European Journal of Medicinal Chemistry 44 (2009) 4367–4375

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Quantitative structure–activity relationship analysis of aryl alkanol piperazine derivatives with antidepressant activities

Ke-Xian Chen [a], Zu-Guang Li [a,*], Hai-Ying Xie [a], Jian-Rong Gao [a], Jian-Wei Zou [b]

[a] College of Chemical Engineering and Materials Science, Zhejiang University of Technology, 18,Chaowang Road, Hangzhou 310014, China
[b] Key Laboratory for Molecular Design and Nutrition Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315104, China

ARTICLE INFO

Article history:
Received 29 May 2008
Received in revised form
25 September 2008
Accepted 20 May 2009
Available online 6 June 2009

Keywords:
Aryl alkanol piperazine derivatives
Antidepressant activities
Genetic function approximation (GFA)
Molecular field analysis (MFA)
Genetic partial least squares (G/PLS)
Quantitative structure–activity relationship (QSAR)

ABSTRACT

Quantitative structure–activity relationship analysis for recently synthesized aryl alkanol piperazine derivatives was studied for their antidepressant activities. The statistically significant 2D-QSAR models ($r^2 > 0.924$, $r^2_{CV} > 0.870$, $r^2_{pred} > 0.890$) were developed using genetic function approximation (GFA) when the number of descriptors in equation was set to four, indicating the descriptors of Atype_C_6, Dipole-mag, S_sssCH and Jurs-PNSA-3 mainly influence the 5-hydroxytryptamine (5-HT) reuptake inhibition activity while the descriptors of HOMO, PMI-mag, S_sssN and Shadow-XZ may chiefly control the noradrenaline (NA) reuptake inhibition activity. The results of the 2D-QSAR models were further compared with 3D-QSAR models generated by molecular field analysis (MFA), investigating the substitutional requirements for the favorable receptor–drug interaction and providing useful information in the characterization and differentiation of their binding sites. The results derived may be useful in further designing novel antidepressants prior to synthesis.

1. Introduction

Depression [1,2] is increasingly becoming familiar and serious spiritual disease by impacting on all aspects of a person's life. The etiology of depression [3–6] is suggested to be the dysfunction of monoamine neurotransmitters in central nervous system (CNS), such as serotonin (5-hydroxytryptamine, or 5-HT), dopamine (DA) and norepinephrine (NE), but the specific etiology of major depression is still far from clear yet [6–8]. The predominant monoamine theory [9] combines depression with lowered concentrations of actual or functional monoamine neurotransmitters at brain synapses. Treatment of depression thus may be achieved by restoring the monoamine levels or action to normal.

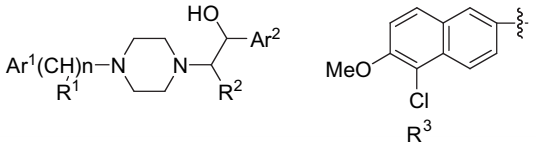
Significant progress in antidepressant chemotherapy has been substantially made to manage the depression over the past half century [4,10]. There are many antidepressant drugs [3] involved in medical therapy, such as tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), selective serotonin reuptake inhibitors (SSRIs) and specific serotonin–noradrenaline reuptake inhibitors (SNRIs), but emerging data have also questioned the safety and efficacy of antidepressants [11]. Recently the new generation of antidepressants with high degrees of selectivity for 5-HT transporter, the selective serotonin reuptake inhibitors (SSRIs), and NA transporter, the serotonin–noradrenaline reuptake inhibitors (SNRIs) have become the most widely prescribed drugs in clinical application [9,12]. Reuptake blockade of monoamine receptors such as α_2-adrenoceptors modulated at the neuronal terminal by antidepressants can increase both noradrenaline (NA) and 5-HT neurotransmission [4]. Unfortunately, the mechanisms of antinociceptive effects of antidepressants are not fully understood [13]. Based on these facts, it may be still a good alternative to further understand the structural specify of antidepressant drugs with neurotransmitters reuptake inhibition activity and identify the phenomenon of how the drugs act and work in intro.

Traditional computer-assisted quantitative structure–activity relationship (QSAR) studies pioneered by C. Hansch et al. [14,15] have been proved to be one of the useful approaches for accelerating the drug design process [16,17], which help to correlate the bioactivity of compounds with structural descriptors [18]. Until now, not more systematical QSAR analysis regarding to serotonin and noradrenaline monoamine neurotransmitters has been reported in the literature [19–21]. Recently synthesized aryl alkanol piperazine derivatives (Table 1) were discovered with the reuptake inhibition activities of 5-hydroxytryptamine (5-HT) and noradrenaline (NA), which may have potential utility in treating

* Corresponding author. Tel./fax: +86 571 88320306.
E-mail address: lzg@zjut.edu.cn (Z.-G. Li).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.05.029

Table 1
Structure and antidepressant activities of aryl alkanol piperazine derivatives.



NO.	Ar^1	Ar^2	R^1	R^2	n	Ki1 (%)[a]	Actual pKi1 (%)	Ki2 (%)[a]	Actual pKi2 (%)
1	Ph	Ph	H	Ph	1	<10	<−1.000	104	−2.017
2	4-Cl-Ph	Ph	H	CH_3	1	82	−1.914	62	−1.792
3	4-F-Ph	Ph	H	H	1	79	−1.898	107	−2.029
4	4-NO_2-Ph	Ph	H	H	1	63	−1.799	80	−1.903
5	Ph	4-Cl-Ph	H	H	1	85	−1.929	51	−1.708
6	4-F-Ph	4-Cl-Ph	H	H	1	81	−1.908	101	−2.004
7	Ph	R^3	H	CH_3	0	60	−1.778	34	−1.531
8	3-Cl-Ph	R^3	H	CH_3	0	99	−1.996	30	−1.447
9	4-CF_3-Ph	R^3	H	CH_3	0	78	−1.892	25	−1.398
10	4-OMe-Ph	R^3	H	CH_3	0	125	−2.097	87	−1.940
11	2-OMe-Ph	R^3	H	CH_3	0	142	−2.152	120	−2.079
12	2,3-Cl_2-Ph	R^3	H	CH_3	0	90	−1.954	0	ND[b]
13	2,3-Me_2-Ph	R^3	H	CH_3	0	100	−2.000	86	−1.934
14	Ph	R^3	H	H	1	54	−1.732	34	−1.531
15	Ph	R^3	H	CH_3	1	114	−2.057	140	−2.146
16	Ph	R^3	H	C_2H_5	1	100	−2.000	133	−2.053
17	Ph	R^3	CH_3	CH_3	1	107	−2.029	58	−1.763
18	Ph	R^3	Ph	CH_3	1	120	−2.079	126	−2.100
19	4-Me-Ph	R^3	H	CH_3	1	90	−1.954	82	−1.914
20	2-Me-Ph	R^3	H	CH_3	1	73	−1.863	49	−1.690
21	3-Me-Ph	R^3	H	CH_3	1	79	−1.898	138	−2.140
22	2-F-Ph	R^3	H	CH_3	1	100	−2.000	19	−1.279
23	4-F-Ph	R^3	H	CH_3	1	91	−1.959	76	−1.881
24	3-F-Ph	R^3	H	CH_3	1	125	−2.097	0	ND[b]
25	3-Cl-Ph	R^3	H	CH_3	1	79	−1.898	0	ND[b]
26	3,4-Cl_2-Ph	R^3	H	CH_3	1	65	−1.813	140	−2.146
27	2,6-Cl_2-Ph	R^3	H	CH_3	1	125	−2.097	0	ND[b]
28	4-NO_2-Ph	R^3	H	CH_3	1	110	−2.041	108	−2.033
29	4-NH_2-Ph	R^3	H	CH_3	1	<10	<−1.000	77	−1.886
30	3,4,5-Me_3-Ph	R^3	H	CH_3	1	116	−2.064	77	−1.886
31	Ph	R^3	(R)CH_3	H	1	108	−2.029	58	−1.763
32	Ph	R^3	(S)CH_3	H	1	107	−2.033	10	−1.000

[a] Ki1 (%) and Ki2 (%) represent 5-HT reuptake inhibition ratio (%) and NA reuptake inhibition ratio (%), respectively. All antidepressant activities are expressed as −log(Ki), which is pKi (%).
[b] ND: not defined.

depression. To gain further insights into the structure–activity relationships of these derivatives and understand the mechanism of their substitutional specificity, the present group of authors have developed some statistically significant QSAR models between antidepressant activities and structural descriptors using genetic function approximation (GFA) and molecular field analysis (MFA) coupled with genetic partial least squares (G/PLS), respectively. The significance of the QSAR models was evaluated using cross-validation tests, randomization tests and external test set prediction. Then, the robust 2D/3D-models were applied to predict the possible activities of 15 newly designed molecules. Thus the results derived can help medicinal chemist to design new candidates as potential antidepressants prior to synthesis.

2. Materials and methods

2.1. Selection of molecules

Data sets of 32 aryl alkanol piperazine derivatives (Table 1) collected from published literature [7] were taken for the present study. The affinity data of antidepressant activities were converted into the corresponding pKi values to get the linear relationship in equation using the following formula: pKi = −log Ki, where Ki value represents reuptake inhibition ratio (%) of 5-hydroxytryptamine (5-HT) and noradrenaline (NA) caused by these molecules at the concentration of 0.1 mmol/L compared with fluoxetine and desipratmine, respectively. Molecules were rationally divided into the training set and test set (Table 2) based on the suggestions given by Oprea et al. [22]. Newly designed 15 molecules with unreported activities i.e., molecules **33–47** (Table 3), captured structural features of the training set molecules, were used for activity prediction.

2.2. Molecular modeling

All computational experiments were performed using QSAR + module of Cerius^2 software (version 4.10) running on Silicon Graphics O2 R5000 workstation [23]. The molecular geometric structures were constructed using a 3D-sketcher in the Cerius^2 Builder option and subjected to an energy minimization procedure of UFF-VALBOND1.1 [24] to generate the lower energy conformation for each molecule. Partial atomic charges were assigned using the Gasteiger method [25]. All the structures were subsequently energy minimized until a root mean square derivation 0.001 kcal/mol was achieved and used in the present study.

2.3. 2D-QSAR analysis

2.3.1. Genetic function approximation (GFA)

Genetic function approximation (GFA) [26] is a useful statistical analysis tool to correlate biological activity or property with chemical characteristics of molecules, and also greatly improves the

Table 2
Actual and predicted activities for 32 molecules based on the best 2D/3D-QSAR models.

NO.	Actual pKi1 (%)	2D-QSAR (model-1)		3D-QSAR (model-3)		Actual pKi2 (%)	2D-QSAR (model-2)		3D-QSAR (model-4)	
		Predicted	Residual[d]	Predicted	Residual[d]		Predicted	Residual[d]	Predicted	Residual[d]
1[a]	<−1.000	−2.304	1.304	−2.203	1.203	−2.017	−2.024	0.007	−2.026	0.009
2[c]	−1.914	−1.920	0.006	−1.908	−0.006	−1.792	−2.173	0.381	−2.162	0.370
3	−1.898	−1.922	0.024	−1.890	−0.008	−2.029	−1.942	−0.087	−2.040	0.011
4	−1.799	−1.825	0.026	−1.807	0.008	−1.903	−1.961	0.058	−1.904	0.001
5	−1.929	−1.888	−0.041	−1.914	−0.015	−1.708	−1.768	0.060	−1.706	−0.002
6	−1.908	−1.894	−0.014	−1.905	−0.003	−2.004	−1.991	−0.013	−2.020	0.016
7[a]	−1.778	−2.282	0.504	−1.992	0.214	−1.531	−1.627	0.096	−1.543	0.012
8	−1.996	−2.009	0.013	−1.996	0.000	−1.447	−1.567	0.090	−1.484	0.007
9	−1.892	−1.901	0.009	−1.914	0.022	−1.398	−1.358	−0.040	−1.423	0.025
10	−2.097	−2.139	0.042	−2.096	−0.001	−1.940	−1.907	−0.033	−1.948	0.008
11[c]	−2.152	−2.112	−0.040	−2.148	−0.004	−2.079	−1.607	−0.472	−2.160	0.081
12[b]	−1.954	−1.982	0.028	−1.973	0.019	ND[e]	–	–	–	–
13[a,c]	−2.000	−2.266	0.266	−2.209	0.209	−1.934	−1.703	−0.231	−1.936	0.002
14	−1.732	−1.770	0.038	−1.724	−0.008	−1.531	−1.641	0.110	−1.498	−0.033
15	−2.057	−2.101	0.044	−2.054	−0.003	−2.146	−2.107	−0.039	−2.151	0.005
16[c]	−2.000	−1.962	−0.038	−1.998	−0.002	−2.053	−1.781	−0.272	−2.115	0.062
17	−2.029	−2.028	−0.001	−2.029	0.000	−1.763	−1.748	−0.015	−1.761	−0.002
18	−2.079	−2.060	−0.019	−2.081	0.002	−2.100	−2.057	−0.043	−2.123	0.023
19[a]	−1.954	−1.933	−0.021	−1.990	0.036	−1.914	−1.750	−0.164	−1.900	−0.014
20[a]	−1.863	−1.973	0.110	−2.135	0.272	−1.690	−1.628	−0.062	−1.680	−0.010
21	−1.898	−1.885	−0.013	−1.910	0.012	−2.140	−2.120	−0.020	−2.104	−0.036
22	−2.000	−1.951	−0.049	−2.008	0.008	−1.279	−1.255	−0.024	−1.280	0.001
23[a]	−1.959	−2.042	0.083	−1.967	0.008	−1.881	−1.909	0.028	−1.874	−0.007
24[a,b]	−2.097	−1.939	−0.158	−1.999	−0.098	ND[e]	–	–	–	–
25[b]	−1.898	−1.924	0.026	−1.895	−0.003	ND[e]	–	–	–	–
26	−1.813	−1.798	−0.015	−1.809	−0.004	−2.146	−2.255	0.109	−2.134	−0.012
27[a,b]	−2.097	−1.928	−0.169	−1.812	−0.285	ND[e]	–	–	–	–
28	−2.041	−2.016	−0.025	−2.027	−0.014	−2.033	−1.977	−0.056	−2.040	0.007
29[a,c]	<−1.000	−2.073	1.073	−1.920	0.920	−1.886	−1.712	−0.174	−1.503	−0.383
30[a]	−2.064	−1.928	−0.136	−2.058	−0.006	−1.886	−1.868	−0.018	−1.880	−0.006
31	−2.029	−2.025	−0.004	−2.026	−0.003	−1.763	−1.831	0.068	−1.758	−0.005
32[c]	−2.033	−2.038	0.005	−2.036	0.003	−1.000	−1.720	0.720	−1.808	0.808

[a] Molecules used in test set to test the predictive ability of the model-1 and model-3.
[b] Molecules discarded in generating model-2 and model-4.
[c] Molecules used in test set to test the predictive ability of the model-2 and model-4.
[d] Residual = actual pKi − predicted pKi.
[e] ND: not defined.

ease of successful model interpretation. GFA is genetically involved in the combination of Fried machs multivariate adaptive regression splines (MARS) and Holland's genetic algorithm (GA) [27–30], which performs as follows: a population of randomly-constructed QSAR models are generated using an error measure that estimates each model's relative predictive ability. The population is "cross-over" evolved by repeatedly replacing the worst-rated models using the better-rated models, then producing 100 new models containing some descriptors of the best model. As evolution proceeds, the population becomes enriched with higher and higher quality models to best fit the training set. The distinctive feature of GFA is that it can produce statistical measures such as Friedman's lack of fit (LOF) score to assess the predictive ability of the generated models.

2.3.2. Calculation of descriptors

Different types of physicochemical descriptors of each molecule were generated in the study table using default setting within

Table 3
Structure and predicted activities of 15 newly designed molecules using the best 2D/3D-QSAR models.

NO.	Ar^1	Ar^2	R^1	R^2	n	Predicted pKi1 (%)		Predicted pKi2 (%)	
						Model-1	Model-3	Model-2	Model-4
33	3,5-Cl_2-Ph	R^3 [a]	H	CH_3	1	−1.897	−1.953	−2.240	−2.079
34	3,4,5-Cl $_3$-Ph	R^3 [a]	H	CH_3	1	−1.851	−1.942	−1.766	−1.507
35	4-Cl-Ph	R^3 [a]	H	CH_3	0	−2.126	−2.077	−1.648	−1.362
36	4-NO_2-Ph	Ph	CH_3	H	1	−2.050	−2.124	−1.748	−1.870
37	2-F-Ph	Ph	H	H	1	−2.046	−2.064	−1.325	−1.571
38	4-Me-Ph	Ph	H	CH_3	1	−2.232	−2.010	−2.076	−2.432
39	3,4-Cl_2-Ph	Ph	H	CH_3	1	−2.005	−2.129	−2.202	−2.422
40	4-OMe-Ph	Ph	H	CH_3	0	−2.374	−1.975	−1.305	−1.461
41	Ph	Ph	CH_3	CH_3	1	−2.348	−2.102	−2.082	−1.911
42	4-NH_2-Ph	Ph	H	CH_3	1	−2.225	−2.018	−1.638	−2.084
43	4-F-Ph	Ph	H	CH_3	1	−2.138	−2.087	−1.831	−1.723
44	Ph	4-Cl-Ph	CH_3	CH_3	1	−2.220	−2.099	−2.118	−2.360
45	Ph	Ph	H	C_2H_5	1	−2.234	−2.101	−2.054	−1.663
46	Ph	Ph	H	CH_3	1	−2.241	−2.046	−2.049	−2.129
47	Ph	R^3 [a]	CH_3	CH_3	1	−2.187	−2.195	−1.919	−1.632

[a] R^3 represents the same structure as shown in Table 1.

Table 4
Remaining descriptors and their types used to build 2D-QSAR models by GFA method.

Type	Descriptors
E-state-indices	Electrotopological-state indices
Spatial	Jurs descriptors, radius of gyration, principal moment of inertia, shadow indices
Electronic	Sum of atomic polarizabilities, dipole moment, energy of highest occupied orbital (HOMO), energy of lowest unoccupied orbital (LUMO), superdelocalizability
Thermodynamic	Molar refractivity, heat of formation, log of the partition coefficient, log of the partition coefficient atom type value, desolvation free energy for water, desolvation free energy for octanol
Structural	Number of chiral centers, number of rotatable bonds, number of hydrogen-bond donors, number of hydrogen-bond acceptors
Topological	Kier and Hall molecular connectivity index, Hosoya index, Balaban indices, molecular flexibility index, molecular shape kappa indices

QSAR+ and descriptor + module of Cerius^2 software. Before generating the 2D-QSAR models, the inter-correlation of 125 descriptors with nonzero values was taken into account and the highly correlated descriptors were removed [30]. A complete list of remaining descriptors used for 2D-QSAR analysis is given in Table 4.

2.3.3. Generation of 2D-QSAR models

2D-QSAR analysis is used to build models (equations) that give one or more dependent properties in terms of independent descriptors. Application of this approach allows construction of high quality predictive models for newly designed molecules with unknown activities and provides other useful information. The length of equation was initially fixed to five terms including a constant. After some preliminary runs for observations, GFA crossover of 5000 and smoothing parameter "d" value of 1.0 were set to give reasonable convergence. Other default settings were maintained. Cross-validated r^2 (r^2_{CV}) was calculated using cross-validated test option in the statistical tools in Cerius^2 software [31].

2.4. 3D-QSAR analysis

2.4.1. Molecular field analysis

Molecular field analysis (MFA) can effectively evaluate the interaction energy between a probe and a set of aligned molecules at a series of points defined by a rectangular grid [32], especially for analysis of the data sets with available activity data but unknown receptor site structure [32–34]. This approach attempts to describe a self-consistent field around the molecules for explaining activity and gain sufficient possible features of a receptor site from the aligned molecules for designing novel analogous molecules. Field values on each point of a 3D-grid are computed using atomic coordinates of binding molecules, and the interaction energies associated with each grid point can be used as input for 3D-QSAR analysis. The charges of the molecules and probes have not been calculated.

2.4.2. Alignment rules

The root mean square (RMS) alignment aligns models [23] via producing the best superposition of two or more models employing the collective RMS of their atomic coordinates; the sum of squares of the distances is functionally minimized between all atoms to be superimposed. In present study, we used the core substructure search (CSS) method coupled with RMS alignment method [30] to rigidly align all of the structures in the analogous series based on a defined substructure of 16 atoms in the active molecule **11** (Fig. 1), so that common features are discernable from any other random arrangement of orientations. Stereoview of aligned molecules in training set and test set is shown in Fig. 2.

2.4.3. Creation of interaction energies

The proton (H^+) and methyl (CH_3) probes were used to calculate steric and electrostatic fields, respectively. The fields were computed at each lattice intersection of a regularly spaced grid of 2.0 Å within defined three-dimensional region. An energy cutoff of ±30 kcal/mol was truncated. The total grid points generated were 936 and 10% of significant columns of H^+ and CH_3 probes with highest variance was automatically selected as independent X variables for subsequent 3D-QSAR analysis.

2.4.4. Genetic partial least squares analysis

Genetic partial least squares (G/PLS) method was employed to linearly correlate the MFA fields with the antidepressant activity values. G/PLS [32–34] is derived from two important regression methods named GFA and partial least squares (PLS), and has been proved as the valuable analysis tool where the number of data set is more than the number of compounds. G/PLS retains the ease of interpretation of GFA by back-transforming the PLS components to the original variables. Each generation has PLS applied to it instead of multiple linear regressions, so each model can have more terms in it without danger of over-fitting. The number of crossovers was set to 5000, the length of equation was fixed at 14 terms including a constant, and the smoothing parameter "d" of 1.0 was maintained. Cross-validation analysis was performed using the leave-one-out (LOO) method by clicking the proper button in the validate control panel.

3. Results and discussion

3.1. 2D-QSAR models

Different sets of 2D-QSAR models (equations) were generated using the genetic function approximation (GFA) in Cerius^2 software. A brute force approach [35] was first employed to investigate the number of descriptors necessary and adequate for the QSAR equation. As the number of descriptors in the QSAR equation increased one by one, the effect of addition of new term was evaluated on the statistical quality of model. Cross-validated r^2 (r^2_{CV}) was selected as the limiting factor for number of descriptors to be used in the final model [30,35,36]. As shown in Fig. 3, the r^2_{CV} value of the best model increases till the number of descriptors in the equation reaches up to 4 and then declines as the number of descriptors increases further. Thus, the number of descriptors was restricted to 4 for the final model. Based upon our experiments, inclusion of more than 4 descriptors in the QSAR model also does not improve the r^2, F-test and LSE. The models with varying number of descriptors for 5-hydroxytryptamine (5-HT) reuptake inhibition activity (pKi1) and noradrenaline (NA) reuptake inhibition activity (pKi2) are shown in Table 5 along with the statistical parameters. The selection of the best model [31] was based on the values of r^2 (square of the correlation coefficient for the training set

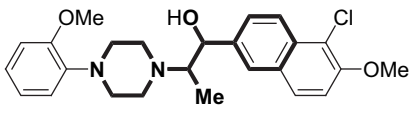


Fig. 1. The bold region in the active molecule **11** includes the 16 atoms used as substructure for alignment.

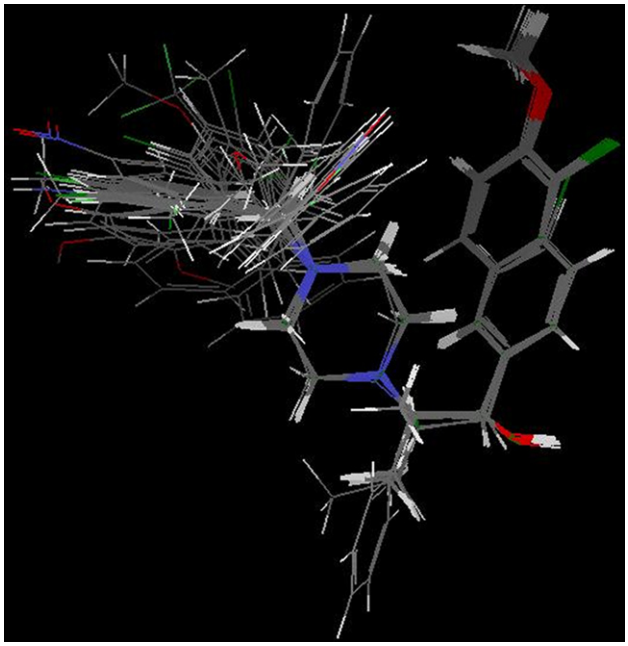


Fig. 2. Stereoview of aligned molecules in training set and test set.

molecules); LOF; F-test; r^2_{-CV} (cross-validated r^2); r^2_{-BS} (bootstrap correlation coefficient); PRESS (predicted sum of deviation squares). The statistically significant 2D-QSAR models for pKi1 and pKi2 are shown as follows.

Model-1

$$\begin{aligned}\text{pKi1} &= -3.09 + 0.150707(\text{Atype\_C\_6}) \\ &\quad + 0.018996(\text{Dipole-mag}) - 0.003689(\text{Jurs-PNSA-3}) \\ &\quad + 0.144186(\text{S\_sssCH})\end{aligned}$$

$$n = 22;\quad \text{LOF} = 0.002;\quad r^2 = 0.926;$$
$$r^2_{-adj} = 0.908;\quad \text{F-test} = 52.825;\quad \text{LSE} = 0.001;$$
$$r = 0.962;$$

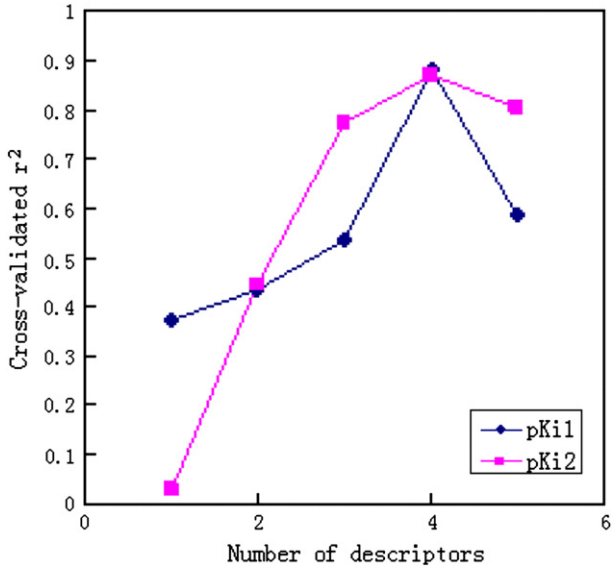


Fig. 3. Plot of cross-validated r^2 (r^2_{-CV}) as a function of number of descriptors.

$$r^2_{-CV} = 0.879;\quad r^2_{-BS} = 0.926 \pm 0.001;$$
$$\text{PRESS} = 0.027;\quad r^2_{-pred} = 0.893$$

Model-2

$$\begin{aligned}\text{pKi2} &= 9.70845 + 0.359573(\text{HOMO}) - 0.000323(\text{PMI-mag}) \\ &\quad - 1.57666(\text{S\_sssN}) + 0.007954(\text{Shadow-XZ})\end{aligned}$$

$$n = 22;\quad \text{LOF} = 0.012;\quad r^2 = 0.925;$$
$$r^2_{-adj} = 0.908;\quad \text{F-test} = 52.618;\quad \text{LSE} = 0.005;$$
$$r = 0.962;$$

$$r^2_{-CV} = 0.871;\quad r^2_{-BS} = 0.926 \pm 0.001;$$
$$\text{PRESS} = 0.027;\quad r^2_{-pred} = 0.972$$

where Ki1 and Ki2 represent 5-HT reuptake inhibition ratio (%) and NA reuptake inhibition ratio (%), respectively, which have already been expressed as −log(Ki); n is the number of compounds in training set; LOF is Friedman's lack of fit score [35,36], which is used to assess the goodness of each progeny equation using the following formula: $\text{LOF} = \text{LSE}/\{1 - (c + dp)/m\}^2$, where LSE is least square error, c is the number of basis function in the model, d is smoothing parameter, p is the number of descriptors and m is the number of observations in the training set; r^2 is the squared correlation coefficient; r^2_{-adj} is square of adjusted correlation coefficient; F-test is the variance related static; LSE is the least square error; r is the correlation coefficient; cross-validated r^2 (r^2_{-CV}) is a squared correlation coefficient generated during the validation procedure; Bootstrap r^2 (r^2_{-BS}) is the average squared correlation coefficient calculated during the validation procedure; predicted sum of deviation squares (PRESS) is the sum of overall compounds of the squared differences between the actual and the predicted values for the dependent variables; r^2_{-pred} is the predictive power of the model.

The inter-correlation of the descriptors appeared in the above best models was taken into account and the descriptors were found to be reasonably orthogonal (Table 6). Descriptor values appeared in the 2D-QSAR models of training set and test set molecules are shown in Table 7.

To determine reliability and significance of these generated models, the leave-one-out (LOO) tests and randomization tests were employed. From the cross-validation tests, r^2_{-CV} of 0.879 and 0.871 for pKi1 and pKi2, respectively, indicated that the results obtained for the above best 2D-QSAR models were not by chance correlation. The randomization tests [30,31] were performed at 95% (19 trials) and 98% (49 trials) confidence levels and carried out by repeatedly permuting the dependent variable set. If the score of the primary QSAR model testified to be better than those from the permuted data sets, the model would be considered statistically significant and predominant. The results of randomization tests showed that none of the permuted data sets produced random r^2 comparable to nonrandom r^2, suggesting that the r^2 values for both pKi1 and pKi2 obtained are significant (data not shown). The predictive power of the model is calculated by $r^2_{-pred} = (\text{SD} - \text{PRESS})/\text{SD}$ [30,33], where SD is the sum of squared deviations between the biological activities of each molecule and the mean activity of the molecules in the training set and PRESS is the sum of squared deviations between the predicted and actual activity values for every molecule in the test set. The r^2_{-pred} values for both the test set molecules were very high, which was promising. The developed 2D-QSAR models thus were robust and were found

Table 5
Statistical evaluation of 2D-QSAR models for pKi1 and pKi2 with varying number of descriptors.

Descriptor	Equation	LOF	r^2	r^2_{adj}	F-test	LSE	r	r^2_{BS}	r^2_{CV}
pKi1									
1	pKi1 = −2.19627 + 0.076723(Atype_C_6)	0.007	0.548	0.500	11.512	0.005	0.740	0.550	0.374
2	pKi1 = −2.57841 + 0.099078(Atype_C_6) + 0.0000889(Jurs-TPSA)	0.006	0.579	0.535	13.086	0.004	0.761	0.581	0.432
3	pKi1 = 3.4858 + 0.013015(S_sCl) + 0.109667(Atype_C_25) − 0.541944(S_sOH)	0.006	0.690	0.639	13.378	0.003	0.831	0.692	0.535
4	pKi1 = −3.09 + 0.150707(Atype_C_6) + 0.018996 (Dipole-mag) − 0.003689 (Jurs-PNSA-3) + 0.144186(S_ssCH)	0.002	0.926	0.908	52.825	0.001	0.962	0.926	0.879
5	pKi1 = −2.2727 − 0.4996(Shadow-XYfrac) + 3.03584(Jurs-RPCG) + 0.10007 (Atype_C_6) + 2.61698(Jurs-FNSA-3) + 0.146392(Atype_Cl_89)	0.007	0.800	0.737	12.768	0.002	0.894	0.802	0.587
pKi2									
1	pKi2 = −1.28783 − 0.298667(Atype_N_68)	0.060	0.211	0.172	5.356	0.050	0.460	0.214	0.032
2	pKi2 = −2.01668 − 0.274064(Rotlbonds) + 0.275286(Shadows-Zlength)	0.036	0.612	0.571	14.995	0.039	0.782	0.613	0.446
3	pKi2 = 5.42244 − 0.000276(PMI-mag) − 1.46401(S_sssN) + 0.265823(S_aaaC)	0.019	0.838	0.811	31.092	0.010	0.916	0.839	0.772
4	pKi2 = 9.70845 + 0.359573(HOMO) − 0.000323 (PMI-mag) − 1.57666(S_sssN) + 0.007954(Shadow-XZ)	0.012	0.925	0.908	52.618	0.005	0.962	0.926	0.871
5	pKi2 = 1.60889 − 1.5997(Jurs-FNSA-2) + 0.001975(Jurs-WNSA-2) − 0.500582(Kappa-3-AM) − 2.6506(JX) + 0.22658(Shadow-Zlength)	0.021	0.899	0.867	28.356	0.006	0.948	0.900	0.805

satisfactory for predicting the activities of the test set (Table 2) and newly designed compounds (Table 3). From Table 2, molecules **1** and **29** turned out to have high residuals because of their low activities in comparison to other compounds. From Table 3, it can be found that most of newly designed compounds have high activities, which still needs to be experimentally validated.

The 2D-QSAR model is important in suggesting important substitutions in the molecules. According to model-1, the observed 5-HT reuptake inhibition activities for aryl alkanol piperazine derivatives are principally influenced by Atype_C_6, Dipole-mag, Jurs-PNSA-3 and S_sssCH, which is confirmed by the usage of these descriptors during the formation of models (Fig. 4(a)). The positive slope of Atype_C_6 as atom type A log P descriptor [35] in this model represents that reuptake inhibition activity decreases with an increase in hydrophobicity (log P) related to C6 atom type in the molecule. The atom type C_6 is C in CH_2RX where X represents any heteroatom (O, N, S, and halogens) [37]. Dipole-mag [23] is an electronic descriptor involved in ligand–receptor interactions and indicates the strength and orientation behavior of a molecule in an electrostatic field. It is positively correlated with biological activity, indicating that the compounds having larger dipole moment may show less reuptake inhibition activity. Jurs-PNSA-3 [38] is the sum of the product of solvent-accessible surface area X partial charge for all negatively charged atoms. Increasing this value in a molecule could increase reuptake inhibition activity because it is negatively correlated with the activity. Molecules **11** and **13** (Table 2) with high value of Jurs-PNSA-3 were more active than molecule **12** with low value of Jurs-PNSA-3 (Table 2). The E-state indices [39] encode information about both the topological environment and the electronic interaction of an atom due to all other atoms in the molecule. S_sssCH as a type of the E-state indices stands for the carbon bonded to one hydrogen atom and three non-hydrogen atoms in the alkane. Increasing the presence of these features contributes less towards 5-HT reuptake inhibition.

An important observation in generating QSAR models for NA reuptake inhibition was the occurrence of HOMO, S_sssN and PMI-mag as frequent descriptors (Fig. 4(b)). From the statistically significant model-2, it was observed that HOMO and S_sssN mainly controlled the NA reuptake inhibition activities of aryl alkanol piperazine derivatives. HOMO (highest occupied molecular orbital) [40] is the highest energy level in the molecule that contains electrons, which are crucially important in governing molecular reactivity and properties. When a molecule acts as a Lewis base (an electron-pair donor) in bond formation, the electrons are supplied from its HOMO. It is positively correlated with the pKi2, indicating the importance of hydrogen bonding interactions. PMI-mag [23] calculates the principal moments of inertia about the principal axes of a molecule for a series of straight lines through the center of mass, indicating the orientation and conformational rigidity of the molecule. Thus the orientation of the aromatic ring bearing small polarizable groups is important for the activity. S_sssN [20] is the sum of the atom level E-state values for all nitrogen atoms with three single bonds in the molecule. The negative correlated coefficient of S_sssN indicated that NA reuptake inhibition activity can be increased in presence of these groups. Shadow-XZ [35] is the area of molecular shadow in the XZ plane (Sxz), which can be calculated by projecting the molecular surface on mutually perpendicular XZ plane. It helps to characterize the shape of the molecules based on their conformation and their spatial orientation. Shadow-XZ shows a positive contribution to pKi2, suggesting that the NA reuptake inhibitory effect can be enhanced by decreasing this value in the molecule.

Table 6
Correlation matrices (a) and (b) of the descriptors appeared in the model-1 and model-2.

(a)

	pKi1	Atype_C_6	Dipole-mag	Jurs-PNSA-3	S_ssCH
pKi1	1				
Atype_C_6	−0.006	1			
Dipole-mag	−0.633	−0.034	1		
Jurs-PNSA-3	0.230	0.122	−0.260	1	
S_ssCH	0.159	−0.249	−0.205	−0.370	1

(b)

	pKi2	HOMO	PMI-mag	S_sssN	Shadow-XZ
pKi2	1				
HOMO	−0.132	1			
PMI-mag	−0.344	0.389	1		
S_sssN	0.119	0.080	−0.323	1	
Shadow-XZ	−0.242	0.402	−0.775	−0.087	1

3.2. 3D-QSAR models

MFA samples the steric and electrostatic fields surrounding a set of ligands and constructs 3D-QSAR models by correlating these 3D fields with the corresponding biological activities. The statistical significant 3D-QSAR models for pKi1 (model-3) and pKi2 (model-4) are given below. In model-3 and model-4, the descriptors H^+/a, H^+/b and H^+/c are the interaction energies between a proton probe and the molecule at the rectangular points a, b and c, respectively [30,32].

Table 7
Descriptor values of training set and test set molecules appeared in the best 2D-QSAR models.

NO.	Descriptors							
	Atype_C_6	S_sssCH	Dipole-mag	Jurs-PNSA-3	S_sssN	HOMO ev	PMI-mag	Shadow-XZ
1	5	−0.563	1.226	−27.094	4.942	−11.6254	1259.866	81.208
2	5	−0.306	10.790	−69.225	4.845	−11.8833	1855.957	79.171
3	6	−0.431	6.467	−55.108	4.678	−12.2511	1582.954	80.672
4	6	−1.456	11.448	−56.606	4.622	−12.2804	1874.431	80.240
5	6	−0.451	8.470	−54.627	4.807	−11.4795	712.442	57.826
6	6	−0.495	6.846	−89.275	4.638	−12.0910	2164.095	82.837
7	4	−0.511	9.931	−64.884	4.774	−10.6927	2095.165	89.627
8	4	−0.564	9.357	−103.404	4.684	−10.8128	2709.861	109.756
9	4	−0.807	14.455	−116.105	4.275	−10.7609	3829.375	98.192
10	4	−0.554	7.650	−76.554	4.732	−10.6664	3609.680	106.579
11	4	−0.575	8.335	−81.338	4.696	−10.7007	2352.300	88.853
12	4	−0.630	10.344	−108.403	4.556	−10.6583	2430.770	89.982
13	4	−0.524	8.926	−68.718	4.848	−10.6077	1765.407	77.525
14	6	−0.513	12.526	−68.367	4.820	−10.6588	2138.119	97.243
15	5	−0.494	12.493	−71.029	4.869	−10.8079	3024.093	91.007
16	5	−0.430	9.959	−67.118	4.938	−10.5935	1761.886	85.002
17	4	−0.102	11.638	−68.467	4.917	−10.6304	1804.067	88.113
18	4	−0.344	9.689	−79.284	4.963	−10.7478	2726.112	101.252
19	5	−0.498	11.711	−67.150	4.882	−10.6339	1885.850	84.466
20	5	−0.500	10590	−61.074	4.896	−10.6033	1314.827	78.066
21	5	−0.499	13.587	−71.980	4.887	−10.7216	3200.689	96.271
22	5	−0.634	8.343	−86.236	4.563	−10.7158	1756.732	81.874
23	5	−0.590	2.662	−94.972	4.700	−10.8222	3233.527	91.524
24	5	−0.609	8.716	−79.982	4.647	−10.7672	1355.869	70.601
25	5	−0.538	6.064	−101.627	4.803	−10.6862	2341.366	88.936
26	5	−02.574	9.091	−121.522	4.752	−10.7849	4186.163	95.353
27	5	−0.600	5.404	−104.455	4.689	−10.7792	2174.262	79.639
28	5	−0.644	4.376	−89.640	4.643	−10.8868	3867.226	100.468
29	5	−0.529	11.688	−75.627	4.821	−10.6707	1839.437	76.937
30	5	−0.614	9.948	−80.248	4.753	−10.7864	3388.506	111.994
31	4	−0.102	12.055	−67.285	4.917	−10.6269	1812.836	77.935
32	4	−0.102	11.602	−66.056	4.917	−10.6252	1666.586	85.763

Model-3

$$\begin{aligned}pKi1 = {} & -2.09387 + 0.002693(H^+/622) + 0.002196(CH_3/836) \\ & + 0.002271(H^+/468) - 0.003839(CH_3/252) \\ & - 0.003123(H^+/619) + 0.003097(H^+/325) \\ & + 0.006558(CH_3/773) + 0.001791(CH_3/626) \\ & - 0.002318(CH_3/702) + 0.000737(H^+/558) \\ & + 0.002443(CH_3/326) - 0.00147(H^+/267) \\ & - 0.002615(H^+/694)\end{aligned}$$

$n = 22;\quad r = 0.996;\quad r^2 = 0.992;\quad r^2_{-CV} = 0.817;$

$PRESS = 0.041;\quad r^2_{-BS} = 0.854;\quad LSE = 0.000;$

$r^2_{-pred} = 0.863$

Model-4

$$\begin{aligned}pKi2 = {} & -1.25823 - 0.001466(CH_3/398) - 0.00548(CH_3/252) \\ & + 0.005139(H^+/475) - 0.002919(CH_3/741) \\ & - 0.00699(CH_3/765) + 0.003645(CH_3/270) \\ & + 0.003501(H^+/622) - 0.006905(H^+/420) \\ & - 0.004784(CH_3/611) - 0.009104(H^+/619) \\ & - 0.005952(CH_3/500) + 0.006988(H^+/604) \\ & - 0.006107(CH_3/542)\end{aligned}$$

$n = 22;\quad r = 0.998;\quad r^2 = 0.996;\quad r^2_{-CV} = 0.998;$

$PRESS = 0.111;\quad r^2_{-BS} = 0.976;\quad LSE = 0.000;$

$r^2_{-pred} = 0.999$

The robust and highly predictive ability of the models is reflected insufficiently only by the cross-validation test, thus the external predictive power of the model was evaluated with the test set molecules after construction of the desired MFA models [31]. The high r^2_{-pred} values for the model-3 and model-4 account for good predictive abilities of generated 3D-QSAR models. The activities of test set were predicted satisfactory by the model-3 and model-4 (Table 2). Then, both of the model-3 and model-4 were used to predict the possible activities of 15 newly designed molecules, further affirming reliability of the results obtained from 2D-QSAR models.

The numbers associated with the MFA fields specify their location in the 3D-grid around the most active molecules **11** and **26** are shown in Fig. 5. The obvious distinction of probes in model-3 and model-4 further proves that the antidepressants have different receptors.

The model-3 explains 99.2% of variances in the activity. The lower PRESS value as well as good r^2_{-CV} value indicates a better predictive credibility compared to corresponding GFA model-1. Molecule **11** with the 3D point grid of model-3 in Fig. 5(a) was taken into account. The presence of steric parameters (CH_3/252 and CH_3/352) close to Ar^1 position indicated the importance of steric interactions and the presence of two electrostatic parameters near the same position namely (H^+/325) with a positive coefficient and (H^+/267) with a negative coefficient describes a subtle balance of electrostatic parameters required at this position [33]. The more presence of electrostatic parameters as moderate electron donating groups in the position of Ar^2, the more reuptake inhibition activities obtained for the antidepressants, which can also be seen in model-3. Hence, the activities of molecules **7**–**32** are generally much higher as compared with molecules **1**–**6** (Table 2).

In model-4, reasonable values of r^2 and PRESS and low least square error (LSE) explain satisfactorily the variances in the activity. Besides,

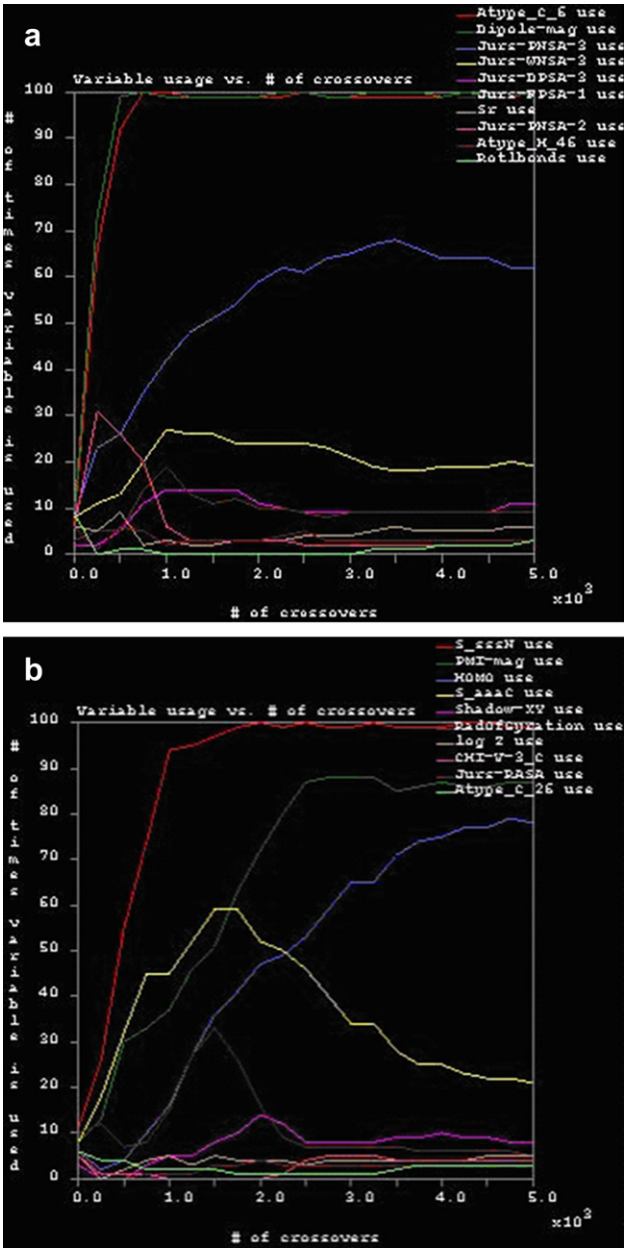


Fig. 4. Descriptor usage graph during generating models.

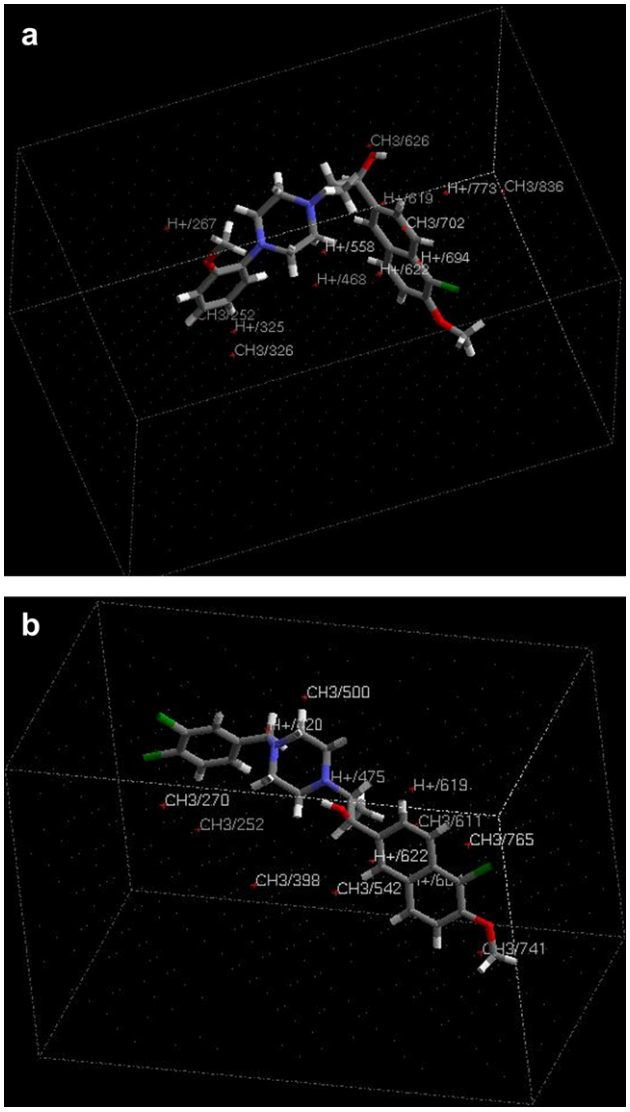


Fig. 5. Molecules **11** (a) and **26** (b) within the 3D point grid of the best 3D-QSAR models, respectively. H^+ represents electrostatic interaction, while CH_3 represents steric interaction.

good r^2_{CV} and r^2_{pred} show that the activities of test set molecules can be predicted with reasonable accuracy (Table 2). Model-4 consists of much more steric parameters (CH_3) than electrostatic parameters (H^+), indicating that moderately bulky substituents are favored for the activities. It is evidently credible from Table 2 that the predicted activities of test set are very close to their actual activities except for molecule **32** for its low activity. Active molecule **26** with the 3D point grid of model-4 can be seen in Fig. 5(b).

4. Conclusions

Statistically significant 2D/3D-QSAR models were generated with the purpose of deriving structural requirements for the antidepressant activities of some aryl alkanol piperazine derivatives. The validation of 2D-QSAR models constructed by GFA was done by the cross-validation tests, randomization tests and external test set prediction. The best 2D-QSAR models indicated that the descriptors of Atype_C_6, Dipole-mag, S_sssCH and Jurs-PNSA-3 mainly influenced the 5-HT reuptake inhibition activity while the descriptors of HOMO, PMI-mag, S_sssN and Shadow-XZ chiefly controlled the NA reuptake inhibition activity. Molecular field analysis (MFA) investigated the substitutional requirements for the favorable receptor–drug interaction and constructed the best 3D-QSAR models by genetic partial least squares (G/PLS) method, providing useful information in the characterization and differentiation of their binding sites. In addition, the conceivable activities of 15 newly designed molecules were predicted and reciprocally validated based on the information provided by the robust 2D/3D-QSAR models, which still need to be verified by actual synthesis. In conclusion, the results derived in present study can provide a preliminary valuable guidance for continuing search for potential antidepressants prior to synthesis.

Acknowledgements

The authors gratefully acknowledge the financial supports for this research by the National Natural Science Foundation of China (NO. 30500339) and the Natural Science Foundation Program of

Zhejiang Province (NO.Y407308) and the Sprout Talented Project Program of Zhejiang Province (NO. 2008R40G2020019).

References

[1] J.J. Crowley, I. Lucki, Curr. Pharm. Des. 11 (2005) 157–169.
[2] M. Köksa, S.S. Bilge, Arch. Pharm. Chem. Life Sci. 340 (2007) 299–303.
[3] F. Henn, B. Vollmayr, A. Sartorius, Drug Discovery Today 1 (2004) 407–411.
[4] S. Gartside, P. Cowen, Psychiatry 5 (2006) 162–166.
[5] D.C. Deecher, C.E. Beyer, G. Johnston, J. Bray, S. Shah, M. Abou-Gharbia, T.H. Andree, J. Pharmacol. Exp. Ther. 318 (2006) 657–665.
[6] C. Sanchez, Drug Discovery Today 3 (2006) 483–488.
[7] J.Q. Li, L.Y. Huang, W.X. Dong, Z.J. Weng, H. Jin, X.L. Ni, S.J. Zhang, C.F. Huang, F.H. Gu, Chin. J. Med. Chem. 16 (2006) 270–276.
[8] S.L. Andersen, M.H. Teicher, Trends Neurosci. 31 (2008) 183–191.
[9] J.C.E. Owen, P.S. Whitton, Brain Res. Bull. 70 (2006) 62–67.
[10] P. Blier, C. de Montigny, Trends Pharmacol. Sci. 15 (1994) 220–226.
[11] A.J. Dean, A. Hendy, T. McGuire, Pharmacoepidemiol. Drug Saf. 16 (2007) 1048–1053.
[12] K.C. Weber, A.B.F. da Silva, Eur. J. Med. Chem. 43 (2008) 364–372.
[13] H. Obata, S. Saito, S. Koizuka, K. Nishikawa, F. Goto, Anesth. Analg. 100 (2005) 1406–1410.
[14] C. Hansch, A. Kurup, R. Garg, H. Gao, Chem. Rev. 101 (2001) 619–672.
[15] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Nature 194 (1962) 178–180.
[16] M.A. Lill, Drug Discovery Today 12 (2007) 1013–1017.
[17] B. Hemmateenejad, R. Miri, M. Akhond, M. Shamsipur, Arch. Pharm. Pharm. Med. Chem. 335 (2002) 472–480.
[18] G.F. Yang, X.Q. Huang, Curr. Pharm. Des. 12 (2006) 4601–4611.
[19] I.C. Muszynski, L. Scapozza, K.A. Kovar, G. Folkers, Quant. Struct.–Act. Relat. 18 (1999) 342–353.
[20] H.J. Kim, H. Choo, Y.S. Cho, H.Y. Koh, K.T. No, A.N. Pae, Bioorg. Med. Chem. 14 (2006) 2763–2770.
[21] P.V. Fish, C. Deur, X.M. Gan, K. Greene, D. Hoople, M. Mackenny, K.S. Para, K. Reeves, T. Ryckmans, C. Stiff, A. Stobie, F. Wakenhut, G.A. Whitlock, Bioorg. Med. Chem. Lett. 18 (2008) 2562–2566.
[22] T.J. Oprea, G.L. Waller, G.R. Marshall, J. Med. Chem. 37 (1994) 2206–2215.
[23] Cerius^2, Version 4.10, A, San Diego, Inc., CA, USA, 2005.
[24] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, J. Am. Chem. Soc. 114 (1992) 10024–10035.
[25] J. Gasteiger, M. Marsili, Tetrahedron 36 (1980) 3291–3228.
[26] Y. Fan, L.M. Shi, K.W. Kohn, Y. Pommier, J.N. Weinstein, J. Med. Chem. 44 (2001) 3254–3263.
[27] D. Rogers, A.J. Hopfinger, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[28] L.M. Shi, F. Yi, T.G. Myers, P.M. O'Connor, K.D. Paull, S.H. Friend, J.N. Weinstein, J. Chem. Inf. Comput. Sci. 38 (1998) 189–199.
[29] J.T. Leonard, K. Roy, Eur. J. Med. Chem. 43 (2008) 81–92.
[30] K.-X. Chen, H.-Y. Xie, Z.-G. Li, J.-R. Gao, Bioorg. Med. Chem. Lett. 18 (2008) 5381–5386.
[31] S. Ramar, S. Bag, N.R. Tawari, M.S. Degani, QSAR Comb. Sci. 26 (2007) 608–617.
[32] A. Hirashima, M. Morimoto, E. Kuwanoa, M. Etoc, Bioorg. Med. Chem. 11 (2003) 3753–3760.
[33] T. Equbal, O. Silakari, M. Ravikumar, Eur. J. Med. Chem. 43 (2008) 204–209.
[34] A. Hirashima, T. Eiraku, E. Kuwano, M. Eto, Internet Electron. J. Mol. Des. 2 (2003) 511–526.
[35] P.C. Nair, M.E. Sobhia, Eur. J. Med. Chem. 43 (2008) 293–299.
[36] S. Deswal, N. Roy, Eur. J. Med. Chem. 41 (2006) 1339–1346.
[37] A.K. Ghose, G.M. Crippen, J. Chem. Inf. Comput. Sci. 27 (1987) 21–35.
[38] R.H. Rohrbaugh, P.C. Jurs, Anal. Chim. Acta 199 (1987) 99–109.
[39] L.H. Hall, B. Mohney, L.B. Kier, Quant. Struct.–Act. Relat. 10 (1999) 43–51.
[40] P. Silakari, S.D. Shrivastava, G. Silakari, D.V. Kohli, G. Rambabu, S. Srivastava, S.K. Shrivastava, O. Silakari, Eur. J. Med. Chem. 43 (2008) 1559–1569.