DAT/SERT selectivity of flexible GBR 12909 analogs modeled using 3D-QSAR methods

Kathleen M. Gilbert,[a,†] Terrence L. Boos,[b] Christina M. Dersch,[c] Elisabeth Greiner,[b,‡] Arthur E. Jacobson,[b] David Lewis,[b,§] Dorota Matecka,[b,§] Thomas E. Prisinzano,[b,¶] Ying Zhang,[b] Richard B. Rothman,[c] Kenner C. Rice[b] and Carol A. Venanzi[a,*]

[a]Department of Chemistry and Environmental Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[b]Laboratory of Medicinal Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA
[c]Clinical Psychopharmacology Section, Intramural Research Program, National Institute on Drug Abuse, National Institutes of Health, Department of Health and Human Services, Baltimore, MD 21224, USA

Abstract—The dopamine reuptake inhibitor GBR 12909 (1-{2-[bis(4-fluorophenyl)methoxy]ethyl}-4-(3-phenylpropyl)piperazine, **1**) and its analogs have been developed as tools to test the hypothesis that selective dopamine transporter (DAT) inhibitors will be useful therapeutics for cocaine addiction. This 3D-QSAR study focuses on the effect of substitutions in the phenylpropyl region of **1**. CoMFA and CoMSIA techniques were used to determine a predictive and stable model for the DAT/serotonin transporter (SERT) selectivity (represented by pK_i (DAT/SERT)) of a set of flexible analogs of **1**, most of which have eight rotatable bonds. In the absence of a rigid analog to use as a 3D-QSAR template, six conformational families of analogs were constructed from six pairs of piperazine and piperidine template conformers identified by hierarchical clustering as representative molecular conformations. Three models stable to y-value scrambling were identified after a comprehensive CoMFA and CoMSIA survey with Region Focusing. Test set correlation validation led to an acceptable model, with $q^2 = 0.508$, standard error of prediction = 0.601, two components, $r^2 = 0.685$, standard error of estimate = 0.481, F value = 39, percent steric contribution = 65, and percent electrostatic contribution = 35. A CoMFA contour map identified areas of the molecule that affect pK_i (DAT/SERT). This work outlines a protocol for deriving a stable and predictive model of the biological activity of a set of very flexible molecules.
© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

Cocaine addiction remains a significant problem around the world. As with many addictive drugs, the exact

* Corresponding author. Tel./fax: +1 973 596 3596; e-mail: venanzi@adm.njit.edu
† Present address: Pestka Biomedical Laboratories, 131 Ethel Road West, Suite 6, Piscataway, NJ 08854, USA.
‡ Present address: df-mp, Dörries Frank-Molnia & Pohlman, Patents, Trademarks, Design, Triftstrasse 13, 80538 Munich, Germany.
§ Present address: Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 20993, USA.
¶ Present address: Division of Medicinal and Natural Products Chemistry, College of Pharmacy, University of Iowa, Iowa City, IA 52242, USA.

mechanism of the addiction has not been identified, although focus has been on the interaction of cocaine with various neurotransmitter systems.[1] The 'dopamine hypothesis'[2,3] implicates the dopamine transporter (DAT) in cocaine abuse and addiction. Structure–activity relationship (SAR) studies for several classes of dopamine (DA) reuptake inhibitors have been reviewed[1,4–7] and document the search for a selective DA reuptake inhibitor that could be used in an agonist-substitution therapy treatment for cocaine addiction.

GBR 12909 (**1**, vanoxerine) is a promising candidate, having demonstrated a decrease in cocaine-maintained responding without affecting food-maintained responding in behavioral studies of rhesus monkeys[8] and having passed Phase I clinical trials.[9] In addition, prior use of

cocaine caused cross-sensitization for several DA reuptake inhibitors, but not for **1**.[10] Compound **1** and its analogs have also been shown to bind relatively selectively to the DAT transporter and cause minimal DA reuptake inhibition, while at the same time attenuating the ability of cocaine and methamphetamine to elevate extracellular DA.[11–13] SAR studies of hundreds of analogs of **1**, based on the scaffold shown in Figure 1, have been summarized in a recent review;[9] additional studies have been carried out.[14,15] Most DA reuptake inhibitors have features common to many ligands that interact with G-protein coupled receptors, such as an aromatic ring located near a basic nitrogen (as seen for example in the relationship between the piperidine nitrogen and aromatic moieties on the A-side in Figure 1) that appear to be necessary for DAT binding. A second aromatic ring on the B-side of the **1** analogs may also be important for their interaction with the DAT transporter. However, photoaffinity labeling studies indicate that **1**-like and tropane-like compounds (such as cocaine) may bind to different regions in the DAT[16] or bind in a somewhat different manner.[17,18] Presence of a 2-[bis-(4-fluorophenyl) methoxy]ethyl substituent (identified as the B-side in Figure 1) usually results in better DAT binding affinity (but in some cases decreased DAT/serotonin transporter (SERT) selectivity) than a 2-benzhydryloxyethyl substituent, although neither effect was significant compared to modifications of the A-side.[9]

In the absence of knowledge of the three-dimensional structure of the DAT, ligand-based three-dimensional quantitative structure–activity relationship (3D-QSAR) techniques such as Comparative Molecular Field Analysis (CoMFA)[20] and Comparative Molecular Similarity Indices (CoMSIA)[21] may be useful in identifying molecular features that improve activity. However, these techniques require the use of a template conformer upon which each 3D-QSAR model is based. For rigid molecules that can adopt only a limited number of conformers, selection of the template conformer is relatively straightforward. Although CoMFA studies have been carried out on other classes of DA reuptake inhibitors, such as tropanes,[22,23] mazindols,[24] methylphenidate,[25]

and benztropines,[26,27] no CoMFA or CoMSIA studies have been done on analogs of **1**.

There is considerable evidence that many ligands do not bind to proteins in their vacuum phase global energy minimum (GEM) conformation;[28–33] however, many pharmacophore models of DA reuptake inhibitors have been based on the GEM structure of the ligand, or on a few structures very close in energy to the GEM.[24,34,35] Other work has shown the necessity of considering conformations other than the GEM in pharmacophore modeling, especially when modeling very flexible molecules.[36–42] For example, the Venanzi group's conformational analysis of the DA reuptake inhibitor methylphenidate[43] identified several local energy minimum conformers (with relative energies within a few kcal/mol of the GEM) which closely match those of a rigid methylphenidate analog that has the same DAT binding affinity as methylphenidate. Since the rigid analog is assumed to contain the important pharmacophore elements in their binding (or bioactive) orientation, this supports the idea that the bioactive conformer need not be the GEM conformer.

Only one other modeling study has been carried out on analogs of **1**. A chemometric quantitative structure–activity relationship (QSAR) study of **1** analogs and mepyramines used only a single minimized structure for each analog.[35] The study modeled 'DAT binding/ DA uptake' selectivity and DAT/SERT selectivity, but did not check the models for stability to random noise. The minimized structures were used as a basis for identifying structural characteristics of the training set compounds; no test set was used. In contrast, the present study uses multiple low-energy, representative conformations as templates for 3D-QSAR studies. No assumption is made that the GEM conformer is the conformation a molecule must achieve (the bioactive conformation) in order to bind to a target. Although the bioactive conformer *may* be one of the representative conformers, computations alone cannot prove that point. Instead, confirmation of the bioactive conformer is frequently derived from pharmacological testing of rigid analogs, none of which are presently available for this SAR series. The goal of the present work is simply to find a stable and predictive model for pK_i (DAT/ SERT).

Since large, flexible molecules such as the **1** analogs can take on a continuum of closely related conformations, they present a challenge to the application of 3D-QSAR techniques. The present work uses the output of a hierarchical clustering study[19] of molecular conformations as input to CoMFA and CoMSIA analyses. Hierarchical clustering has been shown to give similar results to fuzzy relational clustering[44] and singular value decomposition studies[45] of the same data set of conformations. Conformational analysis of **2** and **3** was carried out, and hierarchical clustering was used to select representative conformers to use as templates in 3D-QSAR studies.[19] These templates were chosen to be representative of the regions of three-dimensional space occupied by the analogs. In the present work, each representative con-
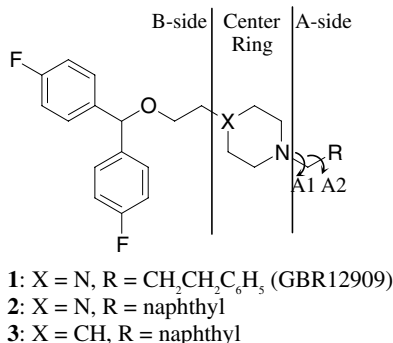

Figure 1. Template for GBR 12909 analogs used in this study. A1 and A2 are torsional angles that represent the relationship of the substituted A-side to the center ring, and were used to cluster conformers in a previous study.[19]

1: X = N, R = CH$_2$CH$_2$C$_6$H$_5$ (GBR12909)
2: X = N, R = naphthyl
3: X = CH, R = naphthyl

former was used as the template upon which a set of 45 analogs was built, constructing a conformational family of analogs, each with a structure similar to that of the template conformer. CoMFA and CoMSIA techniques were then used to evaluate predictive models for each of the families.

2. Computer modeling

2.1. Overview

The 3D-QSAR methodology used here is summarized in Figure 2, which reviews the major steps in a manner applicable to more general 3D-QSAR studies. In the absence of a rigid analog that could serve as a conformational template for the 3D-QSAR studies, six representative conformers[19] were chosen from among the regions of the local energy minima and used as templates for construction of conformational families of analogs. Six sets of 3D-QSAR studies, one for each conformational family, were carried out in parallel. Models which did not meet validation criteria were set aside or reassessed.

CoMFA studies encompassed a range of electrostatic and steric cutoffs for each conformational family. Steric and electrostatic cutoffs impact at what distances (inside the molecule) the CoMFA grid point values are set to a constant value. These distances represent the van der Waals molecular surface of an average molecule. Using a range of steric and electrostatic cutoffs allows these distances to change, optimizing them for a set of molecules, leading to increased q^2 values. CoMSIA studies were done for each of the basic field options: hydrogen donor and acceptor, hydrophobic, and steric and electrostatic. The CoMSIA studies were included to give a higher potential for computing significant models because different field types are available compared to CoMFA.[21] Partial least squares (PLS) analysis was performed on the CoMFA and CoMSIA results, with calculation of the q^2 (predictivity) and r^2 (goodness-of-fit) values and associated statistics. Region Focusing was applied to the best CoMFA or CoMSIA results for each preliminary study. Internal validation (training set y-value scrambling) and external validation (test set correlation validation) methods were used to determine the most predictive and stable model.

2.2. Selection of analogs

A subset of structurally related compounds was selected from over 200 analogs of **1**. The selection criteria were that the analogs contain as few rotatable bonds and chiral centers as possible, either a piperazine or piperidine ring, and the same 2-[bis-(4-fluorophenyl)methoxy]ethyl-B-side structure; and that the corresponding set of pharmacological data covers a wide range of DAT/SERT selectivities. Figure 1 shows the scaffold of the set of analogs of **2** and **3** that were used in this study. Analog **2** was selected because it has fewer rotatable bonds on the A-side than **1**, with a center ring and B-side identical to those of **1**, and no chiral centers. An additional 22 piperazine compounds were selected based on **2**'s scaffold, differing from **2** only in the A-side substituent. Analog **3** is the piperidine analog of **2**, and 21 additional piperidines were selected, differing from **3** only in the A-side substituent. This 3D-QSAR study therefore focuses on the effect of A-side substitutions. A-side moieties include naphthalene, thiophene, furan, and various phenyl substituents. A few of the compounds, such as the naphthyl compounds, have both piperazine and piperidine counterparts. The complete list of analogs, their experimental DAT and SERT binding affinity, as K_i, and pK_i (DAT/SERT) are shown in Tables 1A and B.[15,17,46–49]

The experimental pK_i(DAT/SERT) in the table was calculated from the following equation

$$
\begin{aligned}
&\mathrm{p}K_i(\mathrm{DAT/SERT}) \\
&= -\log[(\mathrm{DAT\ binding\ affinity}K_i) \\
&\quad /(\mathrm{SERT\ binding\ affinity}K_i)]
\end{aligned} \tag{1}
$$

Binding experiments were performed in the Rothman laboratory using identical protocols for all the ana-
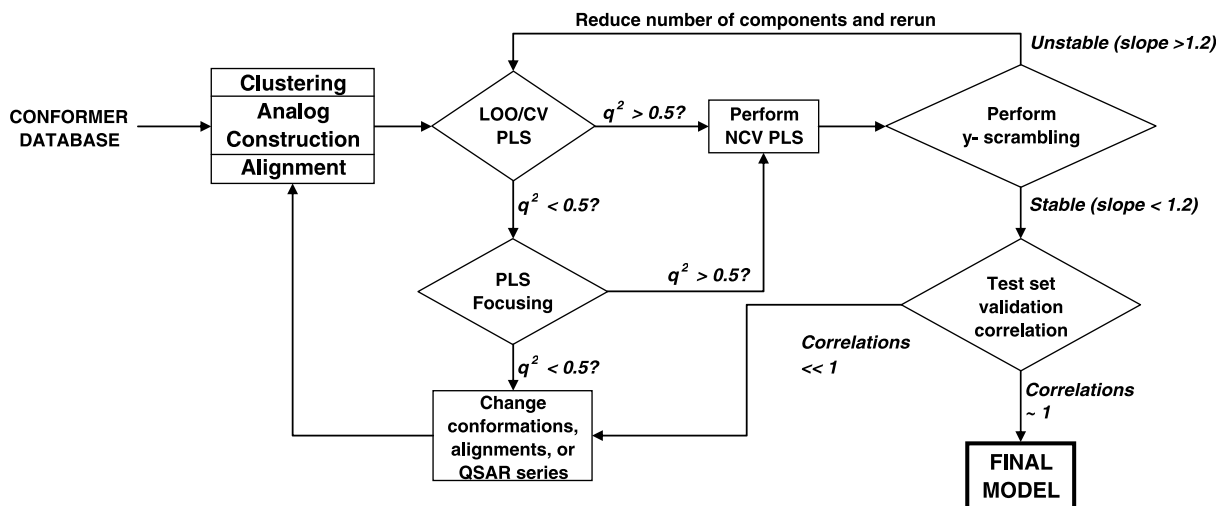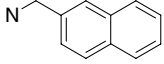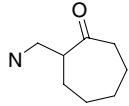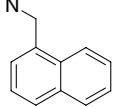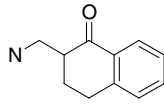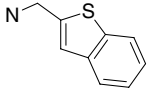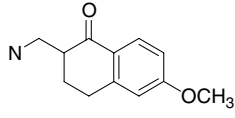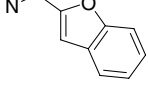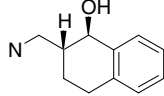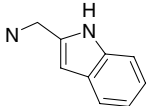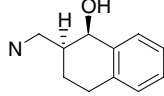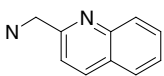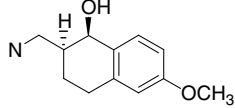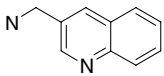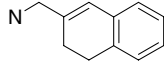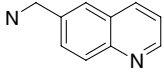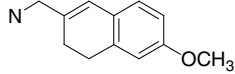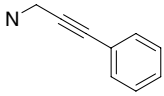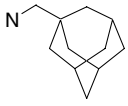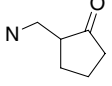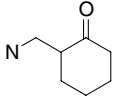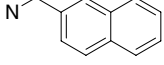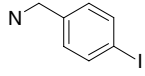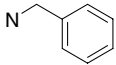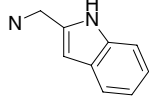

Figure 2. Flow chart for 3D-QSAR study. See text for explanation of LOO/CV (leave-one-out/cross-validation), NCV (non-cross-validation), PLS (partial least squares) focusing, y-scrambling, and test set validation techniques.

Table 1. Structures, binding K_i, and pK_i (DAT/SERT) values for **1** analogs

	N–CH$_2$–R[a]	DAT K_i (nM)[b]	SERT K_i (nM)[b]	pK_i (DAT/ SERT)[c]	Ref.[d]		N–CH$_2$–R[a]	DAT K_i (nM)[b]	SERT K_i (nM)[b]	pK_i (DAT/ SERT)[c]	Ref.[d]
1A Piperazine 1 analogs											
2		8	312	1.59	46	**15**[e]		49	525	1.03	47
4		31	243	0.89	46	**16**[e]		57	112	0.29	47
5		4.1	495	2.08	46	**17**[e]		22	2.96	−0.87	47
6		6.4	286	1.65	46	**18**		18	32	0.25	47
7		0.7	119	2.23	46	**19**		1.3	30	1.36	47
8		56	51	−0.04	46	**20**		14[g]	3	−0.67	47
9		16.1	485	1.48	46	**21**		6.4	261	1.61	47
10		61.8	550.5	0.95	46	**22**		53[h]	57	0.03	47
11		20	266	1.12	49	**23**		64.6	1650	1.41	49
12		100	598	0.78	49	**24**		33.7	2090	1.79	49
13[e]		84	970	1.06	47	**25**		40	888	1.35	49
14[e]		87[f]	1693[f]	1.29	47						
1B Piperidine 1 analogs											
3		0.71	229	2.51	17	**37**		2.1	205	1.99	48
26		9.9	171	1.24	48	**38**		0.73	88	2.08	17

Table 1 (*continued*)

	N–CH$_2$–R[a]	DAT K_i (nM)[b]	SERT K_i (nM)[b]	pK_i (DAT/ SERT)[c]	Ref.[d]		N–CH$_2$–R[a]	DAT K_i (nM)[b]	SERT K_i (nM)[b]	pK_i (DAT/ SERT)[c]	Ref.[d]
27	(4-OCH$_3$ benzyl)	7.9	266	1.53	48	39	(1-naphthylmethyl)	16	370	1.36	17
28	(3-OCH$_3$ benzyl)	13.3	161	1.08	48	40	(benzofuran-2-ylmethyl)	1.01	85	1.93	17
29	(2-OCH$_3$ benzyl)	63	248	0.60	48	41	(benzothiophen-2-ylmethyl)	1.61	246	2.18	17
30	(4-CF$_3$ benzyl)	1.1	442	2.60	48	42	(3-phenylprop-2-ynyl)	5.3	164	1.49	17
31	(4-NO$_2$ benzyl)	1.2	387	2.51	48	43	(tetrahydroisoquinolinyl)	9.32	15.19	0.21	49
32	(4-CH$_3$ benzyl)	5.7	239	1.62	48	44	(tetrahydroisoquinolinyl)	4.77	48.13	1.00	49
33	(4-Br benzyl)	1.1	199	2.26	48	45	(2-phenylcyclopropylmethyl)	0.8[i]	22[i]	1.44	15
34	(4-F benzyl)	2.3	245	2.03	48	46	(2-phenylcyclopropylmethyl)	1.3[j]	136	2.02	15
35	(4-Cl benzyl)	1.2	238	2.30	48						
36	(4-CN benzyl)	1.2	606	2.70	48						

[a] Piperazine (**1A**) and piperidine (**1B**) compounds with X and A-side substituent R as identified in Figure 1. Note that the expression N–CH$_2$–R indicates the nitrogen of the central ring with its –CH$_2$–R substituent.

[b] Measured as displacement of bound [^{125}I] RTI-55 from rat brain synaptosomes.

[c] DAT/SERT selectivity calculated by Eq. 1.

[d] Reference for experimental data.

[e] Compound tested as a racemic mixture.

[f] Values listed are corrections to previously published values.

[g] Modeling performed using value listed; additional experiments refined value to 7.3 nM.

[h] Modeling performed using value listed; additional experiments refined value to 13.5 nM.

[i] Modeling performed using value listed; additional experiments refined values to 0.5 nM for DAT K_i and 16 for SERT K_i.

[j] Modeling performed using value listed; additional experiments refined value to 1.4 nM.

logs.[15,17,46–49] The K_i value is nearly identical to the IC$_{50}$ value under these experimental conditions. DAT binding K_i's range from 0.7 nM for **7** to 100 nM for **12**. SERT binding K_i's range from 2.96 nM for **17** to 2090 nM for **24**, with 74% of the compounds with SERT binding K_i's above 100 nM.

2.3. Random search conformational analysis

Since it is computationally impractical to carry out conformational analysis of all the analogs in Tables 1A and B, **2** and **3** were selected for detailed conformational study. The explicit methodology of the random search conformational analysis of **2** is given elsewhere.[19,44] The same procedure was applied to **3**. A summary of the procedure is given below.

In binding studies of piperidine **1** analogs which had the piperidinyl nitrogen either proximal to the A-side (as shown in Fig. 1) or B-side, Dutta et al.[50] found binding affinity for the DAT to be higher in the former case. HF/ 6-31G* molecular orbital calculations of **1** indicate that protonation of the nitrogen proximal to the A-side is favored over that on the B-side.[51] Therefore, **2** and **3**

were constructed with a protonated nitrogen on the A-side. The side chains were attached in equatorial–equatorial positions, and the center ring treated as an aggregate during the random search procedure to maintain this configuration. The random search algorithm, an option in the SYBYL molecular modeling package,[52] randomly alters the eight non-ring torsional angles, then performs a geometry optimization (i.e., energy minimization) to produce a new conformer. Energy minimization was carried out using the Tripos force field[53] with Gasteiger–Hückel charges.[54] The random search conformational analysis found 728 conformers of **2** and 739 conformers of **3**.

2.4. Clustering

Comprehensive hierarchical clustering was performed on these data sets to determine representative conformers of **2** and **3**. Details of the clustering analysis are given elsewhere.[19] Various feature set/alignment options were examined and the resulting clustering statistics and distance maps were analyzed to determine the most appropriate clustering feature set and clustering level. Six clusters were identified for both **2** and **3**. A representative conformer was selected from each of these clusters. The six distinct representative conformers of **2** had energies ranging from 7.5 to 9.5 kcal/mol above that of its GEM conformer. Those for **3** had energies ranging from 3.5 to 5.1 kcal/mol above its GEM conformer.

2.5. Analog creation and alignment

These representative conformers were used as templates from which the other analogs in Tables 1A and B were constructed. Conformational families of analogs were obtained by changing the substituents of **2** or **3** to produce each new analog and minimizing each resulting structure for 1000 iterations using the Tripos force field[53] with Gasteiger–Hückel charges.[54] This approximation resulted in a significant decrease in computational time by not requiring a random search and clustering study for each of the 48 compounds in the final 3D-QSAR study.

For each representative conformer (or template structure), a SYBYL Programming Language script was used to modify the structure to create the other analogs in Tables 1A and B. Twenty-two additional piperazines were created based on each **2** template, and twenty-one additional piperidines were created based on each **3** template. For each analog, the A-side naphthalene moiety was removed and replaced with the appropriate new substituent R group in Tables 1A and B. Many of the aromatic R group substituents (such as those for **11**, **26**, **27**, etc.) are symmetrical to rotation about the A2 torsional angle (Fig. 1). For those that are not, the orientation of the substituent was set to follow that of the 2-naphthyl substituent in the respective template compound. The new R group substituent was added with the average A1 and A2 torsional angle values for the piperazine **2** or piperidine **3**. The energy of each analog was minimized.

Each analog in the series was then aligned to **24** (R = H), the analog with the smallest substituent, using the four heavy atoms comprising the A1 torsional angle. All analogs in the series were created and aligned together, to ensure that they were aligned in the same manner; additional test set analogs were built on the same template and aligned in the same way. The families of analogs were stored in SYBYL molecular databases. The analogs were divided into a test set and training set for 3D-QSAR studies. The test set was selected to be well-distributed over the range of pK_i (DAT/SERT) values and to be split between piperazine and piperidine analogs. Six analogs were selected for the test set, three piperazines (**5**, **8**, and **13**), and three piperidines (**29**, **32**, and **35**). The substituents were varied: three substituted phenyl groups, a benzothiophenyl group, a cyclopentanonyl group, and a quinolinyl group. These substituents are similar to those of other analogs in the training set. The remaining analogs constituted the training set.

2.6. Prediction of pK_i(DAT/SERT) of training set analogs

Variation of steric and electrostatic cutoff values (Preliminary Models) and Region Focusing (Focused Models) were used to improve the q^2 values obtained from the CoMFA and CoMSIA calculations on the training set analogs. Default CoMFA settings were used for all other parameters. The models with the highest q^2 values were tested for stability to random noise. Of these, the most stable models were subjected to test set correlation validation.

2.6.1. Preliminary models. Preliminary 3D-QSAR studies were performed on each of the six conformational families, consisting of 45 analogs apiece in the training set, in order to identify the optimal steric and electrostatic cutoff (CoMFA) or field (CoMSIA) values for the prediction of the pK_i(DAT/SERT) values in Tables 1A and B. CoMFA and CoMSIA columns were added to the molecular spreadsheet. CoMFA columns were created using electrostatic values of 10, 30, and 50 kcal/mol, in conjunction with steric cutoffs of 10, 30, and 50 kcal/mol, for a total of nine steric/electrostatic cutoff parameter combinations, the 30/30 kcal/mol steric/electrostatic combination being the SYBYL default setting. CoMSIA runs were performed for hydrogen bond donor/acceptor (DA), hydrophobic (HP), and steric/electrostatic (SE) fields. Default settings were used for the remaining CoMFA or CoMSIA parameters for each column. SAMPLS[55] was used to calculate q^2 values for the preliminary 3D-QSAR studies. Leave-one-out cross-validated (LOO/CV) calculations were carried out for each of the 12 (nine CoMFA and three CoMSIA) studies for each family. For each family, the most predictive model was identified as that with the highest value of q^2. The parameter combination (CoMFA) or field choice (CoMSIA) and optimal number of components of the most predictive model were used to calculate a non-cross-validated (NCV), or 'full', model for each family.

2.6.2. Focused models. To refine the models, Region Focusing[56] was performed on the most predictive model from each family. The SYBYL default Region Focusing parameters were used. The focused CoMFA or CoMSIA column was analyzed with SAMPLS. Models with

q^2 greater than 0.5 were tested for stability to random noise.

2.6.3. Stable models. One hundred y-value scramblings using SYBYL's Progressive Scrambling[57] function were performed on each selected model to estimate its stability with respect to random noise in the data. This method 'bins' the activities and scrambles them successively within each bin, thus grouping similar activities to check the sensitivity of the model to randomness. Three key results were produced:

- $Q^2 = 1 - (sSDEP)^2$: the predictivity of the model, using the scaled standard deviation of the error of prediction.
- cSDEP: cross-validated standard deviation of error of prediction
- $dq'^2/dr_{yy'}^2$: the instantaneous slope calculated where the correlation between the original data set and the randomized data set is equal to the critical point (0.85).

Like the q^2 and SEP results noted above, Q^2 should be high, preferably above 0.5,[58] but certainly above the statistical 95% confidence limit of 0.3,[59] and cSDEP should be low and similar to SEP. These results were compared to the recommended cutoffs for stable models (noted in Figure 2), of which the most important is a slope ($dq'^2/dr_{yy'}^2$) under 1.2.[60] In cases where the slope was between 1.2 and 2, the maximum number of components was reduced to five (from the default, six) and the calculations were rerun to find a stable model with a lower optimal number of components. Models that were stable to random noise were used to predict the pK_i (DAT/SERT) of the test set analogs.

2.7. Test set correlation validation

For each stable model, the predicted and experimental pK_i (DAT/SERT) values for each test set analog were used to calculate correlation coefficients as an external validation method for each model.[61] The Predicted pK_i (DAT/SERT) (y) versus Experimental pK_i (DAT/SERT) (x) was plotted to calculate: R^2, the coefficient of determination of the best-fit line; R, the correlation coefficient of the best-fit line; R_0^2, the coefficient of determination of the zero-intercept line; and k, the slope of the zero-intercept line. The reverse relationship, Experimental pK_i (DAT/SERT) (y) versus Predicted pK_i (DAT/SERT) (x), was also plotted to find $R_0'^2$, the coefficient of determination of the zero-intercept line of Experimental versus Predicted, and k', the slope of the zero-intercept line for Experimental versus predicted.

All of the six statistics should be near one, but a model is still considered valid if either R_0^2 or $R_0'^2$ is near 1, and its respective slope, k or k', is near 1.[61] Models that best predict the pK_i (DAT/SERT) of the test set analogs should have all of these statistics near 1.[61] Barring that, the value for R^2 should be near either the value for R_0^2 or $R_0'^2$.[61] These results were used to compare the predicted validity of the selected 3D-QSAR models and to choose the best model.

2.8. Three-dimensional CoMFA contour maps

SYBYL's Map CoMFA function was used to produce a three-dimensional steric and electrostatic contour map based on the conformational family of the best model. The map was analyzed to identify the regions of greatest impact due to changing steric or electrostatic characteristics. Interpretation, in terms of locating important areas, was based on the template molecules **2** and **3**.

3. Results

3.1. Analog creation and alignment

Conformational families of the 45 analogs of **1** were created, minimized, and aligned as noted above. The minimized and aligned six conformational families created from the six representative conformers are shown in the Supplementary Data.

3.2. Prediction of pK_i(DAT/SERT) of training set analogs

The optimal steric and electrostatic cutoff values were determined from CoMFA and CoMSIA studies on the six conformational families. This determined the best preliminary model for each family. Region focusing improved the q^2 value of the best preliminary model for all six families. However, only Families 1, 5, and 6 were found to have region-focused models stable to random noise in the data. Test set correlation validation identified the Family 6 model as the most satisfactory. Therefore, detailed analysis of training and test set predictions is provided below only for the Family 6 model. Analysis of the other models is given in the Supplementary Data.

3.2.1. Focused models. The results of the preliminary set of 3D-QSAR studies led to identification of the modeling parameters which gave the highest q^2 values. These results, along with the results for the default CoMFA cutoffs (30 kcal/mol steric, 30 kcal/mol electrostatic) are provided in the Supplementary Data. The LOO/CV and NCV results from refining the best preliminary model for each family by Region Focusing are given in Table 2. The plots of the predicted versus experimental pK_i (DAT/SERT) values are given in the Supplementary Data. Compared to the preliminary NCV PLS models, all families (except Family 2) show a decrease in the steric contribution and increase in the electrostatic contribution. Family 1 has the largest decrease, 22%; Families 5 and 6 show a 16% and 14% decrease in the steric contributions, respectively. An increase in electrostatic contributions indicates that the weighting of the electrostatic field grid points was increased. This increase is due mainly to the electrostatic effects of the substituents on the A-side aromatic rings. In other words, although the A-side napthyl and phenyl rings have similar hydrophobic characteristics, their substituents exhibit a wide range of electrostatic characteristics. As a result, increasing the weighting factors of the electrostatic grid points magnifies the effect of those points to which the model is the most sensitive. This indicates that the electrostatic characteristics of new substituents

Table 2. 3D-QSAR study—Focused Results

F[c]	LOO/CV PLS Models, SAMPLS[a]			NCV PLS (Full Models)[b]				
	q^2	C[d]	SEP[e]	r^2	SEE[f]	F value[g]	% S[h]	% ES[i]
1	0.512	6	0.635	0.967	0.166	155	53%	47%
1[j]	**0.511**	**2**	**0.599**	**0.685**	**0.481**	**39**	**55%**	**45%**
5	**0.502**	**2**	**0.605**	**0.724**	**0.450**	**47**	**68%**	**32%**
6	0.528	6	0.625	0.967	0.166	155	66%	34%
6[j]	**0.508**	**2**	**0.601**	**0.685**	**0.481**	**39**	**65%**	**35%**

[a] SAMPLS result (equal to column filtering (σ) set to 0 kcal/mol).
[b] No column filtering applied in full models.
[c] Conformational family of analogs developed from representative conformers.
[d] Optimal number of components calculated from focused leave-one-out cross-validated model.
[e] Standard error of prediction associated with the q^2 value.
[f] Standard error of estimate associated with the r^2 value.
[g] Statistical measure of whether the model is significant. F values of models with the same number of components can be compared directly; the model with the higher F value for the same number of components is more significant than a model with a lower F value.
[h] Percent steric field contribution to full model.
[i] Percent electrostatic field contribution to full model.
[j] The six-component models were unstable for these families (Table 3). Therefore, the maximum number of components was reduced to five (from the default of six) and the 3D-QSAR studies were rerun, resulting in two-component models.

should be taken into account when designing new DAT-selective analogs.

All q^2 values increase in the focused models, with substantial increases seen for Families 1, 5, and 6, indicating the usefulness of Region Focusing. These three families meet the criterion of having q^2 values greater than 0.5[58] and were selected for stability testing using Progressive Scrambling. Acceptable r^2 values are seen for most families, but the r^2 values for the Family 1 and Family 6 two-component models are below 0.7. The r^2 values remained approximately the same for the focused versus the preliminary models for all families. However, the r^2 values for the six-component focused models of Family 1 and Family 6 are significantly higher than those for the two-component preliminary models.

3.2.2. Stable models. Table 3 presents the Progressive Scrambling results with the three best models highlighted in boldface type. Q^2 values are lower than the q^2 values in all cases, ranging from 0.154 to 0.405. This is reasonable because Q^2 values are known to be more conservative than LOO/CV PLS q^2 values.[57] Calculated cross-validated standard deviation of error of prediction

(cSDEP) values are slightly larger than the SEP values (Table 2).

The results for Family 5 indicate a stable model. However, the six-component Family 1 and Family 6 models are unstable because the values of $dq^{2\prime}/dr^2_{yy\prime}$ are greater than 1.2.[57] In order to find more stable models, the maximum number of components was reduced to five (instead of the default six) and the 3D-QSAR studies were rerun for Families 1 and 6. These additional focused CoMFA results for Families 1 and 6 are given as the second entry for each family in Table 2 and show that two components are optimal. Table 3 shows that the two-component models for Families 1 and 6 both have $dq^{2\prime}/dr^2_{yy\prime}$ values of approximately 0.65, indicating stability to random noise. Therefore the two-component Family 1, 5, and 6 models were selected for test set correlation validation studies.

3.3. Training set predictions using stable models

Detailed graphical and tabular results for the three stable two-component Family 1, 5, and 6 models are given in the Supplementary Data. The average residuals

Table 3. q^2 values and progressive scrambling results for selected family models

F[a]	C[b]	Q^2[c]	cSDEP[d]	$dq^{2\prime}/dr^2_{yy\prime}$[e]
1	6	0.154	0.832	2.081
1[f]	**2**	**0.405**	**0.660**	**0.645**
5	**2**	**0.386**	**0.671**	**0.730**
6	6	0.266	0.774	1.628
6[f]	**2**	**0.289**	**0.722**	**0.646**

[a] Conformational family of analogs developed from representative conformers.
[b] Number of components at which the progressive scrambling was performed, corresponding to the optimal number of components for the CoMFA model.
[c] Equals $(1 - (sSDEP)^2)$. Predictivity of the model using the scaled standard deviation of error of prediction (sSDEP) instead of the SDEP.
[d] Calculated cross-validated standard deviation of error of prediction.
[e] Slope of q^2 (calculated by SAMPLS using perturbed y-values, therefore denoted $q^{2\prime}$) versus the correlation of the perturbed to the original y-variables (denoted $r^2_{yy\prime}$).
[f] The six-component models were unstable for these families. Therefore, the maximum number of components was reduced to five (from the default of six) and the 3D-QSAR studies were rerun, resulting in two-component models.
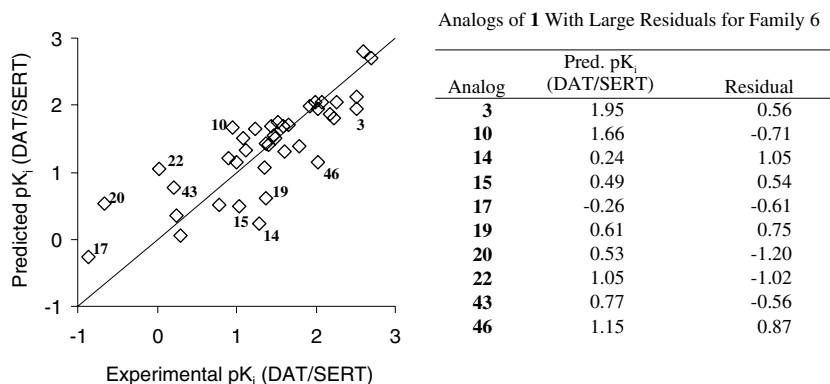
Analogs of **1** With Large Residuals for Family 6

Analog	Pred. pK_i (DAT/SERT)	Residual
3	1.95	0.56
10	1.66	−0.71
14	0.24	1.05
15	0.49	0.54
17	−0.26	−0.61
19	0.61	0.75
20	0.53	−1.20
22	1.05	−1.02
43	0.77	−0.56
46	1.15	0.87

Figure 3. Training set predictions for the Family 6, two-component model. Numbers on the plot identify analogs in the accompanying list of analogs of **1** (GBR 12909) with large residuals. Residual is equal to (Exp. pK_i (DAT/SERT) − Pred. pK_i (DAT/SERT)).

(residual equals the experimental pK_i (DAT/SERT) minus the predicted pK_i (DAT/SERT)) are 0.00 for all three models. Figure 3 shows the plot of the predicted versus the experimental pK_i (DAT/SERT) values derived from the Family 6 model. The r^2 value of 0.685 appears as deviations of data points from the ideal diagonal line in Figure 3. The average of the absolute values of the residuals is 0.34 for Family 6. The analogs with the largest residuals (greater than 0.5) are listed in Figure 3. Of the 10 analogs, half are tetrahydronaphthalene derivatives.

Figure 4 illustrates a possible explanation for why some analogs have poorly predicted pK_i (DAT/SERT) values. Six of the 10 poorly predicted analogs are colored red and extend into regions of space where no other analog lies. In the LOO/CV procedure, each analog is deleted from the training set in turn and its pK_i (DAT/SERT) is predicted from the steric and electrostatic fields of the remaining analogs in the set. If the deleted analog occupies a part of space not common to any other analogs, there will be a lack of steric and electrostatic field information at crucial lattice locations, potentially leading to a poor prediction due to extrapolation. This appears to be the case for the six analogs indicated in the figure. Predictions tend to be most accurate for sets of analogs that occupy very similar regions of conformational space.
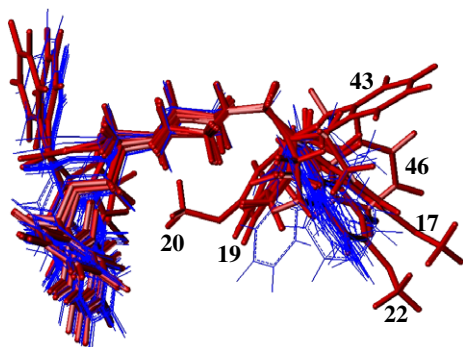

Figure 4. Conformational Family 6 with **17**, **19**, **20**, **22**, **43**, and **46** represented as red capped sticks. Other analogs in Tables 1A and B are shown as blue lines.
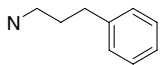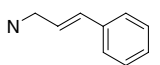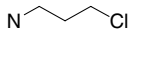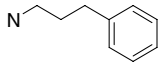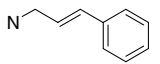
3.4. Test set correlation validation

3.4.1. Original test set. The three stable, two-component Family 1, 5, and 6 models were used to predict the pK_i (DAT/SERT) values of six test set compounds (the piperazines **5**, **8**, and **13**, and the piperidines **29**, **32**, and **35**). The predicted pK_i (DAT/SERT) values and the associated residuals are given in the Supplementary Data. The average residuals are −0.31 for Family 1, −0.32 for Family 5, and −0.25 for Family 6. These numbers are noticeably higher than the 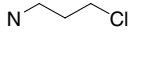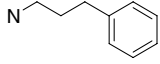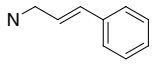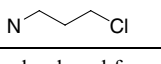average residuals found for the training set and because of this, correlation validation calculations were not carried out for this test set. However, the piperazine **8** is a significant outlier for all three families, predicted to have a much higher pK_i (DAT/SERT) than the experimental pK_i (DAT/SERT) for both models. Interestingly, **8** was previously identified as an outlier among **1** analogs because its experimental pK_i (DAT/SERT) (−0.04) was inconsistent with the pK_i (DAT/SERT) values of similar compounds.[46]

3.4.2. Test sets excluding analog 8. Two additional test sets which excluded **8** were constructed: a 'reduced' test set of five analogs (**5**, **13**, **29**, **32**, and **35**) and a larger, 'extended' test of seven analogs (**1**, **5**, **13**, **32**, **35**, **47**, and **48**). The structures and predicted pK_i (DAT/SERT) values for **1**, **47**, and **48** are given in Table 4. Test set predictions were used to validate the models, according to methods outlined by Golbraikh and Tropsha.[61]

Correlation validation results for both test sets are given in the Supplementary Data and are summarized below. For all three family models, the test set correlation validation values for the extended test set are higher than those for the reduced test set. The coefficient of determination (R^2) of the analogs in the reduced test set are similar to each other (0.264, 0.304, and 0.302, respectively, for Families 1, 5, and 6), and are less than half the goodness-of-fit (r^2) found for the training set (0.685, 0.724, and 0.685, respectively), chiefly because of large residuals for **29**. For the extended test set, Family 6 exhibits the highest coefficients of determination of the three stable models: all values are above 0.7 (e.g., 0.746, 0.735, and 0.722 for R^2, R_0^2, and $R_0'^2$, respectively). This is slight-

Table 4. pK_i(DAT/SERT) predictions for extended test set compounds

$F^{[a]}$	Molecule	X = N N–CH$_2$–R[b]	Exp. pK_i (DAT/SERT)[c]	Pred. pK_i (DAT/SERT)[d]	Residual[e]
1	**1**	N ⟋⟍⟋ phenyl	1.54	1.70	−0.16
	47	N ⟋=⟋ phenyl	2.18	1.71	0.47
	48	N ⟋⟍⟋ Cl	0.91	1.61	−0.70
5	**1**	N ⟋⟍⟋ phenyl	1.54	1.66	−0.12
	47	N ⟋=⟋ phenyl	2.18	1.77	0.41
	48	N ⟋⟍⟋ Cl	0.91	1.14	−0.23
6	**1**	N ⟋⟍⟋ phenyl	1.54	1.61	−0.07
	47	N ⟋=⟋ phenyl	2.18	1.65	0.53
	48	N ⟋⟍⟋ Cl	0.91	0.90	−0.01

[a] Conformational family of analogs developed from representative conformers.

[b] Piperazine compounds, X and A-side substituent R as identified in Figure 1. Note that the expression N–CH$_2$–R indicates the nitrogen of the central ring with its –CH$_2$–R substituent.

[c] pK_i (DAT/SERT) calculated from Eq. 1 using experimental DAT and SERT binding affinities. DAT K_i (nM) and SERT K_i (nM) are: 3.7 and 126 for **1**;[17] 0.9 and 135 for **47**;[47] and 42 and 338 for **48**.[49]

[d] The pK_i (DAT/SERT) predicted by each family's best 3D-QSAR model (highlighted in Tables 2 and 3).

[e] Equals (Exp. pK_i(DAT/SERT) − Pred. pK_i (DAT/SERT)). Residual is calculated from the exact values, and therefore may vary slightly from the apparent difference of the rounded-off values in columns 4 and 5.
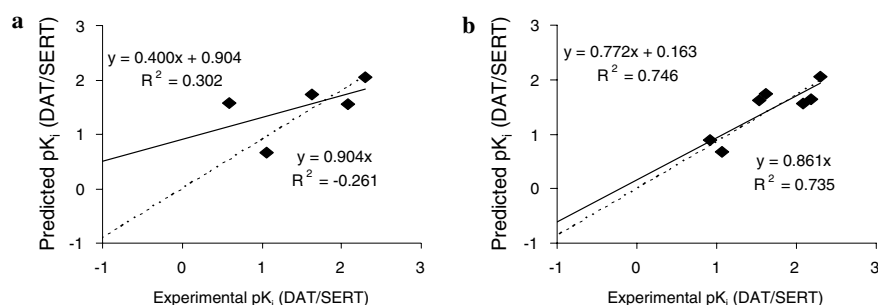
Figure 5. Test set predictions for the Family 6, two-component CoMFA model. (a) Reduced test set. (b) Extended test set. Best-fit (yielding R^2) and zero-intercept (yielding R_0^2 and $R_0'^2$ in each pair of graphs, respectively) lines shown.

ly more than the goodness-of-fit value (0.685) for prediction of the training set compounds by the same two-component Family 6 model. The improved correlation of the extended test set results compared to those of the reduced test set can also be seen by comparing Figure 5a and b. This method of CoMFA model validation indicates that, of the three stable two-component models, the Family 6 model can predict test set compounds the most satisfactorily. In order to gain structural insight based on the CoMFA study, the

Family 6 model was used to calculate the CoMFA contour maps described below.

3.5. Interpretation of three-dimensional CoMFA contour maps

Figure 6 shows the CoMFA steric/electrostatic contour map for the Family 6 model, with **3** shown as a reference molecule. This model is based on the Family 6 conformations, with CoMFA settings of 30 kcal/mol for the
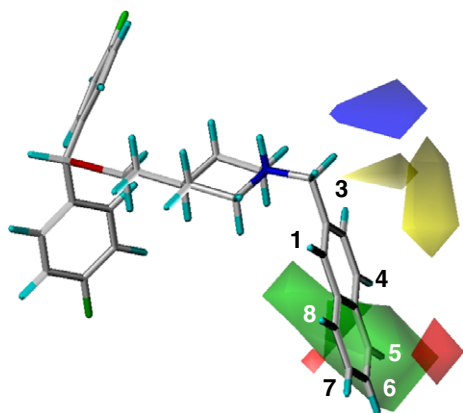
Figure 6. CoMFA steric/electrostatic contour map for Family 6 model. Numbers indicate positions of potential naphthyl substituents. Green (yellow) areas indicate where an increase (decrease) in bulk would lead to a higher pK_i (DAT/SERT). Blue (red) areas indicate where an increase in positive (negative) charge would lead to a higher pK_i (DAT/SERT).

steric cutoff and 10 kcal/mol for the electrostatic cutoff, which gave a q^2 value of 0.508 with two components and a SEP of 0.601. Non-cross-validated results for this model include a r^2 value of 0.685 and a standard error of estimate of 0.481. Steric contribution to the model was 65% and the electrostatic contribution was 35%.

There are several key features of the CoMFA contour map that are predicted to increase DAT/SERT selectivity:

- Less bulk near positions 3 and 4 of the 2-naphthyl substituent
- More bulk near positions 5, 6, and 7 of the 2-naphthyl substituent
- A more positive environment above the naphthyl plane, away from the center ring, extending axially from position 3
- A more negative environment near position 6

Of most interest is the area near positions 5, 6, and 7, where more bulk is predicted to increase pK_i (DAT/SERT). However, the multiple small areas where less bulk leads to a higher pK_i (DAT/SERT) restrict opportunities for increasing pK_i (DAT/SERT) by adding bulky substituents. The red areas near positions 6 and 7 correspond to a *para* substituent on a phenyl analog of **3**, which is supported by the high pK_i (DAT/SERT) values of **30** and **36**. Both the –CF$_3$ and –CN groups impart some bulk and negative electrostatics. The CoMFA contour map elements indicate the potential for using these or related moieties in the development of 3,4-disubstituted phenyl analogs with improved DAT/SERT selectivity.

4. Discussion

4.1. Using multiple fields and field settings for 3D-QSAR development

A range of electrostatic and steric cutoffs for CoMFA columns and different CoMSIA columns were used in this

study. For this data set and alignment method, the q^2 values found for the best settings were from 0.020 to 0.078 higher than those found for the default CoMFA setting of 30 kcal/mol for steric and electrostatic cutoffs. The Venanzi group's CoMFA study of methylphenidate analogs showed up to a 30% increase in q^2 values for steric and electrostatic cutoffs other than the default values, although a larger range of steric and electrostatic cutoffs was used.[25] The amount of time necessary to perform eight extra CoMFA and three extra CoMSIA calculations is minimal compared to the time required to perform conformational analysis and alignment, so it is believed that routinely including these extra field columns and requisite PLS runs may be worth the time spent.

In this study, CoMSIA fields gave the best q^2 values for two of the six families, Family 2 and Family 4. These two families had the third and fourth highest q^2 values of the six families in the preliminary 3D-QSAR studies. After focusing the models, Family 4 (steric/electrostatic field CoMSIA model) had a minimal increase in q^2, making it by far the lowest q^2 value of the six (or eight, including the two-component models for Families 1 and 6) focused models. Family 2 (hydrophobic field CoMSIA model) dropped from third to fourth place when ranking q^2 values. Increase in q^2 values for the CoMSIA families was found to be minimal (0.020) for Family 4, and low (0.107) for Family 2. This is in contrast to the other four families; the four CoMFA models were improved to a greater extent by focusing (with increases in q^2 equal to 0.189/0.188, 0.120, 0.128, 0.140/0.120 for Families 1, 3, 5, and 6, respectively). It is unclear whether this small improvement when applying Region Focusing to the CoMSIA models is somehow related to CoMSIA models themselves, or is an artifact of these particular models. More extensive studies of CoMSIA models would be necessary to determine if this is a general effect.

4.2. Use of same template for piperazine and piperidine analogs

Separate random searches were performed on the piperazine **2** and the piperidine **3**. Because the representative conformers clustered well for both **2** and **3**, 3D-QSAR studies using **2** as the only template, building the piperidine analogs of **1** by substituting a methyne substituent for the B-side nitrogen, were initially tested. Predictivities were poor, so the present study was done using both **2** and **3** as templates. This resulted in different B-side conformations as an alignment issue; most of the B-sides for the **2** templates did not match their **3** counterparts' B-sides. It is noted that the CoMFA map (Fig. 6) has no areas of interest near the B-side. It was found that the positions for the B-sides appeared not to vary enough between each pair of template conformers to affect the CoMFA contour map significantly, and did not have an impact on pK_i (DAT/SERT) prediction.

4.3. Inclusion of racemic compounds in the 3D-QSAR studies

Forty-three of the 48 compounds studied either have no chiral centers or were isolated as a specific stereoisomer.

Five compounds (**13**, **14**, **15**, **16**, and **17**)[47] were synthesized as racemic mixtures. All five are cyclic ketones with one chiral center; three are single ring compounds and two have two fused rings. These compounds were modeled with the chirality that best aligned with the 2-naphthyl substituent in **2**, with the ketone group extending into the same area covered by the naphthyl substituent.

To test the sensitivity of the prediction of training set pK_i (DAT/SERT) values to the inclusion of racemic compounds in the training set, the four racemic compounds in the training set (**14**, **15**, **16**, and **17**) were removed. Using the Family 6 conformational family, a preliminary set of nine CoMFA and three CoMSIA studies were carried out to model the pK_i (DAT/SERT) of the 35 non-racemic compounds. Region Focusing was applied to the model with the highest q^2 from the preliminary studies. The results are listed in the Supplementary Data and show decreased q^2 values for all CoMFA and CoMSIA parameter/field combinations compared to the results when racemic compounds are included. This suggests that the racemic compounds provide useful information that increases the predictivity of the model.

Other 3D-QSAR studies incorporating data from racemic mixtures have been recently carried out. Venanzi et al.[25] used binding affinities for a set of phenyl-substituted methylphenidate analogs to develop a stable and predictive model for DAT binding. The study of Budriesi et al.[62] on selective myocardial calcium channel modulators used racemic data for all but one of the compounds, and made chirality assumptions for both one- and two-chiral center compounds; this data was used as an input for a 3D-QSAR study. Other 3D-QSAR studies, such as that of Gallardo-Godoy et al.[63], do not explicitly mention racemic mixtures, but the structures of certain compounds (e.g., **29** and **30**) implicitly indicate chiral centers.

4.4. 'Direct' versus 'indirect' prediction of pK_i (DAT/SERT) of training set analogs

Since pK_i (DAT/SERT) is a ratio of two binding affinities, one can model DAT/SERT selectivity *directly* by using the experimental values of pK_i (DAT/SERT) (found in Tables 1A and B) in the CoMFA and CoMSIA calculations, or *indirectly* by first modeling the DAT and SERT binding affinities separately, then calculating the ratio according to Eq. 1 using the predicted DAT and SERT binding affinities. It is possible that the direct approach might fortuitously model that ratio with parameters that are unrelated to an accurate representation of the underlying DAT and SERT binding affinities of the analogs. For that reason, the results of the direct and indirect approaches were compared using one of the conformational families. The direct approach was used in the calculations described in the body of this paper. The conformational family which gave the most stable, focused model by the direct approach (i.e., the two-component Family 6 model) was used to construct a model by the indirect approach.

For the indirect approach, CoMFA and CoMSIA studies were performed to separately predict the DAT and SERT binding affinity of all nine test set analogs (**1**, **5**, **8**, **13**, **29**, **32**, **35**, **47**, and **48**). As above, a preliminary set of nine CoMFA and three CoMSIA studies was used to model DAT binding affinity. Region Focusing was applied to the model with the highest q^2 from the preliminary studies. The procedure was repeated to model SERT binding affinity. The focused DAT and SERT binding models were used to predict the DAT and SERT binding affinities of the test set compounds, and their pK_i (DAT/SERT) was calculated from Eq. 1 using these predicted values. These predictions are summarized in the Supplementary Data where they are compared to both the pK_i (DAT/SERT) of the test set analogs predicted from Eq. 1 using the experimental values of their DAT and SERT binding affinities, as well as to the pK_i (DAT/SERT) predicted by the direct method described above. The DAT and SERT binding affinity results compare well to the experimental results, with the exception of the outliers **8** and **29**, whose pK_i (DAT/SERT) was noted above as being poorly modeled by the direct approach. For **47**, the DAT binding and SERT binding affinities were predicted somewhat poorly by the indirect method, but the pK_i (DAT/SERT) calculated from that method is closer to the experimental value than the direct model. Differences between the predicted pK_i (DAT/SERT) values calculated by the direct versus the indirect modeling approach are less than 0.3 for all test set compounds, with five of the nine compounds having differences less than 0.1. The results of indirectly modeling pK_i (DAT/SERT) by first modeling the DAT and SERT binding affinity separately appear to be consistent with the results of the protocol where the pK_i (DAT/SERT) is modeled directly.

5. Conclusion

This work illustrates a computational protocol for a 3D-QSAR study of a highly flexible family of drug-like molecules. Representative conformers were gleaned from large sets of random search conformers using hierarchical clustering. 3D-QSAR runs were performed on six conformational families, which were developed by modifying representative conformers of two template molecules. The process of developing a 3D-QSAR study was carried out in parallel; considering different steric and electrostatic cutoffs for CoMFA and different fields for CoMSIA yielded models of varying predictivity. Comparison of the q^2 values after focusing the 3D-QSAR models was used initially to determine potentially acceptable models. Acceptable models were further screened by internal validation (*y*-value scrambling)[57,60] and external validation (test set correlation validation).[61] The field contour maps were used to identify key regions of the molecule that could be modified for a possible increase in pK_i (DAT/SERT). Direct and indirect protocols for modeling pK_i (DAT/SERT) were shown to give consistent results.

The resulting stable and predictive CoMFA model defines important molecular features of GBR 12909 ana-

logs that can be exploited to develop new compounds with high selectivity for the DAT over the SERT. The CoMFA contour maps indicate that more bulk near positions 5, 6 and 7 of the napthyl ring, as well as more negative electrostatics near positions 6 and 7, are predicted to increase pK_i (DAT/SERT). Substituents such as $-CF_3$ and $-CN$ appear to have the required characteristics. This indicates that there may be the potential for development of 3,4-disubstituted phenyl analogs that have improved DAT/SERT selectivity by substituting these or related moieties at those positions.

Acknowledgments

Supplementary data

Additional data, plots, and molecular graphics are available. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2006.09.070.

References and notes

1. Carroll, F. I. *J. Med. Chem.* **2003**, *46*, 1775.
2. Kuhar, M. J.; Ritz, M. C.; Boja, J. W. *TINS* **1991**, *14*, 299.
3. Kuhar, M. J. *Ciba Found. Symp.* **1992**, *166*, 81.
4. Chen, N. H.; Reith, M. E. A. Structure-Function Relationships for Biogenic Amine Neurotransmitter Transporters. In *Contemporary Neuroscience: Neurotransmitter Transporters: Structure, Function, and Regulation*; Reith, M. E. A., Ed.; Humana Press: Totowa, NJ, 1997; pp 53–109.
5. Carroll, F. I.; Lewin, A. H.; Mascarella, S. W. Dopamine-Transporter Uptake Blockers. In *Contemporary Neuroscience: Neurotransmitter Transporters: Structure, Function, and Regulation*; Reith, M. E. A., Ed.; Humana Press: Totowa, NJ, 1997; pp 381–432.
6. Singh, S. *Chem. Rev.* **2000**, *100*, 925.
7. Dutta, A. K.; Zhang, S.; Kolhatkar, R.; Reith, M. E. A. *Eur. J. Pharmacol.* **2003**, *479*, 93.
8. Glowa, J. R.; Wojnicki, F. H. E.; Matecka, D.; Bacher, J.; Mansbach, R. S.; Balster, R. L.; Rice, K. C. *Exp. Clin. Psychopharmacol.* **1995**, *3*, 219.
9. Prisinzano, T.; Rice, K. C.; Baumann, M. H.; Rothman, R. B. *Curr. Med. Chem.—Central Nervous Syst. Agents* **2004**, *4*, 47.
10. Elmer, G. I.; Brockington, A.; Gorelick, D. A.; Carroll, F. I.; Rice, K. C.; Matecka, D.; Goldberg, S. R.; Rothman, R. B. *Pharmacol. Biochem. Behav.* **1996**, *53*, 911.
11. Baumann, M. H.; Char, G. U.; de Costa, B. R.; Rice, K. C.; Rothman, R. B. *J. Pharm. Exp. Ther.* **1994**, *271*, 1216.
12. Baumann, M. H.; Ayestas, M. A.; Sharpe, L. G.; Lewis, D. B.; Rice, K. C.; Rothman, R. B. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 1190.
13. Villemagne, V. L.; Wong, D. F.; Yokoi, F.; Stephane, M.; Rice, K. C.; Matecka, D.; Clough, D. J.; Dannals, R. F.; Rothman, R. B. *Synapse* **1999**, *33*, 268.
14. Boos, T. L.; Greiner, E.; Calhoun, W. J.; Prisinzano, T. E.; Nightingale, B.; Dersch, C. M.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. *Bioorg. Med. Chem.* **2006**, *14*, 3967.
15. Greiner, E.; Boos, T. L.; Prisinzano, T. E.; De Martino, M. G.; Zeglis, B.; Dersch, C. M.; Marcus, J.; Partilla, J. S.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. *J. Med. Chem.* **2006**, *49*, 1766.
16. Vaughan, R. A.; Gaffaney, J. D.; Lever, J. R.; Reith, M. E. A.; Dutta, A. K. *Mol. Pharmacol.* **2001**, *59*, 1157.
17. Prisinzano, T.; Greiner, E.; Johnson, E. M., II; Dersch, C. M.; Marcus, J.; Partilla, J. S.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. *J. Med. Chem.* **2002**, *45*, 4371.
18. Dutta, A. K.; Davis, M. C.; Reith, M. E. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2337.
19. Gilbert, K. M.; Venanzi, C. A. *J. Comput. Aided Mol. Des.* **2006**, *20*, 209.
20. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
21. Klebe, G.; Abraham, U. *J. Comput. Aided Mol. Des.* **1999**, *13*, 1.
22. Carroll, F. I.; Gao, Y.; Rahman, M. A.; Abrams, P.; Parham, K.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. *J. Med. Chem.* **1991**, *34*, 2719.
23. Carroll, F. I.; Mascarella, S. W.; Kuzemko, M. A.; Gao, Y.; Abraham, P.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. *J. Med. Chem.* **1994**, *37*, 2865.
24. Kulkarni, S. S.; Newman, A. H.; Houlihan, W. J. *J. Med. Chem.* **2002**, *45*, 4119.
25. Venanzi, C. A.; Misra, M.; Gilbert, K. M.; Buono, R. A., New Jersey Institute of Technology, unpublished results.
26. Kulkarni, S. S.; Grundt, P.; Kopajtic, T.; Katz, J. L.; Newman, A. H. *J. Med. Chem.* **2004**, *47*, 3388.
27. Robarge, M. J.; Agoston, G. E.; Izenwasser, S.; Kopajtic, T.; George, C.; Katz, J. L.; Newman, A. H. *J. Med. Chem.* **2000**, *43*, 1085.
28. Nicklaus, M. C.; Wang, S.; Driscoll, J.; Milne, G. W. A. *Bioorg. Med. Chem.* **1995**, *3*, 411.
29. Veith, M.; Hirst, J. D.; Brooks, C. L., III. *J. Comput. Aided Mol. Des.* **1998**, *12*, 563.
30. Boström, J.; Norrby, P.-O.; Liljefors, T. *J. Comput. Aided Mol. Des.* **1998**, *12*, 383.
31. Debnath, A. K. *J. Chem. Inf. Comp. Sci.* **1999**, *38*, 761.
32. Debnath, A. K. *J. Med. Chem.* **1999**, *42*, 249.
33. Perola, E.; Charifson, P. S. *J. Med. Chem.* **2004**, *47*, 2499.
34. Lieske, S. F.; Yang, B.; Eldefrawi, M. E.; MacKerell, A. D., Jr.; Wright, J. *J. Med. Chem.* **1998**, *41*, 864.
35. Benedetti, P.; Mannhold, R.; Cruciani, G.; Pastor, M. *J. Med. Chem.* **2002**, *45*, 1577.
36. Guarnieri, F.; Weinstein, H. *J. Am. Chem. Soc.* **1996**, *118*, 5580.
37. Hopfinger, A. J.; Tokarski, J. S. Three-Dimensional Quantitative Structure–Activity Relationship Analysis. In *Practical Application of Computer-Aided Drug Design*; Charifson, P. S., Ed.; Marcel Dekker: New York, 1997; pp 105–164.
38. Barnett-Norris, J.; Guarnieri, F.; Hurst, D. P.; Reggio, P. H. *J. Med. Chem.* **1998**, *41*, 4861.
39. Barnett-Norris, J.; Hurst, D. P.; Lynch, D. L.; Guarnieri, F.; Makriyannis, A.; Reggio, P. H. *J. Med. Chem.* **2002**, *45*, 3649.
40. Greenidge, P. A.; Merette, S. A. M.; Beck, R.; Dodson, G.; Goodwin, C. A.; Scully, M. F.; Spencer, J.; Weiser, J.; Deadman, J. J. *J. Med. Chem.* **2003**, *46*, 1293.

41. Bernard, D.; Coop, A.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2003**, *125*, 3101.
42. Bernard, D.; Coop, A.; MacKerell, A. D., Jr. *J. Med. Chem.* **2005**, *48*, 7773.
43. Gilbert, K. M.; Skawinski, W. J.; Misra, M.; Paris, K. A.; Naik, N. H.; Deutsch, H. M.; Venanzi, C. A. *J. Comput. Aided Mol. Des.* **2004**, *18*, 719.
44. Misra, M.; Banerjee, A.; Davé, R. N.; Venanzi, C. A. *J. Chem. Inf. Model* **2005**, *45*, 610.
45. Fiorentino, A.; Pandit, D.; Gilbert, K. M.; Misra, M.; Dios, R.; Venanzi, C. A. *J. Comput. Chem.* **2006**, *27*, 609.
46. Matecka, D.; Lewis, D.; Rothman, R. B.; Dersch, C. M.; Wojnicki, F. H. E.; Glowa, J. R.; De Vries, A. C.; Pert, A.; Rice, K. C. *J. Med. Chem.* **1997**, *40*, 705.
47. Lewis, D. B.; Zhang, Y.; Prisinzano, T.; Dersch, C. M.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1385.
48. Greiner, E.; Prisinzano, T.; Johnson, E. M., II; Dersch, C. M.; Marcus, J.; Partilla, J. S.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. *J. Med. Chem.* **2003**, *46*, 1465.
49. Hsin, L. W.; Prisinzano, T. P.; Wilkerson, C. R.; Dersch, C. M.; Horel, R.; Jacobson, A. E.; Rothman, R. B.; Rice, K. C. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 553.
50. Dutta, A. K.; Meltzer, P. C.; Madras, B. K. *Med. Chem. Res.* **1993**, *3*, 209.
51. Skawinski, W. J., New Jersey Institute of Technology, personal communication.
52. SYBYL, Version 6.9; Tripos, Inc.: St. Louis, Missouri, 2000.
53. Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982.
54. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
55. Bush, B. L.; Nachbar, R. B., Jr. *J. Comput. Aided Mol. Des.* **1993**, *7*, 587.
56. Lindgren, F.; Geladi, P.; Rannar, S.; Wold, S. *J. Chemometrics* **1994**, *8*, 349.
57. Clark, R. D.; Sprous, D. G.; Leonard, J. M. Validating models based on large data sets. In *Rational Approaches to Drug Design: 13th European Symposium on Quantitative Structure–Activity Relationships*; Holtje, H. D., Sippl, W., Eds.; Prous Science, S.A: Barcelona, Spain, 2001; pp 475–485.
58. Tripos Bookshelf, SYBYL Version 6.9.1, [QSAR Manual], Tripos, Inc.: St. Louis, MO, 2003.
59. Clark, M.; Cramer, R. D., III *Quant. Struct.–Act. Relat.* **1993**, *12*, 137.
60. Clark, R. D.; Fox, P. C. *J. Comput. Aided Mol. Des.* **2004**, *18*, 563.
61. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Mod.* **2002**, *20*, 269.
62. Budriesi, R.; Carosati, E.; Chiarini, A.; Cosimelli, B.; Cruciani, G.; Ioan, P.; Spinelli, D.; Spisani, R. *J. Med. Chem.* **2005**, *48*, 2445.
63. Gallardo-Godoy, A.; Fierro, A.; McLean, T. H.; Castillo, M.; Cassels, B. K.; Reyes-Parada, M.; Nichols, D. E. *J. Med. Chem.* **2005**, *48*, 2407.