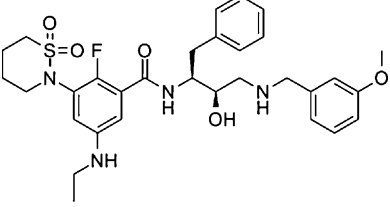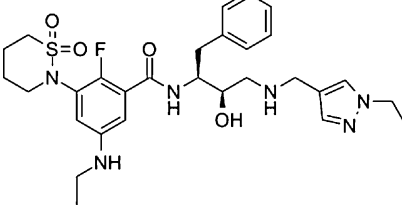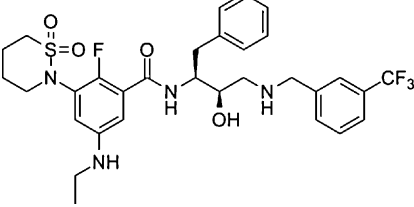
Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
31		6.48	32		6.64
33		6.42	34		6.55
35		6.80	36		7.30
37		8.22	38		8.30
39		8.30	40		7.47
41		7.92	42		8.40
43		8.22	44		8.40

Table 1 (*Continued*)
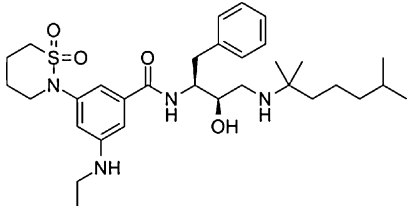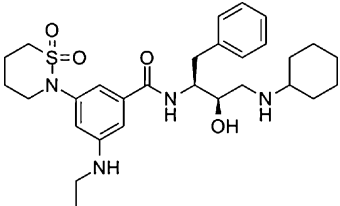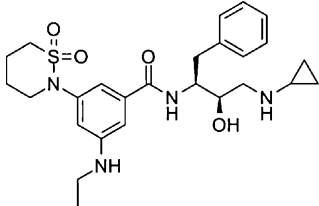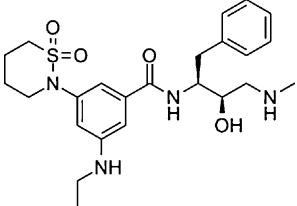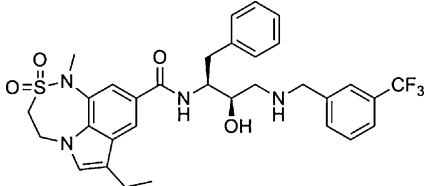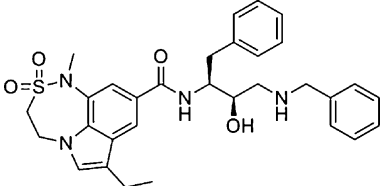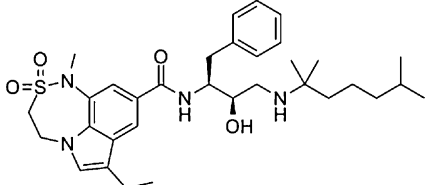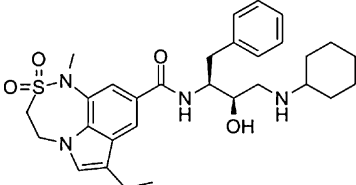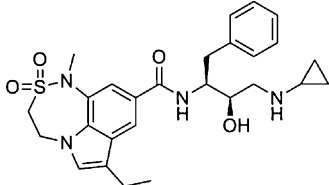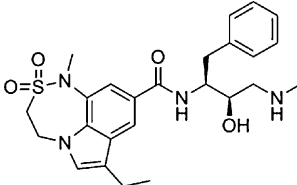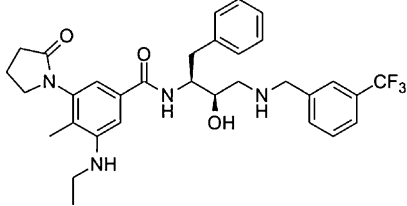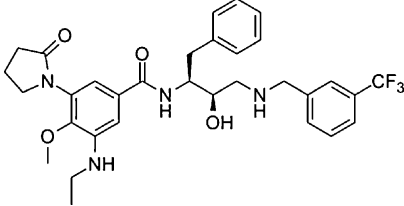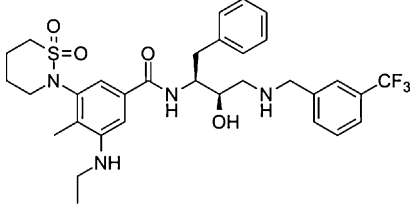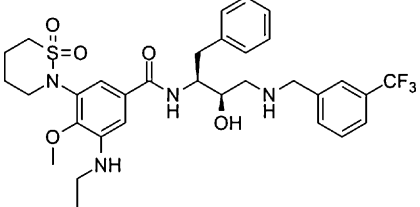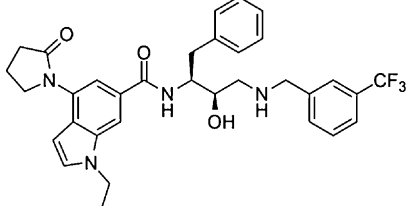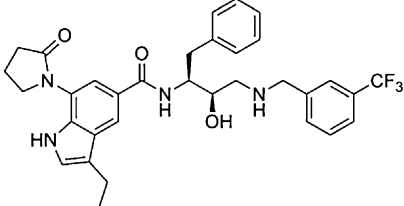
Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
45		7.52	46		7.16
47		6.92	48		6.85
49		8.70	50		7.47
51		8.30	52		8.05
53		7.70	54		7.82
55		7.20	56		6.96
57		7.64	58		7.74
59		7.64	60		7.59

Table 1 (Continued)
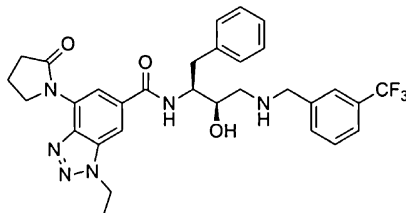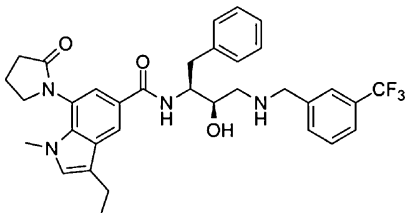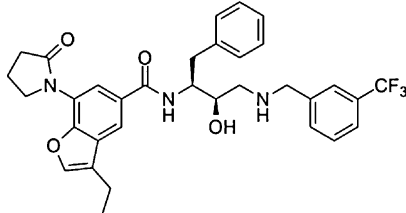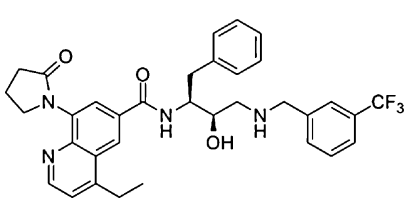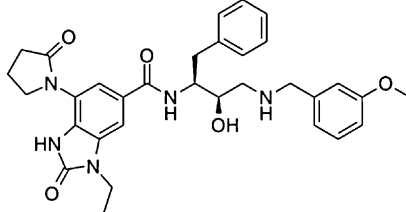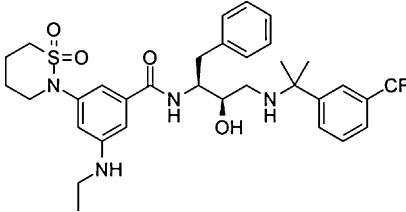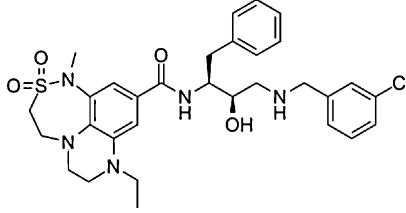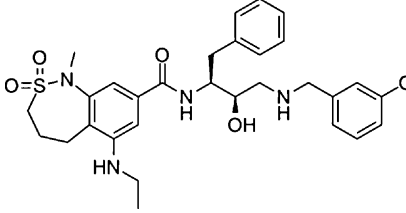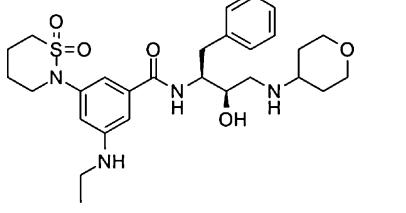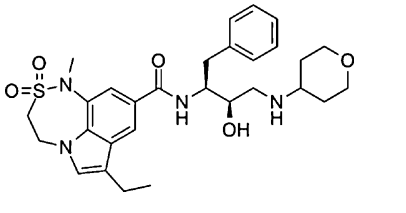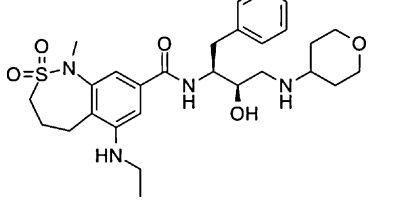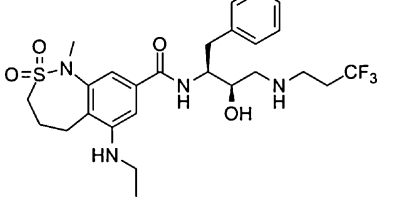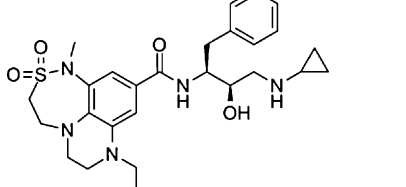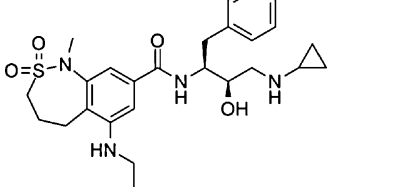
Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
61		7.47	62		7.40
63		7.06	64		6.68
65		6.38	66		8.22
67		7.59	68		7.72
69		7.04	70		7.82
71		6.72	72		7.59
73		6.26	74		6.82

Table 1 (Continued)
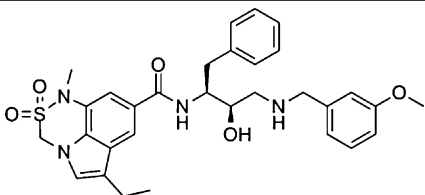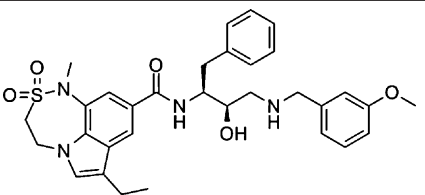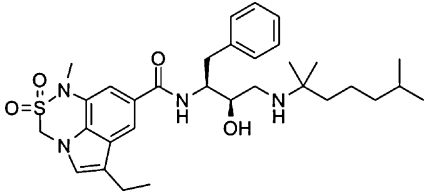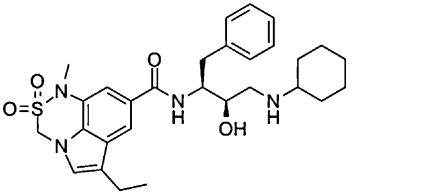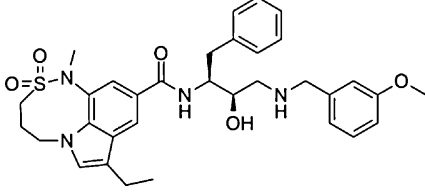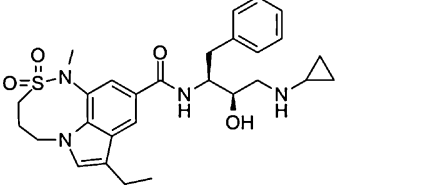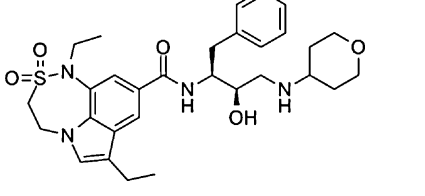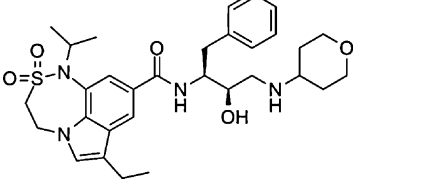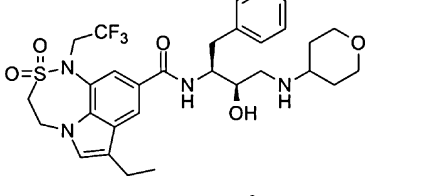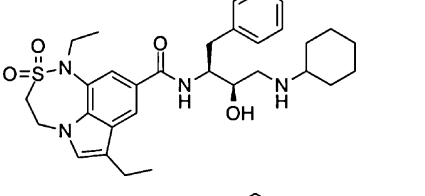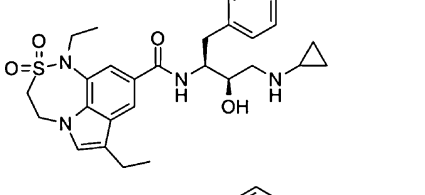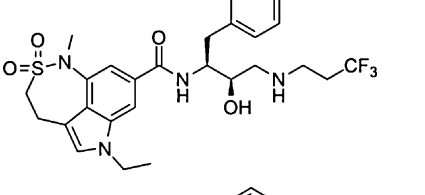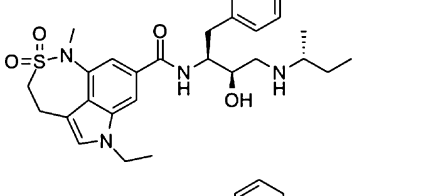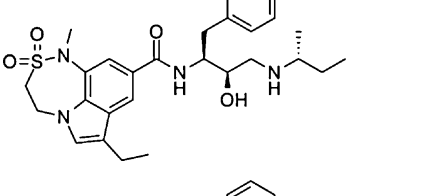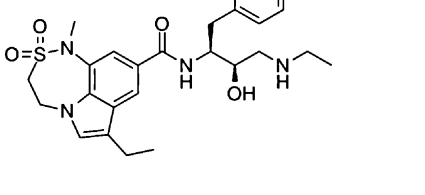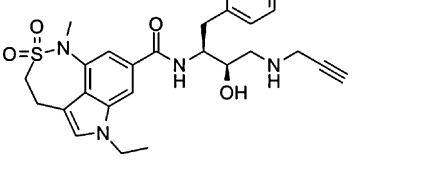
Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
75		8.10	76		8.70
77		7.64	78		7.22
79		7.55	80		6.31
81		8.30	82		7.64
83		7.54	84		8.15
85		7.70	86		8.40
87		8.00	88		7.66
89		7.32	90		7.96

Table 1 (*Continued*)

Training set

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
91		7.14	92		8.05
93		6.60	94		7.07
95		8.10	96		8.10
97		7.30	98		7.40
99		8.30			

Test set

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
100		5.27	101		7.23
102		6.74	103		5.95

Table 1 (*Continued*)
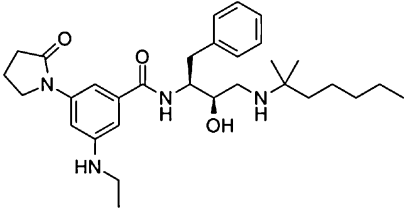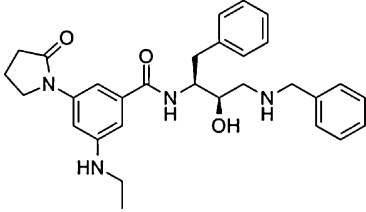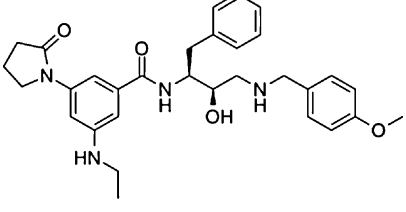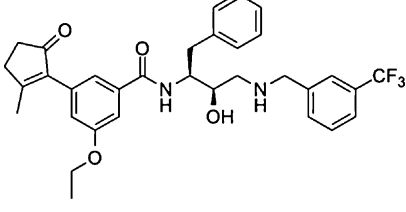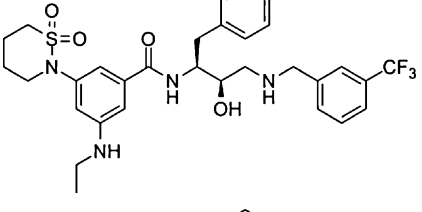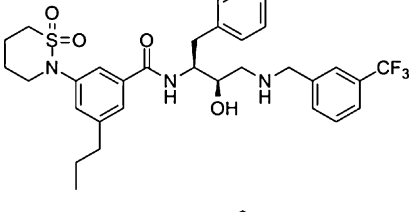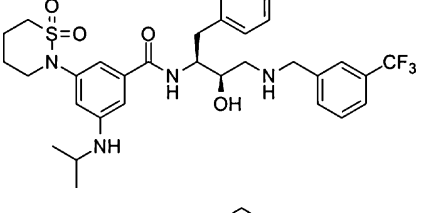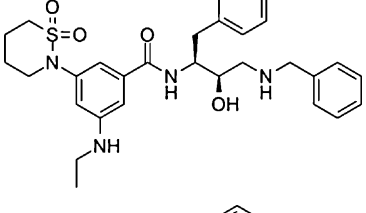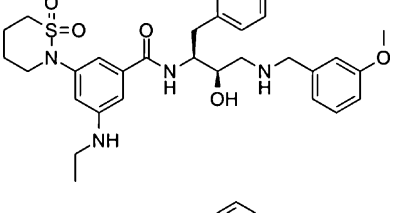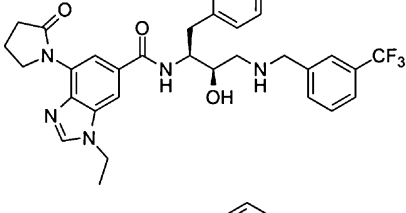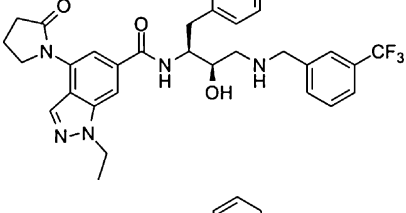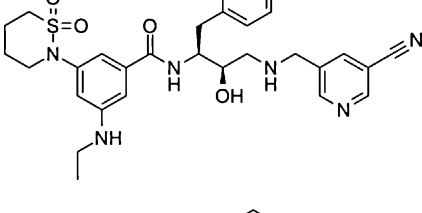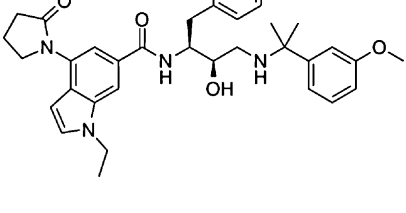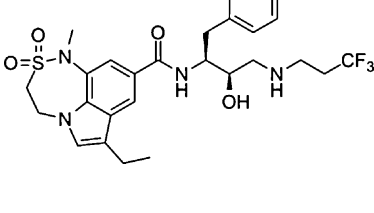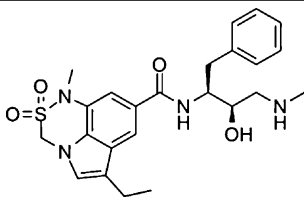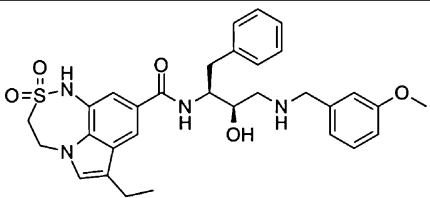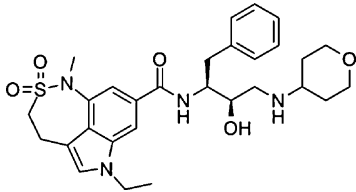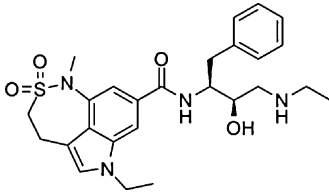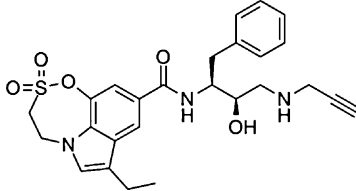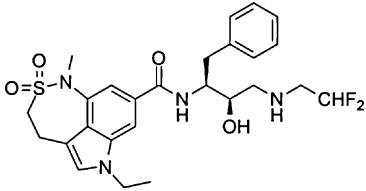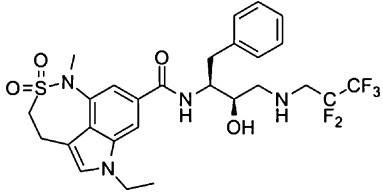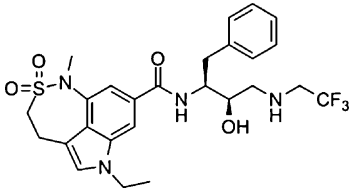
Test set

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
104		6.26	105		6.12
106		5.68	107		6.60
108		8.52	109		8.30
110		7.96	111		6.92
112		8.10	113		7.72
114		7.55	115		7.40
116		7.46	117		8.15

Table 1 (*Continued*)

Training set					
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
118		7.04	119		7.85
120		8.30	121		7.64
122		6.44	123		7.70
124		7.49	125		7.10

3. Results and discussion

3.1. Modeling results and evaluation

The cutting mode of the first sample of training set as a template is displayed in Fig. 2. This cutting mode can effectively guarantee the newly designed inhibitors have same core characteristics as the training samples; meanwhile, it can show the tendency how different groups affect the activity of the designed inhibitors.

A total of 99 training samples (1–99; Table 1) were used to construct a Topomer CoMFA 3D-QSAR model. The multiple correlation coefficient and standard error of fitting modeling, and leave one out cross validation are $R_{fm}^2 = 0.966$, $SE_{fm} = 0.16$, $Q_{cv}^2 = 0.767$ and
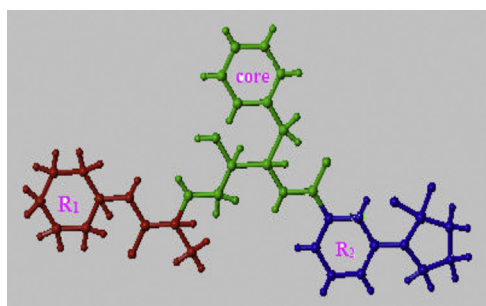
Fig. 2. The cutting style of 3D-QSAR model involving 99 BACE-1 inhibitors based on Topomer CoMFA. (The green, red and blue represent the core group, R$_1$ and R$_2$ in sample 1, respectively.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$SE_{cv} = 0.41$, respectively. The results show that the CoMFA model has good fitting and internal predictive ability. A total of 26 test samples (100–125; Table 1) were used to validate the predictive ability of the model; as a result, the multiple correlation coefficient and standard error of external validation are $Q_{ext}^2 = 0.784$ and $SE_{ext} = 0.53$, respectively, which effectively illustrates the external predictive ability of the mode is robust.

We also performed a Y-randomization response test to assess the robustness of the Topomer CoMFA model. The interceptions of the R_{fm}^2- and Q_{cv}^2-regression lines with the ordinate axis were −0.189 and −0.781, respectively, which are below the limits of $R_{fm}^2 < 0.300$ and $Q_{cv}^2 < 0.050$ [28]. This indicates that our model is not affected by any chance correlation. Fig. 3 displays the linear regression between experimental pIC$_{50}$ and predicted pIC$_{50}$ for training set and test set, indicating that the present model has favorable simulation ability, ant the accurate external predictions can also be achieved by this model.

In order to verify whether the relatively high correlation of the model resulted from the chance correlation by the group method for samples, we randomly split the 125 samples into 99 samples as training set and 26 samples as test set for ten other times. Table 2 lists that the statistical results of different Topomer CoMFA models based on the same cutting mode for R groups. It can be seen that the mean R_{fm}^2, Q_{cv}^2 and Q_{ext}^2 are 0.944, 0.751 and 0.646, respectively, showing that this dataset can be fit for modeling using the Topomer CoMFA method, and the present modeling results are not caused by the chance correlation derived from different group methods for samples. Thus, we can analyze QSAR of BACE-1 inhibitors, and further to design new inhibitors based on the present model (The No. 1 model; Table 2) as following:

Table 2
The statistical results of 10 Topomer CoMFA models.

No.	The number of principal components	R_{fm}^2[a]	SE_{fm}	Q_{cv}^2	SE_{cv}	Q_{ext}^2	SE_{ext}
1[b]	12	0.966	0.16	0.767	0.41	0.784	0.53
2	7	0.877	0.29	0.583	0.52	0.852	0.48
3	6	0.878	0.31	0.671	0.50	0.523	0.52
4	10	0.947	0.20	0.762	0.42	0.740	0.53
5	10	0.949	0.19	0.790	0.40	0.784	0.53
6	12	0.961	0.17	0.790	0.40	0.700	0.39
7	12	0.963	0.17	0.753	0.45	0.769	0.41
8	12	0.964	0.18	0.783	0.43	0.480	0.50
9	11	0.967	0.15	0.822	0.35	0.392	0.70
10	14	0.971	0.15	0.785	0.45	0.437	0.49
11[c]	–	0.944	0.20	0.751	0.43	0.646	0.51

[a] The R_{fm}^2, SE_{fm}, Q_{cv}^2, SE_{cv}, Q_{ext}^2 and SE_{ext} are for the multiple correlation coefficient and standard error of fitting modeling, leave one out cross validation, and external validation, respectively.
[b] The No.1 is the model which we used to analyze QSAR of BACE-1 inhibitors and further to design new inhibitors in this work.
[c] The No.11 is the mean statistical values for the No. 1–10 models.
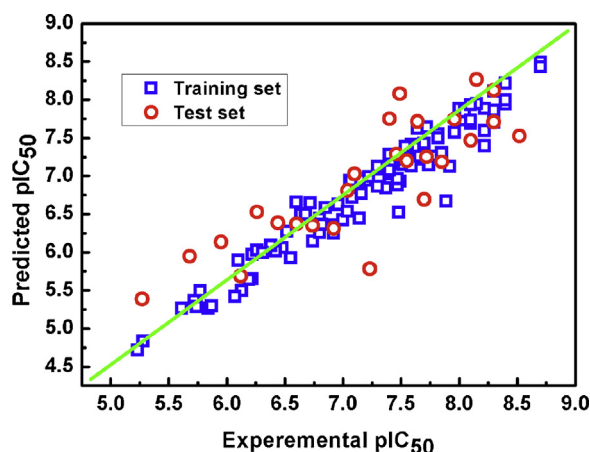
Fig. 3. The linear regression relationship between experimental and predicted activity for training set and test set. (The multiple correlation coefficients of fitting, leave one out cross validation, and external validation are 0.966, 0.767 and 0.784, respectively.)

3.2. QSAR analysis

We explored QSAR of inhibitors using colored contour maps, which was generated from a molecular field analysis when the Topomer CoMFA model is constructed. Contour maps can provide information about how to increase or decrease the biological activity of the molecules studied, and different colors help to correlate the diversified steric and electrostatic field with differences in the activity of the compounds.

The steric field contours are shown in yellow and green, while the electrostatic field contours are shown in red and blue (Fig. 4). In the steric field graph, the larger steric features for groups near the green region are, or the less the ones near the yellow region are, the higher the bioactivity is. In the electrostatic field graph, if the groups

with more electropositivity are connected the blue region, or the group with more electronegativity for the red region, the inhibitor will have higher activity. Compound **1** of training set is used *as* a template to depict the contour map of R_1-group (Fig. 4a and b) since the R_1 structure of this compound is simpler than that of any one of training set; thereby, the QSAR analysis is easily carried out based on its relatively simple structure. Compound **21** is employed to show the contour map of R_2-group (Fig. 4c and d) based on the similarity of chemical structure.

It can be seen from the steric field contour for R_1 (Fig. 4a) there is a large green polyhedron-like region around C_3 and C_5 of the benzene ring. The steric field contour for R_2 (Fig. 4c) shows there is a large green polyhedron-like region around γ-lactam group together with C_1–C_3 of the benzene ring. These observations suggest the substituent groups near these green regions are positive to the activity of the compounds, e.g. by comparing chemical structures and pIC_{50} values of compounds **4**, **6** and **8** (Table 1), we find the larger the volume of hydrocarbonyls connecting with the oxygen atom is, the higher the activity of the compound is. Further, the structure and activity relationship for compounds **17** and **18**, compounds **20** and **22** strengthens this conclusion.

Similarly, by comparing compound **75** with compound **76**, we find the former has a large R_1 group with a 6-ring around the region of C_3 and C_4, while the latter has a 7-ring, whose volume is larger than that of the 6-ring, which result in a higher activity value for compound **76**. This illustrates that increasing the size of the substituent at this position may increase the activity of compounds. Moreover, compounds **36** and **37**, and compounds **13** and **14** are also fit for this case. Besides, yellow regions distribute around C_4 of the ring in Fig. 4a and C_4–C_6 of the ring in Fig. 4c), indicating that the substituent with less volume in these regions are favorable for the improvement of molecular activity.

Except for the influence of the steric field, the potency of inhibitors is closely relevant to the electrostatic characteristic of the substituent groups. The electrostatic contour map describes the relationship between electrical properties of substituents and
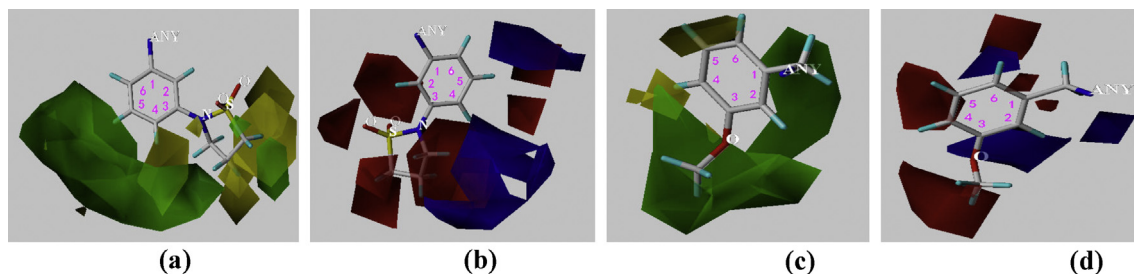
Fig. 4. Three-dimensional contour plot of the steric field map (a and c) and electrostatic field map (b and d).

bioactivities of the corresponding compounds. The potency is higher if more positive charge is near the blue regions and more negative charge is near the red regions. The electrostatic field contour for R_1 group (Fig. 4b) shows the red region stretches around C_2, C_3 and C_5 of the ring, and a small blue region curls up around the C_4 of the ring. And the electrostatic field contour for R_2 group (Fig. 4d) displays the red region occupy the space around C_1, C_4 and C_5 of the ring, and the blue distribute about C_2, C_3 and C_6 of the ring.

Some information is obtained by analysis of electrostatic field contours. An intriguing phenomenon is that compounds **90** and **91** are isomers, but their bioactivity differs by an order of magnitude. By comparison of the two compounds, we find a nitrogen atom attaches to C_5 of the benzene ring of compound **90**, and to C_4 of the benzene ring of compound **91**, respectively. As we all know, the nitrogen atom has 4 sp^3 hybrid orbitals, three of which carry a single electron, thus the nitrogen atom has strong absorption ability for electrons. Fig. 4b indicates the group with positive charge in C_4 of the ring and the ones with negative charge in C_5 of the ring are favorable for the improvement of bioactivities of compounds. Therefore, a nitrogen atom replacing hydrogen atom on C_5 contributes to the improved potency for compound **90**. Moreover, compounds **92** and **93**, compounds **97** and **98** also match this tendency above, respectively.

3.3. Binding mode of BACE-1 inhibitors

We explored the binding mode of BACE-1 inhibitors using molecular docking, which is a crucial tool for structural molecular biology and computer-assisted drug design, and is able to propose structural hypotheses of how the ligand inhibits the target, employing a scoring function to correctly ranks candidate dockings. During this procedure, the protomol (Fig. 5) is generated when the binding pocket is full of 3 molecular probes, i.e., N—H, CH$_4$ and C=O, which represent hydrogen bond donor, hydrogen bond receptor, and hydrophobic sites, respectively. Then, the structure of the ligand (Fig. S1) matches with that of the protomol; meanwhile, the poses are evaluated using scoring function.

The total score containing crash score and polar score is total scoring function for the output pose of the molecules. In general, the output pose qualifies when total score is more than 6.0. Empirical scoring functions includes *D*-score [29], *PMF*-score [30], *G*-score [31], *CHEM*-score [32] and *C*-score [27], of these, the *C*-score is an integration of other scoring functions. Each scoring function has its own limitations. Research shows that, compared with the single scoring function, the use of *C*-score can greatly improve the success rate of the screening [33]. Generally, the total *C*-score is 5.0, and the larger the *C*-score is, the more the possibility for the actual pose of the molecule studied is. A high *C*-score shows that the docking result is in line with the requirements of pose of the molecule studied. In this work, we used the *C*-score method to screen the

optimal pose, combining with a relatively large total score on the pose of each molecule.

We conducted molecular docking for the training ligands and BACE-1, to obtain a total of 198 output poses with full mark of *C*-score, the first 50 of which were extracted according to the descending order of total score for each pose. The detailed docking results are shown in Table S1 in Supplementary data. A comprehensive analysis reveals that there are some hydrogen bonds between the ligands and some key groups of amino acid residues of the protein receptor, e.g. ASP93, GLY95, THR133, GLN134, ASP289, GLY291, THR292, THR293, ASN294, ARG296 and SER386 of BACE-1.

Fig. 6 shows binding mode of inhibitors with BACE-1. Fig. 6a–c represent docking results of the first pose of compound **11**, the first and fifth pose of compound **101**, respectively, and dotted yellow lines indicate hydrogen bonds. It can be seen from Fig. 6a that the first pose of compound **11** interacts with 5 sites of the protein receptor by hydrogen bonds. The 5 sites involve with the oxygen atoms in the side chain of ASP289, GLY291, THR292 and THR293, and the amino group of THR133. By comparison, we find different pose revealed different mode of action, e.g. the structures for the ligands shown in Fig. 6a and b differ from each other; while the different poses in Fig. 6b and c is from the same molecule (compound **101**).

3.4. Theoretical design of new inhibitors with higher inhibitory activity

We used the screened R-groups to construct new molecules with higher inhibitory activity. The contribution values of R-groups
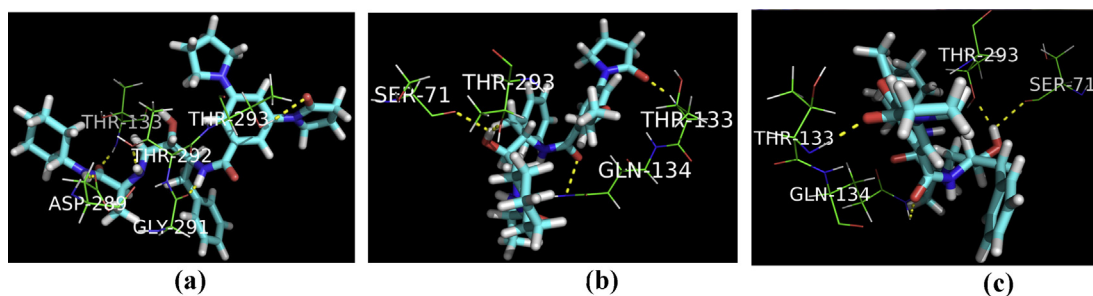
Fig. 6. The mode of action of the training inhibitors (a–c represent docking results of the first pose of compound **11**, the first and fifth pose of compound **101** with 2VIZ, respectively. And dotted yellow lines indicate hydrogen bonds). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Fifty newly designed molecules.
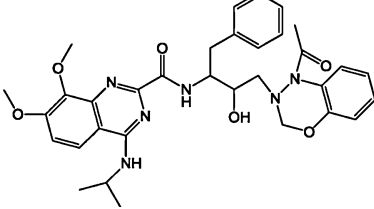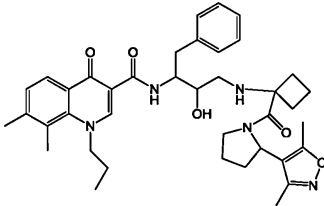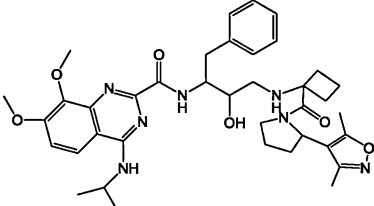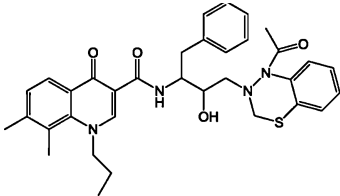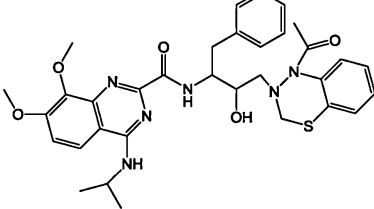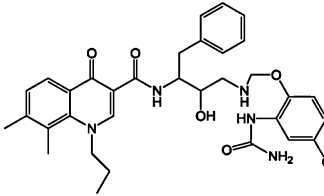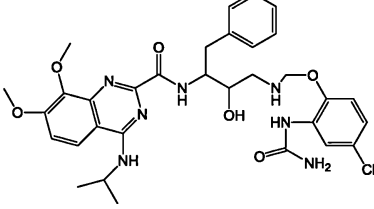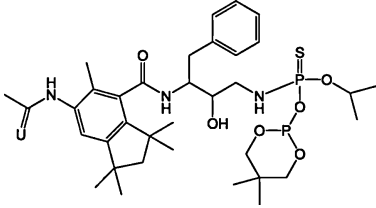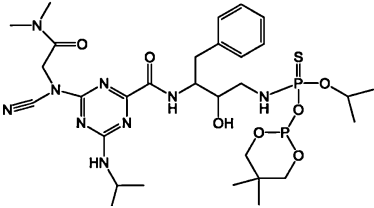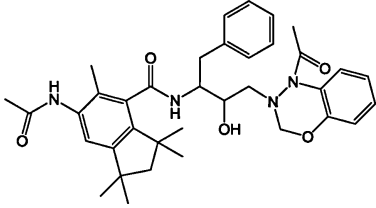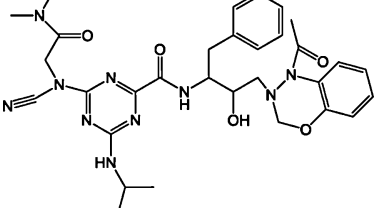
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
01[a]		11.12	26[a]		10.46
02		8.49	27		7.82
03[a]		10.73	28[a]		10.06
04		8.44	29		7.77
05[a]		10.47	30[a]		9.81
06[a]		10.35	31[a]		10.51
07		7.72	32		7.88

Table 3 (*Continued*)
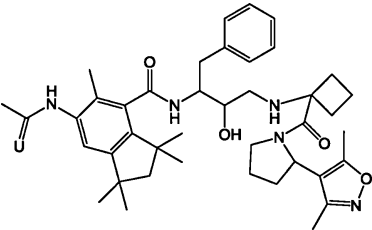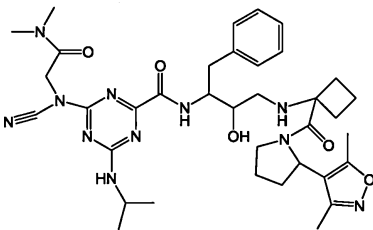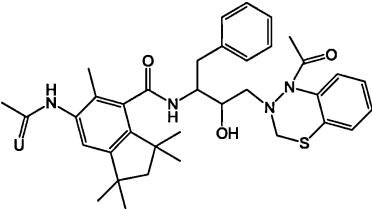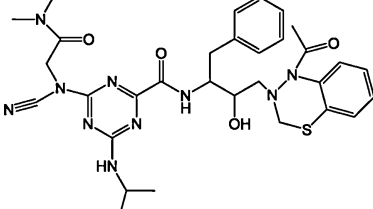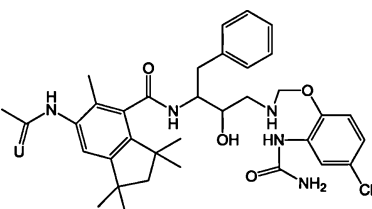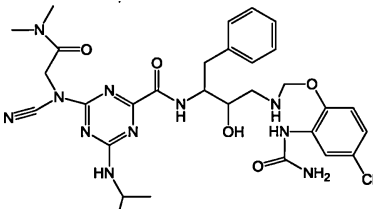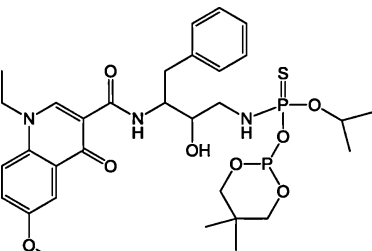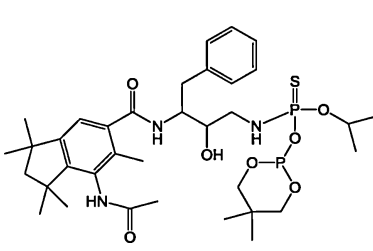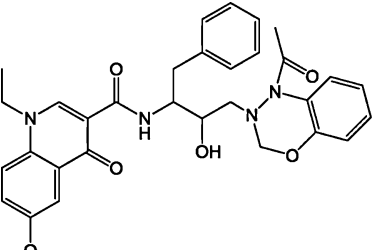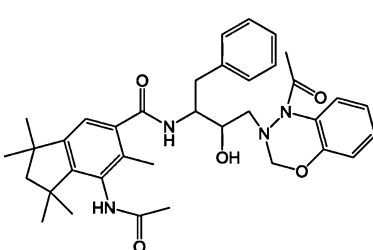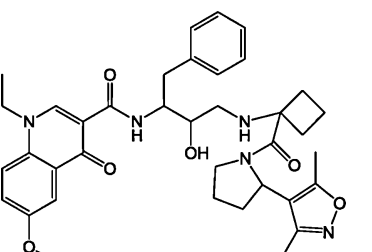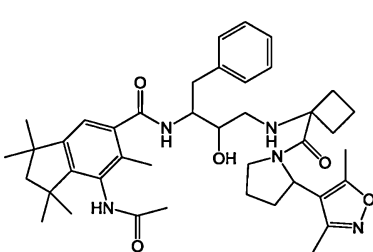
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
08[a]		9.96	33[a]		10.12
09		7.66	34		7.82
10[a]		9.70	35[a]		9.86
11[a]		10.65	36[a]		10.25
12		8.02	37		7.62
13[a]		10.26	38[a]		9.86

Table 3 (*Continued*)

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
14		7.97	39		7.57
15[a]		9.70	40[a]		9.60
16[a]		10.62	41[a]		10.31
17		7.99	42		7.68
18[a]		10.20	43[a]		9.92
19		7.94	44		7.62
20[a]		9.98	45[a]		9.66

Table 3 (*Continued*)

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
21[a]		10.70	46[a]		10.33
22		8.05	47		7.70
23[a]		10.26	48[a]		9.94
24		7.99	49		7.64
25[a]		10.03	50[a]		9.68

[a] Molecules marked with "a" are those with higher predicted activity than that of all inhibitors in training set.

were determined according to the scores predicted by the Topomer CoMFA model. In general, the contribution values played a decisive role in the procedure of selecting R-groups to conduct new molecules in the same scope of the Topomer distance. Different molecules with specific activity were obtained according to the contribution values of the R groups, together with the core of the template compound. Thus, the new molecules have same core with that of the known training compounds, in order to guarantee the category consistency between the newly designed molecules and the training compounds, and have different R-groups to ensure the improvement of bioactivity of inhibitors.

In this study, we virtually screened R$_1$ and R$_2$ groups in the 534,597 lead-like molecules of the ZINC database, respectively. We treated as compound **76** of the training set as a template to extract the top 10 R$_1$-groups and top 5 R$_2$-groups, and further to design 50 new molecules using the rule of permutation and combination (Table 3). And the structures of the newly designed molecules were optimized by adding hydrogen and suitable electric charge

in consistency with what was described previously in establishing Topomer CoMFA model section. The pIC$_{50}$ value for each new molecule was predicted by the QSAR model; finally, we obtained 30 molecules with higher activity than that of all inhibitors in the training set (Table 3).

We comparatively explored binding mode for the newly designed molecules with higher bioactivity and the training molecules using molecular docking. The newly designed 30 molecules with higher activity were docked to the active sites of BACE-1; as a result, 18 qualified poses were obtained according to the scoring functions. We displayed mode of action between small molecular ligand and protein receptor in Table S2 in Supplementary data, in which hydrogen bonds are indicated by yellow dotted lines, and different atoms are displayed by different colors.

Fig. 7 displays the docking results for the newly designed molecule (No. 21) with BACE-1. It can be seen that the obvious hydrogen bonds are formed between the ligand and the oxygen atoms in the side chain of ASP93 and THR292, and the α-amino of
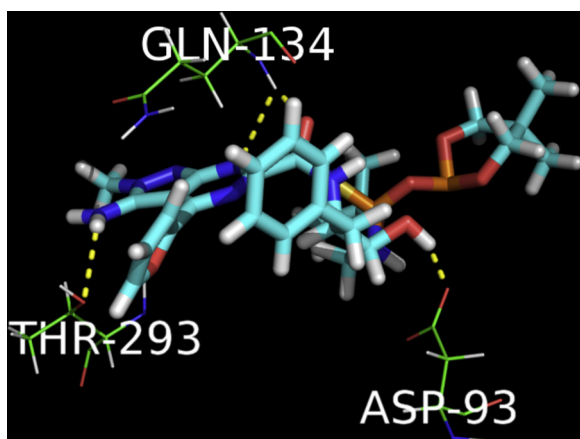
Fig. 7. Docking results of the newly designed compound **21** with 2VIZ (Dotted yellow lines indicate hydrogen bonds). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
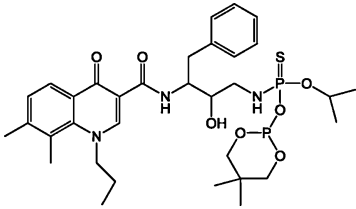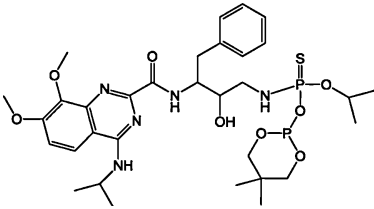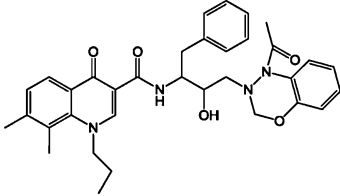
GLN134 of BACE-1. Besides, the other 17 posed pictures for binding mode demonstrates that several hydrogen bonds can be formed between the small ligand and some key sites of BACE-1, e.g. the oxygen atoms in side chain of ASP93, GLY95, ASP289, GLY291 and THR292, the nitrogen atoms in the main chain of THR293, ASN294 and ARG296, the α-amino group of THR133, the α-carboxyl group of TYR132, and the α-amino group of GLN134, etc.

Those intermolecular hydrogen bonds make for the inhibitors to compete with other molecules which make BACE-1 effective, thereby, to inhibit the biological effects of BACE-1. These key sites are agreement with those derived from exploration of binding mode for the training molecules using molecular docking. Therefore, it is indispensable to study the characteristics of GLN134, THR292, THR293 and ASN294 in the crystal structure of BACE-1 in drug design of BACE-1 inhibitors, which was also conformed by the literature [34]. In addition, the roles of the sites, including ASP93, GLY95, THR133, ASP289, GLY291 and ARG296 of BACE-1, should not be ignored.

4. Conclusions

A Topomer CoMFA 3D-QSAR model with good internal and external prediction capability was established for a training set of 99 BACE-1 enzyme inhibitors, and a test set of 26 molecules were employed to validate the external predictive ability of the model obtained. Topomer search, was used to screen R-groups in lead-like compounds of ZINC molecule database, and 30 new molecules with higher bioactivity were successfully designed based on the rule of permutations and combinations. Surflex-dock was employed to comparatively explore mechanism of the inhibitors when acting chemically with BACE-1 enzyme. The results show the inhibitors interact with the sites of ASP93, GLY95, THR133, GLN134, ASP289, GLY291, THR292, THR293, ASN294, ARG296 and SER386 of BACE-1. The present work provides references to drug design for BACE-1 inhibitors. And for the future research, on the one hand, we should explore the interaction between the core and R-groups of inhibitors. On the other hand, it is necessary to verify the activity of the newly designed molecules combining with pharmaceutical chemistry experiments.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm. 2013.08.003.

References

[1] C. Mount, C. Downton, Alzheimer disease: progress or profit? Nat. Med. 12 (7) (2006) 780–784.

[2] I. Melnikova, Therapies for Alzheimer's disease, Nat. Rev. Drug Discov. 6 (5) (2007) 341–342.

[3] E. Mandelkow, Alzheimer's disease: the tangled tale of tau, Nature 402 (6762) (1999) 588–589.

[4] F.S. Esch, P.S. Keim, E.C. Beattie, R.W. Blacher, A.R. Culwell, T. Oltersdorf, D. McClure, P.J. Ward, Cleavage of amyloid beta peptide during constitutive processing of its precursor, Science (New York, NY) 248 (4959) (1990) 1122.

[5] D.J. Selkoe, Translating cell biology into therapeutic advances in Alzheimer's disease, Nature 399 (1999) A23–A31.

[6] D.J. Selkoe, Alzheimer's disease: genes, proteins, and therapy, Physiol. Rev. 81 (2) (2001) 741–766.

[7] C. Dingwall, Spotlight on BACE: the secretases as targets for treatment in Alzheimer disease, J. Clin. Invest. 108 (9) (2001) 1243–1246.

[8] A.K. Ghosh, M. Brindisi, J. Tang, Developing beta-secretase inhibitors for treatment of Alzheimer's disease, J. Neurochem. 120 (Suppl 1) (2012) 71–83.

[9] Y. Luo, B. Bolon, S. Kahn, B.D. Bennett, S. Babu-Khan, P. Denis, W. Fan, H. Kha, J. Zhang, Y. Gong, L. Martin, J.C. Louis, Q. Yan, W.G. Richards, M. Citron, R. Vassar, Mice deficient in BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation, Nat. Neurosci. 4 (3) (2001) 231–232.

[10] S.L. Roberds, J. Anderson, G. Basi, M.J. Bienkowski, D.G. Branstetter, K.S. Chen, S.B. Freedman, N.L. Frigon, D. Games, K. Hu, K. Johnson-Wood, K.E. Kappenman, T.T. Kawabe, I. Kola, R. Kuehn, M. Lee, W. Liu, R. Motter, N.F. Nichols, M. Power, D.W. Robertson, D. Schenk, M. Schoor, G.M. Shopp, M.E. Shuck, S. Sinha, K.A. Svensson, G. Tatsuno, H. Tintrup, J. Wijsman, S. Wright, L. McConlogue, BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics, Hum. Mol. Genet. 10 (12) (2001) 1317–1324.

[11] R. Silvestri, Boom in the development of non-peptidic β-secretase (BACE1) inhibitors for the treatment of Alzheimer's disease, Med. Res. Rev. 29 (2) (2009) 295–338.

[12] K.C. Chou, Structural bioinformatics and its impact to biomedical science, Curr. Med. Chem. 11 (16) (2004) 2105–2134.

[13] K.C. Chou, D.Q. Wei, Q.S. Du, S. Sirois, W.Z. Zhong, Progress in computational approach to drug development against SARS, Curr. Med. Chem. 13 (27) (2006) 3263–3270.

[14] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (18) (1988) 5959–5967.

[15] R.D. Cramer, Topomer CoMFA: a design methodology for rapid lead optimization, J. Med. Chem. 46 (3) (2003) 374–388.

[16] R.D. Cramer, R.J. Jilek, S. Guessregen, S.J. Clark, B. Wendt, R.D. Clark, "Lead hopping". Validation of topomer similarity as a superior predictor of similar biological activities, J. Med. Chem. 47 (27) (2004) 6777–6791.

[17] B. Clarke, E. Demont, C. Dingwall, R. Dunsdon, A. Faller, J. Hawkins, I. Hussain, D. MacPherson, G. Maile, R. Matico, P. Milner, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, D. Riddell, P. Rowland, V. Soleil, K.J. Smith, S. Stanway, G. Stemp, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, J. Ward, G. Wayne, BACE-1 inhibitors part 1: identification of novel hydroxy ethylamines (HEAs), Bioorg. Med. Chem. Lett. 18 (3) (2008) 1011–1016.

[18] B. Clarke, E. Demont, C. Dingwall, R. Dunsdon, A. Faller, J. Hawkins, I. Hussain, D. MacPherson, G. Maile, R. Matico, P. Milner, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, D. Riddell, P. Rowland, V. Soleil, K.J. Smith, S. Stanway, G. Stemp, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, J. Ward, G. Wayne, BACE-1 inhibitors part 2: identification of hydroxy ethylamines (HEAs) with reduced peptidic character, Bioorg. Med. Chem. Lett. 18 (3) (2008) 1017–1021.

[19] P. Beswick, N. Charrier, B. Clarke, E. Demont, C. Dingwall, R. Dunsdon, A. Faller, R. Gleave, J. Hawkins, I. Hussain, C.N. Johnson, D. MacPherson, G. Maile, R. Matico, P. Milner, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, D. Riddell, P. Rowland, J. Skidmore, V. Soleil, K.J. Smith, S. Stanway, G. Stemp, A. Stuart, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, J. Ward, G. Wayne, BACE-1 inhibitors part 3: identification of hydroxy ethylamines (HEAs) with nanomolar potency in cells, Bioorg. Med. Chem. Lett. 18 (3) (2008) 1022–1026.

[20] N. Charrier, B. Clarke, L. Cutler, E. Demont, C. Dingwall, R. Dunsdon, P. East, J. Hawkins, C. Howes, I. Hussain, P. Jeffrey, G. Maile, R. Matico, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, P. Rowland, V. Soleil, K.J. Smith, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, G. Wayne, Second generation of hydroxyethylamine BACE-1 inhibitors: optimizing potency and oral bioavailability, J. Med. Chem. 51 (11) (2008) 3313–3317.

[21] N. Charrier, B. Clarke, L. Cutler, E. Demont, C. Dingwall, R. Dunsdon, J. Hawkins, C. Howes, J. Hubbard, I. Hussain, G. Maile, R. Matico, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, P. Rowland, V. Soleil, K.J. Smith, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, G. Wayne, Second generation of BACE-1 inhibitors. Part 1: the need for improved pharmacokinetics, Bioorg. Med. Chem. Lett. 19 (13) (2009) 3664–3668.

[22] N. Charrier, B. Clarke, E. Demont, C. Dingwall, R. Dunsdon, J. Hawkins, J. Hubbard, I. Hussain, G. Maile, R. Matico, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, P. Rowland, V. Soleil, K.J. Smith, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, G. Wayne, Second generation of BACE-1 inhibitors part 2: optimisation of the non-prime side substituent, Bioorg. Med. Chem. Lett. 19 (13) (2009) 3669–3673.

[23] N. Charrier, B. Clarke, L. Cutler, E. Demont, C. Dingwall, R. Dunsdon, J. Hawkins, C. Howes, J. Hubbard, I. Hussain, G. Maile, R. Matico, J. Mosley, A. Naylor, A. O'Brien, S. Redshaw, P. Rowland, V. Soleil, K.J. Smith, S. Sweitzer, P. Theobald, D. Vesey, D.S. Walter, G. Wayne, Second generation of BACE-1 inhibitors part 3: towards non hydroxyethylamine transition state mimetics, Bioorg. Med. Chem. Lett. 19 (13) (2009) 3674–3678.

[24] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2) (2001) 109–130.

[25] R.D. Cramer, F. Soltanshahi, R. Jilek, B. Campbell, AllChem: generating and searching 10 20 synthetically accessible structures, J. Comput. Aided Mol. Des. 21 (6) (2007) 341–350.

[26] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, ZINC: a free tool to discover chemistry for biology, J. Chem. Inf. Model. 52 (7) (2012) 1757–1768.

[27] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew, Consensus scoring for ligand/protein interactions, J. Mol. Graph. Model. 20 (4) (2002) 281–295.

[28] P.M. Andersson, M. Sjöström, T. Lundstedt, Preprocessing peptide sequences for multivariate sequence-property analysis, Chemom. Intell. Lab. Syst. 42 (1) (1998) 41–50.

[29] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule-ligand interactions, J. Mol. Biol. 161 (2) (1982) 269–288.

[30] I. Muegge, Y.C. Martin, A general and fast scoring function for protein-ligand interactions: a simplified potential approach, J. Med. Chem. 42 (5) (1999) 791–804.

[31] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, J. Mol. Biol. 245 (1) (1995) 43–53.

[32] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, J. Comput. Aided Mol. Des. 11 (5) (1997) 425–445.

[33] P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters, Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins, J. Med. Chem. 42 (25) (1999) 5100–5109.

[34] L.B. Salum, N.F. Valadares, Fragment-guided approach to incorporating structural information into a CoMFA study: BACE-1 as an example, J. Comput. Aided Mol. Des. 24 (10) (2010) 803–817.