New Serotonin 5-HT$_6$ Ligands from Common Feature Pharmacophore Hypotheses

Hye-Jung Kim,[†,‡] Munikumar Reddy Doddareddy,[†] Hyunah Choo,[†] Yong Seo Cho,[†] Kyoung Tai No,[‡] Woo-Kyu Park,[§] and Ae Nim Pae*[,†]

Life Science Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, South Korea, Department of Biotechnology, Yonsei University 220, Sinchon-dong, Seodaemun-gu, Seoul 120-749, South Korea, and Korea Research Institute of Chemical Technology, P.O. Box 107, Yusung-gu, Taejeon 305-343, South Korea

Received May 11, 2007

Serotonin 5-HT$_6$ receptor antagonists are thought to play an important role in the treatment of psychiatry, Alzheimer's disease, and probably obesity. To find novel and potent 5-HT$_6$ antagonists and to provide a new idea for drug design, we used a ligand-based pharmacophore to perform the virtual screening of a commercially available database. A three-dimensional common feature pharmacophore model was developed by using the HipHop program provided in Catalyst software and was used as a query for screening the database. A recursive partitioning (RP) model which can separate active and inactive compounds was used as a filtering system. Finally a sequential virtual screening procedure (SQSP) was conducted, wherein both the common feature pharmacophore and the RP model were used in succession to improve the results. Some of the hits were selected based on druglikeness, ADME properties, structural diversity, and synthetic accessibility for real biological evaluation. The best hit compound showed a significant IC$_{50}$ value of 9.6 nM and can be used as a lead for further drug development.

INTRODUCTION

Serotonin (5-hydroxytryptamine, 5-HT) is an important neurotransmitter that mediates multiple physiological functions by interacting with 14 distinct serotonin (5-HT$_{1-7}$) receptor subclasses.[1,2] The 5-HT$_6$ receptor was the most recently identified and cloned member of the serotonergic receptors.[3,4] It is a G-protein coupled receptor positively coupled to an adenylate cyclase second messenger system[5] and is located primarily in the central nervous system such as the nucleus accumbens, striatum, hippocampus, and olfactory tubercle of the brain.[6,7] Its specific localization and its high affinity for antipsychotics and antidepressants implies a possible role for 5-HT$_6$ receptor antagonists as a promising target for psychiatry.[8,9] There was also some evidence suggesting that the 5-HT$_6$ receptor selective antagonists may be involved in the improvement of cognitive learning performance.[10,11] This cognition enhancement is associated both with Alzheimer's disease and schizophrenia. Recently, the experimental results of the antagonist Ro 04-6790 in rat and 5-HT$_6$ receptor-knockout mice proposed that 5-HT$_6$ receptor antagonists may be implicated in the therapy of obesity.[12]

Regardless of all these expectations, the exact therapeutic significance of 5-HT$_6$ receptor is still being debated because of the lack of selective antagonists having good blood brain permeability of penetration and satisfactory ADME properties. Therefore more potent and selective 5-HT$_6$ receptor ligands are required for further studies. To date most of the 5-HT$_6$ antagonists were mostly provided by high-throughput screening (HTS).

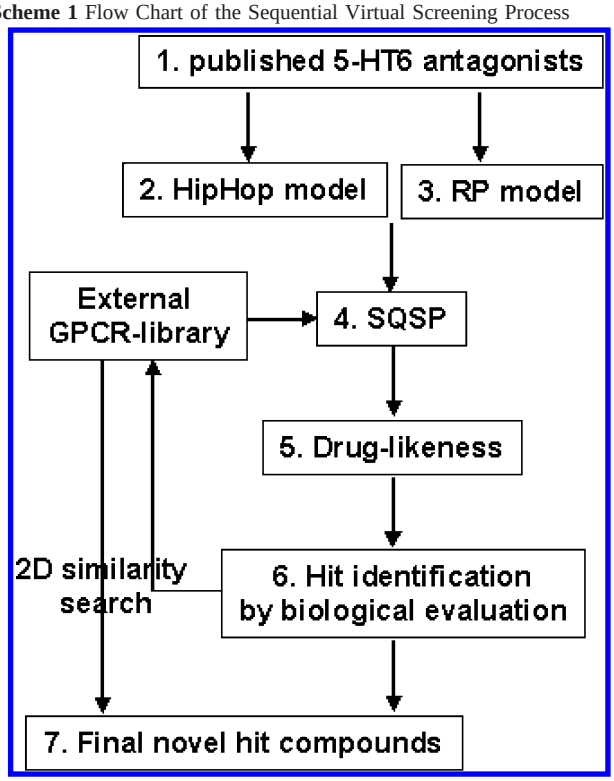
Computer-based virtual screening is a quite useful alternative tool as it is a cost-effective and time saving process. In the case of GPCR ligands, the rationale starting point in virtual screening is to define a pharmacophore model describing the appropriate interactions between a receptor and ligands. In a recently published paper, the authors developed a three-dimensional pharmacophore model from a set of 45 structurally diverse 5-HT$_6$ receptor antagonists using the HypoGen module of Catalyst software,[13] which represents an activity based alignment.[14] The other algorithm relating to automated pharmacophore generation within Catalyst is called HipHop, which is based on alignment of common features present in highly potent compounds.[15] A reasonable HypoGen pharmacophore model is difficult to obtain as the training set should contain ligands with a wide range of activity.[16] Whereas in the case of the HipHop pharmacophore modeling, a small ligand set with highly active compounds is sufficient to get a reasonable model. However, a HipHop model is not able to predict the activity value of the virtual hit, so generally a large number of false positive hits are observed in the case of HipHop pharmacophore-based virtual screening.

In this paper, a HipHop pharmacophore-based virtual screening was conducted to find new 5-HT$_6$ ligands. To overcome its limitation of getting a large number of hits, we adopted a sequential virtual screening procedure (SQSP),[17] wherein the HipHop pharmacophore-based virtual screening was followed by an additional screening using the RP model. Scheme 1 lays out the SQSP strategy: (1) All known 5-HT$_6$ antagonists were collected. (2) The HipHop pharmacophore model was generated with carefully selected small sets of a

* Corresponding author phone: +82-2-958-5185; fax: +82-2-958-5189; e-mail: anpae@kist.re.kr.
† Korea Institute of Science and Technology.
‡ Yonsei University 220.
§ Korea Research Institute of Chemical Technology.

10.1021/ci700160t CCC: $40.75
Published on Web 11/29/2007

Scheme 1 Flow Chart of the Sequential Virtual Screening Process



known 5-HT_6 antagonist. (3) The recursive partitioning (RP) model was built by comparing the known antagonists against a set of decoys. (4) Virtual screening by the models was conducted sequentially using a commercially available database. (5) Some compounds from step 4 were selected based on druglikeness. (6) Initial hits were identified via biological evaluation. (7) Additional hits were found by a two-dimensional similarity search of the initial hits.

In summary, the pharmacophore-based initial virtual screening screens a database by checking the presence of essential functional features in the molecule, and the RP[18]-based virtual screening discriminates the active and inactive compounds by evaluating the molecular topological pattern. The RP methodology is very attractive in that it utilizes property descriptors with continuous value ranges and transforms them into a binary classification scheme and gives the resulting decision tree as output, which makes it easy for nonstatisticians to split active and inactive compounds.

Herein, we described the development of a ligand-based HipHop pharmacophore model for 5-HT_6 receptor antagonists, discussed the comparison between our HipHop and the published HypoGen pharmacophore model, and provided the virtual screening results of a commercial database which yielded a novel potent antagonist with a nanomolar IC_{50} value.

COMPUTATIONAL METHODS

Generation and Validation of a Common Feature Pharmacophore Model. Based on two-dimensional structural similarity, all reported 5-HT_6 receptor antagonists have been clustered into five main groups (indole, indole-like, monocyclic, bicyclic and tricyclic aryl-piperazine, and a miscellaneous group) by Holenz at al.[19] For the pharmacophore generation, five to eight antagonists were carefully selected from each group considering the activity, selectivity, and structural rigidity. A sixth group was additionally defined by extracting 9 diverse structures from the 5 main groups. The selected molecular structures and their activities (K_i, nM) are shown in Figure 1. The sixth group was made of **2, 3, 10, 12, 13, 19, 24, 27**, and **28**. The conformational models of the selected compounds were built using the "best conformer generation" method with a 20 kcal/mol energy cutoff.[20,21] Each of the six groups was submitted to the common feature pharmacophore generation procedure. The HipHop module in Catalyst was used to identify and overlay common features shared by a training set. Four chemical functions, namely hydrogen bonding acceptor, aromatic ring, positive ionizable, and hydrophobic group, were selected as pharmocophoric features based on the chemical features present in the training set. The parameter settings of Maximum Omitted Features, Misses, and Complete Misses were varied to generate multiple hypotheses as some of the compounds may or may not contain all the features.[13] Each hypothesis generation run returns 10 possible pharmacophore hypotheses having a different arrangement of constituent features and sorts them according to the ranking scores. The procedure was repeated with the above parameter variation until no more new pharmacophore models were found in their configuration. Redundant hypotheses that have the same chemical characteristics and nearly the same distances between these functions were deleted, and hypotheses with diverse configurations were selected according to ranking scores and best fit values. To select the best model from the total 6 pharmacophores developed, all known 281 5-HT_6 antagonists obtained from Prous[22] were screened using the best flexible search option given in the catalyst. The hit rates of all the pharmacophores were given in Table 1.

Generation of Recursive Partitioning Model. The recursive partitioning model was developed using the CART algorithm as implemented in the Cerius[2] program.[23] The true positive 281 5-HT_6 antagonists were set as an active class, and the GPCR-focused 3040-member library[24] from Tripos Inc. was selected as an inactive class. The library may include a few active 5-HT_6 antagonists which can be ignored considering the large size of the library. The RP tree was constructed by E-state key[25] and topological descriptors[26,27] based on chemical graph theory. The activity classes were weighted equally, and the splits were scored using the Gini Impurity scoring function. The pruning factor values were varied between 3 and 6. The value of 1/100 of samples was used as the minimum number of samples in any node. The various values were used for maximum tree depth (layers $\leq$10), whereas default values were used for the 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.[28] The statistical results of RP model were given in Table 3.

Comparisons of Enrichments throughout SQSP. Enrichments were calculated to understand whether SQSP provides an improvement compared to simple virtual screening by either the pharmacophore model or the RP model. Two subdatabases containing 50 actives and 450 inactive compounds were chosen randomly from the source used for RP model generation. The RP model was generated with the remaining data set, and the prediction test of the subdatabase was conducted. We also screened the subdatabase using the pharmacophore model only and SQSP. These

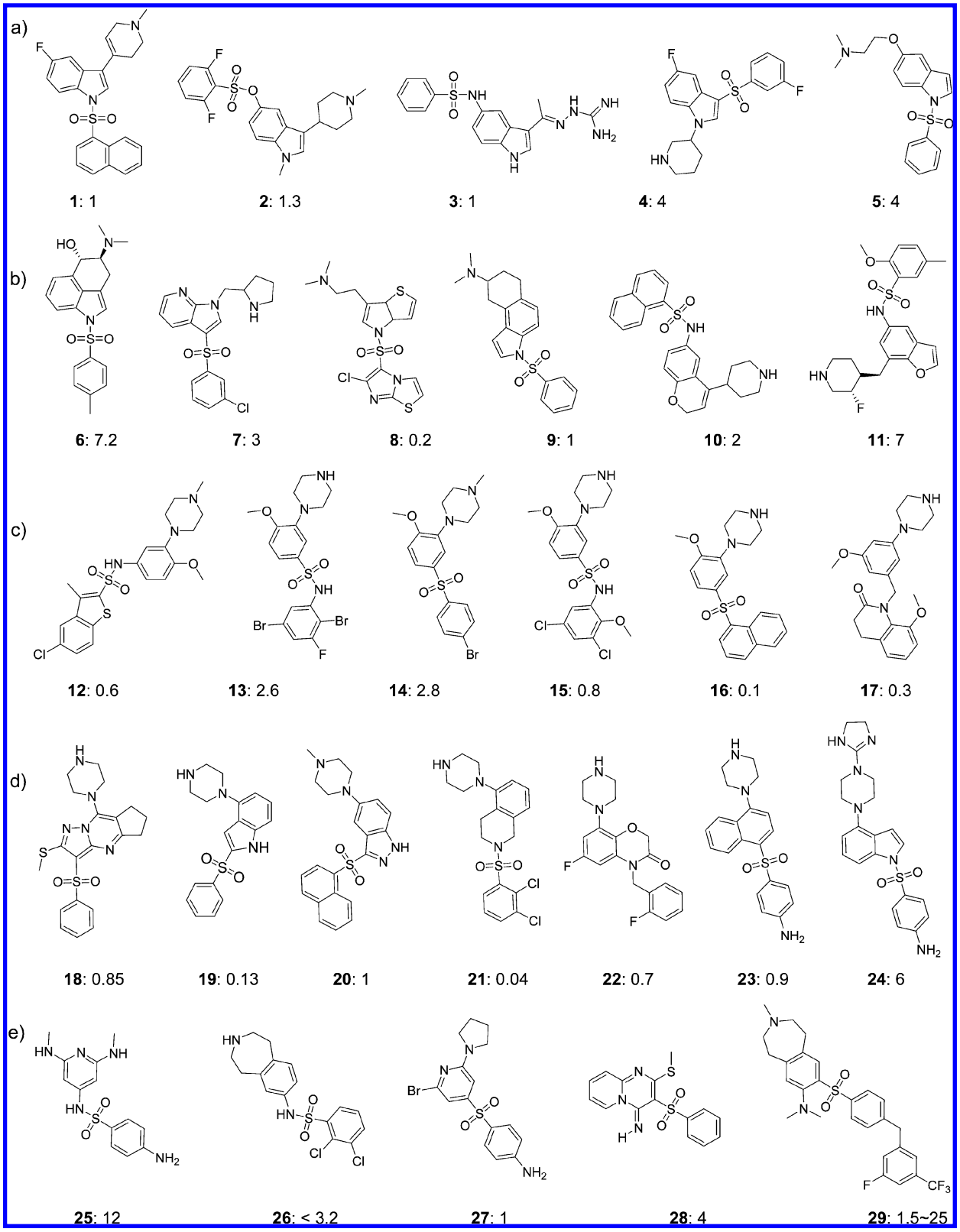


Figure 1. The structures and activities (Ki, nM) of the training 5-HT6 receptor antagonists containing (a) indol, (b) indol-like, (c) monocyclic aryl-piperazine, (d) bicyclic and tricyclic aryl-piperazine, and (e) miscellaneous group. The combination group consists of the compounds **2, 3, 10, 12, 13, 19, 24, 27**, and **28**.

processes were repeated ten times, and the average enrichments were calculated to understand which method is the best (Table 4).

Virtual Screening Experiment Using External GPCR-Library. Virtual screening has been carried out by the sequential virtual screening procedure (SQSP) to obtain new compounds with desired activity profiles. The ChemDiv GPCR-focused library[29] was converted into the Catalyst multiconformational data format in a CHARMM-like force field[21] and was used for screening purposes. The best pharmacophore model was used for a virtual screening experiment by selecting the fast flexible database search

Table 1. Common Feature Hypotheses and Corresponding Hit Rates of 5-HT_6 Antagonists[a]

indole[c]		indol-like[c]	
pharmacophore composition[b]	no. of hits (hit rate %)	pharmacophore composition[b]	no. of hits (hit rate %)
RRPHA	214 (76.2)	RRPHA	212 (75.4)
RPHHA	198 (70.5)	RPHHA	203 (72.2)
RRPA	193 (68.7)	RRPA	200 (71.2)
RPHA	194 (69.0)	RPHA	160 (57.0)
monocyclic aryl-piperazine[c]		**bicyclic and tricyclic aryl-piperazin[c]e**	
pharmacophore composition[b]	no. of hits (hit rate %)	pharmacophore composition[b]	no. of hits (hit rate %)
RRPHHA	200 (71.2)	RRPHA	209 (74.4)
RPAA	124 (44.1)	RRPA	196 (70.0)
RPHA	187 (66.5)	RPHA	203 (72.2)
miscellaneous, arylsulfony[c]l		**combination of five groups[c]**	
pharmacophore composition[b]	no. of hits (hit rate %)	pharmacophore composition[b]	no. of hits (hit rate %)
RRHHA	221 (78.6)	RPHHA	235 (83.6)
RPHHA	182 (64.8)	RRPA	193 (68.7)
RPHA	162 (57.7)	RPHA	214 (76.2)

[a] Diverse hypotheses were acquired from six different training sets, and their hit rates were evaluated by performing virtual screening against all known 281 5-HT_6 antagonists. [b] R: aromatic ring; H: hydrophobic group; A: hydrogen bond acceptor; P: positive ionizable. [c] All known 281 antagonists were collected and clustered into six different groups containing the specific structural motif to generate multiple pharmacophore models.

Table 2. Comparison of Matrix Distances (in Å) of the Chemical Features in Generated HipHop Pharmacophore Model and Reported HypoGen Pharmacophore Model

HipHop Model						HypoGen Model				
	A	H_1	H_2	P	R		A	H	P	R
A						A				
H_1	6.36					H	5.11			
H_2	4.62	8.74				P	6.66	5.18		
P	6.98	5.15	6.33			R	4.05	6.09	4.62	
R	4.14	3.08	7.51	6.54						

Table 3. Statistical Results Produced from the Final RP Model

	no. of compds (%)[a]	Class% Obscorr[b]	Overall% Precorr[c]	enrichment[d]
inactive class	3040 (91.54)	95.95	99.49	1.09
active class	281 (8.46)	94.66	68.38	8.08

[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 (%).

option. Resulting hit lists were then filtered using the RP model acting as a secondary screening model. The first screening figured out new compounds with similar functional and spatial properties defined in 3D pharmacophore query. The subsequent second screening extracted active compounds by comparing topological properties of initial hit lists with those of known compounds. Finally 40 compounds were selected based on druglikeness (Lipinski's rule of 5), structural diversity, ADME properties, and synthetic accessibility for real biological evaluation. The druglikeness and structural diversity were examined using the Hivol tool in the SYBYL program.[30] The in silico ADME properties were calculated by the PreADME program[31] which uses an artificial neural network.[32]

Table 4. Enrichments of Test Set for Each Virtual Screening Method[a]

methods	no. of total hits	no. of true positives (%)[b]	no. of false positives (%)[c]	enrichments[d]
Pharmacophore	289	36 (72)	253 (56)	1.2
RP	102	46 (92)	56 (12)	4.5
SQSP	62	32 (64)	30 (7)	5.2

[a] The test set was composed of 50 active compounds and 450 inactive compounds. [b] % of actives correctly predicted. [c] % of inactives wrongly predicted as actives. [d] (No. true positives/no. total hits) × percentage of actives.

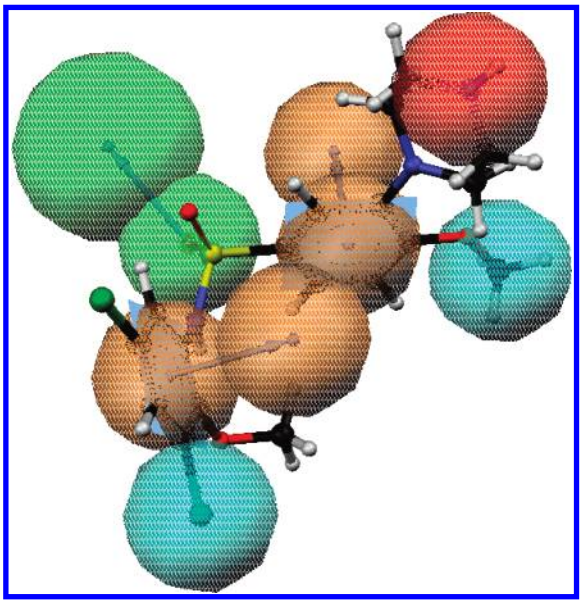


Figure 2. The biggest pharmacophore model generated by a monocyclic aryl-piperazine training set: (cyan: hydrophobic groups; green: hydrogen bond acceptor feature with a vector in the direction of the putative hydrogen donor; orange: aromatic ring; and red: positive ionizable group).

RESULTS AND DISCUSSION

Generation of the Best Common Feature Model. Highly active and selective 5-HT_6 antagonists were chosen for the generation of the HipHop pharmacophore in the computational method section. A total of 5 different groups of compounds were used for pharmacophore generation (Figure 1). A combined group composed of representatives from each of these 5 groups was additionally prepared to generate pharmacophore for diverse compounds. As the HipHop pharmacophore models are pretty sensitive to composition of the training set, the training sets should cover adequate structural diversity to elucidate the common functional features responsible for ligand−receptor binding. The active ligands with different structures may adopt different binding modes to maximize their interaction in the active site, and also all active ligands may not have the same functional groups. For this reason the pharmacophore generation was conducted by altering different parameter options given to get all the possible pharmacophores. In HipHop, the molecules that are going to be used to construct the pharmacophore configuration space can be specified through setting

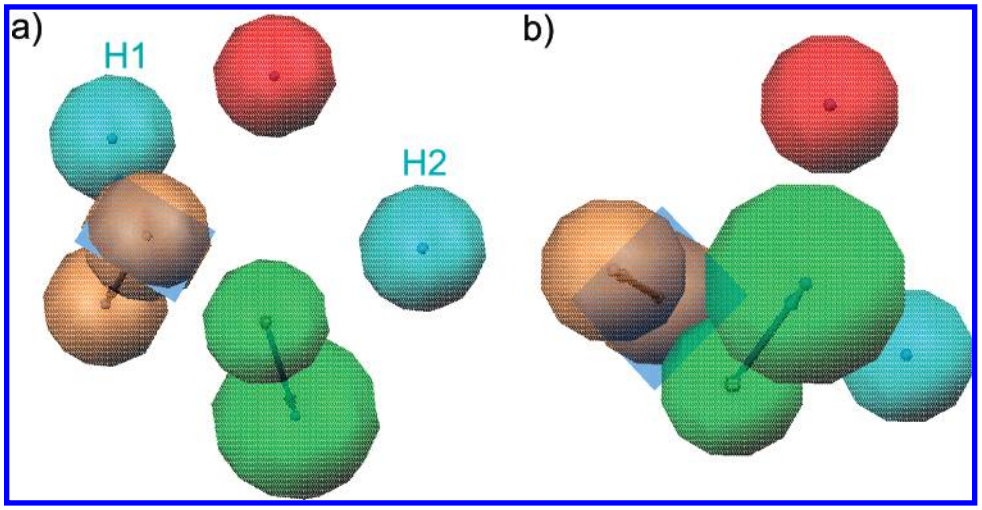


Figure 3. The comparison between (a) the best HipHop pharmacophore model and (b) the HypoGen model of the literature: (cyan: hydrophobic groups; green: hydrogen bond acceptor feature with a vector in the direction of the putative hydrogen donor; orange: aromatic ring; and red: positive ionizable group).

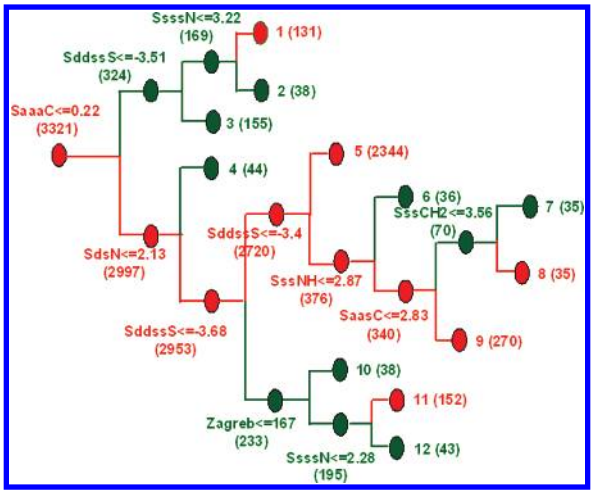


Figure 4. RP tree generated to discriminate active and inactive hits. At each node (decision point), molecules were split into higher or lower responses according to their descriptors and were to reach the 12 terminal nodes. The red color indicates the inactive class, and the active class is plotted using the color green.

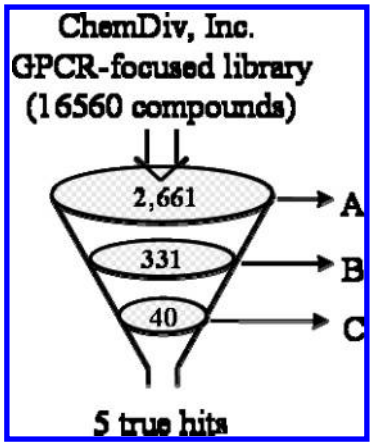


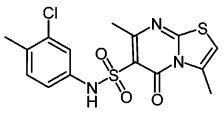
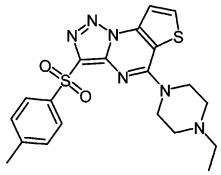
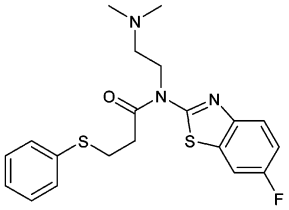
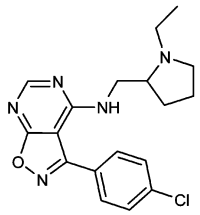
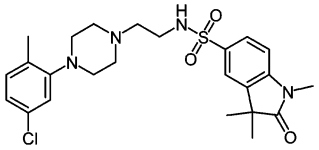
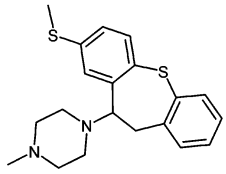
Figure 5. The number of hit compounds reduced for each sequential screening step: A, pharmacophore-based virtual screening; B, filtration by a recursive partitioning tree; and C, final selection based on druglikeness, structural diversity, synthetic accessibility, and favorable in silico ADME properties.

the 'principal' compound parameter. This parameter can adopt the values of 0, 1, or 2. A value of 2 means all the features in this molecule will be considered in building a hypothesis space. A value of 0 means this molecule will be ignored when building the hypothesis. A value of 1 indicates to the program that the molecule either to be used or not depends on the setting of other parameters such as Misses and CompleteMisses. 'Misses' means that hypotheses that fail to map completely to more than one training compound will be disallowed. 'CompleteMisses' specify the number of molecules that do not have to map to any features in the hypothesis. The Misses and CompleteMisses were varied from 1 to 3, respectively, in both the cases. The value for MaxOmitFeat was set to 2 so that a pharmacophore can be kept even if that specific molecule is a 'CompleteMiss' to that pharmacophore. Only four-point pharmacophore models were generated when these parameters were set to 1. Putting less strict parameter values gave a larger size of pharmacophore models.

Table 1 reports the pharmacophore models collected from the repeated trials using different parameters for each training group. The biggest pharmacophore model with 6 features was built by a monocyclic and monocyclic aryl-piperazine training group. Figure 2 shows the overlay of the training set compound **15** on to the generated pharmacophore model. It has two hydrophobic features, two aromatic rings, one hydrogen bond acceptor, and one positive ionizable feature. The other pharmacophore models were configured by the subfeatures of this six feature model. Selecting the best pharmacophore model among them is very difficult for the following limitations: (i) A pharmacophore with more numbers of features is too restrictive and results in a limited number of hits. (ii) On the other hand, a pharmacophore with a fewer number of features is nonselective and tends to capture a large number of false positives without effective discrimination against less active inhibitors. (iii) The ranking score from a HipHop process run also cannot be used as a standard to select the best pharmacophore model, because it describes only the chance as to how well the internal training compounds fit the hypothesis.

To select the best model, the generated hypotheses were tested according to their ability in recognizing known 5-HT$_6$ antagonists. All published 281 5-HT$_6$ antagonists were screened by using the best flexible database search option

Table 5. Hits from Virtual Screening of GPCR-Focused ChemDiv DB Using the Best HipHop Pharmacophore Model

Code Number	Structure	IC_{50} (nM)	Code Number	Structure	IC_{50} (nM)
a		275	b		847
c		1,363	d		640
e		309	Reference Methiothepin		1.1

in Catalyst, and the hits rate for each pharmacophore were shown in Table 1.

The best hit rate of 83.6% was given by a five feature pharmacophore obtained from the combined group (containing representatives of other 5 groups). This pharmacophore is composed of two hydrophobic, one aromatic ring, one hydrogen bond acceptor, and one positive ionizable group (Figure 3). This suggests that the probability of obtaining a good and reasonable common feature pharmacophore model is high when using structurally diverse training set compounds.

Comparison between HipHop and HypoGen Pharmacophore Model. The best HipHop model was compared with the most recently published pharmacophore model. The HypoGen model generated by M. Campillo et al.[14] was reproduced from the same training set and in the same manner as described in the literature. Figure 3 shows both the pharmacophore models, and Table 2 represents the distance matrices of the pharmacophore models. The comparison between these two pharmacophore models shows the following observations: The first and main difference is the absence of a "hydrophobic center (H_1)" in the HypoGen model. The second difference is the small discrepancies in the distances among the hydrophobic, aromatic ring center and positive ionizable features, but the overall configuration of the features is compatible. The hydrophobic group (H_1) may not be essential for good affinity, but most of the known 5-HT_6 antagonists contained this feature. In fact, the four-feature HypoGen model recorded only 25% of the hit rate, whereas the five-feature HipHop model hit 83% of the published antagonists indicating that most of them contained this feature (H_1). Although the four-feature Hypogen model can be used to predict the activity of compounds, the five-feature HipHop query would be more profitable in the case of virtual screening.

Recursive Partitioning Tree for Postfiltering. To classify the hit compounds obtained from the common feature pharmacophore-based virtual screening into an active or an inactive class, a recursive partitioning model (RP) was developed using the two-dimensional descriptors containing molecular shape information. The statistical results for the best RP model were demonstrated in Table 3. Class%ObsCorrect is the measure of the number of compounds of the correctly predicted corresponding class; it gives information about false negatives and false positives. Overall%PredCorrect represents the overall prediction; it gives information about the accuracy of prediction when the whole set is predicted with the model. The enrichment factor for a specific class is the ratio of the Overall%PredCorrect to the percentage of compounds in the total data set.

The statistical results in Table 3 indicate that about an 8-fold improvement over random selection could be achieved by application of the resultant RP model. Figure 4 displays the optimized 12-leaf RP model containing 8 orthogonal descriptors (7 E-state key values and 1 topological index) which can discriminate 5-HT_6 active and inactive compounds. The most important descriptors are the electrotopological values (E-state key) computed for each atom in a molecule, which encode information about both the topological environment of that atom and the electronic interactions due to all other atoms in the molecule. The meaning of the E-state symbols in the Cerius[2] implementation is as follows: S: sum of numerical value for following atom type, s: single bond, d: double bond, t: triple bond, and a: aromatic bond. In this RP model, the chemical environment around an aromatic carbon, nitrogen, and sulfonamide moiety contributed in a major proportion in discriminating active and inactive compounds. The other main topological descriptor splitting the data set is the Zagreb characterizing the degree of atomic branching in a molecule.

Validation of SQSP by Enrichment Comparison. The virtual screening based on just a pharmacophore model may strike a large number of false positives as it does not consider

Table 6. Results from similarity search of the initial virtual hit compounds based on 2D fingerprints

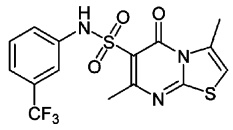
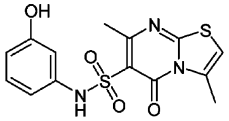
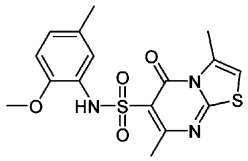
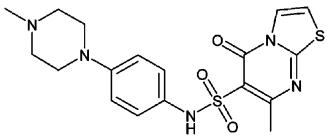
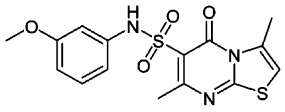
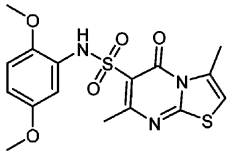
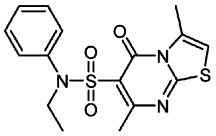
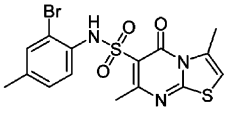
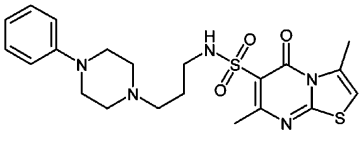
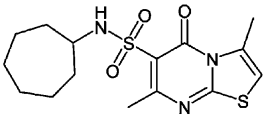
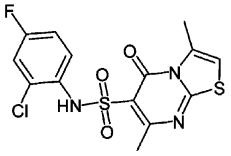
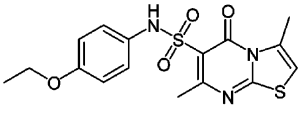
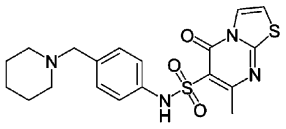
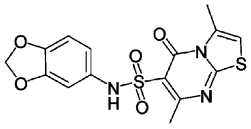
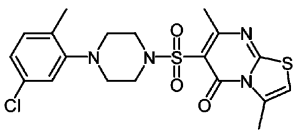
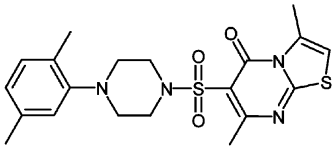
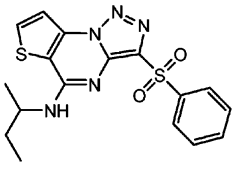
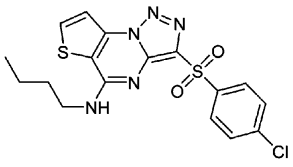
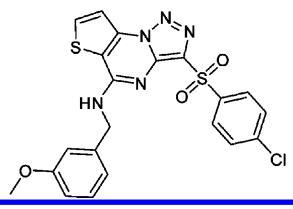
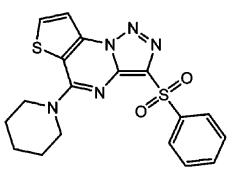
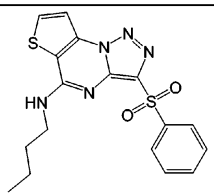
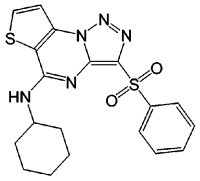
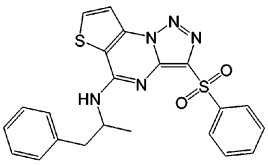
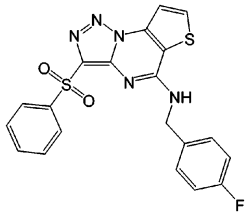
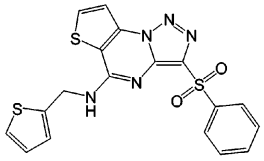
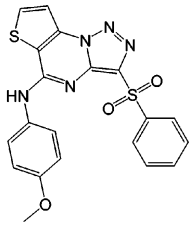
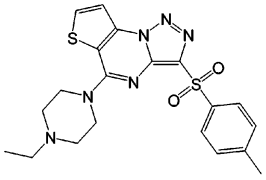
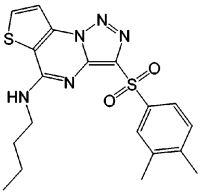
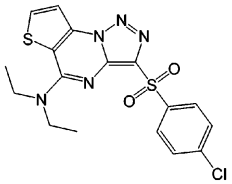
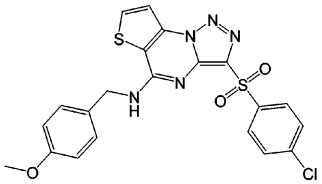
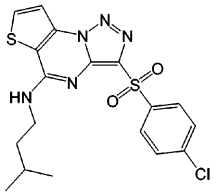
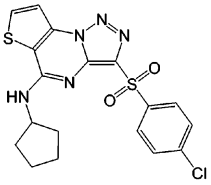
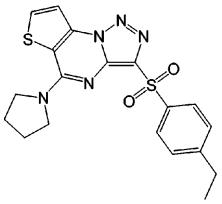
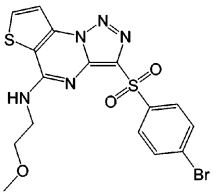
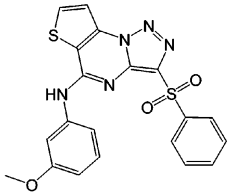
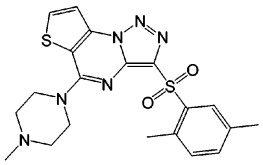
Code Number	Structure	IC_{50} (nM)	Code Number	Structure	IC_{50} (nM)
a-1		58	a-2		62
a-3		63	a-4		84
a-5		96	a-6		133
a-7		202	a-8		276
a-9		1,159	a-10		1,179
a-11		1,937	a-12		2,510
a-13		5,158	a-14		5,546
a-15		2,613	a-16		1,699
b-1		9.6	b-2		48
b-3		323	b-4		371

Table 6. (Continued)

Code Number	Structure	IC_{50} (nM)	Code Number	Structure	IC_{50} (nM)
b-5		42	b-6		36
b-7		161	b-8		31
b-9		63	b-10		5,237
b-11		85	b-12		522
b-13		58	b-14		343
b-15		319	b-16		101
b-17		491	b-18		353
b-19		693	b-20		3,411

molecular shape aspects. The RP model also can generate a lot of misclassified compounds because it distinguishes active and inactive compounds not according to the 3D position of important functional groups required for activity but according to their 2D connectivity properties. The sequential screening procedure (SQSP) which combines both 2D and

3D methods is expected to yield better results as they compensate each other disadvantages.

To compare each method's accomplishment, 50 active compounds were used as a prediction test, and the enrichments of the models were calculated as mentioned in the Method section. Table 4 shows the enrichments for each virtual screening method. The results are drawn by looking at the average values over ten time repeats.

The pharmacophore-based virtual screening achieved a poor level of enrichment (1.2) when the hit compounds were selected without considering the fit value as shown in Table 4. Even when the hit compounds were ranked by fit values and the top 20% were selected, there is little improvement in enrichment. Therefore, it is clear that additional screening tools were required to find real hits. The RP-based virtual screening gave a much higher enrichment (4.5) when compared to pharmacophore-based virtual screening. This may be due to the fact that compounds in the test set with similar structures to the compounds in the training set were predicted well when compared to the others. So it is certain that the enrichment will be decreased in the case of virtual screening (using just the RP model) of the external library containing compounds of many different molecular frameworks. But when both the pharmacophore and the RP model were used in succession (SQSP) the results were better (5.2). A large number of false positive hits from pharmacophore-based virtual screening were filtered out in the RP-based screening process. In fact 52% of inactives were found as false positives in the case of pharmacophore modeling, whereas the RP model showed only 12%. When both methods were used in succession (SQSP), it was further decreased to 7%. This is the advantage of screening using both 3D- and 2D-descriptors or in other words combining both the pharmacophore and the topological descriptor based RP models.

Virtual Screening Experiment Using External GPCR-Library. To identify novel and potent 5-HT$_6$ ligands, a sequential virtual screening was performed first using the fast flexible database search tool in the Catalyst program and then by the RP classification model. The best pharmacophore model (Figure 3a) was used as a query to search a commercial database: the ChemDiv GPCR-focused library (16 560 compounds). The virtual screening procedure and hit reductions in each step were shown in Figure 5. The pharmacophore-based virtual screening gave 2661 compounds as hits. These hits were then put as input to the RP model, and activity was predicted. The RP model eliminated a large number of compounds by categorizing them as inactive. A total of 331 of the remaining compounds were predicted to be active by the RP model. Finally 40 hits were selected based on druglikeness, favorable ADME properties, structural diversity, and synthetic accessibility for real biological evaluation. In these 5 compounds were shown appreciable IC$_{50}$ values (Table 5). Among the true positives, it is observed that compounds **a** and **b** have many analogies with a similar skeleton and different side chains in the ChemDiv database. Finally, the database was again searched for compounds similar to **a** and **b** by a 2D fingerprints-based similarity search. A total of 36 compounds were selected for real biological evaluation; Table 6 shows the structures and binding affinities of the selected compounds. Compound **b-1** showed the best IC$_{50}$ value of 9.6 nm. It is an aryl sulfonyl compound containing a novel tricycle which was not observed in any of the previously known compounds and so can be used as a lead for the discovery of novel 5-HT$_6$ antagonists with favorable pharmacokinetic properties.

CONCLUSIONS

In order to identify new 5-HT$_6$ antagonists, a smart virtual screening strategy (SQSP) was conducted by combining a common feature pharmacophore (HipHop) and a recursive partitioning technique. Three-dimensional distances among pharmacophoric features and two-dimensional topological indices containing molecular shape information were combined as the descriptors in the screening process. Using the common feature pharmacophore commercially available the ChemDiv GPCR focused library was searched for 5-HT$_6$ ligands. The resulting hits were filtered by using the RP model. Finally 40 compounds were selected based on druglikeness, favorable ADME properties, structural diversity, and synthetic accessibility for real biological evaluation. Search for similar compounds to that of true positives yielded a compound with a nanomolar IC$_{50}$ value. This compound contained a novel tricycle which was not observed in any of the previously published 5-HT$_6$ antagonists and so can be used as a lead for future drug development. The successful identification of novel and active 5-HT$_6$ ligands by using SQSP shows the advantage of using a combined methodology instead of just the pharmacophore model or the RP model.

REFERENCES AND NOTES

(1) Barnes, N. M.; Sharp, T. A review of central 5-HT receptors and their function. *Neuropharmacology* **1999**, *38*, 1083−1152.
(2) Hoyer, D.; Hannon, J. P.; Martin, G. R. Molecular, pharmacological and functional diversity of 5-HT receptors. *Pharmacol. Biochem. Behav.* **2002**, *71*, 533−754.
(3) Kohen, R.; Metcalf, M. A.; Khan, N.; Druck, T.; Huebner, K.; Lachowicz, J. E.; Meltzer, H. Y.; Sibley, D. R.; Roth, B. L.; Hamblin, M. W. Cloning, characterization, and chromosomal localization of a human 5-HT6 serotonin receptor. *J. Neurochem.* **1996**, *66*, 47−56.
(4) Glennon, R. A. Higher-end serotonin receptors: 5-HT(5), 5-HT(6), and 5-HT(7). *J. Med. Chem.* **2003**, *46*, 2795−2812.
(5) Sebben, M.; Ansanay, H.; Bockaert, J.; Dumuis, A. 5-HT6 receptors positively coupled to adenylyl cyclase in striatal neurones in culture. *NeuroReport* **1994**, *5*, 2553−2557.
(6) Grimaldi, B.; Bonnin, A.; Fillion, M. P.; Ruat, M.; Traiffort, E.; Fillion, G. Characterization of 5-HT6 receptor and expression of 5-HT6 mRNA in the rat brain during ontogenetic development. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1998**, *357*, 393−400.
(7) Hirst, W. D.; Abrahamsen, B.; Blaney, F. E.; Calver, A. R.; Aloj, L.; Price, G. W.; Medhurst, A. D. Differences in the central nervous system distribution and pharmacology of the mouse 5-hydroxytryptamine-6 receptor compared with rat and human receptors investigated by radioligand binding, site-directed mutagenesis, and molecular modeling. *Mol. Pharmacol.* **2003**, *64*, 1295−1308.
(8) 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.
(9) Roth, B. L.; Craigo, S. C.; Choudhary, M. S.; Uluer, A.; Monsma, F. J.; Jr.; Shen, Y.; Meltzer, H. Y.; Sibley, D. R. Binding of typical and atypical antipsychotic agents to 5-hydroxytryptamine-6 and 5-hydroxytryptamine-7 receptors. *J. Pharmacol. Exp. Ther.* **1994**, *268*, 1403−1410.
(10) Dawson, L. A.; Nguyen, H. Q.; Li, P. In vivo effects of the 5-HT(6) antagonist SB-271046 on striatal and frontal cortex extracellular concentrations of noradrenaline, dopamine, 5-HT, glutamate and aspartate. *Br. J. Pharmacol.* **2000**, *130*, 23−26.
(11) Garcia-Alloza, M.; Hirst, W. D.; Chen, C. P.; Lasheras, B.; Francis, P. T.; Ramirez, M. J. Differential involvement of 5-HT(1B/1D) and 5-HT6 receptors in cognitive and non-cognitive symptoms in Alzheimer's disease. *Neuropsychopharmacology* **2004**, *29*, 410−416.

(12) Woolley, M. L.; Bentley, J. C.; Sleight, A. J.; Marsden, C. A.; Fone, K. C. A role for 5-ht6 receptors in retention of spatial learning in the Morris water maze. *Neuropharmacology* **2001**, *41*, 210−219.
(13) Bronowska, A.; Les, A.; Chilmonczyk, Z.; Filipek, S.; Edvardsen, O.; Ostensen, R.; Sylte, I. Molecular dynamics of buspirone analogues interacting with the 5-HT1A and 5-HT2A serotonin receptors. *Bioorg. Med. Chem.* **2001**, *9*, 881−895.
(14) Lopez-Rodriguez, M. L.; Benhamu, B.; de la Fuente, T.; Sanz, A.; Pardo, L.; Campillo, M. A three-dimensional pharmacophore model for 5-hydroxytryptamine6 (5-HT6) receptor antagonists. *J. Med. Chem.* **2005**, *48*, 4216−4219.
(15) Omoshile, O.; Trope Mehl, C.; Trope Mehl, A. HipHop: Pharmacophores Based on Multiple Common-Feature Alignments. In *Pharmacophore Perception, Development and Use in Drug Design*; International University Line: La Jolla, CA, 2000.
(16) Li, H.; Sutter, J.; Hoffmann, R. HypoGen: An Automated System for Generating 3D Predictive Pharmacophore Models. In *Pharmacophore Perception, Development and Use in Drug Design*; International University Line: La Jolla, CA, 2000.
(17) Engels, M. F.; Thielemans, T.; Verbinnen, D.; Tollenaere, J. P.; Verbeeck, R. CerBeruS: a system supporting the sequential screening process. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 241−245.
(18) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth International Group: Belmont, CA, 1984.
(19) Holenz, J.; Pauwels, P. J.; Diaz, J. L.; Merce, R.; Codony, X.; Buschmann, H. Medicinal chemistry strategies to 5-HT(6) receptor ligands as potential cognitive enhancers and antiobesity agents. *Drug Discovery Today* **2006**, *11*, 283−299.
(20) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285−294.
(21) Brooks, B. R.; Brucolleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. CHARMM: A Program for Macromolecular Energy Minimization, and Dynamic Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(22) Yoshino, T.; Nisijima, K.; Shioda, K.; Yui, K.; Katoh, S. Perospirone, a novel atypical antipsychotic drug, potentiates fluoxetine-induced increases in dopamine levels via multireceptor actions in the rat medial prefrontal cortex. *Neurosci. Lett.* **2004**, *364*, 16−21.
(23) Meyer, J. H.; Cho, R.; Kennedy, S.; Kapur, S. The effects of single dose nefazodone and paroxetine upon 5-HT2A binding potential in humans using [18F]-setoperone PET. *Psychopharmacology* **1999**, *144*, 279−281.
(24) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(25) Kier, L.-B.; Hall, L. H. An Electrotopological State Index for Atoms in Molecules. *Pharm. Res.* **1990**, *7*, 801−807.
(26) Bonchev, D.; Trinajstic, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517−4533.
(27) Kier, L.; Hall, L. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Hertfordshire, U.K., 1986.
(28) Kim, H. J.; Choo, H.; Cho, Y. S.; Koh, H. Y.; No, K. T.; Pae, A. N. Classification of dopamine, serotonin, and dual antagonists by decision trees. *Bioorg. Med. Chem.* **2006**, *14*, 2763−2770.
(29) Marshall, G. R.; Cramer, R. D., 3rd, Three-dimensional structure-activity relationships. *Trends Pharmacol. Sci.* **1988**, *9*, 285−289.
(30) Accelrys, Inc. *Catalyst, Version 4.11*; San Diego, CA, U.S.A., 2003. http://www.accelrys.com (accessed December 2005).
(31) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided. Mol. Des.* **1997**, *11*, 425−445.
(32) Hornik, K.; Stinchombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2*, 359−366.

CI700160T