Classification of Piperazinylalkylisoxazole Library by Recursive Partitioning

Hye-Jung Kim,[†,#] Woo-Kyu Park,[‡] Yong Seo Cho,[†] Kyoung Tai No,[§] Hun Yeong Koh,[#] Hyunah Choo,[†] and Ae Nim Pae[†,*]

[†]*Life Sciences Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, Korea*
[*]*E-mail: anpae@kist.re.kr*
[‡]*Pharmaceutical Screening Research Team, Korea Research Institute of Chemical Technology, Daejeon 305-343, Korea*
[§]*Department of Biotechnology, Yonsei University, Seoul 120-749, Korea*
[#]*Department of Chemistry, Inha University, Incheon 402-751, Korea*
Received September 7, 2007

A piperazinylalkylisoxazole library containing 86 compounds was constructed and evaluated for the binding affinities to dopamine (D_3) and serotonin (5-$HT_{2A/2C}$) receptor to develop antipsychotics. Dopamine antagonists (DA) showing selectivity for D3 receptor over the D2 receptor, serotonin antagonists (SA), and serotonin-dopamine dual antagonists (SDA) were identified based on their binding affinity and selectivity. The analogues were divided into three groups of 7 DAs (D_3), 33 SAs (5-$HT_{2A/2C}$), and 46 SDAs (D_3 and 5-$HT_{2A/2C}$). A classification model was generated for identifying structural characteristics of those antagonists with different affinity profiles. On the basis of the results from our previous study, we conducted the generation of the decision trees by the recursive-partitioning (RP) method using Cerius2 2D descriptors, and identified and interpreted the descriptors that discriminate in-house antipsychotic compounds.

Key Words : Serotonin, Dopamine antagonist, Classification, Recursive partitioning

Introduction

Traditionally, antipsychotics have been considered to act *via* the blockade of the classical 'dopamine D_2 receptor'.[1,2] In early 1990, the discovery of dopamine D_3 and D_4 receptor and their distribution in brain allowed us to consider a new group of antipsychotics devoid of extrapyramidal side-effects.[3]

The second generation antipsychotic drugs, also called atypical antipsychotics serotonin-dopamine dual antagonists (SDAs), were discovered with the fact that clozapine blocks not only dopamine receptor (D_2 and D_4) but also serotonin (5-HT_{2A}) receptor.[4] In comparison with previous antipsychotics, these drugs have been reported to have a reduced propensity to cause extrapyramidal side effects,[5] to be more effective against negative psychotic symptoms,[6] to improve neurocognitive function and functional outcome,[7] and to be possibly effective in patients who are resistant to other treatments.[8] Later, it was proved that the selective serotonin antagonists could be new group of antipsychotics as themselves. Actually, the selective 5-HT_{2A} antagonist MDL 100907 has shown antipsychotic potential without specific affinity for dopamine receptor.[9] Therefore, agents bind to subtype, 5-HT_{2B} and 5-HT_{2C}, as well as to 5-HT_{2A} receptor were established as another class of atypical antipsychotics because of high sequence homology between them.[10] Currently, those three-classes of dopamine antagonists (DA), serotonin antagonists (SA), and serotonin-dopamine dual antagonists (SDA) are being developed for antipsychotics.

In previous report, we have designed and synthesized piperazinylalkylisoxazole library which consists of active dopamine D_3 antagonists through combinatorial method.[11] At this time, we carried out the additional screening to measure affinity for 5-$HT_{2A/2C}$ receptor. Although the library compounds have been constructed on common scaffold, they displayed different biological profiles according to their various substituents and the chain length (Figure 1). Therefore, it looks very interesting to identify the structural characteristics differentiating these three classes of antipsychotics. The ligand-based analysis can be rational strategy, since the three-dimensional structure of aminergic G-protein-coupled receptors (GPCRs) is not experimentally known.

We have previously reported the classification models of dopamine antagonists (DA), serotonin antagonists (SA), and serotonin-dopamine dual antagonists (SDA) collected from the MDDR database.[12] The decision trees from recursive partitioning (RP)[13] has resulted the best predictions as compared with other methods employed, LDA (linear discrimination analysis),[14] SIMCA (soft independent modeling of class analogy),[15] and ANN (artificial neural network).[16] Recursive partitioning (RP) is a family of data analysis techniques dividing a data set into subgroups according to appropriate descriptors. The RP method has many advantages in respect that it overcomes the difficulties of handling non-linear relationships and it can be free from many of the restrictive assumptions of standard linear regression that are associated with error terms. The decision trees resulting from RP can be another good point, which make it easy to understand visually without considering complex statistical analysis.

On that score, we present here a RP model to classify our in-house piperazinylalkylisoxazole library using 2D descriptors and to describe the essential decisive factors to split into three-classes of antipsychotics. Therefore, we would analyze the selectivity profile of compounds visually according to their substitutions. This RP decision tree-based visual

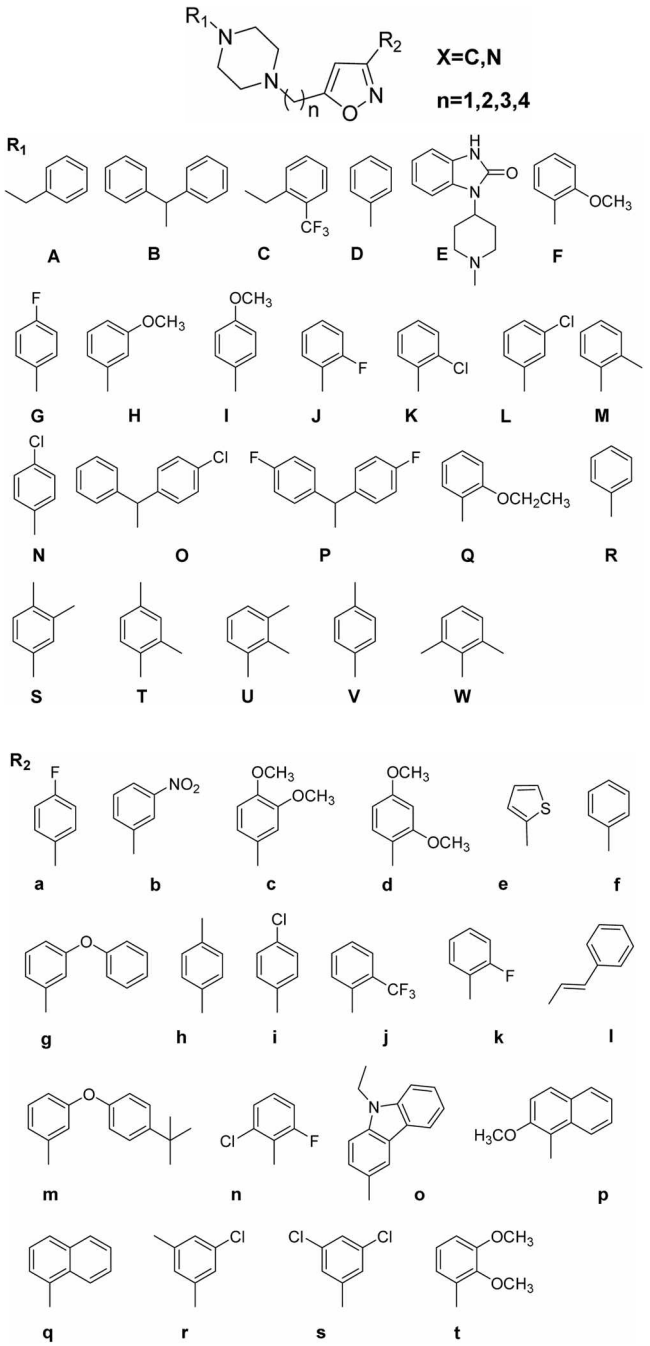

Figure 1. Structure of piperazinalkylisoxazole analogues used in the recursive partitioning analysis.

analysis will utilize for further work to synthesize receptor selective compounds.

Methods

Dataset. A small library of piperazinylalkylisoxazoles was constructed through solution phase combinatorial synthesis. The generated analogues were evaluated *in vitro* for dopamine and serotonin receptors binding affinities by measuring their ability to displace radioligands ([^{3}H]spiperone for D_2, [^{3}H]YM-09151-2 for D_3, [^{3}H]Ketanserin for 5-

Table 1. The structure and activity class of selected dopamine or serotonin active compounds

entry	n	R_1	R_2	inhibitory activity (%) at 1 μM		
				D3	5HT2Aor 2C	class
1	3	Q	b	82	49	DA
2	4	G	a	92	46	DA
3	4	K	f	83	35	DA
4	4	M	f	85	42	DA
5	4	Q	b	103	49	DA
6	4	Q	f	102	42	DA
7	3	W	k	83	49	DA
8	2	B	j	14	94	SA
9	2	R	j	5	89	SA
10	2	O	j	35	86	SA
11	2	E	j	0	72	SA
12	3	R	m	0	82	SA
13	3	J	m	8	68	SA
14	3	R	c	25	95	SA
15	3	J	c	34	88	SA
16	3	K	c	24	86	SA
17	3	L	c	0	90	SA
18	3	N	c	8	82	SA
19	3	W	c	0	65	SA
20	3	S	c	37	100	SA
21	3	U	c	0	92	SA
22	3	V	c	0	90	SA
23	3	R	h	20	79	SA
24	3	J	h	3	78	SA
25	3	K	h	0	86	SA
26	3	L	h	0	95	SA
27	3	N	h	0	90	SA
28	3	W	h	0	63	SA
29	3	T	h	0	67	SA
30	3	S	h	14	91	SA
31	3	U	h	0	93	SA
32	3	V	h	34	83	SA
33	3	R	q	27	91	SA
34	3	J	q	14	85	SA
35	3	K	q	24	85	SA
36	3	L	q	25	82	SA
37	3	N	q	14	77	SA
38	3	W	q	0	61	SA
39	3	U	q	14	84	SA
40	3	V	q	0	89	SA
41	3	B	d	94	95	SDA
42	4	A	a	85	95	SDA
43	4	B	a	110	99	SDA
44	4	B	c	109	100	SDA
45	4	B	d	102	100	SDA
46	4	B	e	108	100	SDA
47	4	B	f	105	96	SDA
48	4	B	i	102	97	SDA
49	4	C	a	103	100	SDA
50	4	C	i	80	96	SDA
51	4	D	a	106	99	SDA
52	4	D	b	96	100	SDA

Table 1. Continued

entry	n	R_1	R_2	inhibitory activity (%) at 1 μM		
				D3	5HT2Aor 2C	class
53	4	D	e	97	99	SDA
54	4	D	d	101	95	SDA
55	4	D	e	91	94	SDA
56	4	D	f	100	95	SDA
57	4	D	i	108	100	SDA
58	4	E	a	88	100	SDA
59	4	F	e	105	90	SDA
60	4	A	g	82	98	SDA
61	3	Q	i	96	85	SDA
62	4	K	l	102	93	SDA
63	4	M	e	96	95	SDA
64	3	M	g	96	96	SDA
65	3	P	e	102	88	SDA
66	3	S	m	84	100	SDA
67	3	U	n	88	100	SDA
68	3	U	s	88	99	SDA
69	3	K	t	100	94	SDA
70	3	K	h	100	98	SDA
71	3	L	h	100	100	SDA
72	3	T	h	97	87	SDA
73	3	S	h	100	100	SDA
74	3	U	h	100	100	SDA
75	3	L	o	82	100	SDA
76	3	S	o	85	100	SDA
77	3	K	k	100	96	SDA
78	3	L	k	94	100	SDA
79	3	S	k	82	100	SDA
80	3	U	k	100	99	SDA
81	3	K	p	85	96	SDA
82	3	S	p	94	100	SDA
83	3	U	p	89	100	SDA
84	3	K	q	89	95	SDA
85	3	L	q	87	100	SDA
86	3	U	q	90	100	SDA

HT_{2A}. [^{3}H]Imipramine for 5-HT_{2C}) from the cloned human receptors which were expressed in CHO cells, respectively.[11] The binding affinity (% inhibition value at 1 μM of test compound) of the ligands on each receptor was used for categorizing activity class. To consider only atypical antipsychotics, the ligands showing binding affinity for dopamine D_2 receptor were deleted in dataset. The ligands that present more than 70% of binding affinity (< 1 μM) for one receptor and report less than 50% for the other one were classified into either SA or DA class. The ligands that show more than 70% binding affinity (< 1 μM) for the both receptors were assigned to SDA class. Table 1 shows the structure and binding data of the selected compounds.

RP Model Generation. Two-dimensional descriptors of Cerius2 were used for RP study. Descriptors with constant values as well as descriptors containing 95% of zero values were removed. The correlation matrices of descriptors were built and some descriptors were deleted on the basis of the correlation threshold of R = 0.9. Finally, total 22 descriptors were obtained as shown in Table 2 and used as independent variables (X) in the all analyses.

The RP method was performed using the CART algorithm implemented in Cerius2 version 4.10.[17] The RP method categorizes objects by deriving a binary decision tree in which descriptors are used to split the data set into smaller, homogeneous subsets. The activity classes were weighted equally, and the splits were scored using Gini Impurity scoring function. The pruning factor values were varied between 0 and 3. The sample number 1 was considered as the minimum number of samples in any node. The various values were used for maximum tree depth (layers < 10) and the default values were accepted for maximum number of generic splits (30), and the number of knots per variable (20). The optimum decision tree was determined by standards described in our previous report.[12]

Results and Discussion

Generation of Recursive Partitioning Trees. To classify the piperazinylalkylisoxazoles library into their own activity class, the recursive partitioning model was developed using the topological descriptors which is based on molecular graph theory. The resulting model was determined by variation of parameters described in experimental section, trying to increase following values: The definition. "Class%Obs-Correct", is a measure of the number of compounds correctly predicted to belong to a class as a percentage of the total number of compounds observed to be each class. The measure of "Overall%PredCorrect" represents the total number of compounds correctly classified divided by the number of compounds predicted to belong to each class. The enrichment factor for a specific class is the ratio of the "Overall%PredCorrect" to the original percentage of compounds belonging to that class. The statistical results of the best RP model are reported in Tables 3 and 4. The entire set composed of 86 antagonists was classified with 90.70% of good classification rates. For DA class, 5 compounds (71.43

Table 2. Descriptors used to develop classification model

descriptor	information
Kappa indices[15] (6)	the shape of a molecule
E_state_keys[16] (12)	the electrotopological interaction for each atom
log Z[17] (1)	the degree of branching
Wiener[18] (1)	the length of chemical bonds existing between all pairs of heavy atoms
structural descriptor (2)	the number of H-bond donor and acceptor

Table 3. Classification results for dopamine antagonists (DA), serotonin antagonists (SA), serotonin-dopamine dual antagonists (SDA)

observed	predicted		
	DA	SA	SDA
DA (7)	5	1	1
SA (33)	0	33	0
SDA (46)	1	5	40

%) among 7 DAs were correctly classified. The number of DA was much smaller than SA and SDA classes. Although current data set seems to be highly unbalanced, the practical HTS experiment consists of very few actives and mostly inactive compounds. As mentioned in introduction section, the RP algorithm offers many advantages for classifying those data.

The classification rate for SA and SDA was 100% and 86.96%, respectively. The number of true positives among the predictions in each activity class is also listed as the term "Overall%PredCorrect". It is noted that 83.33% in the predicted DA class are true hits (DAs), 84.62% of the predicted SA class are true hits, and 97.56% of the predicted SDA class are true hits. The enrichment factor (10.24, 2.21 and 1.82 for DA, SA, and SDA, respectively) also described that the final RP model could be statistically significant for classifying our own library.

Description of Decision Trees. Figure 2 displays the optimized 5-leaf recursive partitioning decision tree by the encouraged descriptors to classify three active classes. The red color indicates DA class, the green color means SA class, and the SDA class is plotted using dark gray color. The structure of decision tree consists of 5 terminal nodes and 4 non-terminal nodes. At each node (decision point), molecules were split into groups, higher and lower responses, according to their descriptors.

One of key descriptors of our RP tree is electrotopological value (E-state key) computed for each atom in a molecule which encode information about both the topological

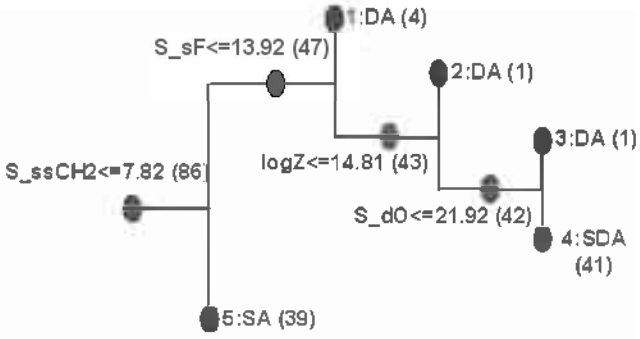

Figure 2. The best RP tree generated with pruning factor (2), and 5 maximum tree depth. The red color indicates dopamine antagonists, the green color means serotonin antagonists, and the serotonin-dopamine dual antagonists are plotted using dark gray color.

environment of that atom and the electronic interactions of all other atoms in the molecule. That is, the information of the electron accessibility at that atom and the degree of adjacency or topological state of the atom were provided by E-state key. The meaning of the E-state symbols in the Cerius2 implementation is as follows: S: sum of numerical value for following atom type, s: single bond d: double bond t: triple bond a: aromatic bond.

Table 5 reports the description and illustration of decision factors that were found to be important. The S_ssCH$_2$ was the first decisive factor which stands for the sum of intrinsic values for -CH$_2$ atom type with two single bonds. This split provided the information that most of SAs have relatively lower availability of -CH$_2$ (sp^3) atoms (alkyl chain length = 2-3) for intermolecular interaction compare to the parts in SDAs and DAs (alkyl chain length = 3-4). The E-state key, S_sF, followed as the second descriptor to classify SDA and DA class. This demonstrates the accessibility at fluorine in some DAs is larger than most SDAs. To characterize the structural difference of DA compounds from the SDAs, the topological descriptor logZ was selected. This means that many SAs have different aspects of the molecular shape

Table 4. Statistical results of recursive partitioning

class	# of compounds (%)[a]	Class%Obscorr[b]	Overall%Precorr[c]	enrichment[d]
DA	7 (8.1)	71.43	83.33	10.24
SA	33 (38.4)	100	84.62	2.21
SDA	46 (53.5)	86.96	97.56	1.82

[a]The number of samples in each class. [b]intraclass prediction. [c]overall prediction. [d]the enrichment factor: Overall%Precorr divided by the original percentage of compounds belonging to that class (%).

Table 5. Summary definition of all descriptors that were found to be important by decision tree with best predictive ability

index	description	illustration
S_ssCH$_2$	Sum of the atom level E-state values for all the group, -CH$_2$, in the molecule	
S_sF	Sum of the atom level E-state values for all fluorine atoms in the molecule	
S_dO	Sum of the atom level E-state values for all oxygen atom with one double bond in the molecule.	
Log Z	the degree of branching of the molecule	

compared to the other classes. The remaining group was classified to subgroups with the E-state key, S_dO, which imply that some DAs and SDAs have different electrotopological environment around all oxygen atom with one double.

Based on these results, we summarized following points. In the first place, the main structural difference between SA and DA or SDA class was decided on the electrotopological value around the -CH_2- atom type. Secondly, the remarkable structural difference between DA and SDA class could be summarized by the availability of oxygen atom with one double, fluorine connected to aromatic ring, and the degree of branching of certain molecule when an antagonist binds to receptor.

We have employed RP methods with 2D structural descriptors to analyze the structure-activity relationships between our DA, SA, and SDA compounds. The classification by above descriptors produced considerable discriminative power in spite of high degree of structural similarity of the library. The decision trees contain useful information that leads to the expedient criteria in predicting selectivity of the compounds. The identified distinctive structural aspects for each class could be guideline to design selective antipsychotic candidates for individual therapy.

Model Validation by Cross-validation Test. To avoid overfitting and to improve generalization of the classification models, ten different trial data sets were validated. Eight-tenths of all compounds were taken as a training set to derive models and the remaining compounds were set to the test set (8:2). The average and standard deviation results of ten different trials are summarized in Table 6. The training sets achieved acceptable classification percent and the test sets also showed reasonably good predictability. The prediction rate of DA class seems to be relatively low. We used seven compounds as training set and only three compounds as test set. Considering the number of limitation in data set, the prediction rate is acceptable. The key descriptors determined in each trial set were consistent with those from total set. The predictive powers prove that our models are valid to classify and predict the active class of new candidates.

Table 6. Mean values and standard deviations of classification obtained in ten different training and test sets

training	DA	SA	SDA	total
Class%Obscorr	65.0 ± 12.9	93.9 ± 3.3	82.6 ± 7.4	86.6 ± 7.1
Overall%Precorr	65.0 ± 1.5	86.1 ± 9.3	88.4 ± 4.7	84.0 ± 4.9
enrichment	7.6 ± 0.0	2.5 ± 0.5	1.6 ± 0.1	.
prediction	56.8 ± 16.4	78.2 ± 13.6	80.0 ± 9.4	71.8 ± 11.3

To validate the robustness of our decision tree, we also generated two-class RP models. As shown in Table 7, all the models produced good statistical values. The key descriptors on two-class (DA/SA and SA/SDA) decision tree generated were almost consistent with decisive factors on the best model of three classes (Figure 3). The decision trees of two-class model showed like sub-tree model of three-class

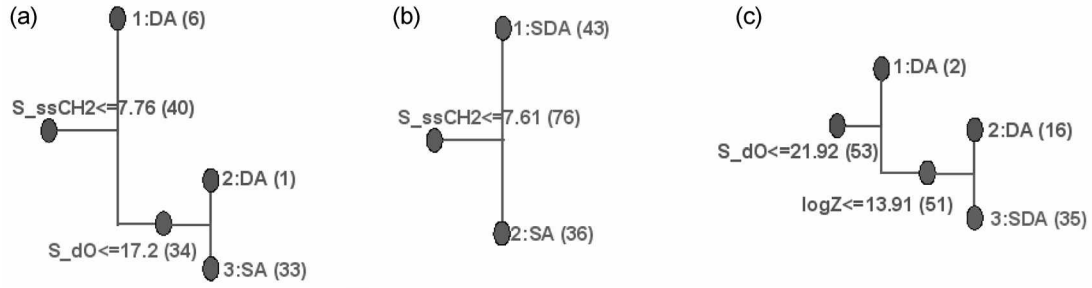

Figure 3. Two-class of RP model. The decisive factors produced by optimum parameter are almost consistent with the three-class of RP model (Figure 2). Decision tree for classifying a) dopamine and serotonin antagonists, b) serotonin and serotonin-dopamine dual antagonists, and c) dopamine and serotonin-dopamine dual antagonists.

Table 7. Statistical results of two-class of recursive partitioning

class	# of compounds (%)	Class%Obscorr	Overall%Precorr	enrichment
SA (33)	41.77	100.0	84.62	2.03
SDA (46)	58.23	86.96	100.0	1.72

class	# of compounds (%)	Class%Obscorr	Overall%Precorr	enrichment
DA (7)	13.21	100	38.89	2.94
SDA (46)	86.79	76.09	100	1.15

class	# of compounds (%)	Class%Obscorr	Overall%Precorr	enrichment
DA (7)	17.50	71.43	100	5.71
SA (33)	82.50	100	94.29	1.14

model. This is the evidence that the best model of three classes was not generated by some chance.

Conclusion

The aim of the present study is to develop discriminative model to predict receptor selective antagonists. The percent activity data was only available instead of continuous activity values such as IC_{50} and we divided the compounds into activity classes responsible for each receptor. We have employed RP classification model because it has advantage for classifying class analogy activity data and for considering appropriate descriptors recursively. We have previously tested other classification methods using published antipsychotics and have obtained the best model with RP method. Here, we have successfully built up the visualized decision tree using the topological descriptors encoding the chemical environment around important functional groups. To design new compounds with specific activity class, it could be used to decide the substituents around main scaffold. Furthermore, current model will be improved continuously with our further product compounds.

Acknowledgment. This work was supported by Ministry of Commerce, Industry and Energy.

Reference and Notes

1. Creese, I.; Burt, D. R.; Snyder, S. H. *Science* **1976**, *192*(4238), 481-3.
2. Seeman, P.; Lee, T.; Chau-Wong, M.; Wong, K. *Nature* **1976**, *261*(5562), 717-9.
3. Bouthenet, M. L.; Souil, E.; Martres, M. P.; Sokoloff, P.; Giros, B.; Schwartz, J. C. *Brain Res.* **1991**, *564*(2), 203-19.
4. Meltzer, H. Y. *J. Clin. Psychiatry* **1994**, *55*(Suppl B), 47-52.
5. Casey, D. E. *Schizophr. Res.* **1999**, *35*(suppl), S61-S66.
6. Moller, H. J. *Eur. Arch. Psychiatry Clin. Neurosci.* **1999**, *249*(suppl 4), 99-107.
7. Green, M. F.; Kern, R. S.; Braff, D. L.; Mintz, J. *Schizophr. Bull.* **2000**, *26*, 119-136.
8. Wahlbeck, K.; Cheine, M.; Tuisku, K.; Ahokas, A.; Joffe, G.; Rimon, R. *Prog. Neuropsychopharmacol Biol. Psychiatry* **2000**, *24*(6), 911-22.
9. Padich, R. A.; McCloskey, T. C.; Kehne, J. H. *Psychopharmacology (Berl)* **1996**, *124*(1-2), 107-16.
10. Roth, B. L.; Craigo, S. C.; Choudhary, M. S.; Uluer, A.; Monsma, F. J., Jr.; Shen, Y.; Meltzer, H. Y.; Sibley, D. R. *J. Pharmacol. Exp. Ther.* **1994**, *268*(3), 1403-10.
11. Cha, M. Y.; Choi, B. C.; Kang, K. H.; Pae, A. N.; Choi, K. I.; Cho, Y. S.; Koh, H. Y.; Lee, H. Y.; Jung, D.; Kong, J. Y. *Bioorg. Med. Chem. Lett.* **2002**, *12*(10), 1327-30.
12. Kim, H. J.; Choo, H.; Cho, Y. S.; Koh, H. Y.; No, K. T.; Pae, A. N. *Bioorg. Med. Chem.* **2006**, *14*(8), 2763-70.
13. Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth: Belmont, CA, 1984.
14. Everitt, B. S.; Dunn, G. *Applied Multivariate Analysis*; Oxford University Press; New York, 1992.
15. Dunn III, W. J.; Wold, S. *Chemometric Methods in Molecular Design*; VCH: New York, 1995; pp 179-193.
16. *Neural Networks* in *Chemistry & Drug Design*, 2nd ed.; Zupan, J., Gasteiger, J., Eds.; Wiley-VCH: Weinheim, 1999.
17. Bronowska, A.; Les, A.; Chilmoneczyk, Z.; Filipek, S.; Edvardsen, O.; Ostensen, R.; Sylte, I. *Bioorg. Med. Chem.* **2001**, *9*(4), 881-95.