Hologram quantitative structure activity relationship studies on 5-HT$_6$ antagonists

Munikumar Reddy Doddareddy, Yeon Joo Lee, Yong Seo Cho, Kyung Il Choi, Hun Yeong Koh and Ae Nim Pae*

Biochemicals Research Center, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130-650, Republic of Korea

Abstract—Predictive hologram quantitative structure activity relationship (HQSAR) models were developed for a series of aryl-sulfonamide compounds acting as specific 5-HT$_6$ antagonists. A training set containing 48 compounds served to establish the model. The best HQSAR model was generated using atoms, bond, and connectivity as fragment distinction and 4–7 as fragment size showing cross-validated $r^2(q^2)$ value of 0.702 and conventional r^2 value of 0.971. The predictive ability of the model was validated by an external test set of 20 compounds giving satisfactory predictive r^2 value of 0.678. The efficiency of HQSAR approach was further evidenced by the generation of predictive models for a training set containing 30 highly diverse, both specific and nonspecific 5-HT$_6$ antagonists. The best HQSAR model for this training set was generated using atoms, bond, and connectivity as fragment distinction and 4–7 as fragment size showing cross-validated $r^2(q^2)$ value of 0.693 and conventional r^2 value of 0.923. This model was also validated by using an external test set of 10 compounds giving satisfactory predictive r^2 value of 0.692. The contribution maps obtained from these models were used to explain the individual atomic contributions to the overall activity.
© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

The 5-HT$_6$ receptor is one of the most recent additions to the 5-hydroxytryptamine super family of receptors consisting of seven classes (5-HT$_1$–5-HT$_7$) that contain totally 14 human subclasses.[1] It was first isolated from rat striatal mRNA[2] in 1993 and subsequently the human 5-HT$_6$ gene was cloned and characterized by Kohen et al.[3] in 1994. The 5-HT$_6$ receptor is a seven trans-membrane 440 amino acid polypeptide, which is positively coupled to the adenylate cyclase secondary messenger system.[4] Evaluation of the expression pattern of the 5-HT$_6$ receptor mRNA and protein revealed that it is selectively expressed in the central nervous system, exhibiting wide spread distribution throughout the brain. This intriguing distribution in the brain, together with its high affinity for a wide range of drugs used in the psychiatry[5,6] stimulated significant interest recently.

Most atypical antipsychotic drugs, which lack extra-pyrimidal side effects, bind with very high affinity to the 5-HT$_6$ receptor. In fact, the prototypic atypical antipsychotic agent clozapine, exhibits greater affinity for the 5-HT$_6$ receptor than for any other receptor subtype. Initial in vivo experiments showed that administration of antisense oligonucleotides (AOs), directed at 5-HT$_6$ receptor mRNA, elicited a behavioral syndrome in rats consisting of yawning, stretching, and chewing, which could be dose dependently blocked by the muscarinic antagonist atropine.[7] This study implies that 5-HT$_6$ receptors modulate cholinergic neurotransmission and hence 5-HT$_6$ receptor antagonists may be useful for the treatment of memory dysfunction. In addition, treatment with AOs significantly inhibited the increase in 5-HT release from the prefrontal cortex produced by conditioned fear stress, suggesting that 5-HT$_6$ receptors may be involved in certain anxiety disorders.[8]

As a part of ongoing work in our lab aimed at the discovery of new 5-HT$_6$ antagonists, we tried to generate predictive QSAR models for different training sets containing both structurally similar and diverse compounds.

Hologram QSAR[9] is a modern QSAR technique developed from unity hashed fingerprint concept,[10] which employs specialized fragment finger prints as predictive variables of biological activity. The premise of HQSAR is that since the structure of a molecule is encoded within its 2D fingerprint and that structure is the key determinant of all molecular properties (including biological activity), then it should be possible to predict the activity from its fingerprint. This assumption is validated, to some extent, by the high degree of success of 2D similarity searching[11] of chemical databases in which the similarity measure is Tanimoto coefficient[12] between finger prints. HQSAR uses an extended form of fingerprint, known as a molecular hologram, which encodes more information, for example, branched and cyclic fragments as well as stereochemistry, than the traditional 2D fingerprint. The key difference, however, is that a molecular hologram contains all possible molecular fragments within a molecule, including overlapping fragments, and maintains a count of the number of times each fragment occurs. This process of incorporating information about each fragment, and each of its constituent sub-fragments, implicitly encodes 3D structural information. Given biological activity data for each of the compounds in a dataset, HQSAR involves a partial least square analysis[13] (PLS) of the molecular holograms to derive a QSAR that can then be used to predict the activity of molecules outside the training set. HQSAR has several potential advantages over existing methods for QSAR. It avoids not only the need for molecular alignment and conformation specification inherent in CoMFA[14] and CoMSIA,[15] but also the selection and calculation or measurement of the physicochemical descriptors required by classical QSAR. In addition, HQSAR is sufficiently rapid in execution and general in approach to enable analyses to be carried out on large datasets, such as combinatorial libraries or database subsets that are not amenable to analysis by existing QSAR methods.

2. Materials and methods

2.1. Datasets

Two datasets were used for the analysis, one comprising of structurally similar arylsulfonamide derivatives (Dataset 1) acting as specific 5-HT$_6$ antagonists obtained from four various papers published by Bromidge and co-workers[16] and another containing structurally highly diverse, both specific and nonspecific 5-HT$_6$ antagonists (Dataset 2) obtained from a review published by Glennon.[17] Dataset 1 contained about 68 compounds (Tables 1 and 2), 48 compounds of which were used as training set and 20 compounds as test set (Compounds shown in bold letters). Dataset 2 (Figs. 1 and 2) contained about 40 highly diverse compounds including tricyclic antipsychotics like chlorpromazine **69** and loxapine **70**, butyrophenones such as spiperone **71**, and haloperidol **72** and selective arylsulfonamide compounds like **76**, **77**, and **78** etc. Total of 30 compounds

served as training set (Fig. 1) and remaining 10 compounds were used as test set (Fig. 2). Sufficient care was taken in the selection of test sets for both the datasets, so that representatives of all compounds were included for prediction. The main purpose of the Dataset 2 was to test the success of HQSAR methodology in generating predictive models when highly diverse training set is used.

2.2. HQSAR analysis

HQSAR analysis involves three main steps: the generation of substructural fragments for each of the molecules in the training set; the encoding of these fragments in holograms; and correlation of the latter with the available biological data. For this purpose we used the novel molecular hologram representation devised by Tripos associates as generated by the HQSAR package.[18]

The input molecule is broken into all possible structural fragments (including branched, cyclic, and overlapping fragments) containing user defined minimum (M) and maximum (N) number of atoms. The exact number and nature of the fragments are determined by the values of M and N. Each unique fragment in the dataset is assigned a specific large integer by means of cyclic redundancy check (CRC) algorithm. Each of these integers corresponds to a bin in an integer array of fixed length L (L is generally in the range 50–500). Bin occupancies are incremented according to the fragments generated. Thus, all generated fragments are hashed[19] into array bins in the range 1 to L. This array is called molecular hologram,[20] and bin occupancies are the descriptor variables. The use of hashing greatly reduces the size of molecular hologram but leads to a phenomenon called 'fragment collision'. During fragment generation, identical fragments are always hashed to same bin, and the corresponding occupancy for that bin is incremented. However, as the hologram length is generally smaller than the total number of unique fragments, different unique fragments can hash to the same bin causing 'collision' between fragments. In order to reduce the probability of identical or similar fragment collisions occurring, values of L are selected to be prime numbers (default values of which are 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401). Computation of the molecular holograms for a dataset of structures yields a data matrix of dimension $R \times L$, where R is the number of compounds in the training set and L is the length of the molecular hologram. Standard PLS analysis is then applied to identify a set of orthogonal explanatory variables (components) that are linear combinations of the original L variables. Leave one out[21] cross-validation is applied to determine the number of components that yield optimally predictive model.

PLS models were selected from cross-validation results on the basis of the first SE$_{CV}$ (cross-validated standard error)-minimum rather than the first q^2 maximum, a procedure that generally favors model parsimony and

Table 1. Structures and activities of the Dataset 1[16]

I   II   III

Series I

Compd	Ar¹	R¹	pK_i
1	phenyl	$-N\_N\cdot Me$	8.00
2	Me-triazolyl	$-N\_N\cdot Me$	6.10
3	Br-phenyl	$-N\_N\cdot Me$	8.30
4	Cl-benzothiophene-Me	octahydropyrrolo-pyridinyl (H)	9.10
5	Cl-benzothiophene-Me	octahydropyrrolo-pyridinyl (H)	9.10
6	Cl-naphthyl	$-N\_N\cdot Me$	9.10
7	Cl-benzothiophene-diMe	$-N\_N\cdot Me$	8.90
8	Cl-benzothiophene-Me	$-N\_NH$	8.90
9	Cl-benzothiophene-Me	$-N\_N\cdot Me$	9.20
10	I-phenyl	$-N\_N\cdot Me$	8.60
11	F,Br,Br,F,Me-phenyl	$-N\_N\cdot Me$	8.50

Series II

Compd	2	3	4	5	6	pK_i
12	Br	F		Br		8.50
13	Br					8.80
14		Cl				8.90
15		Br				8.90
16	Br			Br		9.00
17	OMe			Cl		8.90
18	Me			I		9.10
19	OMe			CF₃		8.40
20	CF₃		Cl			7.60
21		Cl	Cl			8.60
22	F		Br		F	6.80
23	OMe		Cl	Me		8.70
24	Cl		Cl	Cl		8.30
25	Cl	Cl		Cl		8.70
26	Br	Br		Br		8.60
27	Cl	Br		Cl		8.70
28	Cl	Cl		Br		8.40
29	Cl	Br		Br		8.60
30	F	Br		Br		8.60
31	Cl	F		F		7.50
32	i-Pr					9.00
33		I				9.30
34			I			8.70
35		Cl		Cl		9.20
36		Br		Br		9.30
37	Br	Cl		Cl		8.50
38	Cl	F		Br		8.50
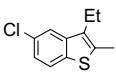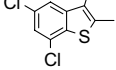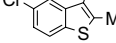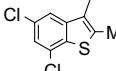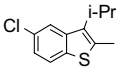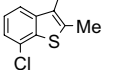
(Note: column headers for Series II R² positions are 2, 3, 4, 5, 6.)

Series III

Compd	Ar²	pK_i
39	I-indoline	9.50
40	Me,F₃C-indoline	8.40
41	Br-tetrahydroquinoline	8.50
42	Cl,Cl-tetrahydroquinoline	9.10
43	OMe,OMe-tetrahydroisoquinoline	8.80
44	I-indoline	8.40
45	Me,Me-indoline	9.10
46	Br-tetrahydroquinoline	9.50
47	Cl-tetrahydroisoquinoline	8.90

Note. Activity data was obtained from four various papers published by Bromidge et al.[16] Compounds in bold letters were used as test set.

which can help prevent training set overfit.[22] An alternative rule of the thumb, the '5% rule', was also applied wherein an additional LV (number of latent variables or optimal number of components) is permitted only where q^2 rises by 5% or more. The automated HQSAR program picks models on the basis of either the overall SE_{CV}-minimum (or overall q^2-maximum), a procedure that can result in the selection of very large numbers of LV leading to rise in q^2 values but not test set predictability. In order to limit the opportunities for chance correlation and training set overfit, the maximum number of LVs (LV_{max}) extracted was limited to $R/4$, where R is the number of training set compounds.[22] For the final model, the QSAR analysis was redone with the number of components set to the optimal number of components identified through cross-validation and the number of cross-validation groups set to zero. Once an optimal model is identified, PLS yields a mathematical equation that relates the molecular hologram bin values to the corresponding biological activity of each compound in the training set.

$$\text{Activity}_i = c_o + \Sigma_L c_{il} x_{il} \qquad (1)$$

where x_{il} is the occupancy value of the molecular hologram of compound i at position or bin l, c_{il} is the coefficient for that bin derived from the PLS analysis, L is the length of the hologram, Activity_i is the biological activity, and c_o is a constant.

Table 2. Structures and activities of the Dataset 1[16]

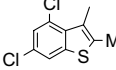|	IV					V			
Compd	Ar³	pK_i	Compd	Ar	pK_i	Compd	R³	R⁴	pK_i
48	(3,5-dichlorophenyl)	7.60	55	(Cl-benzothiophene, Et, Me)	8.90	62	(Me-piperazine)		8.70
49	t-Bu-phenyl	9.20	56	(Cl₂-benzothiophene, Me)	8.20	63	(pyrrolidino-piperazine)		8.70
50	Ph-phenyl	9.30	57	(Cl-benzothiophene, Me, Me)	9.00	64[a]	−N N·Me		7.1
51	(Cl-naphthyl)	9.00	58	(Cl₂-benzothiophene, Me, Me)	8.70	65	−N NH	2-Methyl	8.90
52	(Cl-naphthyl)	8.50	**59**	(Cl-benzothiophene, i-Pr, Me)	8.80	66	(Me-piperazine)	3-Methyl	8.60
53	(Cl-benzofuran, Me)	8.40	**60**	(Cl-benzothiophene, Me, Me)	8.70	**67**	(Me-piperazine)		8.90
54	(Cl-benzothiophene, Me)	8.70	**61**	(Cl₂-benzothiophene, Me, Me)	8.10	**68**	−N N·Me		9.00

Note. Activity data was obtained from four various papers published by Bromidge et al.[16] Compounds in bold letters were used as test set.
[a] Methyl piperazine substituent is at C₃ position.

2.3. HQSAR contribution maps

The results of the HQSAR analysis is graphically displayed as a color-coded structure diagram in which the color of each atom reflects the contribution of that atom to the molecules overall activity. The colors at the red end of the spectrum (red, red-orange, and orange) reflect poor (or negative) contributions, while colors at the green end (yellow, green-blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are colored white.

2.4. Predictive r squared (r^2_{pred})

To validate the derived HQSAR models, biological activities of an external test set were predicted using models derived from the training set. The predictive ability of the models is expressed by predictive r^2 value, which is analogous to cross-validated $r^2(q^2)$ and is calculated by using the formula

$$r^2_{pred} = \frac{SD - PRESS}{SD} \qquad (2)$$

where SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules.

3. Results and discussion

For both the training sets, the analyses were first performed using the default fragment size of 4–7 for different combinations of fragment distinction parameters; atom types (A), bond types (B), connectivity (C), hydrogens (H), and chirality (Ch). The max number of latent variables (LV_{max}) allowed was limited to 12 in case of 48 compound training set (Dataset 1) and to 8 in case of 30 compound training set (Dataset 2) to limit chance correlation and training set overfit. The best hologram length (default values 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401) and optimum number of components (LV) were selected based on the PLS analyses that gave least cross-validated standard error SE_{CV}. The analyses were then repeated with different fragment sizes using the best fragment distinction obtained from previous step to check its influence on key statistical parameters. Finally to get a more robust model and to reduce the noise, the '5% rule' was used and the model with high q^2 value and a lesser number of LV was selected as best.
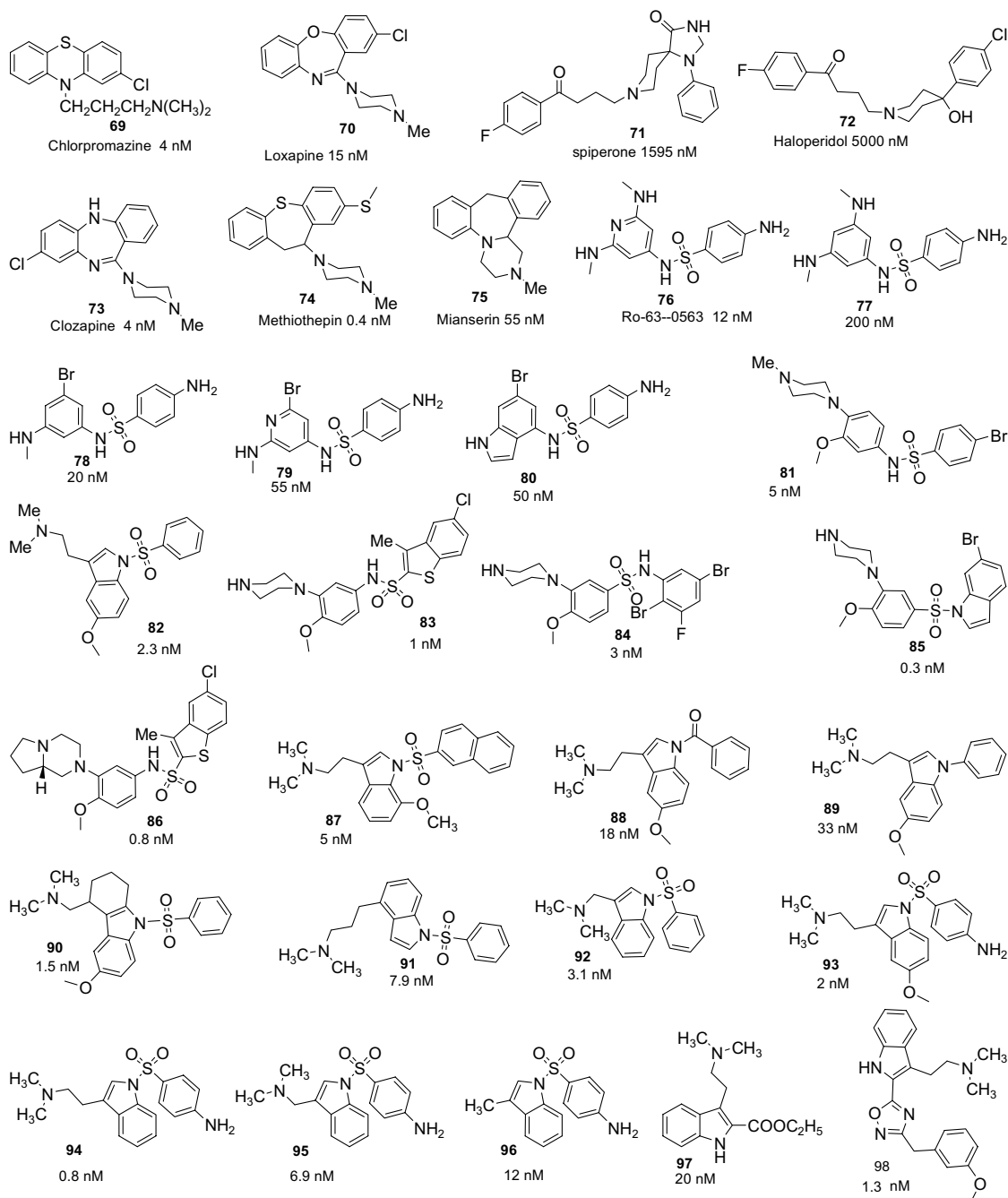
Figure 1. Structures and activities of Dataset 2[17] (training set).

3.1. Model 1 (Dataset 1)

Total of 48 structurally similar arylsulfonamide compounds acting as specific 5-HT$_6$ antagonists served as training set to establish the model. The best HQSAR model for this dataset was generated using atoms, bond, and connectivity as fragment distinction and 4–7 as fragment size showing cross-validated $r^2(q^2)$ value of 0.702 and noncross-validated r^2 value of 0.971 (Tables 3–5). The selected model was validated by an external test set of 20 compounds (compounds shown as bold in Tables 1 and 2) giving satisfactory predictive r^2 value of 0.678. The observed versus predicted activities of both

training set and test set by this model were shown in Table 8. All compounds of the test set were fairly predicted with residual values less than one log unit. Figure 3 shows the graph of observed versus predicted activities of both training set and test set.

3.2. Model 2 (Dataset 2)

Total of 30 structurally highly diverse, both specific and nonspecific 5-HT$_6$ antagonists served as training set to establish the model. The best HQSAR model for this training set was generated using atoms, bond, and

Figure 2. Structures and activities of Dataset 2[17] (test set).
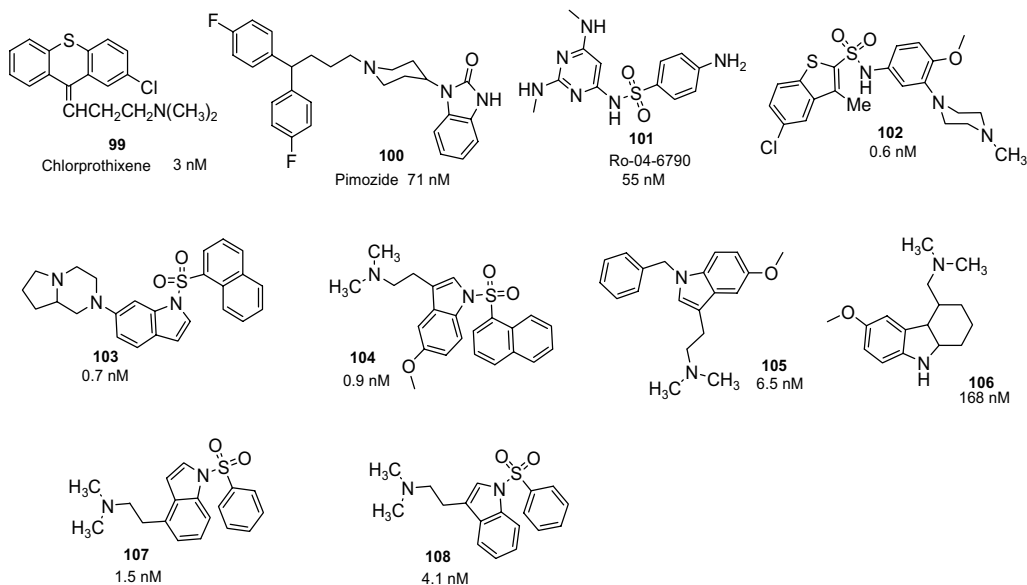
Table 3. Model 1: HQSAR analysis for various *fragment distinction* using default *fragment size* (4–7); ($LV_{max} = 12$)

Fragment distinction	q^2	SE_{CV}	r^2	SE	LV	Length
A/B	0.370	0.551	0.943	0.166	6	257
A/B/C	**0.746**	**0.369**	**0.983**	**0.095**	**10**	**59**
A/B/H	0.582	0.472	0.975	0.114	10	199
A/B/C/H	0.506	0.528	0.986	0.088	12	353
A/C/D&A	0.140	0.629	0.669	0.390	4	61
A/C/Ch	0.505	0.514	0.977	0.111	10	61
A/B/Ch	0.392	0.548	0.952	0.154	7	401
A/B/C/H/Ch	0.486	0.531	0.980	0.105	11	83

q^2—LOO cross-validated correlation coefficient, SE_{CV}—cross-validated standard error, r^2—noncross-validated correlation coefficient, SE—noncross-validated standard error, LV—latent variables or optimal number of components, length—hologram length. *Fragment distinction*: A—atom types, B—bond types, C—connectivity, H—hydrogens, D&A—donor and acceptor, Ch—chirality.

Table 4. Model 1: HQSAR analysis for the influence of various *fragment sizes* using the best *fragment distinction* (A/B/C)

Fragment size	q^2	SE_{CV}	r^2	SE	LV	Length
2–5	0.551	0.477	0.919	0.203	8	71
3–6	0.410	0.547	0.967	0.129	8	199
4–7	**0.746**	**0.369**	**0.983**	**0.095**	**10**	**59**
5–8	0.277	0.583	0.925	0.188	5	151
6–9	0.295	0.576	0.869	0.249	5	61
7–10	0.334	0.596	0.985	0.089	10	151

Table 5. Model 1: Selection of best model with less number of LVs using '5% rule'

LV	q^2	SE_{CV}	r^2	SE	Length	Percentage[a]
10	0.746	0.369	0.983	0.095	59	2.7
9	0.726	0.377	0.979	0.105	59	3.4
8	**0.702**	**0.389**	**0.971**	**0.121**	**59**	**5.1**
7	0.668	0.405	0.966	0.130	59	**20.5**
6	0.554	0.463	0.955	0.148	59	—

[a] Percentage raise in q^2 value with additional LV or optimal number of components.

connectivity as fragment distinction and 4–7 as fragment size showing cross-validated $r^2(q^2)$ value of 0.693 and noncross-validated r^2 value of 0.923 (Tables 6 and 7).

The selected model was validated by an external test set of 10 compounds containing representatives of all training set compounds (Fig. 2) giving satisfactory
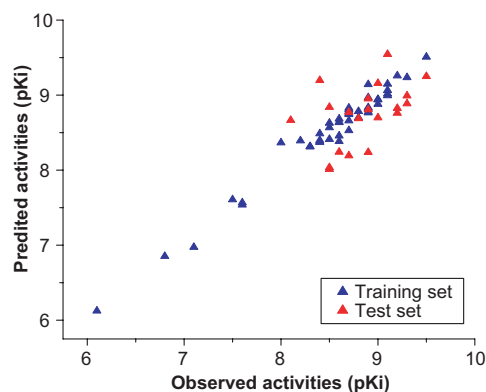
Figure 3. Plot of observed versus predictive activities of Dataset 1 using Model 1.

Figure 4. Plot of observed versus predictive activities of Dataset 2 using Model 2.

predictive r^2 value of 0.692. The observed versus predicted activities of both training set and test set by this model were shown in Table 9. All compounds of th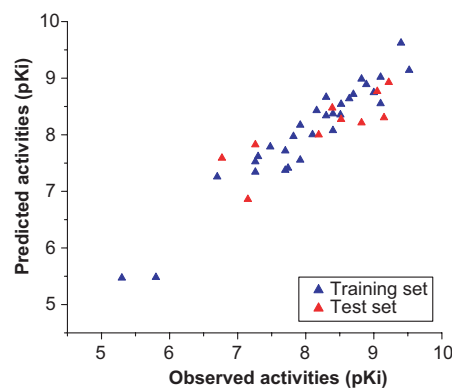e test set were fairly predicted with residual values less than one log unit. Figure 4 shows the graph of observed versus predicted activities of both training set and test set.

3.3. Contribution maps

Two predictive HQSAR models were generated; one using specific 5-HT$_6$ antagonists and another was using both specific and nonspecific antagonists. The results of which were graphically represented in the form of contribution maps in which the color of each atom reflects the contribution of that atom to the molecules overall activity. For discussion point of view, we selected SB-357134,[16b] which is present in both the datasets that is **12** in case of Dataset 1 and **84** in case of Dataset 2. Compound SB-357134 is chemically *N*-(2,5-dibromo-3-fluorophenyl)-4-methoxy-3-piperazin-1yl-benzenesulfonamide acting as a potent, selective, brain penetrant, and orally

active 5-HT$_6$ receptor antagonist showing over 200-fold selectivity versus 72 other receptors and enzymes.[23]

Depending on the training sets used, the two models gave two different contribution maps for SB-357134 (Figs. 5 and 6). Both maps gave some interesting details about the individual contribution of the atoms in the compound toward the overall activity. It is interesting to see that in the contribution map of Model 2 (Fig. 6) pharmacologically important group[24] the basic nitrogen moiety is colored yellow indicating its positive contribution to the activity.

Further more sulfur and nitrogen of sulfonamide group, which is the most common feature present in almost all selective 5-HT$_6$ antagonists[16,17,25] known until now were also colored yellow in both the maps. But as the basic nitrogen is almost invariant in the training set of the Dataset 1, it is colored white in the contribution map of Model 1 (Fig. 5). The contribution map of Model 1 shows the other nitrogen (secondary) as negatively contributing atom and colored orange-red. Interestingly the negative effect of activity due to the presence of electronegative atoms like N and O at this position was

Table 6. Model 2: HQSAR analysis for various *fragment distinctions* using default *fragment size* (4–7); (LV$_{max}$ = 8)

Fragment distinction	q^2	SE$_{CV}$	r^2	SE	LV	Length
A/B	0.646	0.631	0.951	0.236	5	71
A/B/C	**0.693**	**0.564**	**0.923**	**0.283**	**3**	**61**
A/B/H	0.508	0.715	0.823	0.429	3	83
A/B/C/H	0.587	0.655	0.833	0.417	3	61
A/C/D&A	0.472	0.771	0.945	0.248	5	97

Table 7. Model 2: HQSAR analysis for the influence of various *fragment sizes* using the *best fragment* distinction (A/B/C)

Fragment size	q^2	SE$_{CV}$	r^2	SE	LV	Length
2–5	0.610	0.649	0.907	0.317	4	83
3–6	0.565	0.685	0.915	0.303	4	61
4–7	**0.693**	**0.564**	**0.923**	**0.283**	**3**	**61**
5–8	0.593	0.663	0.953	0.225	4	83
5–9	0.520	0.706	0.920	0.288	3	83
7–10	0.320	0.857	0.899	0.330	4	53

Table 8. Observed versus predicted activities of Dataset 1 using Model 1

Compd	Observed pK_i	Predicted pK_i	Residual
Training			
1	8.00	8.366	−0.366
2	6.10	6.124	−0.024
3	8.30	8.316	−0.016
4	9.10	9.017	0.083
5	9.10	9.017	0.083
6	9.10	9.064	0.036
7	8.90	8.967	−0.067
12	8.50	8.412	0.088
13	8.80	8.701	0.099
14	8.90	8.766	0.134
15	8.90	8.835	0.065
16	9.00	8.879	0.121
17	8.90	9.142	−0.242
18	9.10	9.149	−0.049
19	8.40	8.397	0.003
20	7.60	7.538	0.062
21	8.60	8.458	0.142
22	6.80	6.853	−0.053
23	8.70	8.743	−0.043
24	8.30	8.312	−0.012
25	8.70	8.529	0.171
26	8.60	8.641	−0.041
27	8.70	8.828	−0.128
28	8.40	8.375	0.025
29	8.60	8.684	−0.084
30	8.60	8.384	0.216
31	7.50	7.608	−0.108
39	9.50	9.509	−0.009
40	8.40	8.487	−0.087
41	8.50	8.628	−0.128
42	9.10	8.994	0.106
43	8.80	8.784	0.016
48	7.60	7.567	0.033
49	9.20	9.259	−0.059
50	9.30	9.236	0.064
51	9.00	8.944	0.056
52	8.50	8.568	−0.068
53	8.40	8.375	0.025
54	8.70	8.757	−0.057
55	8.90	8.819	0.081
56	8.20	8.392	−0.192
57	9.00	8.939	0.061
58	8.70	8.660	0.040
62	8.70	8.807	−0.107
63	8.70	8.784	−0.084
64	7.10	6.973	0.127
65	8.90	8.775	0.125
66	8.60	8.637	−0.037
Test			
8	8.90	8.952	−0.052
9	9.20	8.825	0.375
10	8.60	8.242	0.358
11	8.50	8.012	0.488
32	9.00	9.158	−0.158
33	9.30	8.991	0.309
34	8.70	8.772	−0.072
35	9.20	8.761	0.439
36	9.30	8.887	0.413
37	8.50	8.839	−0.339
38	8.50	8.035	0.465
44	8.40	9.197	−0.797
45	9.10	9.545	−0.445
46	9.50	9.250	0.250

Table 8 (*continued*)

Compd	Observed pK_i	Predicted pK_i	Residual
47	8.90	8.238	0.662
59	8.80	8.689	0.111
60	8.70	8.193	0.507
61	8.10	8.665	−0.565
67	8.90	8.807	0.093
68	9.00	8.699	0.301

Table 9. Observed versus predicted activities of Dataset 2 using Model 2

Compd	Observed pK_i	Predicted pK_i	Residual
Training			
69	8.398	8.076	0.322
70	7.824	7.972	−0.148
71	5.797	5.482	0.315
72	5.301	5.470	−0.169
73	8.398	8.372	0.026
74	9.398	9.621	−0.223
75	7.260	7.524	−0.264
76	7.921	7.554	0.367
77	6.699	7.257	−0.558
78	7.699	7.375	0.324
79	7.260	7.341	−0.081
80	7.301	7.616	−0.315
81	8.301	8.337	−0.036
82	8.638	8.639	0.001
83	9.000	8.747	0.253
84	8.523	8.540	−0.017
85	9.523	9.141	0.382
86	9.097	9.019	0.078
87	8.301	8.663	−0.362
88	7.745	7.411	0.334
89	7.481	7.787	−0.306
90	8.824	8.986	−0.162
91	8.102	8.002	0.100
92	8.509	8.352	0.157
93	8.699	8.716	−0.017
94	9.097	8.551	0.546
95	8.161	8.430	−0.269
96	7.921	8.172	−0.251
97	7.699	7.717	−0.018
98	8.886	8.889	0.003
Test			
99	8.523	8.273	0.250
100	7.149	6.860	0.289
101	7.260	7.824	−0.564
102	9.222	8.926	0.296
103	9.155	8.303	0.852
104	9.046	8.767	0.279
105	8.187	7.999	0.188
106	6.775	7.588	−0.813
107	8.824	8.212	0.612
108	8.387	8.474	−0.087

also observed in *N*-arylsulfonylindoles acting as 5-HT$_6$ antagonists.[25] Table 10 shows a decreasing trend in activity with the presence of more electronegative atoms in case of *N*-arylsulfonylindole derivatives at the position explained above. The other atom, which is colored orange-red in the contribution map of Model 1 is fluorine atom at 3-position of the aromatic ring. The
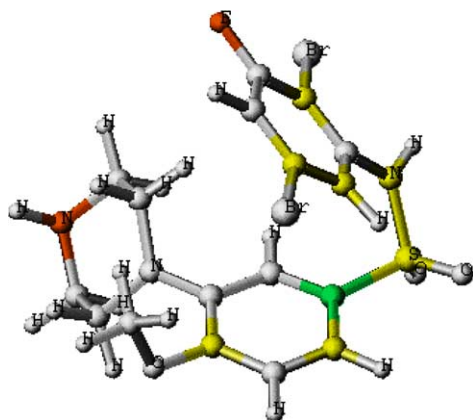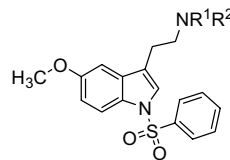
Figure 5. Contribution map of SB-357134 (**12**) from Model 1.

Table 10. *N*-Arylsulfonyl derivatives as 5-HT$_6$ antagonists[25]

Compd	NR^{1}R^2	K_i (nM)^a
1	NMe$_2$	2.3
2	Pyrrolidinyl	44
3	Piperidinyl	350
4	4-Methyl-piperazinyl	490
5	Morpholinyl	1700

^{a} Clozapine was used as control ($K_i = 13$).[25]

activity data of this series of compounds (Table 1: Structure II) clearly shows that the dihalide substitutions at 2,5- or 3,5-positions or just monohalide substitution at 3-position are preferred when compared to 2,3,5-trihalide substitution. This may be the reason why the fluorine atom is colored orange-red in SB-357134, which is a 2,3,5-trihalide substituted compound.

Similarly contribution maps of other compounds also gave some idea about individual contribution of atoms to the overall activity. For example, the imidazole ring of compound **2** was colored completely red indicating its negative effect on the activity. Whereas 5-chloro-3-methyl-benzothiophene ring of compounds **4**, **5**, **7**, **8**, and **9** was colored yellow or green indicating its positive contribution to the activity. The oxygen of the methoxy group (Table 1: Structures I, II, and III) and nitrogen of the quinoline ring system (Table 2: Structures IV and V) both of which act as a hydrogen bond acceptor group were colored white as they are invariant in the training set. The information obtained from these contribution maps can be used in further development of specific 5-HT$_6$ antagonists.

4. Conclusions

Predictive HQSAR models were developed for two datasets acting as 5-HT$_6$ antagonists. Both models were validated by external test sets giving satisfactory Predictive r^2 values. The contribution maps obtained from these models were used to explain the individual contribution of the atoms to the overall activity of the compound.

Acknowledgements

References and notes

1. Hoyer, D.; Martin, G. 5-HT Receptor Classification and Nomenclature: Towards a Harmonisation with the Human Genome. *Neuropharmacology* **1997**, *36*, 419–428; Hoyer, D.; Clarke, D. E.; Fozard, J. R.; Hartig, P. R.;
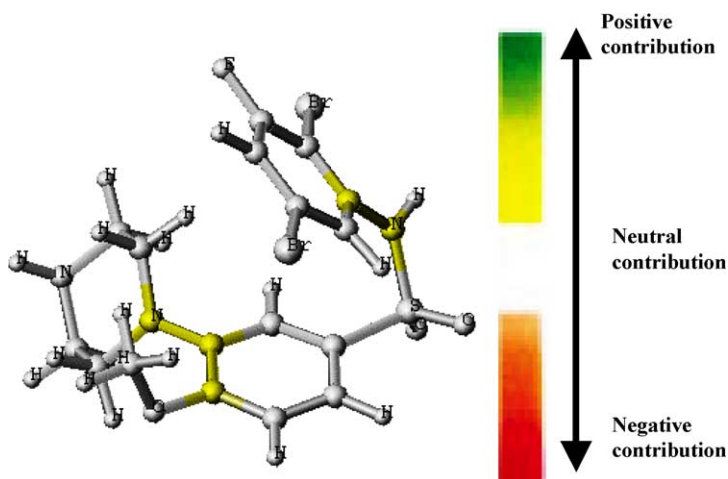

Figure 6. Contribution map of SB-357134 (**84**) from Model 2.

Martin, G. R.; Mylecharane, E. J.; Saxena, P. R.; Humphrey, P. P. A. International Union of Pharmacology Classification of Receptors for 5-Hydroxytryptamine (Serotonin). *Pharmacol. Rev.* **1994**, *46*, 157–204.

2. (a) Monsma, F. J.; Shen, Y.; Ward, R. P.; Hamblin, M. W.; Sibley, D. R. Cloning and Expression of a Novel Serotonin Receptor with High Affinity for Tricyclic Psychotropic Drugs. *Mol. Pharmacol.* **1993**, *43*, 320–327; (b) Ruat, M.; Traiffort, E.; Arrang, J.-M.; Tardivel-Lacombe, J.; Diaz, J.; Leurs, R.; Schwartz, J.-C. A Novel Serotonin (5-HT_6) Receptor: Molecular Cloning, Localisation and Stimulation of cAMP Accumulation. *Biochem. Biophys. Res. Commun.* **1993**, *193*, 269–276.

3. Kohen, R.; Metcalf, M. A.; Druck, T.; Huebner, K.; Sibley, D. R.; Hamblin, M. W. Cloning and Chromosomal Localization of a Human 5-HT_6 Serotonin Receptor. *Soc. Neurosci. Abstr.* **1994**, *20*, 4768.

4. Sleight, A. J.; Boess, F. G.; Bos, M.; Levet-Trafit, B.; Bourson, A. *Exp. Opin. Ther. Patents* **1998**, *8*, 1217.

5. Roth, B. L.; Craigo, S. C.; Choudhary, M. S.; Uluer, A.; Monsma, F. J.; Shen, Y.; Meltzer, H. Y.; Sibley, D. R. *J. Pharmacol. Exp. Ther.* **1994**, *268*, 1403.

6. Glatt, C. E.; Snowman, A.; Sibley, D. R.; Snyder, S. H. *Mol. Med.* **1995**, *1*, 398.

7. Bourson, A.; Borroni, E.; Austin, R. H.; Monsma, F. J.; Sleight, A. J. Determination of the Role of the 5-HT_6 Receptor in the Rat Brain: a Study Using Antisense Oligonucleotides. *J. Pharmacol. Exp. Ther.* **1995**, *274*, 173–180; Sleight, A. J.; Monsma, F. J.; Borroni, E.; Austin, R. H.; Bourson, A. Effects of Altered 5-HT_6 Expression in the Rat: Functional Studies Using Antisense Oligonucleotides. *Behav. Brain Res.* **1996**, *73*, 245–248.

8. Yoshioka, M.; Matsumoto, M.; Togashi, H.; Mori, K.; Saito, H. Central Distribution and Function of 5-HT_6 Receptor Subtype in the Rat Brain. *Life Sci.* **1998**, *62*, 1473–1477.

9. (a) Lowis, D. R. HQSAR a new, highly predictive QSAR technique. http://www.tripos.com/AppNotes/HQSAR/hqsar.html; (b) Heritage, T. W.; Lowis, D. R. Molecular Hologram QSAR. In *Rational Drug Design: Novel Methodology and Practical Applications*; ACS Symposium Series; 1999; Vol. 719.

10. Unity Chemical Information Software Ver 2.3. Tripos Associates: 1699 South Hanley Road, Suite 303, St. Louis, MO 63144. http://www.tripos.com.

11. Ash, S.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. *J. Chem. Inf. Comp. Sci.* **1997**, *37*, 71.

12. Tanimoto coefficient as similarity measure at URL http://www.tripos.com/sciTech/researchCollab/chemCompLib/mdiversity.html.

13. Wold, S.; Abano, C.; Dunn, W. J. Multivariate Data Analysis in Chemistry. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Netherlands, 1984.

14. Crammer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

15. Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130–4146, 102, 1239–1248.

16. (a) Bromidge, S. M.; Brown, A. M.; Clarke, S. E.; Dodgson, K.; Gager, T.; Grassam, H. L.; Jeffrey, P. M.; Joiner, G. F.; King, F. D.; Middlemiss, D. N.; Moss, S. F.; Newmann, H.; Riley, G.; Routledge, C.; Wyman, P. *J. Med. Chem.* **1999**, *42*, 202; (b) Bromidge, S. M.; Clarke, S. E.; Gager, T.; Griffith, K.; Jeffrey, P.; Jennings, A. J.; Joiner, G. F.; King, F. D.; Lovell, P. J.; Moss, S. F.; Newman, H.; Riley, G.; Rogers, D.; Routledge, C.; Serafinowska, H. T.; Smith, D. R. *Bioorg. Med. Chem. Lett.* **2000**, *11*, 55; (c) Bromidge, S. M.; Griffth, K.; Heightman, T. D.; Jennings, A. J.; King, F. D.; Moss, S. F.; Newman, H.; Riley, G.; Routledge, C.; Serafinowska, H. T.; Thomas, D. R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2843; (d) Bromidge, S. M.; Clarke, S. E.; King, F. D.; Lovell, P. J.; Newman, H.; Riley, G.; Routledge, C.; Serafinowska, H. T.; Smith, D. R.; Thomas, D. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1357.

17. Glennon, R. A. *J. Med. Chem.* **2003**, *46*, 2795.

18. HQSAR Software. Tripos Associates: 1699 South Hanley Road, Suite 303, St. Louis, MO 63144.

19. Knuth, D. E. *Sorting and Searching*; Adison-Wesley: Reading, MA, 1973.

20. Flower, D. R. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 379.

21. (a) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. *Quant. Struct. Act. Relat.* **1988**, *7*, 18; (b) Wold, S. *Technometrics* **1978**, *4*, 397.

22. Wold, S.; Johansson, E.; Cocchi, M. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubiniyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 523–550.

23. Stean, T. O.; Hirst, W. D.; Thomas, D. R.; Price, G. W.; Rogers, D.; Riley, G.; Bromidge, S. M.; Serafinowska, H. T.; Smith, D. R.; Bartlett, S.; Deeks, N.; Duxon, M.; Upton, N. *Pharmacol. Biochem. Behav.* **2002**, *71*, 645.

24. (a) Boess, F. G.; Monsma, F. J., Jr; Meyer, V.; Zwingelstein, C.; Sleight, A. J. *Mol. Pharmacol.* **1997**, *52*, 515–523; (b) Kristiansen, K.; Kroaze, W. K.; Willins, D. L.; Gelber, E. I.; Savage, J. E.; Glennon, R. A.; Brian, L. R. *J. Pharmacol. Exp. Ther.* **2000**, *293*, 735–746; (c) Shapiro, D. A.; Kristiansen, K.; Kroaze, W. K.; Roth, B. L. *Mol. Pharmacol.* **2000**, *58*, 877–886; (d) Manivet, P.; Schneider, B.; Smith, J. C.; Choi, D. S.; Maroteaux, L.; Kellermann, O.; Launay, J. M. *J. Biol. Chem.* **2002**, *277*, 17170–17178; (e) Almaula, N.; Ebersole, B. J.; Zhang, D.; Weinstein, H.; Sealfon, S. C. *J. Biol. Chem.* **1996**, *271*, 14672–14675.

25. Russell, M. G. N.; Baker, R. J.; Barden, L.; Beer, M. S.; Bristow, L.; Broughton, H. B.; Knowles, M.; McAllister, G.; Patel, S.; Castro, J. L. *N*-Arylsulfonylindole Derivatives as Serotonin 5-HT_6 Receptor Ligands. *J. Med. Chem.* **2001**, *44*, 3881–3895.