MEDICINAL
CHEMISTRY
RESEARCH   CrossMark

ORIGINAL RESEARCH

Pharmacophore modeling, virtual screening, and 3D-QSAR studies on a series of non-steroidal aromatase inhibitors

Huiding Xie · Kaixiong Qiu · Xiaoguang Xie

Abstract Aromatase inhibitors are the most important targets in treatment of estrogen-dependent cancers. In order to search for potent non-steroidal aromatase inhibitors (NSAIs) with lower side effects and overcome cellular resistance, Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Database was used to derive 3D pharmacophore models. The obtained best pharmacophore model contains one acceptor atom, one donor atom, and two hydrophobes, which was used in effective alignment of dataset. In succession, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed on 84 structurally diverse NSAIs to build 3D-QSAR models based on both pharmacophore and docking alignments. The CoMFA and CoMSIA models based on the pharmacophore alignment show better statistical results (CoMFA: $q^2 = 0.634$, $r^2_{ncv} = 0.986$, $r^2_{pred} = 0.737$; CoMSIA: $q^2 = 0.668$, $r^2_{ncv} = 0.926$, $r^2_{pred} = 0.708$). This 3D-QSAR approach provides significant insights that can be used to develop novel and potent NSAIs. In addition, the best pharmacophore model was used as a 3D query for virtual screening against NCI2000 database. The hit compounds were further filtered by docking, and their biological activities were predicted by the CoMFA and CoMSIA models, and six structurally diverse compounds with good predicted pIC$_{50}$ values were obtained, which are expected to design novel NSAIs with new skeletons.

Keywords Non-steroidal aromatase inhibitors · Pharmacophore · 3D-QSAR · CoMFA · CoMSIA · Virtual screening

Introduction

Aromatase is a cytochrome P-450 dependent enzyme that catalyzes the aromatization of androgens to estrogens. Aromatase inhibitors (AIs) reduce the synthesis of estrogens and offer a therapeutic alternative for the treatment of estrogen-dependent cancers such as breast cancer (Winer *et al.*, 2005; Perez, 2006; Jordan and Brodie, 2007). There are two classes of AIs, steroidal and non-steroidal compounds, which cause potent estrogen suppression (Brueggemeier *et al.*, 2005). The non-steroidal aromatase inhibitors (NSAIs) are mostly azole type compounds such as the clinically used anastrozole and letrozole, which compete with the substrate for binding to the enzyme active site (Recanatini *et al.*, 2002). Among steroidal aromatase inhibitors (SAIs), formestane was widely used during the early 1990s, but it is not used nowadays because of the need to administer it by intramuscular injection. Therefore, the orally active exemestane is the main steroidal inhibitor (Seralini and Moslemi, 2001). These SAIs mimic the natural substrate androstenedione and are converted by the enzyme to reactive intermediates, which bind irreversibly to the enzyme active site, resulting in inactivation of aromatase (Hong *et al.*, 2011). Despite the success of the third-generation NSAIs (anastrozole and letrozole) and SAIs

H. Xie (✉) · X. Xie (✉)
Department of Chemistry, Yunnan University, Kunming 650091, Yunnan, People's Republic of China
e-mail: front701228@gmail.com

X. Xie
e-mail: xgxie@ynu.edu.cn

H. Xie · K. Qiu
Department of Chemistry, School of Pharmaceutical Science & Yunnan Key Laboratory of Pharmacology for Natural Products, Kunming Medical University, Kunming 650500, Yunnan, People's Republic of China

(exemestane), they still have some major side effects, such as increase of bone loss, joint pain, and heart problems (Dutta and Pant, 2008). In addition, after some years of usage they can develop cellular resistance. For these reasons, it is important to search for other potent and specific molecules with lower side effects and which can overcome the resistance phenomenon.

Pharmacophore searches are the best option to find a range of chemical structures with viable features. A pharmacophore model can be considered as the ensemble of steric and electrostatic features of different compounds which are necessary to ensure optimal supramolecular interactions with a specific biological target structure and to trigger or to block its biological response. Thus, pharmacophore modeling is the method of choice for the first round of compound selection. This ability of a pharmacophore model is used to find new classes of inhibitors when one class is known. This is known as 'scaffold hopping' (Bhatt and Patel, 2012).

Three-dimensional quantitative structure–activity relationship (3D-QSAR) methods have been successfully employed to assist the design of new small molecule drug candidates (Honorio et al., 2007; Salum et al., 2007). Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) are two of the most widely used 3D-QSAR methodologies. CoMFA calculates the energies of steric and electrostatic interactions between the compound and the probe atom at various intersections of a regular 3D lattice according to Lennard-Jones and Coulomb potentials. The resulting energies derived from these two potential functions can be contoured to offer a quantitative spatial description of the molecular properties (Cramer et al., 1988). CoMSIA introduces the Gaussian function for the distance dependence between the molecular atoms and the probe atom in order to avoid some inherent deficiencies arising from the Lennard-Jones and Coulomb potential functional forms. CoMSIA is applied to gain an insight into how steric fields, electrostatic fields, hydrophobic fields, hydrogen bond donor (HBD), and hydrogen bond acceptor (HBA) influence the activity of inhibitors (Klebe et al., 1994).

Several structurally diverse NSAIs such as resveratrol analogs, isoflavanone derivatives, and tetrahydropyrroloquinolinone type compounds (Table 1), have been reported in the recent literatures (Sun et al., 2010; Bonfield et al., 2012; Hu et al., 2012; Yin et al., 2013). To understand the structural basis for inhibitory activity and design more potent agents, pharmacophore models were created and 3D-QSAR studies were performed for the fist time for these structurally diverse NSAIs using CoMFA and CoMSIA based on both pharmacophore and docking alignments. In addition, the obtained best pharmacophore model was used as a 3D query for virtual screening against NCI2000 database. The hit compounds were further filtered by docking, and their biological activities were predicted using CoMFA and CoMSIA models.

Materials and methods

Dataset

Compounds **1–26** (resveratrol analogs) (Sun et al., 2010), compounds **27–45**, (isoflavanone derivatives) (Bonfield et al., 2012), compounds **46–61** (Hu et al., 2012), and compounds **62–84** (tetrahydropyrroloquinolinone type) (Yin et al., 2013) were used for this analysis, and their structures and bioactivity values are presented in Table 1. The pIC_{50} ($-\log IC_{50}$) values were used to derive 3D-QSAR models. The whole dataset of 84 compounds was divided into two groups in the ratio of 4:1:a training set with 63 compounds, a test set with 21 compounds (Table 1). The selection of the training and test sets was done manually such that low, moderate, and high activity compounds were present in roughly equal proportions in both sets. The training set was used to build predictive models, while the test set was used to validate the predictive ability of the models.

Computational approach

The 3D-QSAR modeling analyses, calculations, and visualizations were performed using the SYBYL 7.3 molecular modeling package from Tripos Inc., St. Louis, Mo, USA, installed on Red Hat Linux workstations. All the structures were built, and energy was minimized with the following steps: (i) optimization by Steepest Descent with initial optimization of 200 simplex iterations using Tripos force field and Gasteiger–Marsili charges; (ii) optimization by conjugate gradient; and (iii) optimization by BFGS (Hu et al., 2009).

Pharmacophore hypothesis

The pharmacophore hypothesis was generated using Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Database (GALAHAD) module of SYBYL, which operates in two main stages: the ligands are aligned to each other in internal coordinate space, and then the conformations produced are aligned in Cartesian space. The feature considered in developing the pharmacophore model includes HBD atoms, HBA atoms, and hydrophobic and charged centers (Richmond et al., 2006; Shepphird and Clark, 2006; Andrade et al., 2008). In our study, twelve compounds shown in Table 1 were selected to carry out the pharmacophore hypothesis, and the genetic algorithm was used to create conformers for all molecules. The

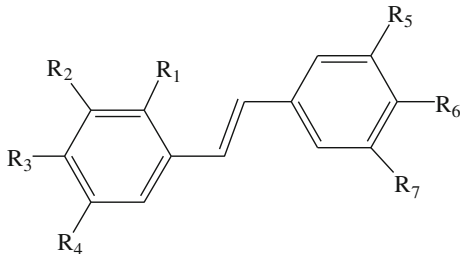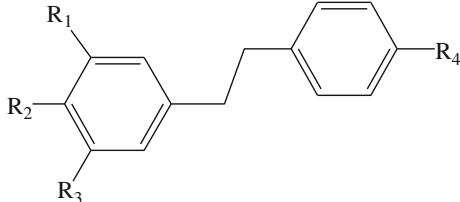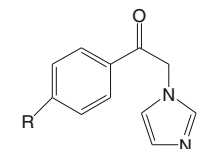**Table 1** Chemical structures and bioactivity values of the non-steroidal aromatase inhibitors in current study
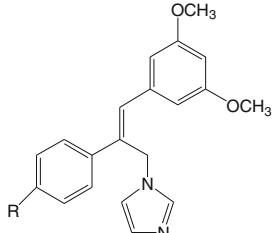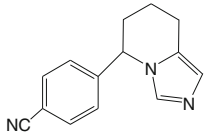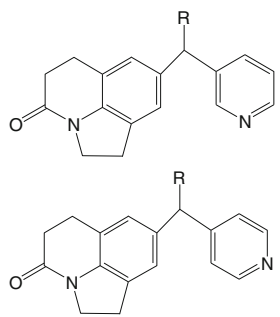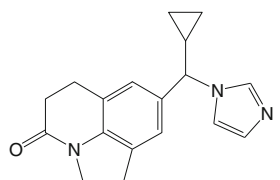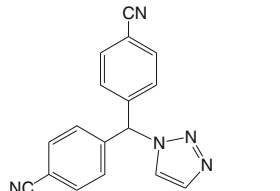
Compound	General structure	Substituents	IC$_{50}$ (µM)
1[b]		1,3,5,7-H; 2,4-OCH$_3$; 6-NH$_2$	0.59
2		2,4,5,7-H; 1,3-OCH$_3$; 6-NH$_2$	0.76
3		2,4,6,7-H; 1,3-OCH$_3$; 5-NH$_2$	11.14
4		2,3,5,7-H; 1,4-OCH$_3$; 6-NH$_2$	1.82
5[a]		3,4,5,7-H; 1,2-OCH$_3$; 6-NH$_2$	0.98
6		1,4,5,7-H; 2,3-OCH$_3$; 6-NH$_2$	14.51
7[a]		1,6,7-H; 2,3,4-OCH$_3$; 5-NH$_2$	16.42
8[b]		1,5,7-H; 2,3,4-OCH$_3$; 6-NH$_2$	0.90
9		4,5,7-H; 1,2,3-OCH$_3$; 6-NH$_2$	3.57
10		3,5,7-H; 1,2,4-OCH$_3$; 6-NH$_2$	2.28
11		1,4,6,7-H; 2,3-OCH$_2$O–; 5-NH$_2$	21.86
12		1,4,5,7-H; 2,3-OCH$_2$O–; 6-NH$_2$	8.49
13[a]		1,2,4,6,7-H; 3,5-NO$_2$	8.33
14		1,3,4,5,7-H; 2-OCH$_3$; 6-NH$_2$	3.08
15		1,3,4,5,7-H; 2,6-NH$_2$	7.46
16		1,3,5,7-H; 2,4,6-NH$_2$	8.51
17[a]		1,3,5,7-H; 2,4-OAc; 6-NH$_2$	2.94
18		1,3,5,7-H; 2,4-OH; 6-NH$_2$	5.0
19		2-H; 1,3-OCH$_3$; 4-NH$_2$	2.76
20		3-H; 1,2-OCH$_3$; 4-NH$_2$	0.67
21[a]		1,2,3-OCH$_3$; 4-NH$_2$	2.79
22		2,3-H; 1,4-NH$_2$	3.21
23		–NO$_2$	0.19
24		–NH$_2$	1.84
25[a]		–NO$_2$	0.07
26[b]		–NH$_2$	0.036

Table 1 continued

Compound	General structure	Substituents	IC$_{50}$ (µM)
27		1,2,3,4,5,6,7,8,9-H	29
28		1,3,4,5,6,7,8,9-H; 2-CH$_3$	44
29		1,2,4,5,6,7,8,9-H; 3-CH$_3$	18
30[a]		1,2,3,4,5,7,8,9-H; 6-CH$_3$	20
31		1,2,3,4,5,6,7,9-H; 8-CH$_3$	41
32		3,4,5,6,7,8,9-H; 1,2-Benzo	20
33		1,2,3,4,5,6,8,9-H; 7-OPh	2.4
34[b]		1,3,4,5,6,7,8,9-H; 2-OCH$_3$	0.26
35		1,2,4,5,6,7,8,9-H; 3-OCH$_3$	41
36[a]		1,2,3,4,6,7,8,9-H; 5-OCH$_3$	97
37		1,2,3,4,5,7,8,9-H; 6-OCH$_3$	99
38		1,2,3,4,5,6,8,9-H; 7-OCH$_3$	24
39[a]		1,2,3,4,5,7,9-H; 6,8-OCH$_3$	11
40		1,3,4,5,6,7,8,9-H; 2-F	32
41		1,3,4,5,6,7,8,9-H; 2-Cl	5.1
42[a]		1,3,4,5,6,7,8,9-H; 2-Br	29
43		1,3,4,5,6,7,8,9-H; 2-^{t}Bu	61
44		3-Thiophene	30
45[a]		3-Pyridyl	5.8
46		1,2-H; 3-Me	0.426
47[a]		1-OH; 2,3-H	3.073
48[a]		1-OMe; 2,3-H	0.447
49		1-OMe; 2-H; 3-Me	0.049
50		1-OEt; 2,3-H	0.488
51[a]		1-OEt; 2-H; 3-Me	0.048
52		1-O-c-Pent; 2,3-H	0.162
53[b]		1-O-c-Pent; 2-H; 3-Me	0.022
54		1-OBz; 2,3-H	0.035
55[b]		1-OBz; 2-H; 3-Me	0.011
56		1,3-H; 2-Br	2.901
57[a]		1,3-H; 2-Ph	0.394
58		1,3-H; 2-(3′-Py)	0.954
59		1,3-H; 2-(3′-Thienyl)	0.275
60[b]			0.038

Table 1 continued

Compound	General structure	Substituents	IC$_{50}$ (μM)
61[b]			0.052
62		H	3.34
63		Ph	2.13
64[a]		2-MeOPh	3.66
65		3-ClPh	1.35
66[b]		4-FPh	0.14
67		OH	2.13
68		=CH$_2$	0.565
69		Me	1.50
70		*i*-Pr	0.889
71		*c*-Hex	0.074
72		Ph	0.105
73[a]		Ph, OH	0.88
74[a]		2-MeOPh	0.081
75[b]		3-MeOPh	0.059
76		4-MeOPh	0.124
77		3-FPh	0.074
78[a]		4-FPh	0.116
79[b]		3-ClPh	0.019
80		4-ClPh	0.055
81[a]		3-CH$_3$Ph	0.032
82		3,5-diCF$_3$Ph	0.246
83[a]			0.228
84[b]			0.036

[a] Test-set compounds

[b] Compounds used to generate pharmacophore models

compounds selected to generate the pharmacophore hypothesis are highly active and structurally diverse.

Molecular docking

The crystal structure of aromatase complexed with NSAIs in present study has not yet been reported. Therefore, the crystal structure of aromatase complexed with ASD (PDB code: 3S79) from the RCSB Protein Data Bank was used (Ghosh et al., 2012). The molecular docking was performed using the Surflex-Dock module of SYBYL. All parameters were set with default values in the whole process. Before docking, the ligand was extracted, all the water molecules were removed, and hydrogen atoms were added to the receptor. The protomol was produced using docking-based method: ligand location in the same coordinate space in the receptor. In our study, each conformer of all 84 inhibitors was docked into the binding site ten times, and the docking score values were used to evaluate the docking analysis. The top ranked conformations for each molecule were extracted and aligned together for the subsequent CoMFA and CoMSIA modeling (Wang et al., 2012). The Surflex-Dock was also used to filter the hit compounds in the virtual screening.

Molecular alignment

The 3D structural alignment is a crucial component in 3D-QSA studies, and affects the outcome of the CoMFA and CoMSIA statistical analysis. There are three main different procedures proposed for aligning molecules for 3D-QSAR: maximum common substructures overlap, pharmacophore overlap and docking-based alignment (Hu et al., 2009). Both of pharmacophore- and docking-based alignment procedures were performed in the present study because of structural diversity of the studied compounds. Pharmacophore-based alignment was done using GALAHAD and docking-based alignment was done using Surlflex-Dock.

CoMFA and CoMSIA models

In CoMFA, the steric fields were calculated using a Lennard-Jones potential, while the electrostatic fields were calculated using a Coulombic potential. To calculate the CoMFA fields, a 3D-cubic lattice with grid spacing of 2.0 Å in X, Y, and Z directions was created automatically by SYBYL. The grid pattern extended 4.0 Å units in all directions beyond the dimensions of each molecule. The steric and electrostatic probe-ligand interaction energies were calculated using a sp^3 carbon probe atom and a +1.0 charge with a distance-dependent dielectric function at each lattice point. The cut-off for energies was set to ±30 kcal/mol, and the electrostatic contributions were ignored at lattice points with maximal steric interactions (Cramer et al., 1988). In CoMSIA, five different similarity fields (steric, electrostatic, hydrophobic, HBD, and HBA) were calculated. CoMSIA models were also derived with the same lattice box, and all five fields were calculated using a probe of charge +1, a radius of 1, hydrophobicity and hydrogen bonding properties of +1, and an attenuation factor of 0.3 for the Gaussian distance-dependent function (Klebe et al., 1994).

Statistical analysis

In order to derive 3D-QSAR models, CoMFA and CoMSIA descriptors were used as independent variables and the pIC_{50} values as the dependent variables. PLS method with cross-validation (leave-one-out) was used in SYBYL to determine the optimal numbers of components using cross-validated coefficient q^2 (r_{cv}^2). After obtaining the optimal numbers of components, a PLS analysis was performed with no validation and column filtering 2.0 to generate the final model with the training set. The obtained final non-cross-validated correlation coefficient (r_{ncv}^2) is a measure of the quality of the model. The predictive capability of the 3D-QSAR models was determined from the predictive

Table 2 The selected models and their statistical values after GALAHAD run

No.	Specificity	N_hits	Features	Pareto rank	Energy	Sterics	H-bond	Mol_Qry
Model_01	3.25	12	4	0	20.71	227.90	34.70	8.86
Model_03	3.20	11	4	0	39.46	250.20	32.10	5.33
Model_07	1.98	8	5	0	44.13	197.60	33.00	11.35
Model_08	1.89	12	4	0	28.18	206.80	32.30	15.77
Model_14	3.08	10	4	0	18.02	233.40	30.20	9.13
Model_18	3.12	9	5	0	12.62	150.70	32.70	5.14
Min[a]	1.89	6	2	0	12.62	150.70	30.20	0.00
Max[a]	4.22	12	7	0	2×10^7	268.00	36.90	15.77

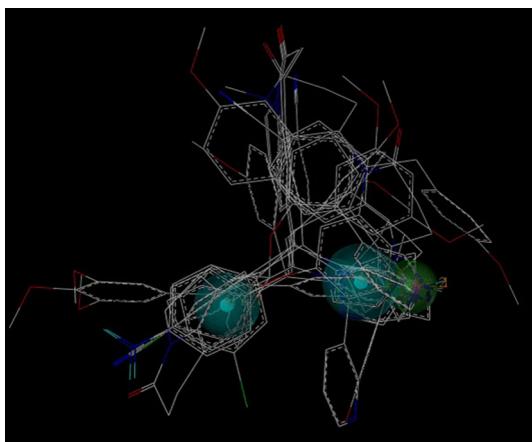[a] Minimum and maximum values between all the obtained 20 models

Fig. 1 The best GALAHAD model includes one acceptor atom (*green*), one donor atom (*magenta*), and two hydrophobes (*cyan*). The *sphere* sizes indicate query tolerances (Color figure online)

correlation (r_{pred}^2). The predicted activities for the test set were obtained from the model produced by the training set.

Results and discussion

Pharmacophore generation

Twenty pharmacophore models were generated with default parameters after GALAHAD run. Each of the obtained models represents a different tradeoff among the conflicting demands of maximizing steric consensus, maximizing pharmacophore consensus, and minimizing energy. All the twenty models had Pareto rank 0, which means no one model is superior to any other one. Some models had very high energy, which is recognized that high energy values are due to steric clashes (Dorfman *et al.*, 2008). Small value of energy and high values of N_hits, Sterics, and Mol_Qry are desired for the best model (Caballero, 2010), so six models were chosen for the analysis, and their statistical values are shown in Table 2. All the selected models were used for the molecular alignment to produce CoMFA and CoMSIA models. Model_08 was considered to be the best model as the best CoMFA and CoMSIA results were obtained when Model_08 was used to align the dataset. This model contains one acceptor atom, one donor atom, and two hydrophobes, which is shown in Fig. 1. The acceptor atom and donor atom are very close to each other. In addition, this model was further converted into a UNITY query for virtual screening studies.

CoMFA and CoMSIA statistical results

The structural alignment of compounds plays a very important role in the development of successful 3D-QSAR models, so all
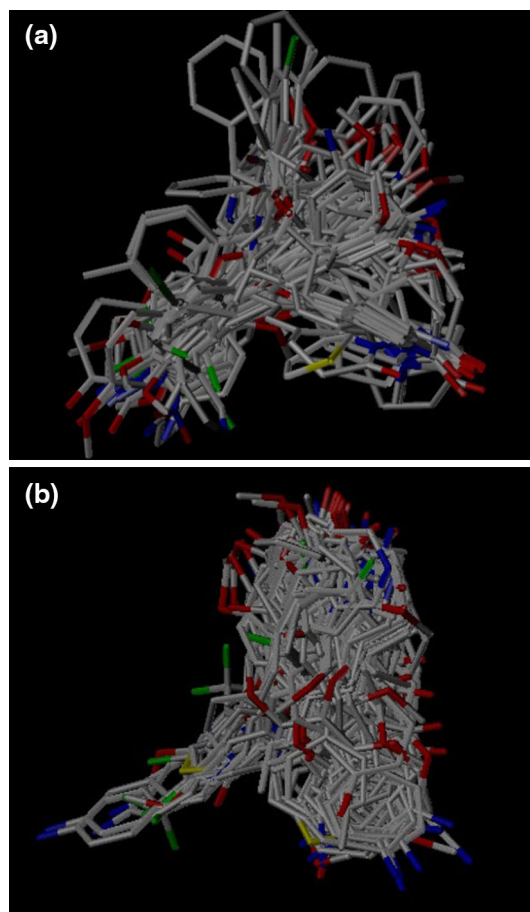

Fig. 2 a pharmacophore-based alignment of NSAIs in the training set. **b** docking-based alignment of NSAIs in the training set

Table 3 Summary of CoMFA and CoMSIA results

Components	Pharmacophore-based model		Docking-based model	
	CoMFA	CoMSIA	CoMFA	CoMSIA
$q^2(r_{cv}^2)$	0.634	0.668	0.563	0.469
r_{ncv}^2	0.986	0.962	0.824	0.761
SEE	0.131	0.215	0.448	0.522
F value	564.951	236.690	140.290	95.754
r_{pred}^2	0.737	0.708	0.532	0.618
No. of compounds	63	63	63	63
No. of optimal components	7	6	2	2
Contribution				
Steric	0.354	0.130	0.277	0.092
Electrostatic	0.646	0.319	0.723	0.287
Hydrophobic	–	0.200		0.158
H-bond donor	–	0.128		0.165
H-bond acceptor	–	0.224		0.298

Table 4 Observed and predicted pIC$_{50}$ of the training and test sets using CoMFA and CoMSIA models

Compound	Observed pIC$_{50}$	CoMFA		CoMSIA	
		Pred.	Res.	Pred.	Res.
1	6.229	6.201	0.028	6.111	0.118
2	6.119	6.120	−0.001	6.029	0.090
3	4.953	5.084	−0.131	4.943	0.010
4	5.740	5.586	0.154	5.911	−0.171
5[a]	6.009	6.164	−0.155	5.965	0.044
6	4.838	5.003	−0.165	5.486	−0.648
7[a]	4.785	4.643	0.142	5.169	−0.384
8	6.046	6.085	−0.039	6.124	−0.078
9	5.447	5.525	−0.078	5.277	0.170
10	5.642	5.560	0.082	5.730	−0.088
11	4.660	4.474	0.186	4.695	−0.035
12	5.071	5.039	0.032	5.184	−0.113
13[a]	5.079	5.497	−0.418	5.646	−0.567
14	5.511	5.555	−0.044	5.335	0.176
15	5.127	5.134	−0.007	5.055	0.072
16	5.070	5.021	0.049	4.921	0.149
17[a]	5.532	5.322	0.210	5.609	−0.077
18	5.301	5.387	−0.086	5.320	−0.019
19	5.559	5.536	0.023	5.613	−0.054
20	6.174	6.107	0.067	5.838	0.336
21[a]	5.554	5.056	0.498	5.394	0.160
22	5.493	5.342	0.151	5.402	0.091
23	6.721	6.690	0.031	6.561	0.160
24	5.735	5.775	−0.040	5.693	0.042
25[a]	7.155	7.378	−0.223	6.157	0.998
26	7.444	7.430	0.014	7.619	−0.175
27	4.538	4.600	−0.062	4.486	0.052
28	4.357	4.518	−0.161	4.581	−0.224
29	4.745	4.753	−0.008	4.807	−0.062
30[a]	4.699	4.744	−0.045	4.784	−0.085
31	4.387	4.452	−0.065	4.574	−0.187
32	4.699	4.735	−0.036	4.628	0.071
33	5.620	5.630	−0.010	5.696	−0.076
34	6.585	6.637	−0.052	6.638	−0.053
35	4.387	4.345	0.042	4.231	0.156
36[a]	4.013	5.072	−1.059	4.522	−0.509
37	4.004	4.037	−0.033	3.904	0.097
38	4.620	4.581	0.039	4.569	0.051
39[a]	4.959	4.142	0.817	4.030	0.929
40	4.495	4.849	−0.354	4.696	−0.201
41	5.292	4.832	0.460	4.718	0.574
42[a]	4.538	4.897	−0.359	4.825	−0.287
43	4.215	4.106	0.109	4.128	0.087
44	4.523	4.495	0.028	4.835	−0.312
45[a]	5.237	6.379	−1.142	6.607	−1.370
46	6.371	6.344	0.027	6.576	−0.205

Table 4 continued

Compound	Observed pIC$_{50}$	CoMFA		CoMSIA	
		Pred.	Res.	Pred.	Res.
47[a]	5.512	5.592	−0.080	5.154	0.358
48[a]	6.350	6.040	0.310	6.246	0.104
49	7.310	7.283	0.027	7.053	0.257
50	6.312	6.317	−0.005	6.398	−0.086
51[a]	7.319	6.809	0.510	6.829	0.490
52	6.790	6.698	0.092	6.655	0.135
53	7.658	7.624	0.034	7.603	0.055
54	7.456	7.689	−0.233	7.534	−0.078
55	7.959	7.844	0.115	7.676	0.283
56	5.537	5.731	-0.194	5.625	−0.088
57[a]	6.405	6.111	0.294	6.345	0.060
58	6.020	5.939	0.081	5.872	0.148
59	6.561	6.526	0.035	6.403	0.158
60	7.420	7.421	−0.001	7.690	−0.270
61	7.284	7.212	0.072	6.908	0.376
62	5.476	5.470	0.006	5.622	−0.146
63	5.672	5.703	−0.031	5.967	−0.295
64[a]	5.437	6.051	−0.614	6.131	−0.694
65	5.870	6.051	−0.181	5.965	−0.095
66	6.854	6.770	0.084	6.816	0.038
67	5.672	5.636	0.036	5.637	0.035
68	6.248	6.311	−0.063	6.113	0.135
69	5.824	5.894	−0.070	6.190	−0.366
70	6.051	6.012	0.039	6.212	−0.161
71	7.131	7.042	0.089	7.052	0.079
72	6.979	7.211	−0.232	7.088	−0.109
73[a]	6.056	6.191	−0.135	6.240	−0.184
74[a]	7.092	6.503	0.589	6.747	0.345
75	7.229	7.417	−0.188	6.980	0.249
76	6.907	6.981	−0.074	7.244	−0.337
77	7.131	6.917	0.214	6.910	0.221
78[a]	6.936	6.807	0.129	6.943	−0.007
79	7.721	7.506	0.215	7.361	0.360
80	7.260	7.172	0.088	7.356	−0.096
81[a]	7.495	6.912	0.583	6.907	0.588
82	6.609	6.597	0.012	6.599	0.010
83[a]	6.642	6.480	0.162	6.539	0.103
84	7.444	7.556	−0.112	7.660	−0.216

[a] Test-set compounds

compounds of the dataset were aligned according to both pharmacophore and docking to derive the CoMFA and CoMSIA models in the present study. Figure 2a, b show pharmacophore-based and docking-based alignments of training set molecules used in 3D-QSAR models, respectively.

To obtain an effective 3D-QSAR model, a number of statistical parameters, $q^2(r^2_{cv})$, r^2_{ncv}, standard error estimate
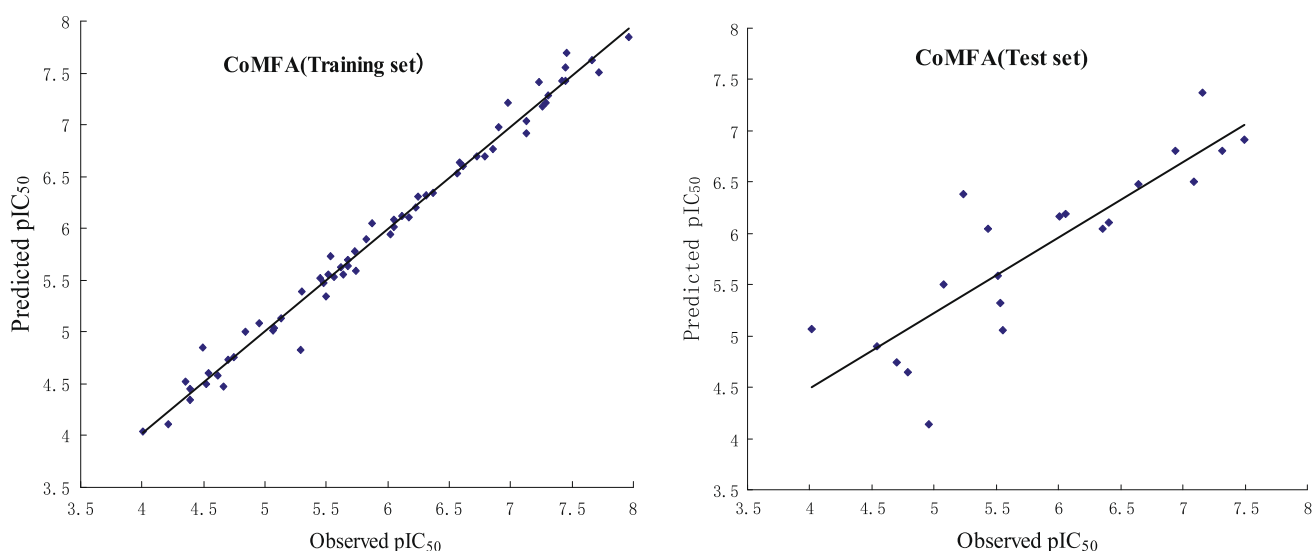
Fig. 3 Plots of observed versus predicted activities of the training set and test set molecules from CoMFA analysis
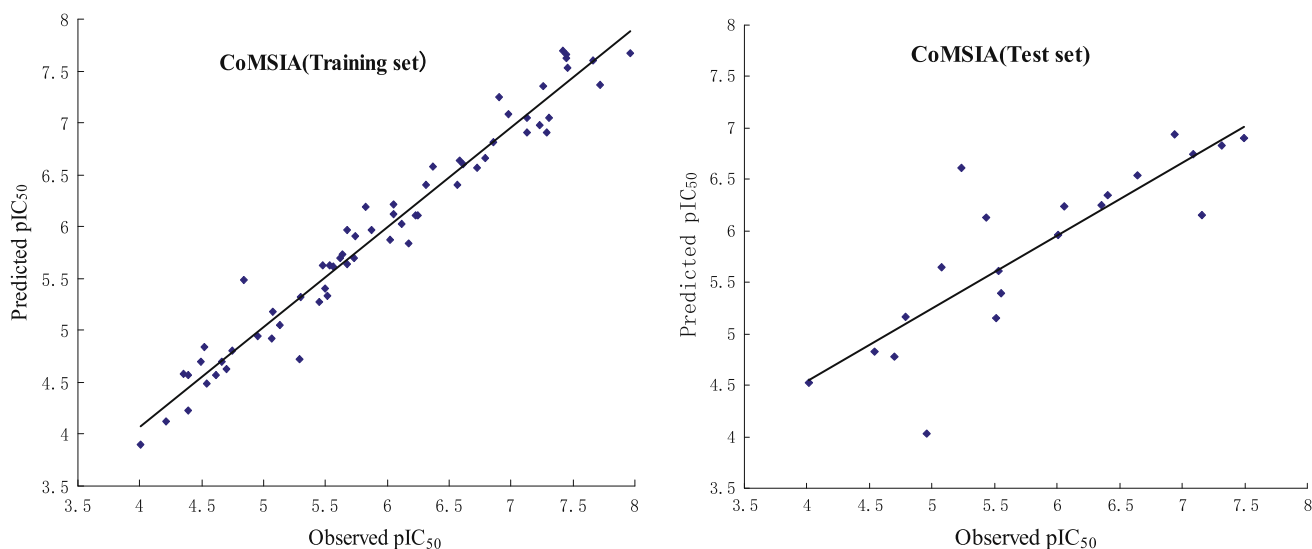

Fig. 4 Plots of observed versus predicted activities of the training set and test set molecules from CoMSIA analysis

(SEE), and F-statistic values (F) were computed as defined in SYBYL. The CoMFA and CoMSIA results obtained from both pharmacophore- and docking-based alignments are shown in Table 3, which show that the pharmacophore-based CoMFA and CoMSIA models exhibit better statistical results than the docking-based CoMFA and CoMSIA models, where the pharmacophore-based modeling yielded $q^2(r_{cv}^2) = 0.634$, $r_{ncv}^2 = 0.986$ for CoMFA model and $q^2(r_{cv}^2) = 0.668$, $r_{ncv}^2 = 0.962$ for CoMSIA model, while the docking-based modeling gave $q^2(r_{cv}^2) = 0.563$, $r_{ncv}^2 = 0.824$ for CoMFA model, $q^2(r_{cv}^2) = 0.469$, $r_{ncv}^2 = 0.761$ for CoMSIA model, respectively.

We mainly focus on the CoMFA and CoMSIA obtained from pharmacophore-based alignment due to its

more satisfactory statistical results. As shown in Table 3, the CoMFA model has a high q^2 (r_{cv}^2) of 0.634 with seven optimal components. This CoMFA model has r_{ncv}^2 of 0.986, SEE of 0.131 and F value of 564.951. The corresponding field contributions of steric and electrostatic are 0.354 and 0.646, respectively, which means the electrostatic field gives more contribution to activity than the steric field. The CoMSIA model ($q^2 = 0.668$, $r_{ncv}^2 = 0.962$, $F = 236.690$, SEE = 0.215) was obtained using the combination of steric, electrostatic, hydrophobic, HBD, and HBA fields with six optimal components. The corresponding field contributions are 0.130, 0.319, 0.200, 0.128, and 0.224, respectively. Both CoMFA and CoMSIA models obtained from pharmacophore-based
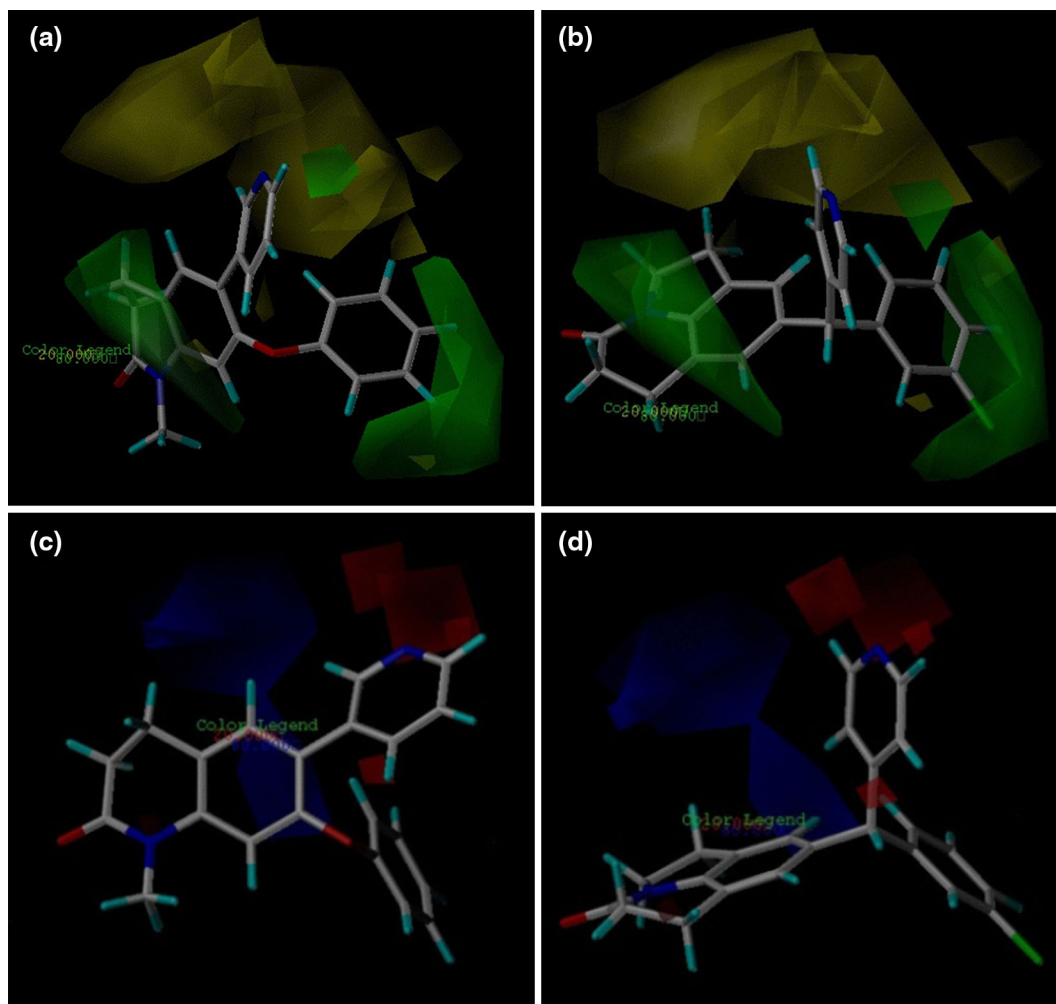
Fig. 5 CoMFA contour maps (standard deviation × coefficient) in combination with compounds **55** and **79**. **a, b** Steric contour maps: *green contours* (80 % contribution) refer to sterically favored regions; *yellow contours* (20 % contribution) indicate sterically disfavored regions. **c, d** Electrostatic contour maps: *blue contours* (80 % contribution) refer to regions where positively charged substituents are favored; *red contours* (20 % contribution) indicate regions where negatively charged substituents are favored (Color figure online)

alignment are fairly good as their q^2 values are more than 0.6.

In order to validate the 3D-QSAR models, r_{pred}^2 was used to determine the predictive abilities of the CoMFA and CoMSIA models from the 21 compounds (test set) which were not included in the generation of the models. The obtained r_{pred}^2 of the test set is 0.737 and 0.708 for the CoMFA and CoMSIA model, respectively, which indicates that both models have high predictive ability. The observed and predicted pIC$_{50}$ by the CoMFA and CoMSIA model of the training and test sets are given in Table 4, and the correlations between the observed and predicted pIC$_{50}$ of training and test sets are depicted in Fig. 3 for CoMFA model, Fig. 4 for CoMSIA model, respectively.

CoMFA contour maps

The steric contour maps for the CoMFA model with two most active inhibitor compounds **55** and **79** are shown in Fig. 5a, b, respectively. In these figures, the green contours represent regions of high steric tolerance (80 % contribution), while the yellow contours represent regions of low steric bulk tolerance (20 % contribution). Figure 5a combined with compound **55** shows that a large green contour near the –OBz group attached to the 2-position of benzene ring indicates that a bulky group in this position is favorable to bioactivity. It is confirmed by the fact that compounds **52–55** and **60** with bulky substitution in that position have higher bioactivity while compounds **56–59**
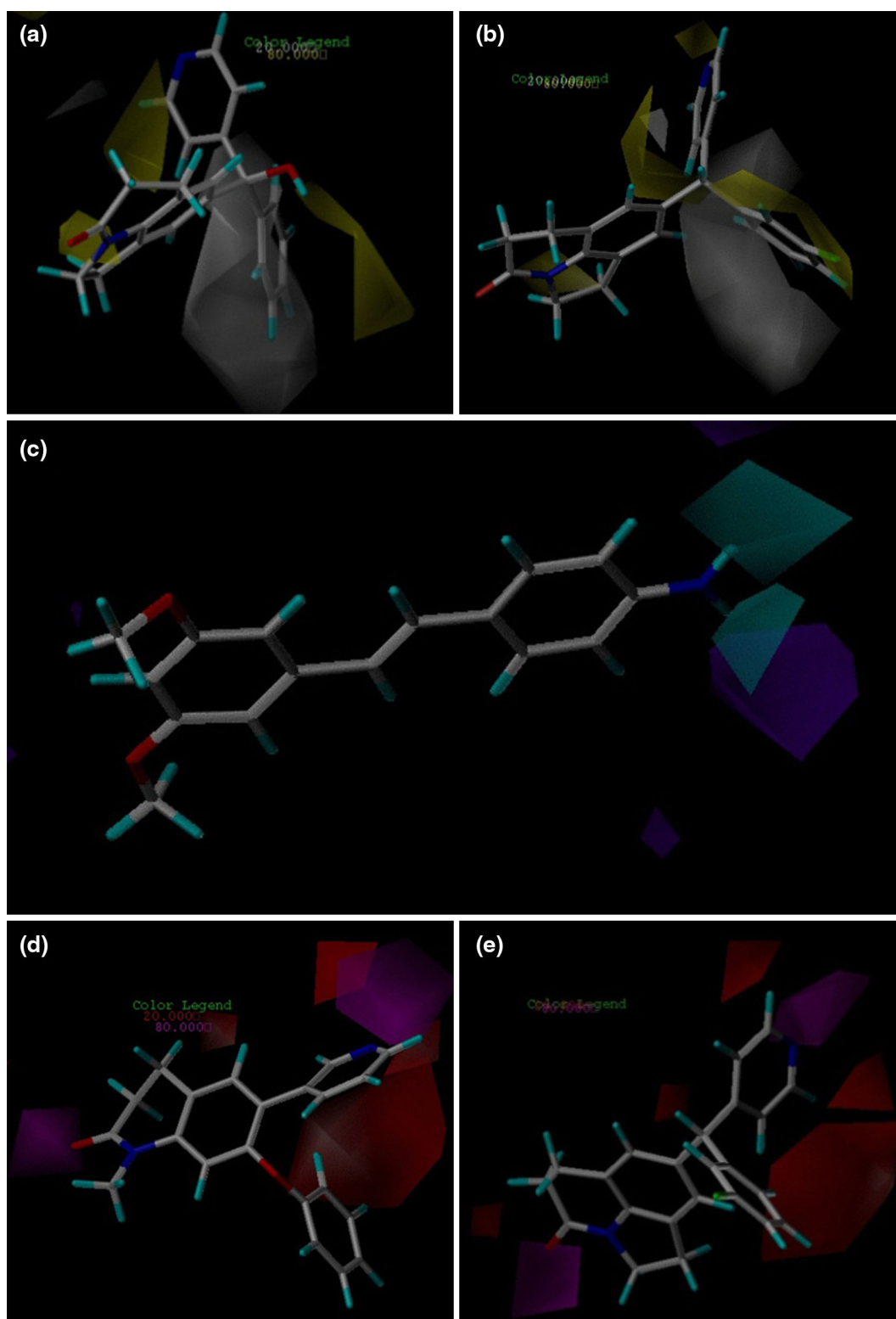
Fig. 6 CoMSIA contour maps (standard deviation × coefficient). **a, b** Hydrophobic contour maps in combination with compounds **73** and **79**: *white contours* refer to regions where hydrophilic substituents are favored; *yellow contours* indicate regions where hydrophobic substituents are favored. **c** HBD contour map in combination with compound **1**: *cyan contours* indicate HBD substituents in this region are favorable to activity; *purple contours* represent that HBD groups in this area are unfavorable. **d, e** HBA contour maps in combination with compounds **55** and **79**: *magenta contours* show regions where HBA substituents are expected; *red contours* refer to areas where HBA substituents are unexpected (Color figure online)

Fig. 7 Plots of C_score values versus biological activity (pIC$_{50}$ values) of 84 inhibitors
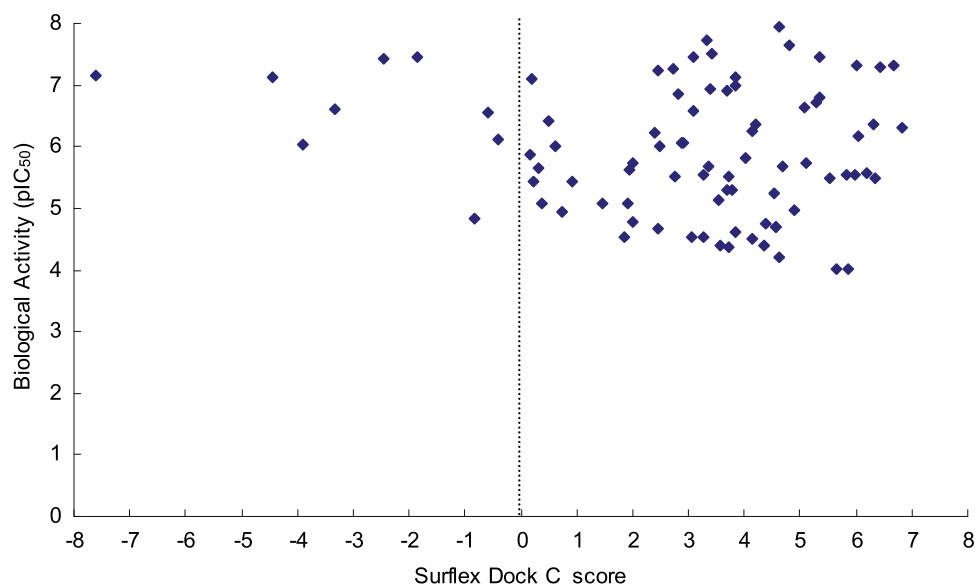
with no substitution in that position have lower bioactivity. A large green contour near the 3-ClPh group can also be observed in Fig. 5b combined with compound **79**, which suggests that a bulky group in this region will increase inhibitory activity. This is supported by the higher activity of compounds **74–82**, which have large substituents in that position, compared with the lower activity of compounds **68–70**, which have small substituents in that position.

The CoMFA electrostatic contour maps in combination with compounds **55** and **79** are shown in Fig. 5c, d, respectively. The red areas are the regions where a negative potential is favorable to activity, while a negative potential is unfavorable in the blue areas. Figure 5c combined with compound **55** displays a red contour near the nitrogen atom of the pyridine ring, which indicates that the bioactivity can be enhanced if an electronegative atom is present. It is clear that most of the compounds **46–60** with a nitrogen atom in that position have high inhibitory activity. Similarly, Fig. 5d combined with compound **79** also shows a red contour near the nitrogen atom of the pyridine ring, which means that an electronegative atom in this position can increase the bioactivity. This can be confirmed by the higher activity of compounds **72, 74,** and **79,** compared with the lower activity of compounds **63, 64** and **65,** respectively.
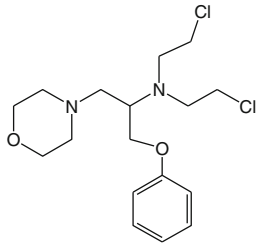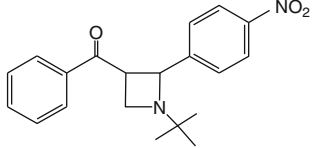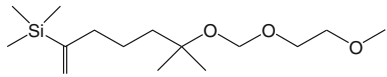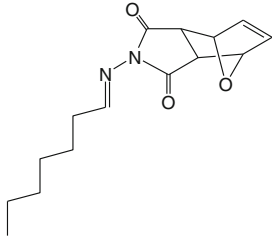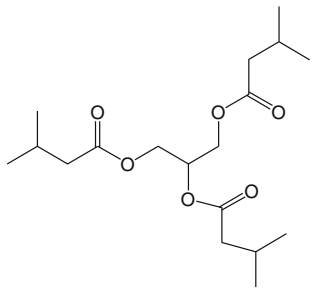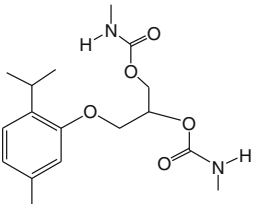
CoMSIA contour maps

CoMSIA not only calculates steric and electrostatic fields as in CoMFA, but also additionally computes hydrophobic, HBD, and HBA fields. The steric and electrostatic contour maps of CoMSIA are consistent with those of CoMFA. The contour maps of CoMSIA hydrophobic, HBD, and HBA

fields are presented in Fig. 6. For each field, the favorable and disfavored contours represent 80 and 20 % level contributions, respectively. The hydrophobic contour maps in combination with compounds **73** and **79** are shown in Fig. 6a, b, respectively, in which yellow contours indicate regions where hydrophobic groups are favorable to bioactivity while white contours represent areas where hydrophilic groups are favorable. Figure 6a combined with compound **73** shows that Ph group is near the white region and –OH group is near the yellow region, while Fig. 6b combined with compound **79** shows that –ClPh group is near the yellow region. This is supported by the case that the activity of compound **73** is much lower than that of compound **79**.

The HBD contour map in combination with compound **1** is present in Fig. 6c, in which cyan contours indicate that HBD substituents in this region are favorable to activity while purple contours represent that HBD groups in this area are unfavorable. Two cyan contours are shown near the –NH$_2$ group of the compound **1**, which indicates the necessity of the hydrogen atoms at this position for high bioactivity. It is confirmed by the fact that the compounds **2, 8,** and **12** with –NH$_2$ group at 6-position have higher bioactivity compared with the compounds **3, 7,** and **11** with –NH$_2$ group at 5-position have lower bioactivity, respectively.

Figure 6d, e show the HBA contour maps in combination with compounds **55** and **79**, respectively, where magenta and red contours represent areas where HBA substituents are favored and disfavored, respectively. Figure 6d combined with compound **55** displays that one magenta contour is near the carbonyl group, and the other magenta contour is near the nitrogen atom of the pyridine ring, which indicates that the activity can be enhanced if the electronegative

Table 5 Chemical structures and their predicted activity values of screened hit compounds

Hit compound	Structure	pIC$_{50}$ (predicted by CoMFA)	pIC$_{50}$ (predicted by CoMSIA)
NCI 170394		7.748	7.619
NCI 150322		7.277	6.880
NCI 617389		6.743	7.669
NCI 60353		6.987	6.731
NCI 4048		6.400	6.833
NCI 170454		6.540	6.401

atoms are in these two positions. It is supported by the case that most of the compounds **46–60,** which have nitrogen and oxygen atoms in these two positions, show high bioactivity. Similarly, Fig. 6e combined with compound **79** also shows that one magenta contour is near the carbonyl group, and the other magenta contour is near the nitrogen atom of the pyridine ring, which represents that the electronegative atoms in these two positions can increase the bioactivity. This is confirmed by the higher activity of compounds **72, 74,** and **79** compared with the lower activity of compounds **63, 64,** and **65,** respectively.

Docking analysis

In order to validate the docking model in the virtual screening, the C_score values of all 84 inhibitors were tested using Surflex-Dock, and the correlations between the C_score values and pIC$_{50}$ values are depicted in Fig. 7. It can be seen that 75 of 84 inhibitors have positive C_score values, so the hit compounds should have positive C_score values in the virtual screening.

Virtual screening

The obtained best GALAHAD model (Fig. 1) was converted into a UNITY query, which was screened against NCI2000 database. The "flexible database search" option was implemented to perform virtual screening. Primary filters such as Lipinski's rule of five, Van der Waals bumps, and QFIT (pharmacophoric match between query and the hit compound) were applied to reduce the dataset (Kothandan et al., 2013). The screening of the pharmacophore query yielded 336 hits that met the specific requirements. The 336 hit compounds were further subjected to molecular docking using the Sulflex-Dock module of SYBYL. 174 compounds were selected based on the docking score values (C_score value > 0). The pIC50 values of the selected 174 compounds were predicted using CoMFA and CoMSIA models generated on the basis of pharmacophore alignment. Finally, six structurally diverse compounds with good pIC$_{50}$ values predicted by both CoMFA and CoMSIA models (both pIC$_{50}$ > 6.40) are listed in Table 5. Thus, these potential hit compounds are expected to design novel NSAIs with new skeleton.

Conclusion

Aromatase inhibitors have proven to be the most important targets for treatment of estrogen-dependent cancers. In order to search for more potent NSAIs with lower side

effects and overcome the drug resistance, pharmacophore modeling, virtual screening, and 3D-QSAR studies were performed. Pharmacophore model was derived from twelve highly active and structurally diverse compounds using GALAHAD. The best pharmacophore model includes one acceptor atom, one donor atom, and two hydrophobes, which was used as a query to search NCI2000 database. 336 hit compounds were obtained after the screening of the pharmacophore query, and 90 compounds were selected by molecular docking. Finally, six structurally diverse compounds with good predicted pIC$_{50}$ values were obtained, which are expected to design novel NSAIs with new skeletons. The 3D-QSAR techniques based on both pharmacophore and docking alignments, CoMFA and CoMSIA, were applied to 84 NSAIs with different skeletons. The CoMFA and CoMSIA models obtained from pharmacophore-based alignment shows better statistical results, $q^2 = 0.634$, $r^2_{ncv} = 0.986$, $r^2_{pred} = 0.737$ for CoMFA and $q^2 = 0.668$, $r^2_{ncv} = 0.962$, $r^2_{pred} = 0.708$ for CoMSIA, which indicate that these structurally diverse NSAIs must bind to the same active site of aromatase, and it is feasible to drive CoMFA and CoMSIA models based on structurally diverse compounds. The present pharmacophore modeling, virtual screening, and 3D-QSAR approach provides useful information to design and synthesize novel NSAIs.

References

Andrade CH, Salum LB, Pasqualoto KFM, Ferreira EI, Andricopulo AD (2008) Three-dimensional quantitative structure-activity relationships for a large series of potent antitubercular agents. Lett Drug Des Discov 5:377–387

Bhatt HG, Patel PK (2012) Pharmacophore modeling, virtual screening and 3D-QSAR studies of 5-tetrahydroquinolinylidine aminoguanidine derivatives as sodium hydrogen exchanger inhibitors. Bioorg Med Chem Lett 22:3758–3765

Bonfield K, Amato E, Bankemper T, Agard H, Steller J, Keeler JM, Roy D, McCallum A, Paula S, Ma L (2012) Development of a new class of aromatase inhibitors: Design, synthesis and inhibitory activity of 3-phenylchroman-4-one (isoflavanone) derivatives. Bioorg Med Chem 20:2603–2613

Brueggemeier RW, Hackett JC, Diaz-Cruz ES (2005) Aromatase inhibitors in the treatment of breast cancer. Endocr Rev 26:331–345

Caballero J (2010) 3D-QSAR (CoMFA and CoMSIA) and pharmacophore (GALAHAD) studies on the differential inhibition of aldose reductase by flavonoid compounds. J Mol Graph Model 29:363–371

Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967

Dorfman RJ, Smith KM, Masek BB, Clark RD (2008) A knowledge-based approach to generating diverse but energetically

representative ensembles of ligand conformers. J Comput Aided Mol Des 22:681–691

Dutta U, Pant K (2008) Aromatase inhibitors: past, present and future in breast cancer therapy. Med Oncol 25:113–124

Ghosh D, Lo J, Morton D, Valette D, Xi J, Griswold J, Hubbell S, Egbuta C, Jiang W, An J, Davies HM (2012) Novel Aromatase Inhibitors by Structure-Guided Design. J Med Chem 55:8464–8476

Hong Y, Rashid R, Chen S (2011) Binding features of steroidal and nonsteroidal inhibitors. Steroids 76:802–806

Honorio KM, Garratt RC, Polikatpov I, Andricopulo AD (2007) 3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators. J Mol Graph Model 25:921–927

Hu RJ, Barbault F, Delamar M, Zhang RS (2009) Receptor- and ligand-based 3D-QSAR study for a series of non-nucleoside HIV-1 reverse transcriptase inhibitors. Bioorg Med Chem 17:2400–2409

Hu QZ, Yin L, Hartmann RW (2012) Selective dual inhibitors of CYP19 and CYP11B2: targeting cardiovascular diseases hiding in the shadow of breast cancer. J Med Chem 55:7080–7089

Jordan VC, Brodie AMH (2007) Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. Steroids 72:7–25

Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146

Kothandan G, Madhavan T, Gadhe CG, Cho SJ (2013) A combined 3D QSAR and pharmacophore-based virtual screening for the identification of potent p38 MAP kinase inhibitors: an in silico approach. Med Chem Res 22:1773–1787

Perez EA (2006) Appraising adjuvant aromatase inhibitor therapy. Oncologist 11:1058–1069

Recanatini M, Cavalli A, Valenti P (2002) Nonsteroidal aromatase inhibitors: recent advances. Med Res Rev 22:282–304

Richmond NJ, Abrams CA, Wolohan PRN, Abrahamian E, Willett P, Clark RD (2006) GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D. J Comput Aided Mol Des 20:567–587

Salum LD, Polikarpov I, Andricopulo AD (2007) Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. J Mol Graph Model 26:434–442

Seralini GE, Moslemi S (2001) Aromatase inhibitors: past, present and future. Mol Cell Endocrinol 178:117–131

Shepphird JK, Clark RD (2006) A marriage made in torsional space: using GALAHAD models to drive pharmacophore multiplet searches. J Comput Aided Mol Des 20:763–771

Sun B, Hoshino J, Jermihov K, Marler L, Pezzuto JM, Mesecar AD, Cushman M (2010) Design, synthesis, and biological evaluation of resveratrol analogues as aromatase and quinone reductase 2 inhibitors for chemoprevention of cancer. Bioorg Med Chem 18:5352–5366

Wang JN, Wang FF, Xiao ZT, Sheng GW, Li Y, Wang YH (2012) Molecular simulation of a series of benzothiazole PI3 K alpha inhibitors: probing the relationship between structural features, anti-tumor potency and selectivity. J Mol Model 18:2943–2958

Winer EP, Hudis C, Burstein HJ et al (2005) American society of clinical oncology technology assessment on the use of aromatase inhibitors as adjuvant therapy for postmenopausal women with hormone receptor-positive breast cancer: status report 2004. J Clin Oncol 23:619–629

Yin L, Hu QZ, Hartmann RW (2013) Tetrahydropyrroloquinolinone type dual inhibitors of aromatase/aldosterone synthase as a novel strategy for breast cancer patients with elevated cardiovascular risks. J Med Chem 56:460–470