2D QSAR Modeling and Preliminary Database Searching for Dopamine Transporter Inhibitors Using Genetic Algorithm Variable Selection of Molconn Z Descriptors

Brian T. Hoffman, Theresa Kopajtic,† Jonathan L. Katz,† and Amy Hauck Newman*

Medicinal Chemistry and Psychobiology Sections, National Institute on Drug Abuse−Intramural Research Program, Baltimore, Maryland 21224

Received September 16, 1999

In light of the chronic problem of abuse of the controlled substance cocaine, we have investigated novel approaches toward both understanding the activity of inhibitors of the dopamine transporter (DAT) and identifying novel inhibitors that may be of therapeutic potential. Our most recent studies toward these ends have made use of two-dimensional (2D) quantitative structure−activity relationship (QSAR) methods in order to develop predictive models that correlate structural features of DAT ligands to their biological activities. Specifically, we have adapted the method of genetic algorithms−partial least squares (GA-PLS) (Cho et al. *J. Comput.-Aided Mol. Des.*, submitted) to the task of variable selection of the descriptors generated by the software Molconn Z. As the successor to the program Molconn X, which generated 462 descriptors, Molconn Z provides 749 chemical descriptors. By employing genetic algorithms in optimizing the inclusion of predictive descriptors, we have successfully developed a robust model of the DAT affinities of 70 structurally diverse DAT ligands. This model, with an exceptional q^2 value of 0.85, is nearly 25% more accurate in predictive value than a comparable model derived from Molconn X-derived descriptors ($q^2 = 0.69$). Utilizing activity-shuffling validation methods, we have demonstrated the robustness of both this DAT inhibitor model and our QSAR method. Moreover, we have extended this method to the analysis of dopamine D_1 antagonist affinity and serotonin ligand activity, illustrating the significant improvement in q^2 for a variety of data sets. Finally, we have employed our method in performing a search of the National Cancer Institute database based upon activity predictions from our DAT model. We report the preliminary results of this search, which has yielded five compounds suitable for lead development as novel DAT inhibitors.

Introduction

Drug abuse continues to remain one of the most difficult and costly issues of modern society, and cocaine is among the most heavily abused and devastating illicit substances. Although the factors influencing cocaine addiction and treatment are complex, converging evidence suggests a prevalent role of the dopaminergic system in the mechanism of action of cocaine. Research aimed at testing this "dopamine hypothesis" in cocaine abuse has yielded evidence implicating the dopamine transporter (DAT) as the monoamine transporter most closely associated with the reinforcing effects of cocaine.[2,3] The resultant concept of the mechanism of action of cocaine is that binding of cocaine at the DAT blocks the translocation of dopamine from the synaptic cleft into the nerve terminal, potentiating its effect at postsynaptic receptors. Although numerous cocaine analogues have been reported with exceptional affinity for the DAT, many of these compounds demonstrate a cocaine-like behavioral profile in animal models of drug abuse.[4] Recent evidence suggests that certain classes of dopamine uptake inhibitors structurally distinct from cocaine also exhibit a non-cocaine pharmacological profile and therefore may provide better leads toward a cocaine-abuse therapeutic.[5−7] Hence, there is a need for compounds with structures divergent from those that have been tested to date, as a novel arrangement of the pharmacophore or other structural properties may be required to yield the desired pharmacological profile.

To further our drug design efforts, we have developed molecular models that could help to both understand pharmacological data and predict novel biologically active compounds. In doing so, we have relied upon the method of quantitative structure−activity relationship (QSAR) studies, as it remains the major approach for developing predictive correlations between ligand structure and activity. Numerous QSAR approaches have been developed over the years. The accumulation of experimental three-dimensional (3D) structural information about drug molecules[8,9] has led to the development of 3D structural descriptors and associated 3D QSAR methods. Examples of such descriptors include 3D shape descriptors used in molecular shape analysis[10,11] and the steric and electrostatic field sampling implemented in comparative molecular field analysis (CoMFA;[12] a review of 3D QSAR has been published recently[13]). One common characteristic of these methods, as opposed to traditional QSAR, is a dramatic increase in the number of descriptors, and as this number increases, multiple regression methods become inadequate. Advances in mathematical constructs such

* To whom correspondence should be addressed. Tel: 410-550-1455. Fax: 410-550-1648. E-mail: anewman@intra.nida.nih.gov.
† Psychobiology Section.

10.1021/jm990472s
Published on Web 09/26/2000

as principal component analysis,[14] partial least squares (PLS),[15] and machine learning algorithms (e.g., neural network[16,17]), however, have provided researchers with adequate statistical tools to deal with this problem.

As perhaps the most popular example of 3D QSAR, CoMFA, developed by Cramer et al.,[18] has elegantly combined the power of molecular graphics and PLS technique and has found numerous applications in medicinal chemistry and toxicity analysis.[19−23] The CoMFA procedure requires that the scientist devise alignment precepts for the series that overlap the purported pharmacophore for each molecule; the active conformation and alignment rules are subjective, unless derived from some definitive source. Once aligned, the electrostatic and steric components for each molecule are sampled via calculation of the molecular interaction with a probe atom (e.g., a charged sp^3 carbon atom). The probe atom is systematically positioned throughout a 3D grid, and its interaction with the molecule is calculated at intersecting lattice points within the grid. The resulting parameters are introduced into equations that relate them to the activities of each compound, as indicated:

$$\text{Activity}_1 = \text{Constant}_1 + A_1(\text{steric}_{xyz}) + B_1(\text{electrostatic}_{xyz}) + \ldots + A'_1(\text{steric}_{xyz}) + B'_1(\text{steric}_{xyz}) + \ldots$$

$$\text{Activity}_2 = \text{Constant}_2 + A_2(\text{steric}_{xyz}) + B_2(\text{electrostatic}_{xyz}) + \ldots + A'_2(\text{steric}_{xyz}) + B'_2(\text{steric}_{xyz}) + \ldots$$

$$\text{Activity}_n = \text{Constant}_n + A_n(\text{steric}_{xyz}) + B_n(\text{electrostatic}_{xyz}) + \ldots + A'_n(\text{steric}_{xyz}) + B'_n(\text{steric}_{xyz}) + \ldots$$

Traditional regression methods are limited in that the number of parameters must be considerably smaller than the number of compounds in the data set. Since CoMFA analysis typically results in many more parameters than compounds, PLS regression, an algorithm that circumvents the limitation of traditional regression methods, is used to derive the coefficients for each steric and electrostatic term.

Although the use of methods such as CoMFA has extended the use of QSAR methods beyond that of data sets limited to congeneric compounds, the restrictions of issues of conformational choice and alignment in CoMFA make the use of data sets composed of widely divergent molecular structures daunting, if at all possible. One of the main benefits of 2D QSAR methods is of the circumvention of an underlying assumption of 3D QSAR methods: the proposed conformation is the bioactive one. Effectively, 2D QSAR methods eliminate the task of generation of a large number of possible conformations and the "rational" choice of the most likely conformation that binds to the receptor. Additionally, 2D QSAR methods may be more easily adapted to the task of database searching or similarity/dissimilarity comparisons, since their fairly automated nature and lack of computational complexity suit these 2D methods for such tasks. These considerations led to the development of variable selection QSAR algorithms such as genetic algorithm-partial least squares (GA-PLS).[1] As opposed to utilizing 3D steric and electrostatic descriptor values, the GA-PLS approach relies upon 2D topological descriptors of chemical structures that eliminate the conformational and alignment ambiguities inherent within the CoMFA process. Additionally, the GA-PLS method is less computationally intensive, is practically automated, and has been used to produce highly predictive models that were comparable to, or better than, those obtained with traditional CoMFA.[1,24]

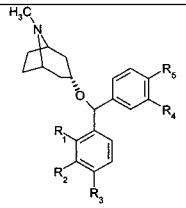
Our 2D QSAR method makes use of both topological and electrotopological state (E-state) descriptors, as calculated by the commercially available software Molconn Z.[25] Molconn Z extends upon the descriptor set produced by its predecessor, Molconn X, via its ability to calculate E-state hydrogen and bond-type descriptors; whereas Molconn X generated 462 descriptors, Molconn Z calculates an additional 287 descriptors (when using default parameters). Due to the relatively large number of descriptors produced by Molconn Z − not all of which make positive contributions in correlative value to a predictive model − a procedure is required to rationally select those descriptors that are meaningful to a particular data set. Such optimization procedures as the genetic algorithm have been the subject of extensive investigations in the computational science realm[26−28] and have recently been incorporated into QSAR studies by a number of research groups.[29−32] In developing our QSAR studies of DAT inhibitors, we have utilized both Molconn X- and Molconn Z-derived descriptor sets, providing an opportunity to gauge the contribution of the additional descriptors provided by Molconn Z to model enhancement. These additional descriptors, the E-state indices, are a combination of electronic and topological information in which an index is calculated for each atom in the molecule, and this index is representative of an atomic-level context with regard to the whole molecule.[33] Additionally, Molconn Z calculates bond-type E-state indices, although we have not yet incorporated those descriptors in our QSAR experiments. Bond-type E-state descriptors are calculated by assigning a valence-state bond type to each bond in the molecule, grouping all bonds of the same type, and adding together the E-state values of each bond type.[33] Inclusion of the bond-type E-state descriptors in our analyses may lead to further enhancement of our QSAR models, and efforts to this end are currently underway.

The crowning accomplishment of any QSAR effort is the identification of novel leads, and in pursuit of such, we have devised and implemented an ongoing 2D QSAR-based database search program. While our screening efforts are ongoing, we have already identified several active leads by utilizing predictions from our model of DAT inhibitor activity and a structural similarity metric aimed at maintaining prediction accuracy. Our preliminary success in database searching has emphasized the utility of our QSAR method.

Results and Discussion

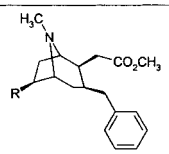
The methods employed for this study are described in detail in the Experimental Section. Tables 1−6 list predicted vs actual biological values for each compound, as determined by GA-PLS-Z analysis. The reference number for the biological data for each compound is also

Table 1. Actual vs Predicted Activity Values for DAT Inhibitors **1**−**25** of 70 Analyzed by GA-PLS-Z

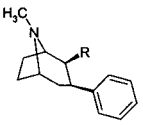


ID	R_1	R_2	R_3	R_4	R_5	Actual activity ($-\log K_i$)	Predicted activity ($-\log K_i$)	Residual	Ref.
1	H	H	H	H	H	8.10	7.44	0.66	35
2	CH_3	H	H	H	H	6.51	6.82	-0.31	39
3	Cl	H	H	H	H	6.64	7.19	-0.55	39
4	F	H	H	H	H	7.30	7.20	0.10	39
5	NH_2	H	H	H	H	5.73	5.80	-0.07	39
6	H	Cl	H	H	H	7.66	7.28	0.38	39
7	H	CF_3	H	H	H	6.72	6.42	0.30	39
8	H	F	H	H	H	7.16	7.33	-0.17	39
9	H	H	F	H	H	7.49	7.34	0.15	40
10	H	F	F	H	H	7.55	7.59	-0.04	40
11	H	F	H	F	H	7.32	7.38	-0.06	39
12	H	F	H	H	F	7.63	7.40	0.23	40
13	H	H	F	H	F	7.93	7.94	-0.01	40
14	H	Cl	Cl	H	H	7.67	7.63	-0.04	40
15	H	Cl	Cl	H	F	7.72	7.79	-0.07	40
16	H	H	Cl	H	H	7.52	7.67	-0.15	39
17	H	H	Cl	H	Cl	7.70	7.88	-0.18	40
18	H	H	OCH_3	H	H	7.10	6.67	0.43	40
19	H	H	OCH_3	H	OCH_3	5.70	5.99	-0.29	40
20	H	H	CF_3	H	H	6.20	6.37	-0.17	39
21	H	H	CH_3	H	H	6.73	6.85	-0.12	40
22	H	H	CN	H	H	6.71	6.71	-0.00	40
23	H	H	OH	H	H	6.53	6.69	-0.16	40
24	H	H	Br	H	H	7.42	7.18	0.24	40
25	H	H	Br	H	Br	7.04	7.18	-0.14	40

Table 2. Actual vs Predicted Activity Values for DAT Inhibitors **26**−**39** of 70 Analyzed by GA-PLS-Z



ID	R	Actual activity ($-\log K_i$)	Predicted activity ($-\log K_i$)	Residual	Ref.
26	H	7.04	6.89	0.15	38
27	methyl	6.67	6.60	0.07	38
28	ethyl	6.51	6.12	0.39	38
29	*n*-propyl	5.38	5.85	-0.47	38
30	*n*-butyl	5.07	5.59	-0.52	38
31	benzyl	5.51	5.49	0.02	38



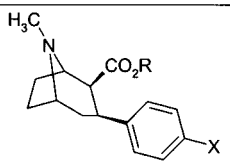
ID	R	Actual activity ($-\log K_i$)	Predicted activity ($-\log K_i$)	Residual	Ref.
32	$CH{=}CHCO_2CH_3$	7.66	7.80	-0.14	37
33	$CH_2CH_2CO_2CH_3$	7.64	7.80	-0.16	37
34	$CH_2CH_2CH_2OH$	7.96	7.85	0.11	37
35	$CH_2CH_2CH{=}CHCO_2CH_3$	7.70	7.58	0.12	37
36	$(CH_2)_4\ CO_2CH_3$	7.52	7.42	0.10	37
37	$CH{=}CHCH_2OH$	7.79	7.82	-0.03	37
38	$CH_2CH_2C(O)C_6H_5$	7.55	7.77	-0.22	37
39	$(CH_2)_3C_6H_5$	7.79	7.60	0.19	37

listed. The use of biological data from one source is of particular significance, preventing interlaboratory variations in the values of dependent variable for the QSAR analysis. Our GA-PLS-Z (GA-PLS based upon descriptors derived from Molconn Z) analysis yielded a q^2 value of 0.85, whereas a comparable analysis of the same data set using GA-PLS with Molconn X descriptors yielded a q^2 of 0.69. In comparing the GA-PLS-Z and GA-PLS-X QSAR methods, the same number of genetic algorithm crossover iterations (10 000) was performed in each case. The statistical data for these analyses, as well as for the D_1 antagonist and serotonin data sets, is summarized in Table 7. These analyses were performed to validate this method with other data sets and to further demonstrate its utility.

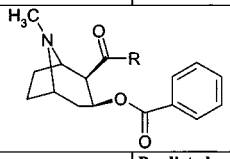
Clearly, the additional descriptors generated by Molconn Z vs Molconn X are very valuable in developing accurate and predictive QSAR models. Figure 1 illustrates the evolution of the maximum q^2 value for the GA-PLS-Z vs GA-PLS-X experiments. The optimal GA-PLS-Z model for the 70 DAT inhibitors investigated in our studies utilized 78 descriptors, as selected by the genetic algorithm component of our analysis from the 270 non-zero variance descriptors generated by Molconn Z (although Molconn Z, under default parameters, generates a total of 749 descriptors for each data set member, 479 of these descriptors were determined to have no variance for our data set). Table 8 lists the identification labels for the 78 optimal descriptors following 10 000 iterations of the GA-PLS-Z procedure (descriptions and methods of calculation for these descriptors are provided online via the Molconn Z web page[25]). Also listed in Table 8 are the correlation coefficients and relative contributions (as a percentage) of each descriptor.

The relative influence of the additional descriptors provided by Molconn Z vs Molconn X is apparent not only in Figure 1 but also in Table 8, by observation of the contribution of each descriptor. Seven of the top 10 descriptors with regard to relative contribution are descriptors calculated produced by Molconn Z, and overall, Molconn Z descriptors account for 49% of the relative contributions of individual descriptors to the model. For instance, one of the descriptors provided by Molconn Z and selected by the genetic algorithm for use in our QSAR model is "SHsNH2", which is the sum of the atom-type E-state indices for all of the amino ($-NH_2$) hydrogens in a molecule. Interestingly, some of the most heavily contributing descriptors are ones not originally intended for use in heterogeneous data sets. Those descriptors with the label "HesX" are the hydrogen electrotopological state index values for atoms in predetermined position X within a common molecular scaffold. For instance, the Hes1 descriptor values represent the hydrogen electrotopological state index value for a hydrogen atom located on the nitrogen common to all of our compounds in the data set. However, Hes6 corresponds to a methylene hydrogen in the benztropine **1** and a methyl hydrogen in bupropion (compound **67**). Intrigued by the inclusion of these descriptors following variable selection with our generally diverse data set, we speculated that the homologous subpopulations (such as tropane analogues) that comprised our data set

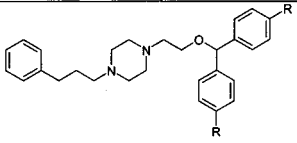
Table 3. Actual vs Predicted Activity Values for DAT Inhibitors **40**−**49** of 70 Analyzed by GA-PLS-Z



ID	X	R	Actual activity (-logK$_i$)	Predicted activity (-logK$_i$)	Residual	Ref.
40	H	CH_3	7.06	7.91	-0.85	35
41	Cl	CH_3	8.70	8.08	0.62	35
42	F	CH_3	7.66	7.91	-0.25	35
43	I	CH_3	8.68	8.58	0.10	35
44	CH_3	CH_3	8.30	7.69	0.61	35
45	I	isopropyl	8.40	8.26	0.14	35



ID	R	Actual activity (-logK$_i$)	Predicted activity (-logK$_i$)	Residual	Ref.
46	CH_3	6.72	6.87	-0.15	35
47	ethyl	6.51	6.49	0.02	35

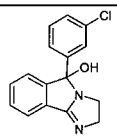
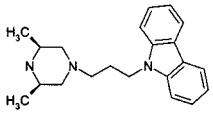
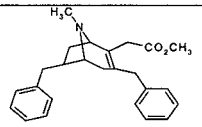
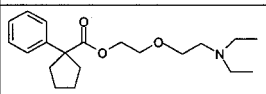
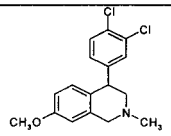
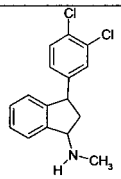


ID	R	Actual activity (-logK$_i$)	Predicted activity (-logK$_i$)	Residual	Ref.
48	H	7.89	7.89	0	35
49	F	7.92	8.11	-0.19	35

afforded use of these descriptors in the QSAR equation. To test this hypothesis, we derived separate models of strictly homologous (compounds **1**−**25**) and heterogeneous (compounds **50**−**70**) compounds using only the electrotopological state index values and hydrogen electrotopological state index values for atoms in predetermined molecular positions. As expected, the QSAR model for the strictly heterogeneous data set was vastly less predictive than that for the homologous data set (q^2 values of −0.01 and 0.80, respectively). Thus, we cannot advocate the use of this subset of Molconn Z descriptors with strictly heterogeneous data sets. However, we have noticed that data sets are frequently composed of one or more congeneric sets of analogues of high-activity compounds and assorted other compounds of varying activities. In this circumstance, we advocate the concomitant use of both variable selection and rigorous model validation (i.e. activity-shuffling experiments) in order to ensure that spurious correlations are not being made between structural features and activity.

Table 9 summarizes the results of our activity-shuffling validation experiments, in which we randomly reassign activity values for data set members. As expected of a statistically valid QSAR method, such activity reassignment experiments resulted in markedly diminished q^2 values, none of which approached the 0.50 q^2 threshold considered to be of predictive value. The method of GA-PLS (as performed in the past with Molconn X descriptors) has been demonstrated to be a

Table 4. Actual vs Predicted Activity Values for DAT Inhibitors **50**−**55** of 70 Analyzed by GA-PLS-Z

ID	Compound	Actual activity (-logK$_i$)	Predicted activity (-logK$_i$)	Residual	Ref.
50		7.80	7.77	0.03	35
51		6.98	7.09	-0.11	34
52		5.32	5.19	0.13	38
53		5.51	5.45	0.06	34
54		8.15	8.21	-0.06	35
55		8.10	8.10	0	35

highly competitive QSAR analytical technique by its application to several data sets.[1,24] This method has certain inherent advantages over 3D methods, such as circumvention of alignment issues and the elimination of the possibility of inappropriate conformation selection. Furthermore, the employment of variable selection procedures by GA-PLS in order to improve the quality of QSAR analysis is essential in dealing with the relatively large number of descriptors produced by Molconn Z analysis. Obviously, these advantages inherent within the GA-PLS-Z method have been instrumental in the resultant highly predictive q^2 value for our data sets. Moreover, the utility of our method is emphasized by our preliminary results from database searching of the National Cancer Institute (NCI) database, comprising more than 237 000 compounds. Table 10 lists the pharmacological data for the 25 compounds identified by our model as the most likely to possess high activity for the DAT. We screened the database by predicting the activity and calculating a similarity metric for each compound in the database. Our reasoning in employing this similarity metric was that prediction accuracy might be best maintained by placing an emphasis on those database compounds located in the closest chemical descriptor space to our training set. It may seem paradoxical that, while our long-term goal is to identify structurally divergent DAT inhibitors with novel pharmacological profiles, we have focused our search on database compounds chemically similar to our existing DAT inhibitors. However, this similarity is calculated in a manner that has led to preliminary success in our long-term goal; we have identified from

Table 5. Actual vs Predicted Activity Values for DAT Inhibitors **56**−**62** of 70 Analyzed by GA-PLS-Z

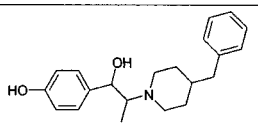
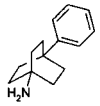
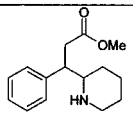
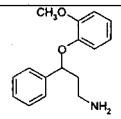
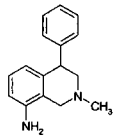
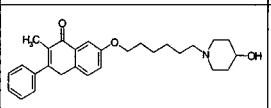
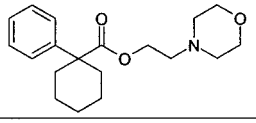
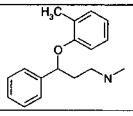
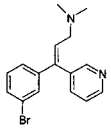
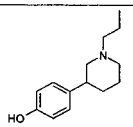
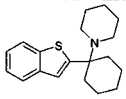
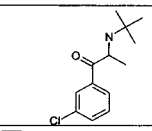
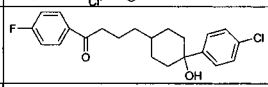
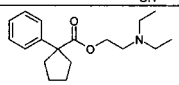
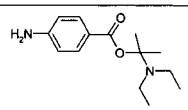
ID	Compound	Actual activity (-logK$_i$)	Predicted activity (-logK$_i$)	Residual	Ref.
56		6.44	6.38	0.06	34
57		7.08	7.30	-0.22	35
58		6.73	6.73	0	35
59		6.25	6.13	0.12	35
60		7.37	7.35	0.02	35
61		5.29	5.26	0.03	34
62		5.27	5.02	0.25	34

Table 6. Actual vs Predicted Activity Values for DAT Inhibitors **63**−**70** of 70 Analyzed by GA-PLS-Z

ID	Compound	Actual activity (-logK$_i$)	Predicted activity (-logK$_i$)	Residual	Ref.
63		6.15	6.18	-0.03	35
64		4.95	4.76	0.19	35
65		6.25	6.39	-0.14	34
66		8.00	7.89	0.11	35
67		6.43	6.46	-0.03	35
68		5.28	5.14	0.14	34
69		5.28	5.41	-0.13	34
70		6.21	6.32	-0.11	35

our first screening five compounds somewhat divergent in structure from our training set that are suitable for development as novel DAT inhibitors. These compounds, with IC_{50} values of approximately 1 μM or better, are depicted in Table 11. Furthermore, a full concentration curve (9 concentrations ranging from 0.01 nM to 100 μM, each performed 2−3 times, with each curve performed in triplicate) for the most potent compound, NSC 64540, was obtained resulting in a $K_i = 114 \pm 7.51$. Because the NCI database is, by nature, composed of compounds geared toward biological targets very dissimilar to our own, the choice of limiting screening candidates with a similarity metric is an appropriate measure. Additionally, each successive screening iteration yields biological data that we may incorporate into our original QSAR model, expanding the sampling of descriptor space and, hence, increasing the threshold of similarity for screening candidates.

Conclusions and Prospectus

We have developed a particularly robust QSAR model for 70 diverse inhibitors of the DAT, constituting the largest QSAR treatment of such compounds in the reported literature. The wide range of structural variations of the DAT ligands comprising our data set makes a successful QSAR analysis by the more traditional method of CoMFA unlikely, due to the complexities of conformational assignment, identification of a common pharmacophore, and rational alignment. Our methodology has been bolstered by the use of data from our laboratory, eliminating variability in data introduced by differences in assay conditions. Moreover, the validity of our method has been illustrated by the poor q^2 values exhibited by data sets in which activity has been randomized. These activity-shuffling experiments discount the possibility that our PLS implementation has "overfit" the structure−activity correlation − a distinct possibility when utilizing more dependent variables than data set members in a QSAR study. In addition to the verification of the existence of an intrinsic activity among the diverse DAT inhibitors in our data set, our studies have also provided the opportunity to evaluate the benefits of use of the additional descriptors provided by Molconn Z. The inclusion of these E-state index descriptors present in our GA-PLS-Z model resulted in substantially higher q^2 values for comparable Molconn X-derived models.

Perhaps the most limiting feature of the GA-PLS-Z method is its insensitivity to stereochemistry of members of the training and prediction data sets. We have dealt with this limitation by choosing the activity of the most active enantiomer as the dependent variable. This choice is based upon our intended application of this model; we have developed a reliable model with the aim of utilizing it in searching databases of existing drugs to identify new DAT drug candidates. The choice of the activity of the most active enantiomer for inclusion in our model ensures that an inaccurate prediction due to

Table 7. Statistical Data for QSAR Method Results

data set	QSAR method	no. of compounds	no. of iterations	no. of variables	q^2	optimal no. of components	standard error of estimate	R^2	F value
DAT	GA-PLS-X	70	10000	53	0.69	4	0.45	0.80	66.16 ($n = 70$, $k = 5$)
DAT	GA-PLS-Z	70	10000	78	0.85	5	0.26	0.93	174.2 ($n = 70$, $k = 5$)
D_1 Ant	GA-PLS-X	29	10000	55	0.68	2	0.57	0.78	45.08 ($n = 29$, $k = 2$)
D_1 Ant	GA-PLS-Z	29	10000	108	0.95	5	0.18	0.98	225.6 ($n = 29$, $k = 5$)
5-HT	GA-PLS-X	14	10000	97	0.68	3	0.15	0.80	14.21 ($n = 14$, $k = 3$)
5-HT	GA-PLS-Z	14	10000	62	0.82	3	0.10	0.91	32.30 ($n = 14$, $k = 3$)

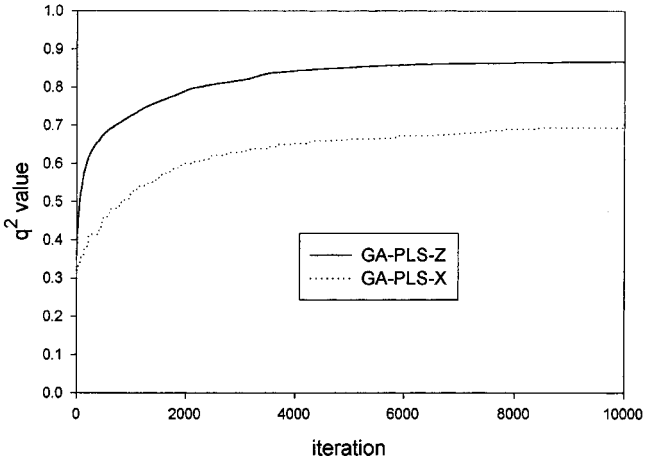


Figure 1. Depiction of the maximum q^2 value yielded per iteration of Molconn Z-derived vs Molconn X-derived GA-PLS models of DAT inhibitor activity.

"noise" introduced by such selection will result in a "false positive" score in the prediction of activity. A common hypothesis in database search theory is that "false negatives", or low predictions of database members that are actually of high activity, are more costly than "false positives", since "false positives" are eventually tested and lead to model improvement while "false negatives" are discarded and no longer considered. While the undesirable exclusion of structural data regarding stereochemistry is currently necessary (since Molconn Z does not support generation of descriptors for the SMILES representations of stereoisomers), we are currently investigating tailored approaches that will extend the capabilities of GA-PLS-Z to that of stereoisomeric analysis.

Another limitation of the 2D QSAR approach can be the lack of easily interpretable information useful for the design of new highly active drugs. For instance, CoMFA analysis yields graphical representations of CoMFA fields – regions in which chemical modifications may lead to an alteration in a compound's pharmacological nature. An analogous interpretation of 2D descriptor values has been reported for congeneric data sets.[33] With the anticipated success of our lead discovery program, we will shift our research focus to that of drug optimization and design and the attendant process of decoding individual Molconn Z descriptors for use in such a program. However, our fundamental goal has been that of utilizing our QSAR efforts in the task of database searching, a process in which interpretation of individual descriptor values and their corresponding contribution toward activity (although heuristically interesting) is unnecessary. A prerequisite to effective database searching to yield a lead divergent in structure from contemporary DAT inhibitors is the use of a data set with heterogeneous components.

Table 8. Descriptor Labels, Correlation Coefficients, and Relative Contributions of the 78 Descriptors Utilized in the Optimized GA-PLS-Z QSAR Model[a]

descriptor label	coefficient	contrib (%)	descriptor label	coefficient	contrib (%)
xp9	−0.0594	0.48	hes6	0.4670	3.32
xp10	−0.1420	0.90	hes7	0.1710	1.06
xvp4	0.0267	0.44	hes8	−0.2420	1.95
xch5	2.0200	1.62	hes10	0.1250	1.23
xch9	−0.6040	0.06	hes11	−0.1920	2.22
xch10	10.3000	0.42	hes19	0.0713	0.82
xvpc4	0.1170	1.08	hes28	0.3710	1.55
xvch5	2.2600	1.54	hes29	0.5210	2.07
xvch10	27.0000	0.31	hes30	0.6680	2.23
dxvp3	0.1360	2.09	hes31	0.6540	2.23
dxvp4	0.1210	1.79	hes32	0.1890	1.06
k2	−0.0628	2.02	idc	−0.0002	1.04
ka3	−0.0022	1.29	Wp	−0.0053	0.80
phia	−0.0920	1.96	knotp	−0.0054	0.86
SHsNH2	−0.2580	1.84	knotpv	−0.0070	0.84
SHother	−0.0349	2.45	numHBd	−0.1310	1.67
Gmax	0.0092	0.60	nxp4	−0.0026	0.58
SssCH2	0.0082	0.37	nxp5	−0.0005	0.17
SsssCH	0.0995	1.89	nxp8	−0.0007	0.42
SdssC	−0.1380	1.08	nxc3	−0.0340	0.97
SsNH3p	−0.0375	0.60	nxch5	0.1230	1.04
SsssN	0.0681	0.79	nxch9	−0.0346	0.20
SdO	−0.0093	1.19	nxch10	0.1170	0.45
SsF	0.0067	1.24	ntpath	−0.0001	0.74
CHsNH2	−0.2940	1.68	tg3	0.1330	1.22
CHCHnX	0.0681	2.07	nasH	−0.0266	2.06
CGmax	0.0092	0.60	nasF	0.0847	1.17
CHmin	1.6000	2.05	nasCl	0.1020	1.11
CssCH2	−0.0095	0.39	nasI	1.0200	3.24
CdsCH	0.1210	1.53	nd1	−0.0355	0.78
CtsC	0.1280	0.29	nd4	−0.1870	1.47
CdssC	−0.0874	1.15	ne12	−0.2050	2.68
CaasC	−0.0020	0.05	ne24	−0.0549	0.91
CtN	0.1280	0.29	n2pe24	−0.0489	1.54
CaaN	−0.6220	1.40	n3pe14	0.1490	1.06
CsF	0.0847	1.17	n3pe24	−0.0800	2.03
hes1	−0.4670	2.16	n4pe11	0.0555	1.00
hes2	−0.6470	4.11	n4pe23	0.0115	1.73
hes3	−0.1650	1.11	n4pe33	−0.0041	0.35

[a] Intercept = 9.39.

Table 9. Summary of Activity-Shuffling Validation Experiments

random data set	q^2 after 10 000 iterations	random data set	q^2 after 10 000 iterations
1	0.21	**11**	0.20
2	0.24	**12**	0.22
3	0.23	**13**	0.11
4	0.22	**14**	0.26
5	0.16	**15**	0.21
6	0.29	**16**	0.11
7	0.02	**17**	0.21
8	0.20	**18**	0.30
9	0.20	**19**	0.22
10	0.12	**20**	0.33

The features of 2D QSAR make it an especially powerful tool in searching databases for novel drug leads, and the advantages of our 2D QSAR approach

Table 10. High-Throughput Pharmacological Analysis of DAT Binding Affinity: Evaluation at Concentrations of 1 μM for Competition Binding Analysis with [^{3}H]WIN 35,428[a]

NSC no.[46]	% specific bound		NSC no.[46]	% specific bound	
	mean	SEM		mean	SEM
5921	93.40	4.15	64540*	8.43	0.26
7465	81.60	3.97	72839	84.91	7.13
9111*	49.67	3.15	80563	90.50	9.12
16377*	42.13	2.62	83393	78.28	0.78
17606	91.02	5.39	99417	77.67	1.81
19471	89.93	6.73	102852	87.89	0.79
22169	88.53	5.05	127721	95.70	4.94
25543	66.02	2.92	142430	86.47	6.44
32475	74.60	2.10	167160	87.17	5.29
33090	88.08	3.86	237796	85.33	3.82
40318	81.92	4.00	289757*	54.32	3.05
40320	78.22	3.62	354657*	55.77	1.30
51830	88.25	4.40			

[a] Compounds indicated by an asterisk may be suitable candidates for further development.

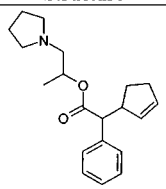
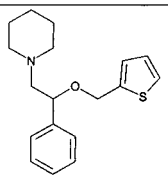
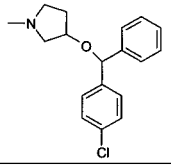
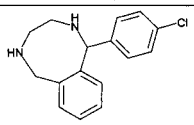
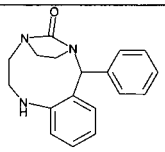
are apparent in the favorable q^2 value from these studies. Given the highly predictive quality of our models and the automated nature of our method, we are optimistic that our efforts will be fruitful not only in assisting in the direction of our synthetic drug design program but also in our ongoing program of database searching. In fact, our preliminary search of the NCI database has been highly successful, with a 20% hit rate in accurately identifying compounds suitable for lead optimization. For these reasons, we consider the GA-PLS-Z method an exceptional tool in developing and utilizing QSAR, especially in the case of structurally unrelated data sets.

Experimental Section

Biological Activity Data. For this work, we have selected 70 previously reported chemically diverse DAT inhibitors (Tables 1−6) whose activities were measured by Dr. Sari Izenwasser.[34−40] This data set includes a wide variety of tropane analogues, as well as divergent representatives of several other series, such as sigma receptor ligands that possess low affinity for the DAT. The competition binding activity of each compound is expressed as the $-\log(K_i)$ vs [^{3}H]-WIN 35,428 binding in rat caudate-putamen. Methods for the measurement of K_i values of DAT inhibitors using these conditions have been reported previously[35,41]

GA-PLS-Z Routine. The GA-PLS-Z program suite was programmed in C language using GNU[42] development software and an SGI Octane R10000 workstation. The algorithm of the GA-PLS method[1] is implemented as follows. Step 1: The Molconn Z[25] program is applied to the simplified molecular line entry system (SMILES)[43] representation of each molecule in order to generate 749 molecular descriptors automatically. Step 2: All descriptors with zero variance are removed. Step 3: All applicable descriptors are numbered arbitrarily, and this enumeration is maintained throughout the entire analysis. A population of 100 different random combinations of these descriptors is generated. To apply genetic algorithm methodology, each descriptor combination is considered to be a parent. Each parent represents a binary string of digits, either "one" or "zero"; the length of each string is the same and is equal to the total number of descriptors (indices). The value of "one" implies that the corresponding descriptor is included for the parent, and "zero" means that the descriptor is excluded. Step 4: Using each parent combination of descriptors, a QSAR equation is generated for the whole data set using the PLS algorithm; thus for each parent an initial value of q^2 is obtained. In the manner of Cho et al.[1] the $[1 - (n - 1)(1 - q^2)/(n - c)]$ expression (in which q^2 is cross-validated R^2, n is the number of compounds, and c is the optimal number of components) is used as the fitting function to guide the GA optimization. Step 5: Two parents are selected randomly. Step 6: The population is evolved by performing a crossover between two randomly selected parents, producing two offspring. Step 7: Each offspring is subjected to a random single-point mutation; one descriptor is included/excluded. Step 8: The fitness of each offspring is evaluated as described above (cf. step 4). Step 9: If the resulting offspring are characterized by a higher value of the fitness function, then they replace less fit parents; otherwise, the parents are kept. Step 10: Steps 5−9 are repeated until a predefined maximum number of crossovers are reached. Further details of this method are described elsewhere.[1]

Table 11. Five Compounds Identified by Database Screening of the NCI Database Determined To Have IC_{50} Values of ~1 μM or Better by High-Throughput Pharmacological Analysis

NSC number[46]	Structure
9111	
16377	
64540	
289757	
354657	

The optimal number of components (ONC) in the final PLS model was determined by the cross-validated R^2 (q^2) and standard error of prediction (SDEP) values, as obtained from the leave-one-out cross-validation technique. The q^2 value was calculated from the following standard equation:

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \tilde{y})^2}$$

in which y_i and $\hat{y}_i$ are the actual activity and the predicted activity of the ith compound, respectively, and $\bar{y}$ is the average activity of all the compounds in the training set. Both summations are inclusive of all compounds in the training set. The number of components with the lowest SDEP value was selected as the ONC.

QSAR Method and Model Validation. As part of our validation procedures aimed at verifying the merit of the GA-PLS-Z method, we have performed activity-shuffling experiments. Using an algorithm to randomly shuffle the activity values for the members of the original data set, we obtained 20 alternate data sets, each with the original structures of our 70 compounds but with different dependent variable assignments. The activity-shuffling algorithm was based upon a roulette wheel selection for 1000 iterations. Data sets for these

alternative activity assignments were then analyzed by the GA-PLS-Z method for 10 000 iterations each.

To gauge the generalization of improvement in q^2 for models developed using Molconn Z vs Molconn X descriptors, we have applied our analyses to two additional data sets. The first, a series of D_1 dopamine antagonists, has been recently investigated via a variety of QSAR techniques.[24] The second data set is composed of serotonin ligands, for which a CoMFA model based upon measurements of efficacy has been reported.[32]

Database Searching. The NCI database was obtained in the SMILES format from the NCI Developmental Therapeutics Program (DTP) web site.[45] For each compound in the database, a prediction of activity at the DAT site was calculated via the QSAR equation derived by our model; the values of those descriptors determined by the genetic algorithm to be most predictive were utilized with the slope intercept to yield a numerical prediction of DAT affinity. A similarity metric was also calculated to provide an additional screening threshold. This measurement of the database compound's similarity to the training set was determined by calculating the cumulative number of standard deviations from the mean of each relevant descriptor. In this manner, the standardized Euclidean distance in *n*-dimensional descriptor space is calculated based upon the mean values of each descriptor's value for the training set compounds. The cumulative sum of this standardized distance for each descriptor type yields a singular numerical metric of similarity that may be used to maintain prediction accuracy. Based upon the concept that the QSAR equation derived by our model most accurately predicts those compounds with descriptor values near or within the range of those observed within the training set, we have used this similarity metric as a second threshold for determining criteria for subsequent pharmacological analysis. From the database of >240 000 compounds, 35 of those with both high activity prediction and similarity were requested from the NCI DTP administrators as the first screening batch. Twenty-five of the requested compounds were available for pharmacological screening.

High-Throughput Pharmacological Analysis of DAT Binding Affinity. To rapidly determine the relative DAT affinity for each compound obtained from the NCI database, we chose an evaluation at concentrations of 1 μM for competition binding analysis with the radioligand [^{3}H]WIN 35,428. Therefore, for this screening, brains from male Sprague−Dawley rats weighing 200−225 g (Taconic Labs) were removed, striatum dissected and placed on ice. Membranes were prepared by homogenizing tissues in 20 volumes (w/v) of ice-cold modified Krebs-HEPES buffer (15 mM HEPES, 127 mM NaCl, 5 mM KCl, 1.2 mM $MgSO_4$, 2.5 mM $CaCl_2$, 1.3 mM NaH_2PO_4, 10 mM D-glucose, pH adjusted to 7.4) using a Brinkman Polytron (setting 6 for 20 s) and centrifuged at 20000*g* for 10 min at 4 °C. The resulting pellet was resuspended in buffer, recentrifuged and resuspended in buffer to a concentration of 10 mg/mL. Ligand binding experiments were conducted in assay tubes containing 0.5 mL modified Krebs-HEPES buffer for 60 min on ice. Each tube contained 1.5 nM [^{3}H]WIN 35428 (specific activity 84 Ci/mmol) and 2.5 mg striatal tissue (original wet weight). Nonspecific binding was determined using 0.1 mM cocaine·HCl. For determination of binding affinity, triplicate samples of membranes were preincubated for 5 min in the presence or absence of 1 μM of the compound being tested. Incubations were terminated by rapid filtration through Whatman GF/B filters, presoaked in 0.1% BSA, using a Brandel R48 filtering manifold (Brandel Instruments Gaithersburg, MD). The filters were washed twice with 5 mL cold buffer and transferred to scintillation vials. Beckman Ready Safe (3.0 mL) was added and the vials were analyzed the next day using a Beckman 6000 liquid scintillation counter (Beckman Coulter Instruments, Fullerton, CA). Data were analyzed by using GraphPad Prism software (San Diego, CA).

Acknowledgment. B.T.H. was supported by a National Institutes of Health Intramural Research Training Award fellowship. We also acknowledge the data provided by the Developmental Therapeutics Program (DTP), NCI, and the drug samples provided by the Drug Synthesis and Chemistry branch of the DTP. This study utilized the high-performance computational capabilities of the SGI Origin 2000 system at the Center for Information Technology, National Institutes of Health, Bethesda, MD.

References

(1) (a) Cho, S. J.; Cummins D.; Bentley J.; Andrews C. W.; Tropsha, A. An Alternative to 3D QSAR: Application of Genetic Algorithms and Partial Least Squares to Variable Selection of Topological Indices. *J. Comput.-Aided Mol. Des.*, submitted. (b) The GA-PLS method is described in detail and can be executed on the QSAR Web server available at http://mmlin1.pha.unc.edu/~jin/QSAR.

(2) Kuhar, M. J.; Ritz, M. C.; Boja, J. W. The Dopamine Hypothesis Of The Reinforcing Properties Of Cocaine. *Trends Neurosci.* **1991**, *14*, 299−302.

(3) Self, D. W.; Nestler, E. J. Molecular Mechanisms Of Drug Reinforcement And Addiction. *Annu. Rev. Neurosci.* **1995**, *18*, 463−495.

(4) Cline, E. J.; Terry, P.; Carroll, F. I.; Kuhar, M. J.; Katz, J. L. Stimulus Generalization from Cocaine to Analogues with High In Vitro Affinity for Dopamine Uptake Sites. *Behav. Pharmacol.* **1992**, *3*, 113−116.

(5) Newman, A. H.; Allen, A. C.; Izenwasser, S.; Katz, J. L. Novel 3-Alpha-(diphenylmethoxy)tropane Analogues: Potent Dopamine Uptake Inhibitors without Cocaine-like Behavioral Profiles. *J. Med. Chem.* **1994**, *37*, 2258−2261.

(6) Newman, A. H.; Kline, R. H.; Allen, A. C.; Izenwasser, S.; George, C.; Katz, J. L. Novel 4′-Substituted and 4′,4″-Disubstituted 3Alpha-(diphenylmethoxy)tropane Analogs as Potent and Selective Dopamine Uptake Inhibitors. *J. Med. Chem.* **1995**, *38*, 3933−3940.

(7) Newman, A. H. Novel Dopamine Transporter Ligands: The State of the Art. *Med. Chem. Res.* **1998**, *8*, 1−11.

(8) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday: A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge Crystal Data Centre: Computer-based Search, Retrieval, Analysis, and Display of Information. *Acta-Crystallogr. Sect B: Struct., Crystallogr. Cryst. Chem.* **1979**, *B35*, 2331−2339.

(9) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The Development of Versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187−204.

(10) Simon, Z.; Badileuscu, I.; Racovitan, T. Mapping of Dihydrofolate-reductase Receptor Site by Correlation with Minimal Topological (Steric) Differences. *J. Theoret. Biol.* **1977**, *66*, 485−495.

(11) Rhyu K. B.; Patel H. C.; Hopfinger A. J. A 3D-QSAR Study of Anticoccidial Triazines Using Molecular Shape Analysis. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 771−778.

(12) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.* **1989**, *291*, 161−165.

(13) Kubinyi, H.; Folkers, G.; Martyn, Y. *3D QSAR in Drug Design*; Kluwer/ESCOM: Leiden, 1998; Vol. 2−3.

(14) Camilleri, P.; Livingstone, D. J.; Murphy, J. A.; Manallack, D. T. Chiral Chromatography and Multivariate Quantitative Structure−Property Relationships of Benzimidazole Sulphoxides. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 61−69.

(15) Geladi, P.; Kowalski, B. R. Partial Least Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.

(16) Jansson, P. A. Neural Networks: An Overview. *Anal. Chem.* **1991**, *63*, 357A-362A.

(17) So, S. S.; Richards, W. G. Application of Neural Networks: Quantitative Structure−Activity Relationships of the Derivatives of 2,4-Diamino-5-(substituted-benzyl)pyrimidines as DHFR Inhibitors. *J. Med. Chem.* **1992**, *35*, 3201−3207.

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

(19) Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure−Activity Relationship of Angiotesin-Converting Enzyme and Thermolysin Inhibitors. II. A Comparison of CoMFA Models Incorporating Molecular Orbital Fields and Desolvation Free Energies Based on Active-Analogue and Complementary-Receptor-Field Alignment Rules. *J. Med. Chem.* **1993**, *36*, 2390−2403.

(20) Oprea, T. I.; Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure-Activity Relationship of Human Immunodeficiency Virus (I) Protease Inhibitors. 2. Predictive Power Using Limited Exploration of Alternate Binding Modes. *J. Med. Chem.* **1994**, *37*, 2206−2215.
(21) Waller, C. L.; Minor, D. L.; McKinney, J. D. Using Three-Dimensional Quantitative Structure-Activity Relationships to Examine Estrogen Receptor Binding Affinities of Polychlorinated Hydroxybiphenyls. *Environ. Health Perspect.* **1995**, *103*, 702−707.
(22) Oprea, T. I.; Garcia, A. E. Three-Dimensional Quantitative Structure-Activity Relationships of Steroid Aromatase Inhibitors. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 186−200.
(23) Cho, S. J.; Tropsha, A.; Suffness, M.; Cheng, Y. C.; Lee, K. H. Antitumor Agents. 163. Three-Dimensional Quantitative Structure−Activity Relationship Study of 4′−O-Demethylepipodophyllotoxin Analogues Using the Modified CoMFA/q2-GRS Approach. *J. Med. Chem.* **1996**, *39*, 1383−1395.
(24) Hoffman, B.; Cho, S. J.; Zheng, W.; Wyrick, S.; Nichols, D. E.; Mailman, R. B.; Tropsha, A. Quantitative Structure−Activity Relationship Modeling of Dopamine D1 Antagonists Using Comparative Molecular Field Analysis, Genetic Algorithms-Partial Least-Squares, and K Nearest Neighbor Methods. *J. Med. Chem.* **1999**, *42*, 3217−3226.
(25) (a) MOLCONN-Z; Hall Associates Consulting, Quincy, MA (information on Molconn Z is available at http://www.eslc.vabiotech.com/molconn/). (b) Hall, L. H.; Kier, M. B. *Molecular Structure Description: The Electrotopological State*; Academic Press: New York, 1999.
(26) Goldberg, D. E. *Genetic Algorithm in Search, Optimization, and Machine Learning;* Addison-Wesley: Reading, MA, 1989.
(27) Holland, J. H. Genetic Algorithms. *Sci. Am.* **1992**, *267*, 66−72.
(28) Mitchell, M. *An Introduction to Genetic Algorithms;* MIT Press: Cambridge, MA, 1996.
(29) Leardi, R. Application of a Genetic Algorithm to Feature Selection Under Full Validation Conditions and to Outlier Detection. *J. Chemom.* **1994**, *8*, 65−79.
(30) Kubinyi, H. Variable Selection in QSAR Studies. 1. An Evolutionary Algorithm. *Quant. Struct.-Act. Relat.* **1994**, *13*, 285−294.
(31) Kubinyi, H. Variable Selection in QSAR Studies. 2. A Highly Efficient Combination of Systematic Search and Evolution. *Quant. Struct.-Act. Relat.* **1994**, *13*, 393−401.
(32) Hasegawa, K.; Kimura, T.; Funatsu, K. GA Strategy for Variable Selection in QSAR Studies: Application of GA-Based Region Selection to a 3D-QSAR Study of Acetylcholinesterase Inhibitors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 112−120.
(33) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological and Valence State Information *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039−1045.
(34) Izenwasser, S.; Newman, A. H.; Katz, J. L. Cocaine and several sigma receptor ligands inhibit dopamine uptake in rat caudate-putamen. *Eur. J. Pharmacol.* **1993**, *243*, 201−205.
(35) Izenwasser, S.; Terry, P.; Heller, B.; Witkin, J. M; Katz, J. L. Differential Relationships Among Dopamine Transporter Affinities and Stimulant Potencies of Various Uptake Inhibitors. *Eur. J. Pharmacol.* **1994**, *263*, 277−283.
(36) Chen, Z.; Izenwasser, S.; Katz, J. L.; Zhu, N.; Klein, C. L.; Trudell, M. L. Synthesis and Dopamine Transporter Affinity of 2-(Methoxycarbonyl)-9-methyl-3-phenyl-9-azabicyclo[3.3.1]-nonane Derivatives. *J. Med. Chem.* **1996**, *39*, 4744−4749.
(37) Xu, L.; Kelkar, S. V.; Lomenzo, S. A.; Izenwasser, S.; Katz, J. L.; Kline, R. H.; Trudell, M. L. Synthesis, Dopamine Transporter Affinity, Dopamine Uptake Inhibition, and Locomotor Stimulant Activity of 2-Substituted 3-Beta-Phenyltropane Derivatives. *J. Med. Chem.* **1997**, *40*, 858−863.
(38) Lomenzo, S. A.; Izenwasser, S.; Katz, J. L.; Terry, P. D.; Zhu, N.; Klein, C. L.; Trudell, M. L. Synthesis, Structure, Dopamine Transporter Affinity, and dopamine Uptake Inhibition of 6-Alkyl-3-benzyl-2-[(methoxycarbonyl)methyl]tropane Derivatives. *J. Med. Chem.* **1997**, *40*, 4406−4414.
(39) Kline, R. H.; Izenwasser, S.; Katz, J. L.; Joseph, D. B.; Bowen, W. D.; Newman, A. H. 3′-Chloro-3-alpha-(diphenylmethoxy)-tropane But Not 4′-Chloro-3-alpha-(diphenylmethoxy)tropane Produces a Cocaine-like Behavioral Profile. *J. Med. Chem.* **1997**, *40*, 851−857.
(40) Katz, J. L.; Izenwasser, S.; Kline, R. H.; Allen, A. C.; Newman, A. H. Novel 3-Alpha-diphenylmethoxytropane Analogues: Selective Dopamine Uptake Inhibitors with Behavioral Effects Distinct from Those of Cocaine. *J. Pharmacol. Exp. Ther.* **1999**, *288*, 302−315.
(41) Kelkar, S. V.; Izenwasser, S.; Katz J. L.; Klein, C. L.; Zhu, N.; Trudell, M. L. Synthesis, Cocaine Receptor Affinity, and Dopamine Uptake Inhibition of Several New 2-Beta-substituted 3-beta-phenyltropanes. *J. Med. Chem.* **1994**, *37*, 3875−3877.
(42) GNU software is available through the Free Software Foundation at http://www.gnu.ai.mit.edu/.
(43) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(44) Agarwal, A.; Taylor, E. W. 3-D QSAR for Intrinsic Activity of 5-HT_{1A} Receptor Ligands by the Method of Comparative Molecular Field Analysis. *J. Comput. Chem.* **1993,** *14*, 237−245.
(45) The NCI database is available at http://dtp.nci.nih.gov/.
(46) Structures in the NCI database may be browsed via the following link: http://cactus.cit.nih.gov/ncidb/frame.html.

JM990472S