ABSTRACT

MODELING OF FLEXIBLE DRUG-LIKE MOLECULES: QSAR OF GBR 12909 ANALOG DAT/SERT SELECTIVITY

by
Kathleen Mary Gilbert

The dopamine reuptake inhibitor GBR 12909 and related dialkyl piperazine and piperidine analogs have been studied as agonist substitution therapies acting on the dopamine transporter (DAT) to treat cocaine addiction. Undesirable binding to the serotonin transporter (SERT) can vary greatly depending on the specific substituents on the molecule. This study uses Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices (CoMSIA) techniques to determine a stable and predictive model for DAT/SERT selectivity for a set of flexible GBR 12909 analogs.

Families of analogs were constructed from six pairs of naphthyl-substituted piperazine and piperidine templates identified by hierarchical clustering as representative conformers. Three-dimensional quantitative structure-activity relationship (3D-QSAR) studies led to focused models that were stable to y-value scrambling. Test set correlation validation led to one acceptable model ($q^2 = 0.508$, two components, $r^2 = 0.685$, average residual = 0.00 for the training set, 0.22 for the extended test set). DAT/SERT selectivities higher than that of the most active compound in the QSAR series were predicted for nine novel compounds.

This is the first CoMFA/CoMSIA study of the highly flexible GBR 12909 class of dopamine reuptake inhibitors. Previously, molecular modeling was based on more rigid dopamine reuptake inhibitors, and often only on global energy minimum (GEM) structures. Flexible molecules like GBR 12909 have multiple possible binding

conformations, distributed across the potential energy surface in key torsional angle space, which can vary from the GEM by as much as 20 kcal/mol or more. The significance of this study lies in the combining of a clustering technique for identifying representative conformers from a set of low-energy (less than 20 kcal/mol from the GEM) conformers with an extensive 3D-QSAR analysis based on each representative conformer and analogs in a similar potential bioactive conformation.

MODELING OF FLEXIBLE DRUG-LIKE MOLECULES:
QSAR OF GBR 12909 ANALOG DAT/SERT SELECTIVITY

by
Kathleen Mary Gilbert

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Chemistry

Department of Chemistry and Environmental Science

May 2005

APPROVAL PAGE

MODELING OF FLEXIBLE DRUG-LIKE MOLECULES:
QSAR OF GBR 12909 ANALOG DAT/SERT SELECTIVITY

Kathleen Mary Gilbert

Dr. Carol A. Venanzi, Dissertation Advisor   Date
Distinguished Professor of Chemistry, NJIT

Dr. Joseph P. Bozzelli, Committee Member   Date
Chair, Department of Chemistry and Environmental Science
Distinguished Professor of Chemistry, NJIT
Ada C. Fritz Professor of Environmental Engineering and Science

Dr. Tamara M. Gund, Committee Member   Date
Professor of Chemistry, NJIT

Dr. Sanjay V. Malhotra, Committee Member   Date
Assistant Professor of Chemistry, NJIT

Dr. Christopher C. Van Dyke, Committee Member   Date
Tripos, Inc., St. Louis, MO

BIOGRAPHICAL SKETCH

Author: Kathleen Mary Gilbert

Degree: Doctor of Philosophy

Date: May 2005

Undergraduate and Graduate Education:

- Doctor of Philosophy in Chemistry
  New Jersey Institute of Technology, Newark, NJ, 2005

- Master of Science in Computational Biology
  New Jersey Institute of Technology, Newark, NJ, 2002

- Master of Science in Environmental Engineering
  Syracuse University, Syracuse, NY, 1992

- Bachelor of Science in Engineering (Bioengineering)
  University of Pennsylvania, Philadelphia, PA, 1990

Major: Chemistry

Presentations and Publications:

Gilbert, K.M., Skawinski, W.J., Misra, M., Paris, K.A., Naik, N., Deutsch, H.M., Buono, R.A., and Venanzi, C.A., "Conformational Analysis of Methylphenidate: Comparison of Molecular Orbital and Molecular Mechanics Methods", Journal of Computer-Aided Molecular Design, 18, 719-738, 2004.

Gilbert, K.M., Matecka, D., Prisinzano, T., Rice, K.C., and Venanzi, C.A., "Clustering Analysis of Flexible GBR 12909 Dialkyl Piperazine and Piperidine Analogs", National Meeting of the American Chemical Society, Philadelphia, PA, August 2004.

To Kelly, Anna, and the boys

ACKNOWLEDGEMENT

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

Table   **Page**

LIST OF FIGURES

Figure   **Page**

CHAPTER 1

INTRODUCTION

Cocaine addiction remains a significant problem around the world.[1] As with many addictive drugs, the exact mechanism of the addiction has not been identified, although focus has been on the interaction of cocaine with various neurotransmitter systems.[2] The "dopamine hypothesis"[3,4] implicates the dopamine transporter (DAT) in cocaine abuse and addiction. Structure-activity relationship (SAR) studies for several classes of dopamine (DA) reuptake inhibitors (shown in Figure 1.1) have been reviewed[2,5-8] 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**) is a promising drug candidate, having demonstrated a decrease in cocaine-maintained responding without affecting food-maintained responding in behavioral studies of rhesus monkeys[9] and

1
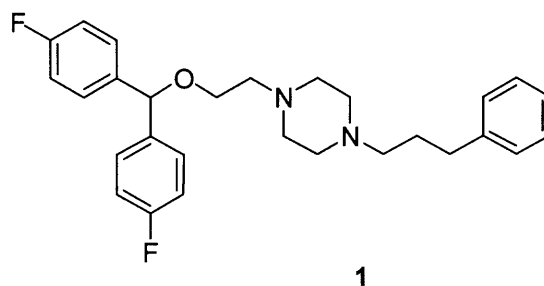
having passed Phase I clinical trials.[10] In addition, prior use of cocaine caused cross-sensitization for several DA reuptake inhibitors, but not for **1**.[11] Compound **1**, also known as vanoxerine (CAS No. 67469-78-7), and related dialkyl piperazines and piperidines have also been shown to bind selectively to the DAT cocaine binding site and cause minimal DA reuptake inhibition, without causing other of cocaine's deleterious side effects. SAR studies of hundreds of analogs of **1**, based on the scaffold shown in Figure 1.2, have been summarized in a recent review.[12] Most DA reuptake inhibitors
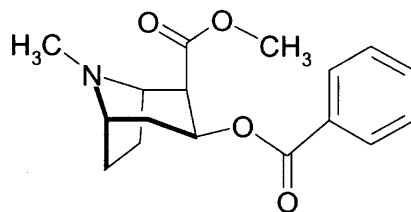
cocaine
(CAS No. 50-36-2)
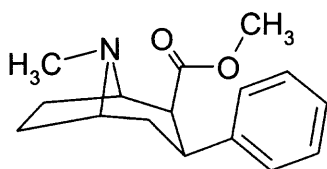
tropane (WIN 35065-2 or 2-CPT)
(CAS No. 50372-80-0)
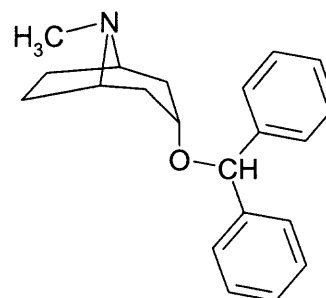
benztropine
(CAS No. 132-17-2)
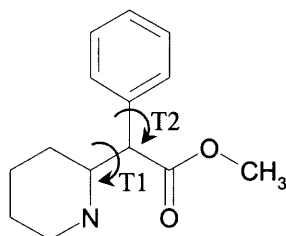
methylphenidate
(CAS No. 113-45-1)
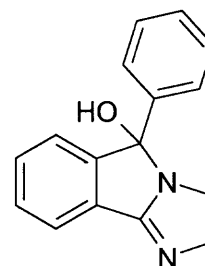
mazindol
(CAS No. 22232-71-9)

Figure 1.1 Cocaine and representatives of four classes of dopamine reuptake inhibitors.
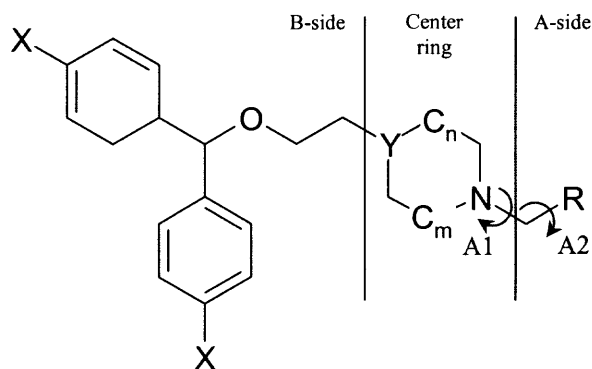
Figure 1.2 Template for GBR 12909 analogs. X can be either H or F, Y is a nitrogen for piperazine analogs, and a carbon for piperidine analogs. The R group is a naphthalene for analog **2** and analog **3**, and can be a variety of aromatic or non-aromatic substituents. n and m are usually equal to 1; however, n=1 and m=0 for five-membered rings and n=1 and m=2 for seven-membered rings.

have common structural features, such as an aromatic ring located near a quaternary nitrogen (as seen for example in the relationship between the center ring nitrogen and A-side in Figure 1.2) that appear to be necessary for DAT binding. However, photoaffinity labeling studies indicate that 1-like and tropane-like compounds (such as cocaine) may bind to different regions in the DAT[13] or bind in a somewhat different manner.[14,15] Compared to a 2-benzhydryloxyethyl substituent, presence of a 2-[bis-(4-fluorophenyl)methoxy]ethyl- substituent (identified as the B-side in Figure 1.2) usually results in better DAT binding affinity, but in some cases decreased DAT/serotonin transporter (SERT) selectivity, although neither effect was significant compared to modifications of the A-side.[16-20] Most SAR studies have focused on either the piperazine or piperidine analogs of **1**, although the trend has been towards piperidines because **1** analog piperazines could bind to the "piperazine acceptor site" found in the brain and liver, identified as a cytochrome P450.[21] In order to focus on the activity of piperazine and piperidines, the present three-dimensional quantitative structure-activity relationship

(3D-QSAR) studies are carried out on analogs of the piperazine **2**, and the piperidine **3**, all of which have the same 2-[bis-(4-fluorophenyl)methoxy]ethyl- B-side but differ in their A-side substituents. Activities that were tested for these compounds include DA reuptake inhibition, DAT binding affinity, and SERT binding affinity.

2

3
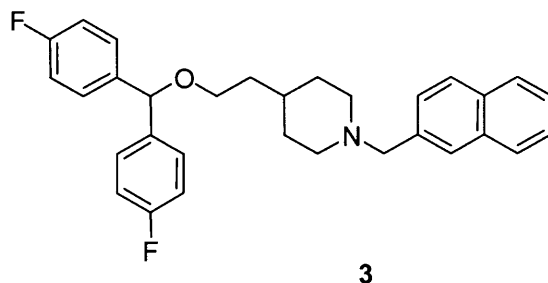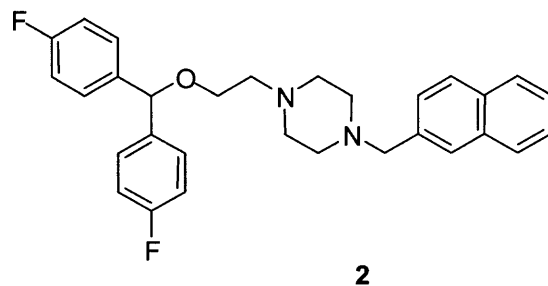
In the absence of knowledge of the three-dimensional structure of the DAT, ligand-based 3D-QSAR techniques such as Comparative Molecular Field Analysis (CoMFA)[22] and Comparative Molecular Similarity Indices (CoMSIA)[23] may be useful in identifying molecular features that improve activity. However, these techniques require the use of a template conformer, i.e., a putative bioactive 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. To date, molecular modeling studies have focused on fairly rigid classes of DA reuptake inhibitors, such as tropanes,[24-36] piperidine-based cocaine analogs,[37] benztropine, [36,38,39] BTCP,[40] bupropion,[41] mazindol,[42] methylphenidate,[43,44] novel piperadinols,[45] and more rigid bicyclic **1** analogs,[46] some examples of which are shown in Figure 1.1.

Although there is considerable evidence that many ligands do not bind to proteins in their vacuum phase global energy minimum (GEM) conformation,[47-52] 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.[28,42,53] Other work has shown the necessity of considering conformations other than the GEM in pharmacophore modeling, especially when modeling very flexible molecules.[54-59] For example, the Venanzi group's conformational analysis of the DA reuptake inhibitor methylphenidate[60] identified several conformations within a few kcal/mol of the GEM structure that are potential bioactive conformers. Subsequent CoMFA analysis and comparison to a rigid methylphenidate analog with the same DAT binding affinity as methylphenidate supported the idea that the bioactive conformation need not be the GEM conformer.[44]

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 significance of this study lies in the combining of a clustering technique for identifying representative conformers from a set of low-energy (less than 20 kcal/mol from the GEM) conformers with an extensive 3D-QSAR analysis based on each representative conformer and analogs in a similar potential bioactive conformation. Conformational analysis of **2** and **3** was carried out,[61] with hierarchical clustering used to select representative conformers to use as templates in 3D-QSAR studies.[62] These templates were chosen to be representative of the regions of three-dimensional space occupied by the analogs. In this research, CoMFA and CoMSIA techniques were used to analyze conformational families of forty-five **1** analogs (23 analogs of **2** and 22 analogs of **3**) which differ only in their A-side substituents. Six conformational families were constructed from six pairs of template conformers identified by hierarchical clustering as representative molecular conformations of **2** and **3**.

CoMFA and CoMSIA studies were performed on the selected molecules. CoMFA studies encompassed a range of electrostatic and steric cutoffs for each conformational family. 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 as compared to CoMFA.[23,63] Six sets of CoMFA/CoMSIA studies were carried out, one for each conformational family. Partial least squares (PLS) analysis was performed on the CoMFA and CoMSIA results, with the q^2 (predictivity) and r^2 (goodness-of-fit) values and associated statistics being calculated. PLS Region Focusing was applied to the best CoMFA or CoMSIA results for each preliminary study. Internal validation (y-value scrambling) and external validation (test set correlation validation) methods were used to determine the most predictive and stable model. The present work is the first CoMFA and CoMSIA study of multiple potential bioactive conformations of highly flexible analogs of **1**.

CHAPTER 2

BACKGROUND

2.1 Overview and Objectives

Currently, there are no medications that can be prescribed in the U.S. to treat cocaine addiction, although cocaine addiction remains a significant problem. As for many addictive drugs, the exact mechanism of cocaine addiction has not been identified. This work is based on one significant hypothesis of addiction, the "dopamine hypothesis",[3,4] which applies to cocaine along with other addictive drugs and certain addictive behaviors. According to this hypothesis, cocaine's effect on the DAT is implicated in the addictive effects of cocaine. The DAT releases DA into the synaptic junction of certain nerve cells and reuptakes DA from the synaptic junction. Among its modes of action, cocaine binds to the DAT, inhibiting the reuptake of DA (Figure 2.1). A compound that binds to the cocaine binding site on the DAT, causing some DA reuptake inhibition, but having a minimal effect other neurotransmitter systems, could block the negative actions of cocaine on the body. These selective DA reuptake inhibitors are being explored as a potential treatment for cocaine addiction.

Mutagenesis studies have been performed on the DAT to elucidate the effects of cocaine at the DAT,[64,65] but the three-dimensional structure of the DAT is not known.[5] Therefore, ligand-based 3D-QSAR techniques such as CoMFA[22] and CoMSIA[23] may be useful to help identify molecular features that improve DAT binding or DAT/SERT selectivity. The DAT is a Na^+Cl^- ion-dependent transporter, and the effect of Na^+ concentration on DAT ligand binding has been studied.[66] Key residues involved in
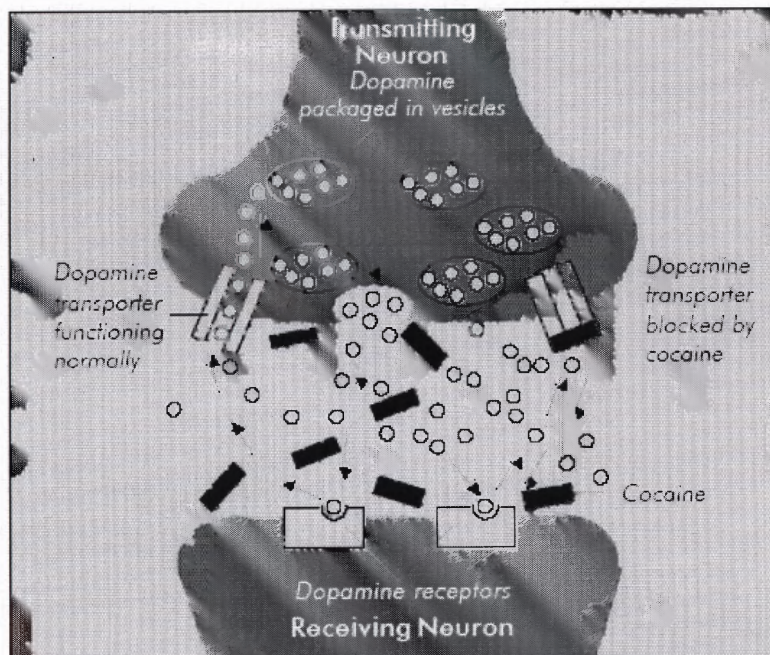
Figure 2.1 DAT at the synaptic junction. The influence of cocaine binding is shown. From National Institutes of Health (NIH) Publication Number 99-4342 Cocaine Abuse and Addiction, May 1999.

cocaine analog binding have been identified, including tryptophans and aspartates, although it appears that different residues affect cocaine-like and dissimilar DA reuptake inhibitors.[65] Selectivity of DA reuptake inhibitors is also an issue, especially relative to other Na^+Cl^- ion-dependent transporters like the SERT; binding to the SERT and other transporters may cause undesirable side effects and diminish the relative effect of a dose of a DA reuptake inhibitor, part of the dose binding other transporters instead of only the DAT. The majority of the compounds have an acceptable low nanomolar DAT binding K_i, with a wider range of SERT binding K_i's; therefore, the DAT/SERT selectivity of the compounds was modeled.

Although CoMFA studies have been carried out on other classes of DA reuptake inhibitors, such as tropanes,[24,26] mazindols,[42] and benztropines,[36,38] no CoMFA or CoMSIA studies have been done on **1** analogs. A chemometric QSAR study was done on **1** analogs, but used only a single minimized structure for each analog.[53] The study used GRID[67] independent descriptors (GRIND),[68] a method that takes the molecular interaction fields (for example, hydrophobic or hydrogen-bond donor) calculated by GRID to create several variables that represent regions where energetically favorable interactions can occur, to model DAT/SERT selectivity and found predictive models. In contrast, the present study looks at multiple low energy conformations, and no assumption is made that the GEM is the conformation a molecule must achieve (the bioactive conformation) in order to bind to a target. This project is the first detailed CoMFA and CoMSIA study on multiple potential bioactive conformers of **1** analogs.

The objectives of this study are:

- To use conformational analysis and clustering methods to identify low-energy conformations that represent the wide range of conformations attainable by the piperazine **2** and the piperidine **3**

- To use CoMFA and CoMSIA 3D-QSAR techniques to identify a model for DAT/SERT selectivity

- To provide the modeling results as a contribution to the iterative cycle of synthesis, pharmacological testing, and modeling with a goal of development of a treatment for cocaine abuse

2.2   Cocaine Addiction and the Dopamine Hypothesis

DA reuptake inhibitors target the "cocaine binding site" on the DAT. Cocaine's addictive properties are thought to be at least in part due to an increase in synaptic DA achieved by reuptake inhibition of DA at the DAT. The "dopamine hypothesis"[3,4]

implicates cocaine's inhibition of DA reuptake as the reason for cocaine addiction. However, it is noted that cocaine and methylphenidate (the active ingredient in the common attention-deficit/hyperactivity disorder medication Ritalin®) both bind to the DAT and cause DA reuptake inhibition, yet methylphenidate is not addictive like cocaine and does not cause the same level of negative side effects.[69] Using methylphenidate or **1** analogs to treat cocaine abuse is dependent upon the complex nature of cocaine's effect on the body; the methylphenidate and **1** analogs are dissimilar enough from cocaine to not have the multi-target binding profile of cocaine. The "reward response" to cocaine use mediated by the DAT is mimicked by methylphenidate analogs and **1** analogs, without some of the negative side effects such as addiction.[70] Use of **1** for agonist substitution therapy has been proposed, because **1** binds to the DAT, but does not substantially increase extracellular DA.[71]

Chief attributes of the best candidates for treating cocaine addiction via agonist substitution are expected to be:

- high binding affinity at the "cocaine binding site" on the DAT,

- partial (i.e., minimal) DA reuptake inhibition,

- high selectivity for the DAT versus the SERT, and

- desirable kinetics of slow on-rate and slow off-rate when binding to the DAT.

The first three attributes can be predicted from models using the following activities:

- *DAT binding affinity,* measured by the displacement of radiolabeled cocaine analog [^{125}I]RTI-55 from the DAT,

- *DA reuptake inhibition,* measured by the decrease of synaptic DA in response to binding, and

- *SERT binding affinity*, measured by the displacement of radiolabeled cocaine analog [^{125}I]RTI-55 from the SERT.

Binding affinities are equal to K_i, calculated from:[72]

$$K_i = IC_{50} / (1 + S/K_d) \qquad (2.1)$$

where the K_i is the inhibition constant (the dissociation constant of the test compound), IC_{50} is the concentration of the radiolabeled cocaine analog [^{125}I]RTI-55 that is displaced from the DAT or SERT, S is the concentration of free [^{125}I]RTI-55, and K_d is the binding affinity constant of [^{125}I]RTI-55, calculated from saturation binding experiments. It can be seen that if the amount of free radioligand is small compared to the binding affinity of the radioligand, the K_i is nearly identical to the IC_{50} value. This is the case for all of the experimental data provided for this study by Richard Rothman of the National Institutes of Health (NIH).

DAT/SERT selectivity is calculated from the following equation:

DAT/SERT selectivity = log[(DAT binding activity) / (SERT binding activity)] (2.2)

where:

DAT binding activity = 1 / (DAT binding affinity K_i) (2.3)

SERT binding activity = 1 / (SERT binding affinity K_i) (2.4)

therefore:

DAT/SERT selectivity = log[(SERT binding affinity K_i) / (DAT binding affinity K_i)]

(2.5)

DAT/SERT selectivity is modeled in this study because selectivity for the DAT over the SERT increases specificity for the DAT and decreases potential side effects from SERT binding. Most of the DAT binding affinities of the analogs studied are in an acceptable

low nanomolar range of DAT binding affinity. The direct DAT/SERT selectivity was used instead of scaling against cocaine's DAT/SERT selectivity because experimental values for cocaine's DAT/SERT selectivity are near 1.[7] DAT/SERT selectivity has been a target for the NIH and other researchers as well.[14,20,46,53,73,74]

2.3 Classes of Dopamine Reuptake Inhibitors

There are five major classes of DA reuptake inhibitors being studied for their cocaine-blocking properties: tropanes, benztropines, mazindols, methylphenidates, and 1 analogs. The structure of cocaine and typical structures of the first four classes are shown in Figure 1.1. A recent review of DAT binding compounds has been published,[8] and a slightly older, but extremely comprehensive review of DAT binding compounds oriented towards treatment of cocaine abuse is available.[7]

2.3.1 Tropanes and Other Cocaine-Like Dopamine Reuptake Inhibitors

Tropanes were developed based on the bicyclic cocaine backbone. Like cocaine, they have the bicyclic backbone with a quaternary nitrogen and acetyl side chain, but in position 3, various substituents replace the benzoyl moiety.[75] Three tropanes are commonly used for binding studies; each has a phenyl substituent instead of the benzoyl substituent. WIN35065-2 (CPT)[76] has a phenyl substituent, WIN35428 (CFT)[76,77] has a 4-fluorophenyl substituent, and RTI-55 (CIT)[78] has a 4-iodophenyl substituent.[79] Cocaine and related tropanes bind to other proteins, especially the SERT and the norepinephrine transporters, and thus may cause side effects or dilution of activity. The tropanes CPT, CFT, and CIT have been used extensively for radiolabeling DAT and SERT; this QSAR study is based on [^{125}I]RTI-55-labeled DAT and SERT, using

techniques to isolate one transporter or the other.[80] Many other tropane compounds that act on the DAT have been synthesized and analyzed.[24-27,29,30,77,79,81-123]

Piperidine-based cocaine analogs have also been developed, with a piperidine ring in place of the tropane ring. These compounds have shown varying levels of activity.[79,124] A recent QSAR study of piperidine-based cocaine analogs identified two potential binding modes with different 3α- substituent binding characteristics.[37]

2.3.2 Benztropine Dopamine Reuptake Inhibitors

Benztropines are similar to cocaine and other tropanes, but lack the acetyl side chain and have a diphenyl methoxy or similar substituent in position 3, which makes them more similar to **1** analogs than any other DAT binding class.[36,38,39,125-133] Benztropine is currently used to treat Parkinson's disease, as well as drug-induced Parkinsonian symptoms.

2.3.3 Mazindol Dopamine Reuptake Inhibitors

Mazindol has been used as an appetite-suppressant with actions similar to amphetamine. Although mazindol analogs contain a quaternary nitrogen and aromatic moieties, the resemblance to cocaine, tropanes, and benztropines ends there. Mazindol itself consists of three fused rings, with a phenyl ring attached to the center ring.   There has been considerable SAR work on the series, including accompanying 3D-QSAR studies in some cases.[42,134-142]

2.3.4 Methylphenidate Dopamine Reuptake Inhibitors

Methylphenidate also has a different structure than cocaine, but includes the quaternary nitrogen and aromatic substituent present in the other DAT binding families.[43] Methylphenidate is familiar as the active ingredient in Ritalin®, although that prescription drug is a mixture of the active 2R, 2'R (D-threo) and the much less active 2S, 2'S (L-threo) enantiomers.[143,144] SAR studies and analyses have been completed on methylphenidate analogs,[145-153] including pyran analogs that are among the first DAT binders developed to have high affinity for the DAT without containing a nitrogen atom.[154] The Venanzi group has completed detailed conformational analysis of methylphenidate[60] and CoMFA studies on phenyl-substituted methylphenidates.[155,156] These CoMFA studies will be published along with comparisons to novel rigid methylphenidates.[157] The conformational analysis and CoMFA work on methylphenidate is summarized below.

2.3.4.1 Conformational Analysis of Methylphenidate. To identify possible bioactive conformers, the Venanzi group performed detailed conformational analysis of methylphenidate and related rigid methylphenidates using multiple molecular modeling methods.[60] Hartree-Fock 6-31G*[158] basis set calculations as well as density functional theory (B3LYP/6-31G*),[159] AM1/SM5.4,[160-162] and Tripos force field[163] calculations were performed using the Gaussian 98[164] (available from Gaussian, Inc., Wallingford, CT), Spartan (available from Wavefunction, Inc., Irvine, CA), and SYBYL® (available from Tripos, Inc., St. Louis, MO) software. These techniques were used to explore the potential energy surface (PES) of neutral and protonated methylphenidate, as well as that of several rigid methylphenidate invertamers. In addition, methods were compared to

determine whether the less computationally-intensive force field molecular mechanics method gave similar results to molecular orbital methods. All methods gave local minima in approximately the same regions of torsional angle space. The Tripos force field molecular mechanics random search found at least one conformation in each local minimum area of the methylphenidate PES, thus validating the utility of this technique compared to more computationally-intensive ab initio methods.

2.3.4.2 Methylphenidate CoMFA Studies. CoMFA studies were performed on representative templates of neutral[155] and protonated[156] phenyl-substituted methylphenidate analogs The DAT binding affinities for methylphenidate and 29 analogs were modeled using CoMFA at a comprehensive set of steric cutoff/electrostatic cutoff/column filtering (σ) combinations. Template conformers of the conformational families with the best CoMFA results were compared to the conformations of rigid methylphenidate analogs that have the same binding affinity for the DAT, leading to three potential bioactive conformations whose results matched well for both neutral and protonated studies.

2.3.5 GBR 12909 Dopamine Reuptake Inhibitors

2.3.5.1 Discovery of GBR 12909 and Applications. Compound 1 is a 1,4-disubstituted piperazine that was first synthesized in the 1970's at Gist-Brocades N.V., a Netherlands-based food product and biotechnology company. It was initially investigated in Europe as an antidepressant, but was targeted as a potential treatment for cocaine abuse in the late 1980's.[165] Researchers at the NIH have studied **1** and related compounds for their potential use to modulate cocaine addiction.

Towards this end, over 300 related compounds were synthesized and their DAT and SERT binding affinities tested by the Rice and Rothman groups at NIH,[12,14,16-18,20,166,167] exploring many aspects of the complex structure of **1**. The data has been provided to the Venanzi group for QSAR and pharmacophore modeling. The present study is the first step in the computational analysis of these compounds. Forty-five analogs of **1** that are based on the structures of **2** and **3** were selected for this study because they have fewer rotatable bonds than **1** and thus are easier to model. Three additional compounds, including **1** and two similarly flexible **1** analogs, were added as additional test set compounds. The range of DAT/SERT selectivities (equal to the logarithm of the ratio of DAT binding affinity to SERT binding affinity) covers more than three log units, a recommended range for QSAR studies.[168]

2.3.5.2 Structure-Activity Relationships of GBR 12909 Analogs. The three major areas of **1**, noted above in Figure 1.2, have been subjected to many changes.[12] Elimination of the B-side resulted in extremely poor DAT binding ($K_i = 30,000$ nM for both DA uptake and DAT binding). Addition of single atom *para* substituents to both B-side phenyl rings has been well tolerated. Adding substituents to only one phenyl ring has been studied, with position 3' substitutions being better tolerated than 4'-substitutions.[169] Selectivity studies have generally not been performed for B-side structures other than those with *para* hydrogen and *para* fluorine phenyl ring substituents.

The center ring has been modified in various ways, including conversion to a tropane-like bicyclic ring,[170] and addition or removal of a carbon atom. Changing the piperazine to a seven-membered diazepane ring resulted in no detectable DAT binding by

measuring [^{125}I]RTI-55 displacement. Other variations on the center ring include a five-membered 1,2-substituted azolidine ring with a nitrogen in the second position from the center ring. This compound showed approximately one-fifth the DAT binding affinity of **1**. DAT binding affinity was also poor when the center ring was changed to an acyclic *N,N*-dimethyl-1,2-diamine ethane; no displacement of [^{125}I]RTI-55 from the DAT was detected at all. Removal of the A-side resulted in no detectable DAT binding for [^{125}I]RTI-55 studies, and poor DA uptake (K_i = 71 nM). Methyl substituents at the 2- and 5- positions have been added, improving DAT activity but these compounds performed poorly in studies on laboratory animals. A similar trend was seen when adding a two-carbon bridge to compound **1** across the A-side nitrogen. Other researchers have explored center rings with oxygen instead of nitrogen[171] and constrained bicyclics as well.[46] The modification from a piperazine ring to a piperidine ring has resulted in better DAT binding affinity and DAT/SERT selectivity;[74,172] for this reason, both piperazine and piperidine **1** analogs are included in this study.

A-side substitutions have been numerous. The original 3-phenylpropyl substituent has been modified to include a double bond (slightly improved DAT binding) or a bisphenyl substituent (better DAT binding in bench experiments but performs poorly in animal models). Removal of flexible bonds from the A-side has been a focus of studies; substituents include methylnaphthyl substituents and other double ring systems (for example, thiophene, benzofuran, and quinoline). Different attachments of the double ring systems have been explored. Adamantyl and ethynylbenzyl substituents have also been used as A-side substituents. Work on piperidine analogs has resulted in many *para*-substituted phenyl rings with high DAT binding affinities, with the cyano and

trifluoromethyl substituents showing especially good DAT binding. These multiple studies have shown that the SAR characteristics of **1** are complex indeed, with hundreds of data points available on as many permutations of the structure. This study focuses on a sample (forty-five **1** analogs) of the hundreds of compounds studied, to highlight the effect of A-side substitution on the DAT/SERT selectivity of **1** analogs.

2.3.6 Shared Pharmacophore Elements of Dopamine Reuptake Inhibitors

As noted, all of the above classes of DA reuptake inhibitors share at least two pharmacophore elements: a quaternary nitrogen and an aromatic ring. Figure 1.1 shows that the torsional angles T1 and T2 in methylphenidate determine the relative orientation of these two pharmacophore elements. Figure 1.2 shows that A1 and A2 provide a similar relationship for **2** and **3**. The Venanzi group has performed conformational analysis of methylphenidate, **2**, and **3** in order to investigate whether these two classes of DA reuptake inhibitors might share a common pharmacophore. Figure 2.2 shows conformational energy minima for methylphenidate, plotted in (T1, T2) torsional angle space, and **2** and **3**, plotted in their counterpart (A1, A2) conformational space. Comparison of the local minima of methylphenidate to those of **2** and **3** indicates that these two very different classes of DA reuptake inhibitor appear to attain a similar orientation of their quaternary nitrogen and aromatic substituents. The similarity of the T1 and T2 torsional angles in methylphenidate to the A1 and A2 torsional angles in **2** and **3**, and the fact that these compounds all displace [^{125}I]RTI-55 bound to the DAT supports the hypothesis that the pharmacophores for the two classes of partial DA reuptake inhibitors are similar.
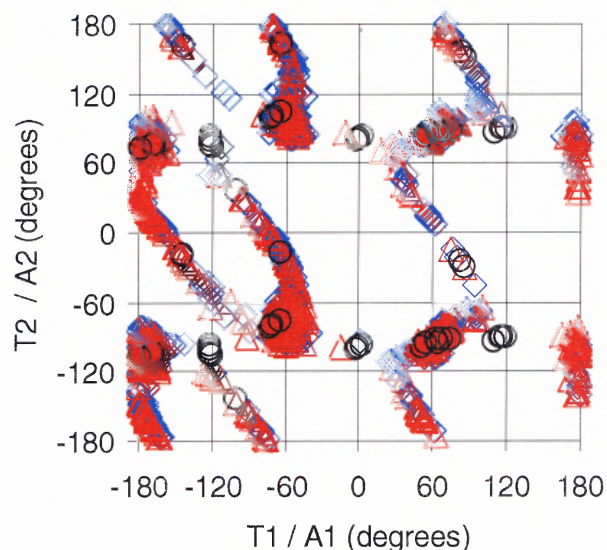
Figure 2.2 Local conformational minima for protonated methylphenidate and GBR 12909 analogs. T1 and T2 values for protonated methylphenidate (black circles), and A1 and A2 values for analog **2** (red triangles), and analog **3** (blue diamonds). Results for protonated methylphenidate are shown with both positive and negative T2 values because of the symmetry of the phenyl ring. Local minima are SYBYL random search conformations found within 20 kcal/mol of the GEM for each molecule.

2.4 Synthesis and Pharmacology

Synthetic work was carried out by Dr. Kenner C. Rice's group at the National Institute for Diabetes and Disorders of the Kidneys, part of NIH. Binding and uptake experiments were conducted by Dr. Richard B. Rothman's laboratory at the National Institute on Drug Abuse at NIH.[12,14,16-20,71,73,80,166,167,170,173-193] Four experimental results are of interest to this project: DAT binding, DA uptake, SERT binding, and serotonin (SER) uptake. DAT binding was measured by loading synaptosomes with radiolabeled [^{125}I]RTI-55 and measuring the amount of radiolabel remaining after incubation with a test compound. DA uptake inhibition was measured by determining the decrease in the amount of DA in the synaptosomes after incubation with a test compound. SERT binding and SER uptake

inhibition were measured in a similar way, using blocking techniques to isolate SERT binding from DAT binding.[80]

2.5 Molecular Modeling and Conformational Analysis

With the advent of the computer in the 1980's, computerized molecular modeling began to progress rapidly as well, being previously relegated to large institutions with extensive computing resources and costing inordinate amounts of computational time. Computerized molecular modeling is based on a set of approximations of the interactions of atoms and how they form molecules. Molecular mechanics uses empirical force field parameters to define the shape, potentials, and interactions of atoms in a given molecule. The molecular modeling package SYBYL uses the Tripos force field[163] to calculate the attributes of molecules, including atom locations and relationships. Classical mechanical forms, such as springs with differing elasticity constants, are used to represent different atoms. Force field parameters are derived by fitting equations to experimental results.

Molecular mechanics methods use a series of equations to derive the locations of atoms in a molecular structure. All molecular mechanics methods use a total energy equation calculated as a result of different intra- and inter-atomic forces in the molecule. The Tripos force field uses the following equation to calculate the total energy of a molecule:

$$E = \Sigma E_{str} + \Sigma E_{bend} + \Sigma E_{oop} + \Sigma E_{tors} + \Sigma E_{vdw} (+ \Sigma E_{ele}) \qquad (2.6)$$

where: E_{str} is the energy of a bond stretched or compressed from its natural bond length,

E_{bend} is the energy of bending bond angles from their natural values,

E_{oop} is the energy of bending planar atoms out of the plane,

E_{tors} is the energy due to twisting (torsion) about bonds, and

E_{vdw} is the energy due to van der Waals forces between non-bonded atoms.

E_{ele} is the energy due to electrostatic interactions, and is optional depending on the user's selection. Absolute values of the energies found when using empirical molecular mechanics calculations are meaningless. Relative energies for different conformations of the same molecule, and conformational energy differences between different molecules (that is, relative to their own GEM) do have meaning. These techniques lend themselves to mathematical comparison of molecules in order to predict relative activities of the compounds.

Conformational analysis of molecules with many rotatable bonds is computationally intensive. A full grid search, by systematically changing all eight key torsional angles and subsequent minimization of each resulting structure, would produce over 400 million conformations for 30° increments of the torsional angles. A wider grid spacing of 60° for each torsional angle would produce almost 1.7 million conformations. For this reason, the random search method[194] was selected as an appropriate way to sample the energy minima on the PES of these flexible molecules. SYBYL's implementation of random search uses random variation of selected torsional angles and minimization of the randomized structures with the Tripos force field. An interconformational root mean square deviation (RMSD) cutoff is used to remove duplicate conformers. The RMSD is calculated from the atom locations of the new conformer and each existing conformer; a new conformer that is within the RMSD cutoff

of any existing conformer is discarded. The exhaustiveness of a random search can be estimated by the following equation:[194]

$$\text{Probability of finding all conformers} = 1 - (0.5)^n \qquad (2.7)$$

where n = the minimum number of times any conformer is found. Finding almost all low-energy conformations is probable with small and less flexible molecules, within a reasonable search time. However, for molecules like **1**, **2**, and **3**, it is probable that not all energy minima will be found, or that conformations must be limited by a high RMSD between atoms in similar conformations (equivalent to a wider grid spacing for a grid search). Random search combined with clustering of the conformations can be a computationally efficient method to facilitate representation of all potential low-energy binding conformations, instead of using a high interconformational RMSD cutoff value which may arbitrarily remove important conformations based on search order.

2.6 Clustering

CoMFA results are notoriously dependent upon the conformer used as the template structure. Highly flexible molecules have a continuum of conformers covering the many changes in their torsional angles. In the present work, a series of CoMFA and CoMSIA calculations are carried out on different template structures, which are representative conformers identified using hierarchical clustering. Clustering is a method that takes many data points (or sets of data points) and finds groupings based on the similarities of one or more features of the data set. Items that share a cluster are assumed to be similar; items that are not in the same cluster are assumed to be dissimilar. Clustering can be

applied to many systems, but chemists have found it especially useful to reduce large sets of conformations to a few representative conformers.[195]

2.6.1 Types of Clustering

Clustering techniques can range from visual inspection to complex computational techniques. All clustering techniques involve a feature set selected either manually by the user, or automatically by the computer program. For molecules, a feature set might consist of atom locations, distances between atoms, angles, or torsional angles. Visual inspection can be performed on graphs created by selection of two features, one for each axis. Automatic and exhaustive searching of possible feature sets (realizing that a feature set can have any combination of features, from one to the maximum in the set) is not available for general chemical systems. However, clustering based on root-mean-square deviations (RMSDs) of all key atoms in a data set, such as carbon backbones in proteins, has been used for very large data sets, such as those in the Protein Data Bank.[196]

Two basic types of clustering methods exist: hierarchical and non-hierarchical.[197] Hierarchical methods can be perceived as "joining" clusters if proceeding from individual data points to agglomerated clusters, or "splitting", if clusters are separated from one cluster comprising the entire data set. Many hierarchical clustering algorithms join clusters, where data points are joined into clusters by the highest similarity (for example, the lowest RMSD), going from N clusters, where N = number of data points, to one cluster. Identification of the optimal clustering level can be performed automatically or manually to meet the specifications of the user, such as number of clusters desired or size of average cluster. A further division of hierarchical clustering methods is between "single-chain" linking and "multiple-chain" linking. Single-chain linking hierarchical

clustering programs, such as XCluster,[198] successively look for pairs of data points which have the highest similarity in terms of the feature set. The relationship between a single data point and the other members of the cluster it is joining, other than its nearest neighbor in the cluster, is immaterial, although this nearest neighbor was obviously found earlier to link to another member of the cluster. Multiple-link clustering can drastically increase computational time, as each member in a cluster needs to be compared to a potential new member. A compromise is to compare the feature set of a potential new member to the centroid of the feature set of the existing cluster, the centroid being made up of the average value of each feature.

Non-hierarchical clustering methods generally require selection of the desired number of clusters as an input, in addition to identifying a feature set where necessary. Although the computational time is decreased compared to hierarchical clustering because the number of clusters is usually set low (less than 10), the possibility of partial membership in a cluster (e.g., fuzzy clustering) increases computational time. k-means clustering is non-hierarchical,[199] finding the n most dissimilar data points, where n equals the target number of clusters, then adding data points to each cluster by similarity to this initial set.

Clustering data points is not usually an end in itself in conformational analysis of a single molecule; rather, finding the centroid or data point nearest the centroid in each cluster is the goal. Versatile clustering programs allow identification of data points representing clusters that are inside (a member of) or outside (not an actual member) of the data set, based on average values of features or other criteria.

2.6.2 Hierarchical Clustering as Implemented in XCluster

As noted above, XCluster is a single-link hierarchical clustering method which was designed specifically for clustering of molecular conformations. As such, issues such as circularity of torsional angle data and molecular symmetry are handled automatically with some user options available. Circularity of torsional angle data addresses the fact that torsional angles are polar coordinates, and need to be treated differently than scalar values. For scalar values, the shortest distance between two points is the simple difference of their values. The scalar distance between torsional angles may or may not be the correct shortest distance, because $0° = 360° = 720°$ etc. all represent the same torsional angle. For example, the shortest distance between $350°$ and $5°$ torsional angles is $15°$ degrees, not the simple difference $345°$. Symmetry issues include not considering equivalent carbons, such as those in the *ortho* positions on symmetrical phenyl rings, as different atoms.

Based on the selected feature set/alignment option, XCluster performs clustering by first calculating an intermolecular distance matrix of a set of superimposed molecules. For feature sets consisting of atoms only, XCluster allows the user to select a new superposition, or to use the molecular database's existing superposition. In the case of atom superpositions, each interconformational distance is a RMSD average between locations of the atoms in the feature set. Superpositions are not applicable to torsional angle-based clusterings, since in that case, the distances are calculated based on torsional angle values, as opposed to atom locations. For torsional angles, the distance is the average of the differences between the torsional angle values, and is independent of atom locations.

Each distance value d_{ij} in the distance matrix represents the similarity between conformers i and j by comparison of the items in the feature set. The critical distance d^* is incremented from 0 at clustering level 1 (where the number of clusters equals N, the number of conformers) to the minimum distance d_{ij} that is more than zero, to create clustering level 2 with N-1 clusters. As the d^* value increases (in accordance with the increases in the actual d_{ij} in the intermolecular distance matrix), additional clusters are formed by combining two clusters found at the previous clustering level. Eventually, all of the conformers would be in the same cluster; the d^* value would reach the highest d_{ij} value (representing the two most dissimilar conformers by the feature set criteria) in the intermolecular distance matrix, and the conformers would become one data set. XCluster is hierarchical; each cluster is an agglomeration of clusters formed on the previous level, and no cluster is disassembled at any succeeding level. Figure 2.3 shows an example data set, along with its distance matrix and the input-ordered and generic-ordered distance maps. The color-coding of the distance maps is discussed in the next section.

Conformational space coverage and population of clusters can be compared. The radius of gyration (r_g) of the conformational point cloud is used to measure the occupation of conformational space. Large values indicate more coverage or "sweep" through conformational space, and small values indicate less coverage.

Representatives from each cluster can be cluster members, or calculated conformations that derive from the centroids of the members of each cluster. This study uses representative conformers that are the members of each cluster that most closely resembles the calculated "center" (average feature set atom locations) of each cluster, to

(a)

(b)

	1	2	3	4	5
1	0	4	1	3	3.162
2	4	0	4.123	1	1.414
3	1	4.123	0	3.162	3
4	3	1	3.162	0	1
5	3.162	1.414	3	1	0

(c)

(d)
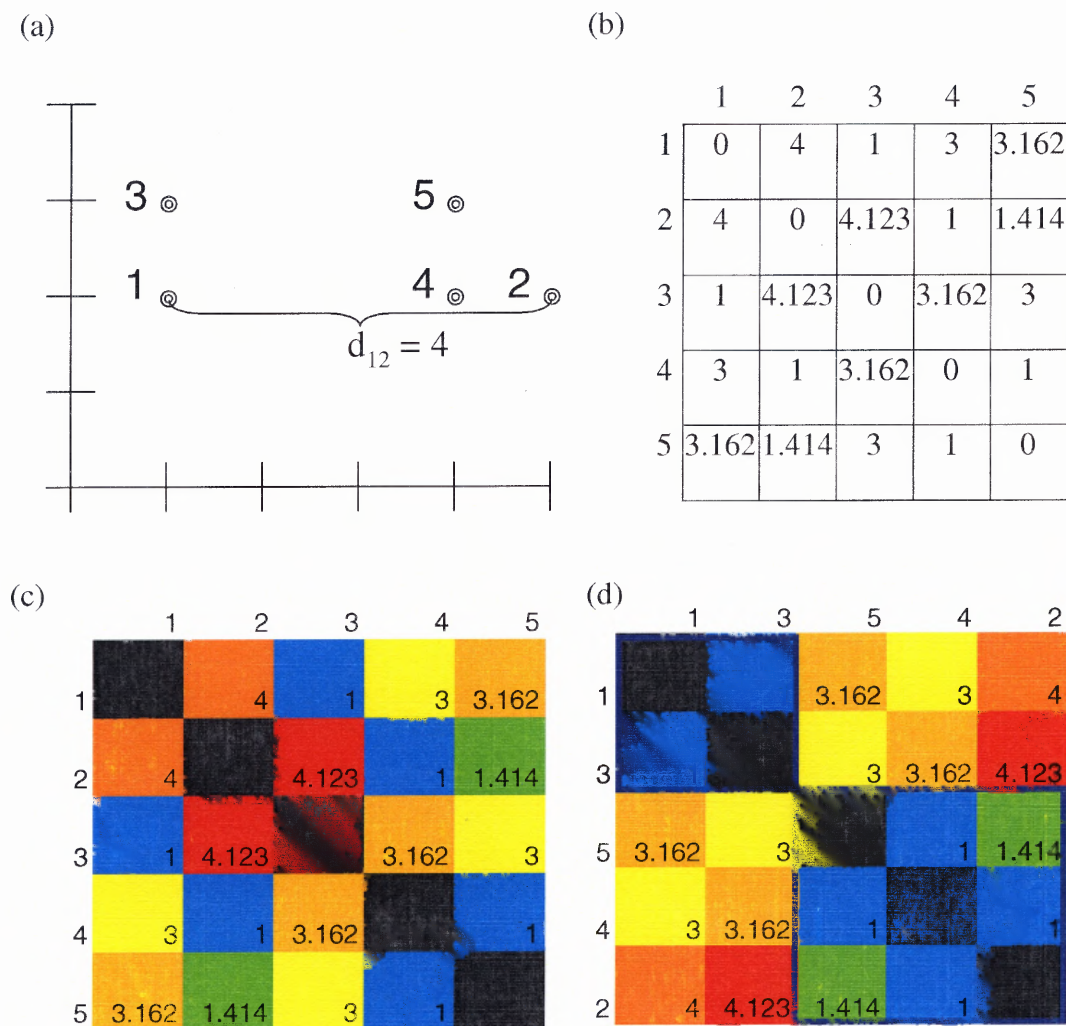
Figure 2.3 Example of a simple clustering study. (a) Five conformers, with feature set of x-y coordinates. (b) Distance matrix. (c) Input-ordered distance map. (d) Generic-ordered distance map.

ensure that the representative conformer used in the CoMFA and CoMSIA studies is a local energy minimum.

2.6.2.1 Distance Maps. A first step in the analysis of an XCluster study is review of its distance map. Distance maps are a color-coded display of the RMSD values for pairs of molecules in each analog's set of conformers. The distance map therefore summarizes the many different RMSD values into a small number of RMSD ranges indicated by color

on a grid. Distance maps for different combinations of feature set/alignment options can be compared, or distance maps can be evaluated based on their appearance. In this study, the input order of the conformations was not meaningful, so a generic ordering was used to display the distance maps. This procedure orders the conformers such that clusters form from adjoining conformers; multiple generic orderings are possible, depending on the first conformer selected. The generic ordering displayed in the distance map is not the only one possible; it is the one generic ordering that starts with conformer 1 of each analog's data set on the left side.

2.6.2.2   XCluster Basic Statistics. The XCluster run produces simple statistics by default; there are additional statistics which may be requested when the run is completed. The seven basic statistics are:[198]

- Clustering level, L : corresponding to N-1-L clusters, where N is the number of conformers.

- Critical distance, d^*_L: the critical threshold distance between the clustering level L and the previous clustering level, L-1. The critical distances are derived directly from the distance matrix, and match the exact distance matrix entries. Thus, a sorted list of all distance matrix values would match a sorted list of critical distances for a particular clustering study.

- S_{re}: the reordering entropy. At clustering levels 1 and N, there are $N!$ reorderings. Each clustering level between 1 and N has a smaller number of reorderings, limited by forbidden reorderings that break up existing clusters. The reordering entropy is high when there is a small number of large clusters and low when there is a large number of small clusters.

- Number of clusters, k, equals N-1-L, as noted above. This can also be called the actual number of clusters, to differentiate it from the effective number of clusters. The actual number of clusters is the sum of all clusters, large and small.

- Effective number of clusters, k^*, corresponds to the "number of large clusters", calculated as $k^* = \exp(S_{cl})$, where:

$$S_{cl} = -\sum_{i=1}^{k} x_i \ln x_i \tag{2.8}$$

and x_i = fraction of the data set in the cluster and k = the actual number of clusters at the clustering level.

- Minimum separation ratio and maximum separation ratio. The separation ratio is the ratio between the critical threshold distance of the current level, and the critical threshold distance of the next highest clustering level (the level with one less cluster):

$$R_L = d^*_L / d^*_{L+1} \qquad (2.9)$$

The effective number of clusters is a measure of how many large clusters exist at a particular clustering level. The effective number of clusters reaches its maximum when all clusters are of equal size. The maximum effective number of clusters for a given level is the actual number of clusters, which can be achieved at individual levels, but not maintained across the entire range of clusters.

For example, if at clustering level N-1, the actual number of clusters equals the effective number of clusters, which equals two, it can be seen that at clustering level N-2, the effective number of clusters could not reach its theoretical maximum of three (because one of the clusters with 50% of the conformers remains, with the other cluster splitting, at the most evenly distributed, into two clusters containing 25% of the conformers). This concept is illustrated in Figure 2.4. Qualitatively, the effective number of clusters measures how many "large clusters" are present at a particular clustering level for a clustering study, but the more uneven the sizes of the clusters are, the lower the effective number of clusters. The minimum separation ratio shows how far apart the most similar clusters are in a set of clusters; this value is important when it rises above 2,[198] indicating a clear separation between the most similar clusters. The maximum separation ratio shows how far apart the most dissimilar clusters are in a set of clusters. These basic statistics are provided as part of the standard XCluster output file.

Set of conformations   Set of conformations

$k = 2, k^* = 2$ $k = 3, k^* = 2.83$
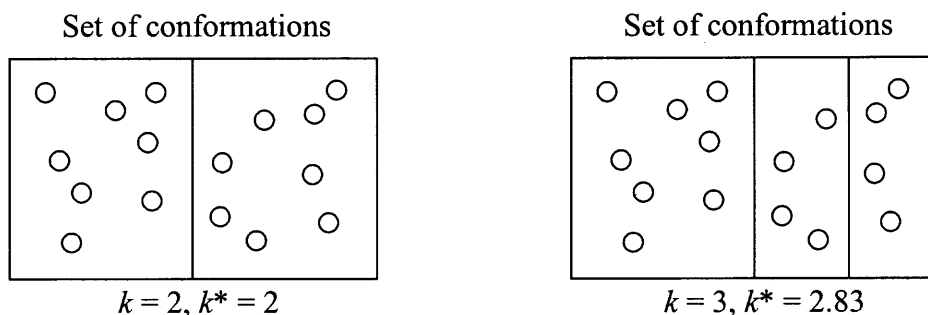
Figure 2.4 Diagram of successive clustering levels. Actual (k) versus effective (k^*) number of clusters is demonstrated. Hierarchical clustering prevents splitting of more than one cluster at a step, thus $k = k^*$ cannot be maintained.

2.6.2.3 XCluster Derived Statistics. Although the minimum separation ratio is noted by Shenkin and McDonald to be a figure of merit,[198] it may not be sufficient to select the best clustering study or clustering level for all data sets.[198,200,201] Large data sets can have a continuum of conformations, such that the minimum separation ratio never rises much above 1. For this reason, the present work defines the novel derived statistic, percentage change in the effective number of clusters (%ΔEff), which may be a more useful statistic for large data sets. The %ΔEff at level L is calculated as:

$$\%\Delta\text{Eff} = \text{Abs}[(\text{Eff}_L) - \text{Eff}_{L-1}]/\text{Act}_L * 100 \tag{2.10}$$

where:

%ΔEff = percentage change in the effective number of clusters

Eff_L = effective number of clusters at clustering level L

Eff_{L-1} = effective number of clusters at clustering level $L\text{-}1$

Act_L = actual number of clusters at clustering level L

Consecutive application of the distance map and clustering statistics criteria outlined above leads to determination of the best feature set/alignment option and the

optimal number of clusters. Identification of the representative conformer for each cluster is discussed below.

2.6.2.4 Determining Representative Conformers. Once the optimal feature set/alignment option and clustering level are chosen, the representative conformer for each cluster is determined as the conformer with feature set atom locations closest to the average feature set atom locations within the cluster.

2.7 QSAR studies

2.7.1 Review of CoMFA and CoMSIA Methods

This study uses ligand-based QSAR methods because there are no available three-dimensional structures of the DAT.[5] The ligand-based QSAR technique CoMFA was first used to predict the binding of steroid molecules to carrier proteins.[22] Since then, CoMFA has been used in many applications to determine the relationship between the structures and electrostatics of a set of compounds and one or more properties of the compounds. The steric bulk and electrostatic potential fields at various grid points, as seen by a probe atom, are used to develop a model to predict the biological activity of each compound. Figure 2.5 shows a representation of the CoMFA procedure, with a sample grid identified. Each SYBYL CoMFA column shown in the molecular spreadsheet (i.e., QSAR table) contains a volume estimate of each molecule based on the number of lattice points contained in the molecule. The columns shown represent the
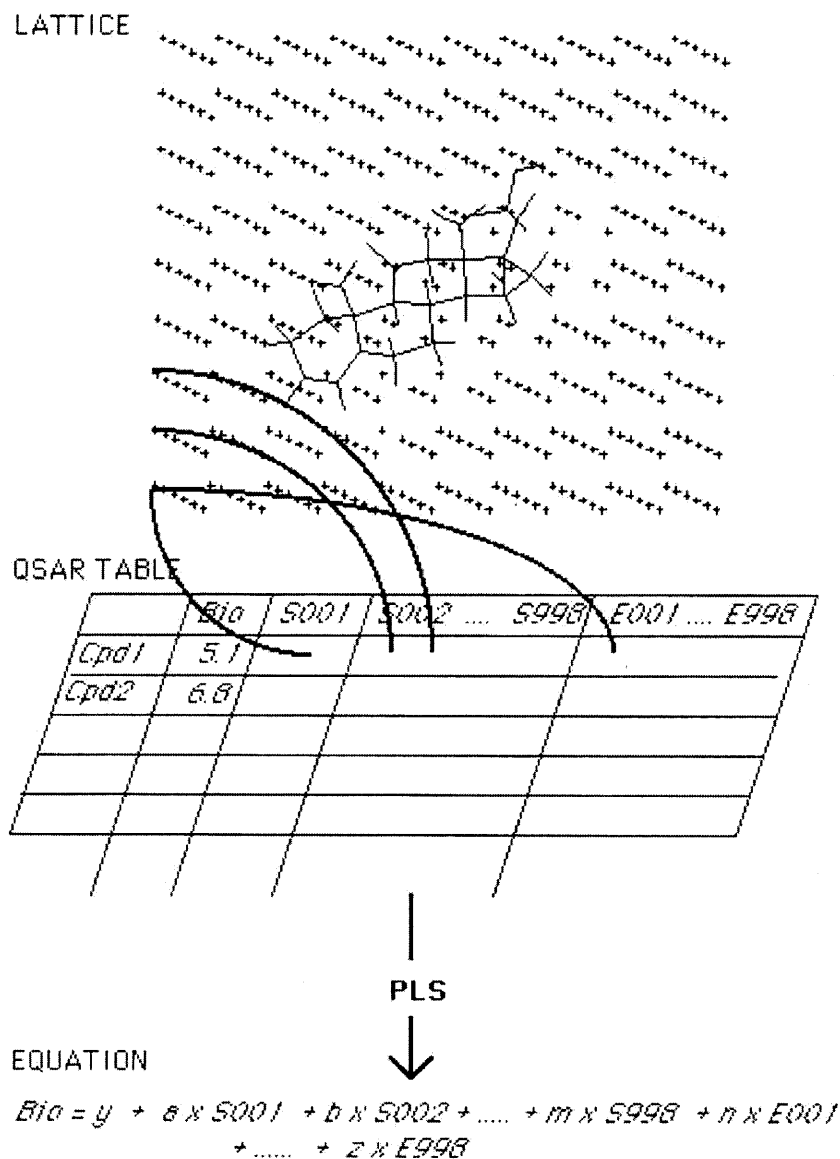
Figure 2.5 Representation of CoMFA procedure.[22] Three-dimensional grid of points is shown around a sample compound. Steric and electrostatic potentials are calculated at each grid point. Partial least squares analysis relates the ligand's molecular properties (S001, ...S998, E001,...E998) to its biological activity ("Bio"), creating a predictive model. (Reprinted from *SYBYL QSAR Manual*, Tripos, Inc.)

thousands of steric or electrostatic potential values that make up the CoMFA field. The individual values for steric bulk and electrostatic potential fields at grid points are calculated using Lennard-Jones and Coulombic potentials. Columns are produced in a SYBYL molecular spreadsheet that represent a specific CoMFA run. Cutoffs are used to ensure that large values for steric and electrostatic potentials are not weighted too heavily. The cutoff tells the program that any value above the cutoff is set to the cutoff value. There is a steric potential cutoff, an electrostatic potential cutoff, and a column filtering value (σ) to even out variations within a CoMFA column.

CoMSIA columns are calculated in a similar manner, with field values calculated at each grid point; however, Gaussian functions are used to approximate the shape of each molecule, and properties derived from atom-based descriptors are associated with the Gaussian functions.[202] CoMSIA is based on the molecular alignment program, SEAL,[203] which uses rigid bodies to find the maximum overlap of electrostatic and steric fields. The effect of using the Gaussian function is that the distance-dependence response of field variables is smoother and singularities possible with Lennard-Jones and Coulombic potentials used in CoMFA can be avoided. CoMSIA also has the advantage of a selection of different field types: hydrogen bond donor/hydrogen bond acceptor, hydrophobic/hydrophilic, and steric/electrostatic.

CoMSIA studies were performed to supplement the CoMFA studies. Since CoMSIA addresses hydrophobicity and hydrogen donor and acceptors, as well as electrostatic and steric effects, in addition to the different calculation methods noted above, the CoMSIA results might provide a different insight into the molecular requirements of the higher activity **1** analogs. CoMSIA studies can also produce more

meaningful contour maps because molecular similarity is used to align ligands, allowing easier identification of areas which can be substituted to improve activity.

Both CoMFA and CoMSIA columns are converted to models through use of PLS. PLS, also known as projection to latent structures, is used because traditional least-squares methods cannot be used for the large number of data points created in a CoMFA or CoMSIA study. The use of the PLS method, developed by Wold et al.,[204] allows manipulation of thousands of x-values, using a modified least squares algorithm to create a model. An iterative process is used where new variables t_a are formed from the x variables, and used to predict y variables. Each step produces a component, consisting of a set of t_a variables known as "scores" or "latent variables". As the number of components increases, so does the complexity of the model. The limit for the number of components is one less than the number of objects (molecules) in the study, essentially modeling the variability in response due to each molecule. Components are therefore limited to either the SYBYL default of six components, or the rule-of-thumb of one-third of the number of molecules in the training set (compounds used to form the model).

PLS requires the use of a cross-validation method to determine where the optimal number of components occurs. Various cross-validation methods exist, all consisting of leaving out a group of compounds for each run within the cross-validation analysis. The leave-one-out (LOO) approach has as many groups as molecules; each molecule is left out once, for a total number of runs equal to the number of compounds in the training set for each cross-validation run. A correlation coefficients is calculated for the LOO cross-validation (LOO/CV) run:

$$q^2 = 1 - [PRESS/SS(Y)] \qquad (2.11)$$

where q^2 equals predictivity and is dimensionless, generally between 0 and 1, but can be less than zero; PRESS equals predictive residual sums of squares formed from all of the sums of squares of residuals for each cross-validation run (residual = predicted y-value - experimental y-value); and SS(Y) equals sum of the squares of the y variables. The optimal number of components is identified as the number of components where the q^2 still increases, and the standard error of prediction (based on the residuals found) does not increase significantly.

The optimal number of components is used for a non-cross-validated (NCV) PLS study, also known as a "full model" because all molecules in the training set are used to create it. A second correlation coefficient is created from the NCV PLS analysis, known as the "goodness-of-fit", r^2, calculated by:

$$r^2 = 1 - [SS(F)/SS(Y)] \qquad (2.12)$$

where r^2 equals the goodness of fit and is dimensionless, generally between 0 and 1; SS(F) equals the sum of the squares of the y residuals, and SS(Y) equals the sum of the squares of the y variables. Many CoMFA studies show r^2 values in the range of 0.9 or above,[205] although Wold et al.[204] indicate that q^2 and r^2 should be within 0.15 of each other.

Two standard errors are also calculated; one for LOO/CV PLS analyses (standard error of prediction, SEP) and the other for NCV PLS analyses (standard error of estimation, SEE). SYBYL also performs a statistical F-test on the NCV PLS results. The F value has a general definition of the normalized results of a model divided by one minus the normalized results of the model. F values for NCV PLS analyses with the

same optimal number of components can be compared directly, but models with different numbers of components cannot.

CoMFA and CoMSIA studies using PLS have recently been augmented by the use of PLS Region Focusing.[206] Region Focusing allows refinement of a CoMFA or CoMSIA model through increasing the weighting of more important lattice points and decreasing the weighting of less important points, as indicated by a NCV PLS model.

Use of the PLS method avoids the necessity of x variable selection or x variable reduction in CoMFA or CoMSIA studies. Although other methods of statistical treatment of CoMFA x values have been developed, including genetic algorithms,[207] K nearest neighbor,[208] and neural networks,[209] the application of PLS to CoMFA has allowed it to become a widely used method.

2.7.2 QSAR Studies on Training Set

Dividing data into a training set and test set allows external validation of a QSAR model. The QSAR model is based on part of the total data set (the training set), and the remaining compounds are left out of the model for external validation (the test set). In the present study, preliminary 3D-QSAR studies were performed on the training set, the best results focused, and then supported by internal validation using y-value scrambling. One hundred y-value scramblings using SYBYL's Progressive Scrambling[210] function were performed on each focused model to estimate their stabilities. 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. Ten times the default number of scramblings were used to minimize the effects of the random seed. Three key results are 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.

Q^2 is limited by the selected critical point (0.85), and is more conservative than the cross-validated q^2 calculated in the LOO/CV CoMFA PLS runs. cSDEP can be compared to the SDEP and the actual error in the experimental samples. Like the q^2 and SEP results above, Q^2 should be high, preferably above 0.5, but certainly above 0.3, and cSDEP should be low and similar to SEP. These results were compared to the recommended cutoffs for stable models, of which the most important is a slope $(dq'^2/dr_{yy'}^2)$ under 1.2.[211]

2.7.3 Prediction of Test Set DAT/SERT Selectivities

Experimental and predicted test set DAT/SERT selectivities were used to calculate correlation coefficients for each of the selected CoMFA models as an external validation method.[212] The Predicted DAT/SERT selectivity (y) vs. Experimental DAT/SERT selectivity (x) was plotted to calculate:

- R^2, the coefficient of determination, of the best-fit line,

- R, the correlation coefficient, R, 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 DAT/SERT selectivity (y) vs. Predicted DAT/SERT selectivity (x), was also plotted to find:

- $R'_0{}^2$, the coefficient of determination, of the zero-intercept line of Experimental vs. Predicted, and

- k', the slope of the zero-intercept line for Experimental vs. 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. Models that better predict the test set have all of these statistics near 1. Barring that, the value for R^2 should be near either the value for $R_0{}^2$ or $R'_0{}^2$. These results were used to compare the predicted validity of the selected QSAR models.

2.8 Prediction of Novel Compounds

A pharmacophore is a set of required, but not necessarily sufficient, molecular characteristics that represent a binding (or bioactive) conformation of a set of compounds at a binding site. The geometry of a molecule determines its conformational elements. Qualitative pharmacophore mapping was performed using SYBYL's Map CoMFA function. This feature uses the CoMFA model to highlight areas of the molecules that would lead to higher activity; areas where varying electrostatic and steric effects by chemical group substitution leads to more active compounds are identified. This information was used to predict the DAT/SERT selectivities of novel compounds.

This project resulted in identification of a predictive and stable QSAR model that was used to identify novel compounds with predicted high DAT/SERT selectivities. Molecular modeling of synthesized compounds is part of the iterative process of finding the best compounds to meet the needs of the research. These results will be provided to the Rice and Rothman groups at NIH as a contribution to ongoing GBR 12909 research.

CHAPTER 3

METHODS

The 3D-QSAR methodology is summarized in Figure 3.1, which reviews the major steps in a manner applicable to more general QSAR studies. Validation methods can be applied comparatively or individually, depending upon whether multiple conformations and/or alignments are chosen for parallel analysis. This project compares six different QSAR studies which go through the validation steps in parallel; results which do not meet validation criteria are set aside or reassessed.

3.1 Assumptions

All modeling projects have some level of approximation. In QSAR work, the primary goal is to develop a model based on the structure and/or conformation of molecules that yields accurate activity estimates. Assumptions are of three main types: limitations in scope, accuracy of source data, and software/algorithm approximations.

3.1.1 Limitations in Scope

Thousands of compounds from different classes of DA reuptake inhibitors have been assessed for activity at the DAT. This project focuses on one subset of the GBR 12909 class of DA reuptake inhibitors, due to Drs. Rice and Rothman's research on the use of **1** and its analogs for agonist substitution therapies for cocaine addiction. Initially, forty-five compounds based on **1** were selected from over three hundred **1** analogs, and three compounds, including **1**, were added as additional test set compounds. Piperazine and

39
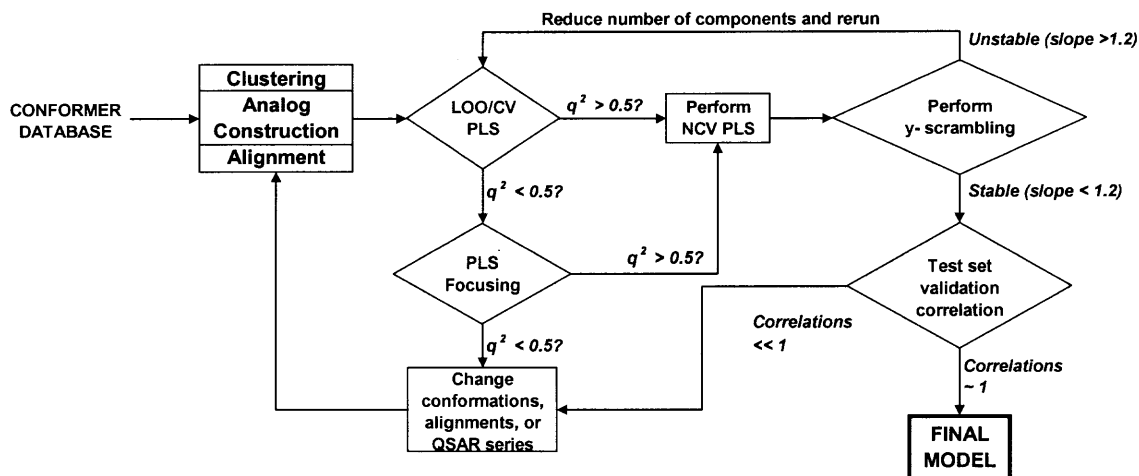
Figure 3.1 Flow chart for QSAR study.

piperidine compounds were selected based on varying A-side structures, and activities were reviewed to ensure an acceptable range of activities (more than three log units).[168]

Conformational analysis is expected to yield a subset of conformations of all of the possible low energy conformations of the **1** analogs. This subset is assumed to have one or more conformations that adequately represent the bioactive conformation of these compounds in a QSAR model. Detailed conformational analyses were performed only on **2** and **3**; it is assumed that the local minima found for these analogs are significant for the other 43 analogs. Selection of random search and clustering parameters was done with this goal in mind. Further approximations are that a static binding conformation can adequately represent the dynamic changes in a ligand that occur during binding, and that the bioactive conformation is one of the local minima on the conformational PES of each molecule.

3.1.2 Accuracy of Source Data

The Venanzi group has access to several different sets of data on DA reuptake inhibitors. To minimize propagation of errors and allow comparison of data, this research project is

based on activity data provided by the Rice group from their colleagues in the Rothman group, both at NIH. The pharmacological tests were performed on the same species of rat using the same methodology for each type of test.[184] By using binding data from the same laboratory, the relative accuracy of the calculations is assured.

3.1.3 Software/Algorithm Approximations

All calculations were performed on a SGI Origin 2000 with twenty 300 MHz processors, 20 GB of memory, using the IRIX® 6.5.18m UNIX®-based operating system. Molecular modeling was carried out with SYBYL Version 6.9 (available from Tripos, Inc., St, Louis, MO) and Macromodel 8 (available from Schrödinger, Inc., Portland, OR). Appendix A lists additional software specifications.

A discussion of the basis of the modeling software and algorithms is found in the Background section of this work. Molecular mechanics methods model atoms and bonds using macroscopic-scale relationships. Modeling results for molecules not found in the set of molecules that validated the Tripos force field, such as the 1 analogs, are approximated from values for individual atom types.

The CoMFA and CoMSIA methods are further approximations, both in the sense of being based on the Tripos force field, and that the methods both are based on responses to probe atoms. Default probe atoms were used for all calculations. Grid spacing and cutoff values for CoMFA further approximate the results. Although studies have been performed on these factors, no definitive protocols other than use of the default values have been recommended.

Recording and presentation of CoMFA and CoMSIA results is subject to the limitations of the modeling software as well. Estimated activities are reported to four or
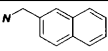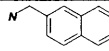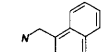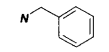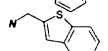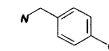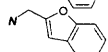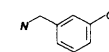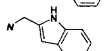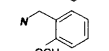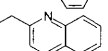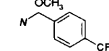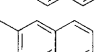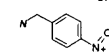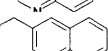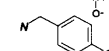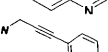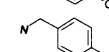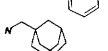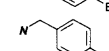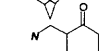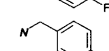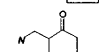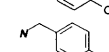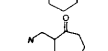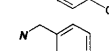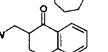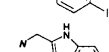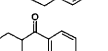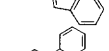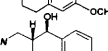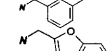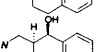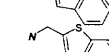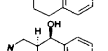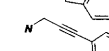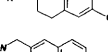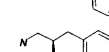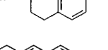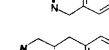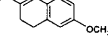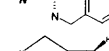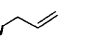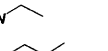
more significant figures, yet the expected accuracy of "good" models can be as poor as one full log unit off or as good as less than 0.1 log unit off from the experimental value. The CoMFA convention of three digits past the decimal point was used for q^2, SEP, r^2, and SEE.

3.2 Selection of Analogs

A subset of structurally-related compounds was selected from over 300 analogs in the Rice/Rothman series. The selection criteria included each analog having as few rotatable bonds and chiral centers as possible, while maintaining either a piperazine or piperidine ring; having the same 2-[bis-(4-fluorophenyl)methoxy]ethyl- B-side structure; and pharmacological data covering a wide range of DAT/SERT selectivities. Analog **2** was selected because it has fewer rotatable bonds than **1** (two on the A-side for **2** versus four on the A-side for **1**), an identical center ring and B-side as compared to **1**, but 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. DAT and SERT binding affinity and DAT/SERT selectivity are shown in Table 3.1 for the 45 analogs of **1** in the study. Binding experiments were performed in the Rothman laboratory using identical protocols.[12,14,17] 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.

Table 3.1 Structures, Binding K_i Values and DAT/SERT Selectivities for **1** Analogs[a]

	X = N N-CH$_2$-R[b]	DAT K_i (nM)[c]	SERT K_i (nM)[c]	DAT /SERT Select.[d]	Ref.[e]		X = C(H) N-CH$_2$-R[f]	DAT K_i (nM)[c]	SERT K_i (nM)[c]	DAT /SERT Select.[d]	Ref.[e]
2	[structure]	8	312	1.59	*17*	**3**	[structure]	0.71	229	2.51	*14*
4	[structure]	31	243	0.89	*17*	**26**	[structure]	9.9	171	1.24	*72*
5	[structure]	4.1	495	2.08	*17*	**27**	[structure]	7.9	266	1.53	*72*
6	[structure]	6.4	286	1.65	*17*	**28**	[structure]	13.3	161	1.08	*72*
7	[structure]	0.7	119	2.23	*17*	**29**	[structure]	63	248	0.60	*72*
8	[structure]	56	51	-0.04	*17*	**30**	[structure]	1.1	442	2.60	*72*
9	[structure]	16.1	485	1.48	*17*	**31**	[structure]	1.2	387	2.51	*72*
10	[structure]	61.8	550.5	0.95	*17*	**32**	[structure]	5.7	239	1.62	*72*
11	[structure]	20	266	1.12	*h*	**33**	[structure]	1.1	199	2.26	*72*
12	[structure]	100	598	0.78	*h*	**34**	[structure]	2.3	245	2.03	*72*
13[g]	[structure]	84	970	1.06	*20*	**35**	[structure]	1.2	238	2.30	*72*
14[g]	[structure]	87	169	1.29	*20*	**36**	[structure]	1.2	606	2.70	*72*
15[g]	[structure]	49	525	1.03	*20*	**37**	[structure]	2.1	205	1.99	*72*
16[g]	[structure]	57	112	0.29	*20*	**38**	[structure]	0.73	88	2.08	*14*
17[g]	[structure]	22	2.96	-0.87	*20*	**39**	[structure]	16	370	1.36	*14*
18	[structure]	18	32	0.25	*h*	**40**	[structure]	1.01	85	1.93	*14*
19	[structure]	1.3	30	1.36	*20*	**41**	[structure]	1.61	246	2.18	*14*
20	[structure]	14	3	-0.67	*20*	**42**	[structure]	5.3	164	1.49	*14*
21	[structure]	6.4	261	1.61	*20*	**43**	[structure]	9.32	15.19	0.21	*h*
22	[structure]	53	57	0.03	*20*	**44**	[structure]	4.77	48.13	1.00	*h*
23	[structure]	64.6	165	1.41	*h*	**45**	[structure]	0.8	22	1.44	*h*
24	[structure]	33.7	209	1.79	*h*	**46**	[structure]	1.3	136	2.02	*h*
25	[structure]	40	888	1.35	*h*						

[a] Compounds **2** and **4** through **25** are piperazines; compounds **3** and **26** through **46** are piperidines.
[b] Piperazine compounds, X and A-side substituent R as identified in Figure 3.2.
[c] Measured as displacement of bound [^{125}I] RTI-55 from rat brain synaptosomes.
[d] DAT/SERT selectivity defined in Equation (2.5)
[e] Reference for experimental data; all data is from the Rice and Rothman labs at NIH.
[f] Piperidine compounds, X and A-side substituent R as identified in Figure 3.2.
[g] Compound tested as a racemic mixture.
[h] Unpublished data from Rice and Rothman groups, National Institutes of Health, 2005.
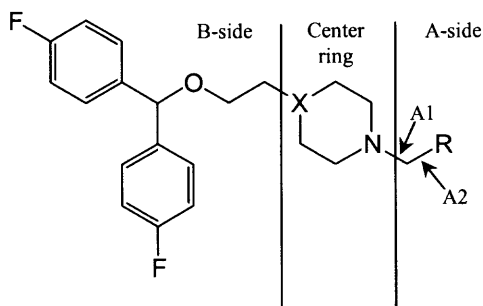
Figure 3.2 GBR 12909 analog scaffold. Analog **2**: X = N, R = 2-naphthyl. Analog **3**: X=C(H), R = 2-naphthyl.

Figure 3.2 shows the scaffold of the set of analogs of **2** and **3** that were used in this study. To keep the variations to a minimum, the compounds are either piperazine- or piperidine-based, have fluorine atoms in the *para* positions of the B-side phenyl substituents, and have only one carbon atom between the A-side nitrogen and the A-side substituent. This QSAR study therefore focuses primarily on the effect of A-side substitutions. A-side moieties include naphthalene, thiophene, furan, and various phenyl substituents. There are a few compounds, such as the naphthyl compounds, that have piperazine and piperidine counterparts. This may be because the piperidine compounds were developed later than most of the piperazine compounds, so lessons learned from earlier SAR work on the piperazines were applied. Many of the aromatic substituents are symmetrical about the A2 torsional angle (e.g. **11, 24, 26,** etc.). In cases where the substituents were not symmetrical, the orientation of the substituent was created to follow the orientation of the naphthalene substituent in the respective template compound. 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 DAT/SERT selectivities and to be split between piperazine and piperidine analogs.

3.3 Conformational Analysis and Clustering

Since it is computationally impractical to carry out conformational analysis of all the analogs in Table 3.1, **2** and **3** were selected for detailed conformational study. Two random searches were performed, one on the piperazine **2** and one on the piperidine **3**. Comprehensive hierarchical clustering using XCluster was performed on several hundred conformations of each analog. A representative conformer in each conformational cluster was used as a template for building the other **1** analogs by changing the substituents as necessary, and minimizing each resulting structure. 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.

3.3.1 State of Protonation

Studies on DA binding have shown that it most likely binds to the DAT in the protonated state.[213] Cocaine is thought to exist as a 90/10 mixture of protonation states, favoring the protonated form, but having similar conformations in the neutral and protonated forms.[27] HF/6-31G* molecular orbital calculations performed on **1** showed that protonation of the A-side nitrogen is favored over the B-side nitrogen.[214] This agrees with binding studies of piperidine analogs of **1** that show that the binding affinity for the DAT is substantially higher when the A-side nitrogen rather than the B-side nitrogen is present.[172] The A-side nitrogens in **2** and **3** were therefore modeled with an associated proton.

B-side   **center ring**   **A-side**

☐ *Key center ring atoms.*
○ *Key A-side atoms.*
◇ *Key B-side atoms.*
● *B-side plane atoms.*
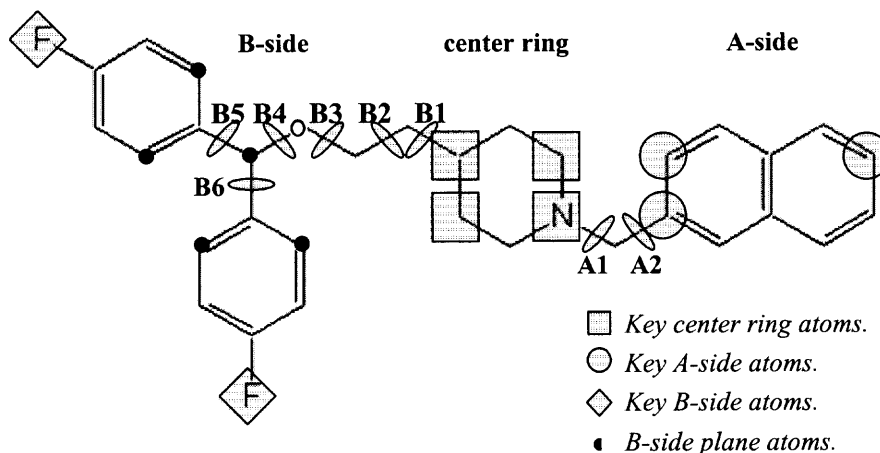
Figure 3.3 Diagram of **3** and its features. Identified are potential pharmacophore elements, key torsional angles and key atoms.

3.3.2 Random Search Conformational Analysis

Two random search conformational analyses were performed on the minimized structures of **2** and **3** by Milind Misra and Deepangi Pandit from the Venanzi group, using the eight torsional angles (A1, A2, B1, B2, B3, B4, B5, and B6) noted in Figure 3.3. Each search was performed for 10,000 cycles over an energy range of 20 kcal/mol from the GEM; additional parameters are found in Appendix A – Software Specifications and Input Parameters. SYBYL Version 6.9 with the Tripos force field was used to calculate conformations and energies. The detailed protocol for the random search conformational analysis of **2** has been published elsewhere.[215] The same protocol was repeated for **3**. A full analysis of the conformational profiles of **2** and **3** will be given in a future publication.[61] The conformers were converted from SYBYL mol2 to Macromodel mae multiple molecule file format for analysis by XCluster.

3.3.3 Clustering

Hierarchical clustering using XCluster[198] (Schrödinger, L.L.C., Portland, OR) was performed to determine representative conformers of **2** and **3** from the large data sets produced by the random searches. Various feature sets were examined and the resulting clustering statistics and distance maps were analyzed to determine the most appropriate clustering feature set and clustering level. A representative conformer was selected from each of the major clusters found at the selected clustering levels. These representative conformers were then used as templates from which the other analogs in Table 3.1 were constructed, creating conformational families. The clustering techniques used are summarized in three flowcharts, found in Figures 3.4, 3.5, and 3.6, and described below.

3.3.3.1 Initial Clustering Studies. Clustering studies were performed on the sets of conformers produced by random search as noted above. The multiple feature set/alignment options explored are shown in Table 3.2 and their use is described in Figure 3.4. Distance matrices were constructed separately for analogs **2** and **3** based on these options. Each entry in the distance matrix consists of the RMSD of the atoms or torsional angles in the feature set calculated for pairs of conformers superimposed according to the chosen alignment option. Table 3.2 shows the combinations of atoms or torsional angles used in the clustering studies and the superpositions where applicable. Atoms and angles used in the feature set/alignment options are identified in Figure 3.3. The feature sets range from all heavy atoms (options *a* and *b*) to key atoms that contain the DAT pharmacophore elements (quaternary nitrogen and aromatic ring on the A-side) (options *c* − *e*) to key atoms that contain the bisphenyl substituent required for **1** analog binding (options *f* − *j*). These are coupled with alignment options which superimpose the

Table 3.2 Feature Set and Alignment Options

Option	Feature Set[a]	Alignment[b]
a	All heavy atoms	☐ Center ring[c]
b	All heavy atoms	All heavy atoms
c	☐ Center ring and ○ A-side key atoms	☐ Center ring[c]
d	☐ Center ring and ○ A-side key atoms	☐ Center ring and ○ A-side key atoms
e	A1 and A2	Five heavy atoms defining A1 and A2[d]
f	☐ Center ring and ◇ B-side key atoms	☐ Center ring[c]
g	☐ Center ring and ◇ B-side key atoms	☐ Center ring and ◇ B-side key atoms
h	◇ B-side key atoms	Oxygen and neighboring carbons[c]
i	◇ B-side key atoms and ● B-side plane atoms	Oxygen and neighboring carbons[c]
j	B1 through B6	Nine heavy atoms defining B1 through B6[d]

[a] Feature set: The atoms or angles used to calculate the intermolecular distances. Key atoms (identified by symbols noted) and torsional angles are shown in Figure 3.3. RMSD calculations for the distance matrix were carried out on the feature set atoms in the corresponding alignment.
[b] Alignment performed by XCluster unless otherwise noted.
[c] Alignment performed in SYBYL.
[d] As noted above in the text, alignment of conformers for the torsional angle studies is not necessary to calculate RMSD values, but alignment was carried out to properly visualize clusters.

molecules by the center ring atoms or all heavy atoms (options *a*, *b*, *c*, and *f*) or focus on the A-side DAT pharmacophore elements (options *d* and *e*) or on the B-side atoms (options *g* – *j*). XCluster[216] was used to calculate intermolecular atomic or torsional RMSD values. Symmetry options were set to the default, which handles basic symmetry issues, such as symmetry of some aromatic groups around torsional angle A2.
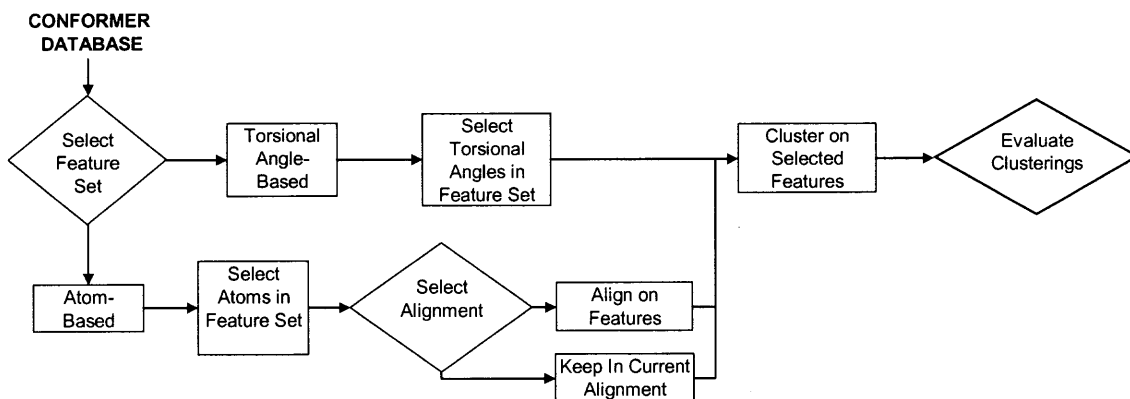
Figure 3.4 Flowchart of how to perform clustering with XCluster.
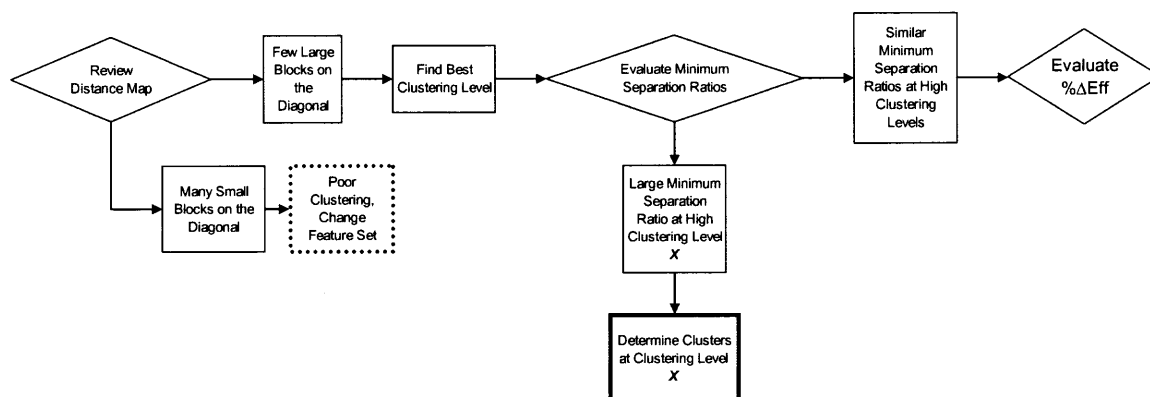
Figure 3.5 Flowchart of how to evaluate XCluster clustering studies.
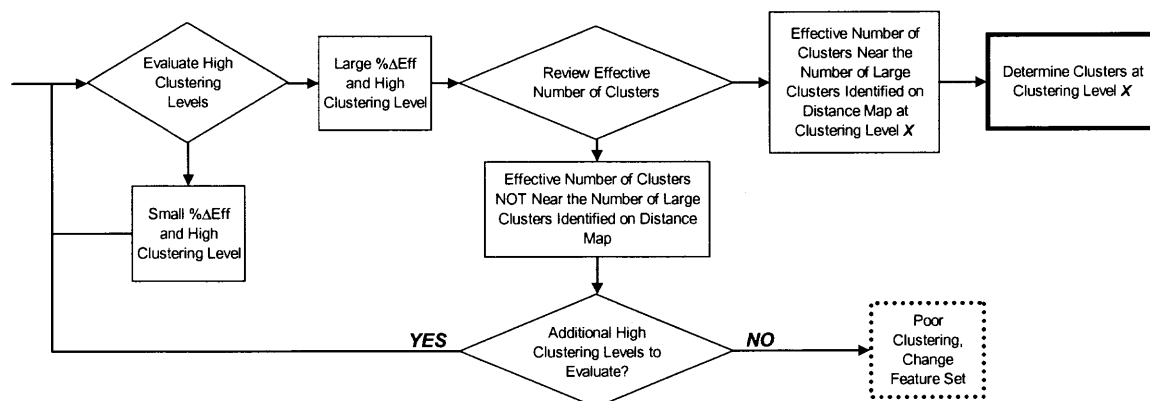
Figure 3.6 Flowchart of how to select appropriate clustering level.

3.3.3.2 Selection of the Optimal Number of Clusters. Two levels of review were necessary to select the optimal number of clusters. First, the distance maps for each feature set/alignment option were reviewed for both analogs, and those that showed small numbers of well-defined large clusters were selected for further analysis, as shown in Figure 3.5. Review of distance maps focused on the number of large squares present on the diagonal, and the relationship between these squares. For the goal of finding a manageable number of clusters in order to identify representative conformations for a future CoMFA study, maps with 10 or fewer large squares on the diagonal (with few or no small squares) were deemed to be most appropriate. In other words, the best distance maps would show relatively few large boxes centered on the black diagonal line that represents each conformer being compared to itself. This type of distance map identified the feature/set alignment option that was selected for detailed analysis (calculation of clustering statistics). Maps with many small squares along the diagonal, indicating many small clusters, were not analyzed further. For individual clustering studies, this initiated the selection of a new set of features and rerunning XCluster (see Figure 3.5).

In the second phase, the percentage of conformers included in the large clusters was calculated and the maps which had the largest percentages were selected for calculation of clustering statistics (minimum separation ratio, actual and effective number of clusters, percentage change in the effective number of clusters). These statistics were used to determine the optimal clustering level and associated optimal number of clusters, as shown in Figures 3.5 and 3.6. For example, clustering statistics were calculated for the 20 highest clustering levels for each selected feature set/alignment option. If the minimum separation ration was above two for a particular clustering level compared to

neighboring levels with values around one, that level was identified as the optimal level, as shown in Figure 3.5. If all the clustering levels had about the same minimum separation ratio (around one), then the percentage change in the effective number of clusters, along with a comparison of the effective versus actual number of clusters, was used to identify the optimal clustering level and the optimal number of large clusters, as shown in Figure 3.6. In addition, the radius of gyration was calculated to determine the coverage of conformational space by the selected clusters.

3.3.4 Representative Conformers

Once the optimal feature set/alignment option and clustering level were chosen, the representative conformer for each cluster was determined as the conformer with feature set atom locations closest to the average feature set atom locations within the cluster. The representative conformers of analogs **2** and **3** in each cluster were compared by calculating the RMSD values for the fit of the feature set atoms as well as all heavy atoms. This was accomplished by converting each conformer of **3** into a *pseudo* **2** structure by replacing the carbon atom with a nitrogen and removing the "extra" hydrogen. This allowed Database RMSD Fit calculations using SYBYL, which can only be done on a database of identical molecules. A visual comparison of the pairs of closest related analog **2** and **3** representative conformers was also made, using superimposed structures.

Templates for the conformational families to be used in the QSAR studies were selected by determining which conformers of the piperazine and piperidine **1** analogs had the lowest RMSD among pairs of conformers. Conformational families were numbered

by succeeding feature set values (in this study, by ascending A1 and A2 torsional angle values) of their template conformers.

3.4 Creation of Conformational Families

Each piperazine (or piperidine) representative conformer was used as a template from which to build the remaining 22 (or 21) analogs shown in Table 3.1. Each of the conformational families identified had one template for piperazine **1** analogs, and another template for piperidine **1** analogs. These templates were used as a basis for creation of the sets of **1** analogs to use in the CoMFA and CoMSIA studies. A SYBYL Programming Language (SPL) script was used to modify each pair of template structures to create the other analogs. SYBYL's Log Session command was used to record the creation of a set of **1** analogs. The recorded commands were saved as a SPL file and modified to make a script that created the other forty-three **1** analogs from the **2** and **3** templates. A sample script is given in Appendix B. The sets of **1** analogs were placed in separate databases, one for each cluster. The script was also used to reset the torsional angles A1 and A2 to the values found in the cluster's template, minimize each structure, and fit the atoms to align the analogs.

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 from Table 3.1. In cases where the aromatic substituents were not symmetrical, the orientation of the substituent was made to match that of the 2-naphthyl substituent in the respective template compound. Thus, the second ring in fused ring substituents would overlap that of the naphthalene substituent in **2** or **3** as much as

possible. The new substituent was added with the average A1 and A2 torsional angle values for the piperazine **2** and piperidine **3**. Each molecule was minimized for 1,000 iterations using the Tripos force field[163] with Gasteiger-Hückel charges.[217] The additional test set compounds (all piperazines) were constructed individually from analog **2**, minimized, and aligned in the same manner as the other compounds.

3.5 Molecular Alignment

Each analog in the series was aligned to analog **24** (R = H), the analog with the smallest substituent, using the four atoms comprising the A1 torsional angle. An atom-based alignment had been used in the Venanzi group's previous methylphenidate study and other[28] studies of the DAT's cocaine binding site. This alignment is based on the atoms that are proposed to be included in the DAT cocaine binding site pharmacophore. Alignments reviewed included all heavy atoms and the five atoms comprising the A1 and A2 torsional angles. The selected alignment is based on the four atoms in the torsional angle A1. This aligns the key part of the nitrogen-based pharmacophore. Carbon atom 25, the final atom in torsional angle A2, could not be included because the smallest **1** analog has only a methyl substituent and no atom 25, and therefore could not be aligned on the A1 and A2 atoms.

3.6 CoMFA and CoMSIA Studies

3.6.1 CoMFA and CoMSIA Study-Specific Concerns

Many published articles on CoMFA[22] use SYBYL's default settings. Work by the Venanzi group has indicated that improved models can be found by varying the CoMFA parameters.[44] Although the column filtering value (σ) was also changed, the most significant effect was in response to changing the electrostatic and steric energy cutoffs from the default settings of 30 kcal/mol each. The research conducted on modifying electrostatic and steric energy cutoffs used a grid of electrostatic cutoffs from 10 to 60 kcal/mol and steric cutoffs from 10 to 250 kcal/mol. The CoMFA studies were performed at nine different sets of electrostatic and steric cutoffs to decrease the computation time but still cover a range of settings.

3.6.2 CoMFA and CoMSIA Preliminary and Focused Studies

Preliminary QSAR studies were performed on each of the six conformational families, consisting of 45 analogs apiece. The procedure followed for the QSAR studies is outlined in Figure 3.1. A molecular spreadsheet was created from the SYBYL molecular database of each conformational family of analogs. Energies and activities were added as columns, with DAT binding affinity and SERT binding affinity being entered directly and the DAT/SERT selectivity being calculated from the two binding affinities. CoMFA and CoMSIA columns were added to the molecular spreadsheet.

CoMFA runs were performed 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 kcal/mol steric and 30 kcal/mol

electrostatic combination being the SYBYL default setting. CoMSIA runs were performed for hydrogen bond donor/acceptor, hydrophobic, and steric/electrostatic fields. Default settings were used for the remaining CoMFA or CoMSIA parameters in each run. After the columns were added, the six rows for the test set compounds were "hidden" in the molecular spreadsheet to perform QSAR on the remaining 39 training set compounds. Sample-distance Partial Least Squares (SAMPLS)[218] was used to calculate q^2 values for the preliminary QSAR studies. LOO/CV calculations were carried out for each of the twelve studies (nine CoMFA and three CoMSIA) 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 NCV PLS model for each family in this preliminary series of calculations.

To refine the models, Region Focusing[206] was performed on the most predictive model for each family identified through the preliminary studies above using the parameter combination for CoMFA or field choice for CoMSIA that gave the most predictive NCV PLS model. The SYBYL default Region Focusing parameters were used. The focused CoMFA or CoMSIA column was analyzed with both $\sigma = 0$ and $\sigma = 2$ kcal/mol cutoffs, after initial SAMPLS screening, to ensure the highest q^2 values were obtained.

The q^2 (predictivity), standard error of prediction (SEP), r^2 (goodness-of-fit), and standard error of estimate (SEE) results of the focused models were compared, and the results validated using two different methods. Activities of the training set compounds were predicted using the optimal number of components from the NCV PLS region-

focused models. The CoMFA and CoMSIA results were analyzed, and the models derived were subjected to training set (internal) validation and test set (external) validation methods. After CoMFA and CoMSIA columns were added to the molecular spreadsheets, PLS analyses were performed to determine the best models. LOO/CV SAMPLS[218] calculations were run on the DAT/SERT selectivity/QSAR column combinations. For each cluster, the highest q^2 was selected for further analysis. LOO/CV full calculations with the default σ (column filtering value) of 2 kcal/mol was used to calculate q^2 again. The full NCV PLS calculation was carried out with the default of no σ applied (σ=0 kcal/mol, column filtering off). The q^2 values, standard errors of estimate, r^2 values, and standard errors of predictions were recorded. The best molecular field combination was then rerun using SYBYL's Focus CoMFA function, which can be applied to any field type. LOO/CV PLS and NCV PLS calculations were performed for the focused model, and the model statistics recorded. The models with the highest q^2 and lowest standard error of prediction were selected as the most predictive models. Both the initial best models and focused models were used to determine the overall best models of the cluster set. Models with q^2 values above 0.5 were checked for stability to random noise.

3.7 Interpretation of Modeling Results

To avoid depending on q^2 and r^2 values alone for identification of a predictive model,[212] the compounds selected were divided into a training set and a test set to determine the accuracy of predicting the activities of known compounds. The test set consisted of six compounds selected to span the range of activities and the two types of **1** analogs

included in this study. In addition, y-value scrambling[210,211] (in this study, the y-values are the DAT/SERT selectivities) was used to estimate the stability (the effect of randomization of the y-values, or the resistance to "noise") of the best models using the training set.

The selected region-focused QSAR model for each family was then used to predict the test set DAT/SERT selectivities. The test set compounds were displayed and the Predict Activity command used to find the DAT/SERT selectivity of each compound. The predicted DAT/SERT selectivities for each region-focused model were used to calculate residuals and to validate the model using test set correlation validation.

3.8 Prediction of Novel Compounds

SYBYL's Map CoMFA function was used to produce a three-dimensional steric and electrostatic contour map. The map was analyzed to determine where the greatest impact of changing steric or electrostatic characteristics of a **1** analog would be. Interpretation, in terms of locating important areas, was based on the template molecules **2** and **3**. The SYBYL Optimize QSAR procedure was used to replace hydrogen atoms in key locations identified on the CoMFA contour map with different substituents, and predict the DAT/SERT selectivities of the new molecules.

CHAPTER 4

RESULTS

4.1 Random Search Conformational Analysis

Seven hundred and twenty-eight conformations were found for the piperazine **2**, and 739 conformations were found for the piperidine **3**. These conformations were used as the input data to the clustering studies. Figure 4.1 shows the conformer sets for **2** and **3**. The large number of conformers required use of clustering to identify a small number of representative conformers from the larger data sets.

4.2 Clustering of Conformers

Clustering was carried out using the Macromodel module XCluster. The optimal number of clusters was selected by application of the distance map and clustering statistics criteria described in the Methods section.

4.2.1 Clustering Distance Maps

Distance maps created from the 10 clustering feature set/alignment combinations are shown in Figures 4.2 and 4.3. Small RMSDs, representing high similarity, are shown by the color black (where the RMSD = 0) or blue. The colors graduate through green, yellow, orange and finally red for the highest RMSDs, representing the least similar conformations. The scale varies based on the RMSD range of the study; blue might represent a range from 0 to 0.5 Å in one clustering study, and 0 to 2 Å in another
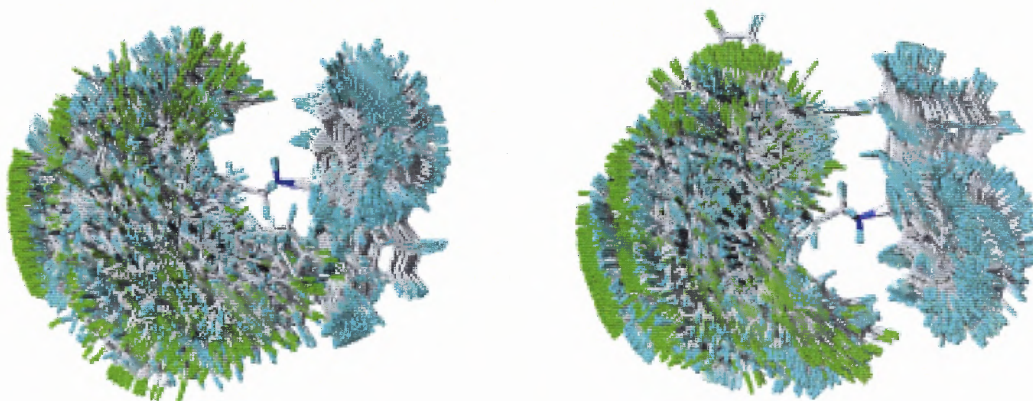
Figure 4.1 Random search results for **1** analogs. (a) The piperazine **2** and (b) the piperidine **3**. Conformers are superimposed on center piperazine (**2**) or piperidine (**3**) ring.
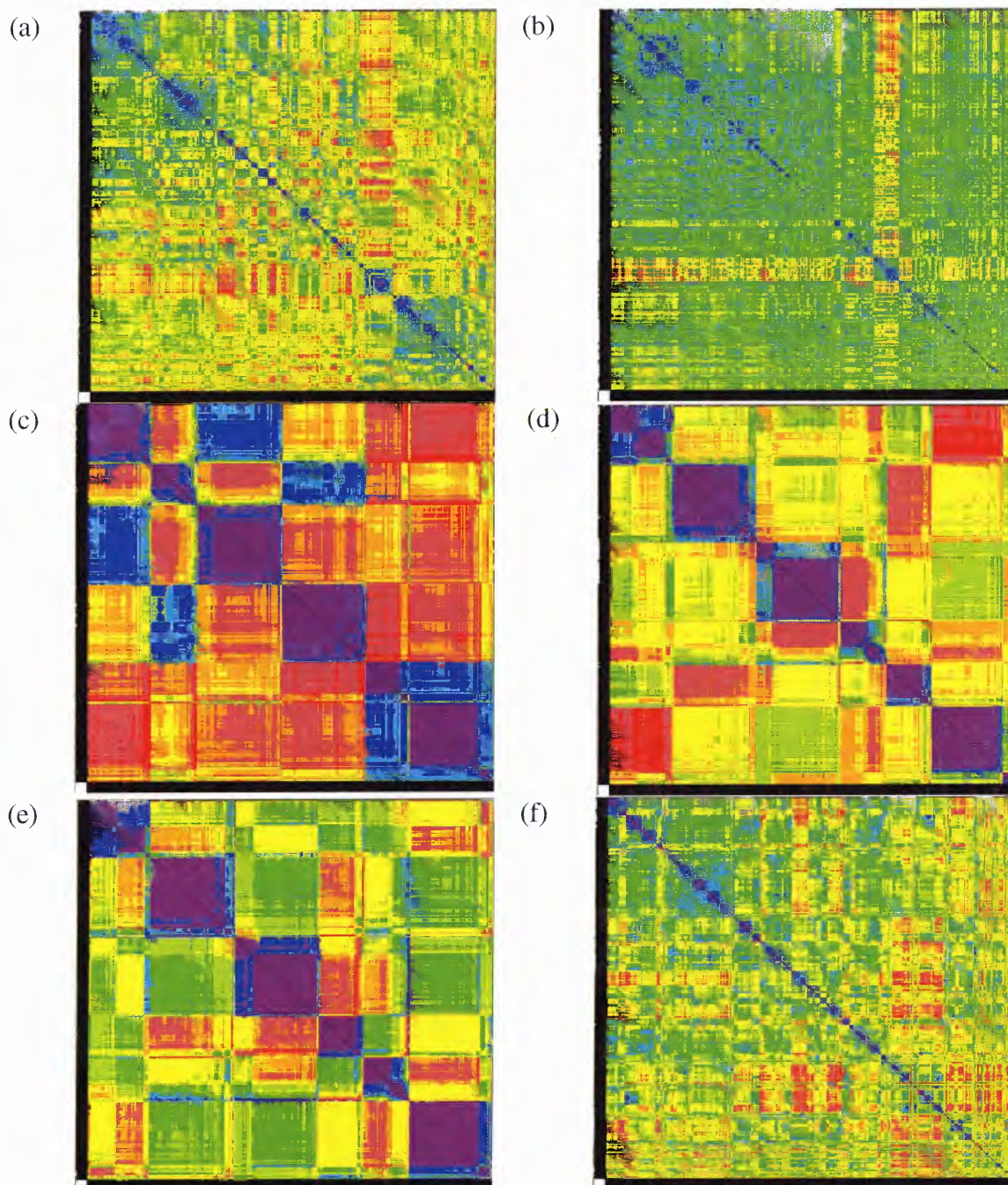
Figure 4.2 Distance maps for analog **2** clustering studies. Black borders on left and bottom sides represent generic-ordered conformers. Distance maps are color-coded by interconformational distances, from black through blue to red. Blue boxes along the diagonal represent clusters of conformers; large boxes contain many conformers, and small boxes contain only a few conformers. Distance maps are lettered by the feature set/alignment options identified in Table 3.2.

Figure 4.2 Distance maps for analog **2** clustering studies. Black borders on left and bottom sides represent generic-ordered conformers. Distance maps are color-coded by interconformational distances, from black through blue to red. Blue boxes along the diagonal represent clusters of conformers; large boxes contain many conformers, and small boxes contain only a few conformers. Distance maps are lettered by the feature set/alignment options identified in Table 3.2. (Continued)

clustering study. The distance maps were reviewed to select those with a small number of large clusters for further analysis. The results of the review of the distance maps are shown in Figure 4.4.

4.2.1.1 All Heavy Atoms Feature Sets. The all heavy atom feature set distance maps are shown in Figures 4.2(a), 4.2(b), 4.3(a), and 4.3(b) and correspond to feature set/alignment options *a* and *b*, as noted previously in Table 3.2. All four distance maps
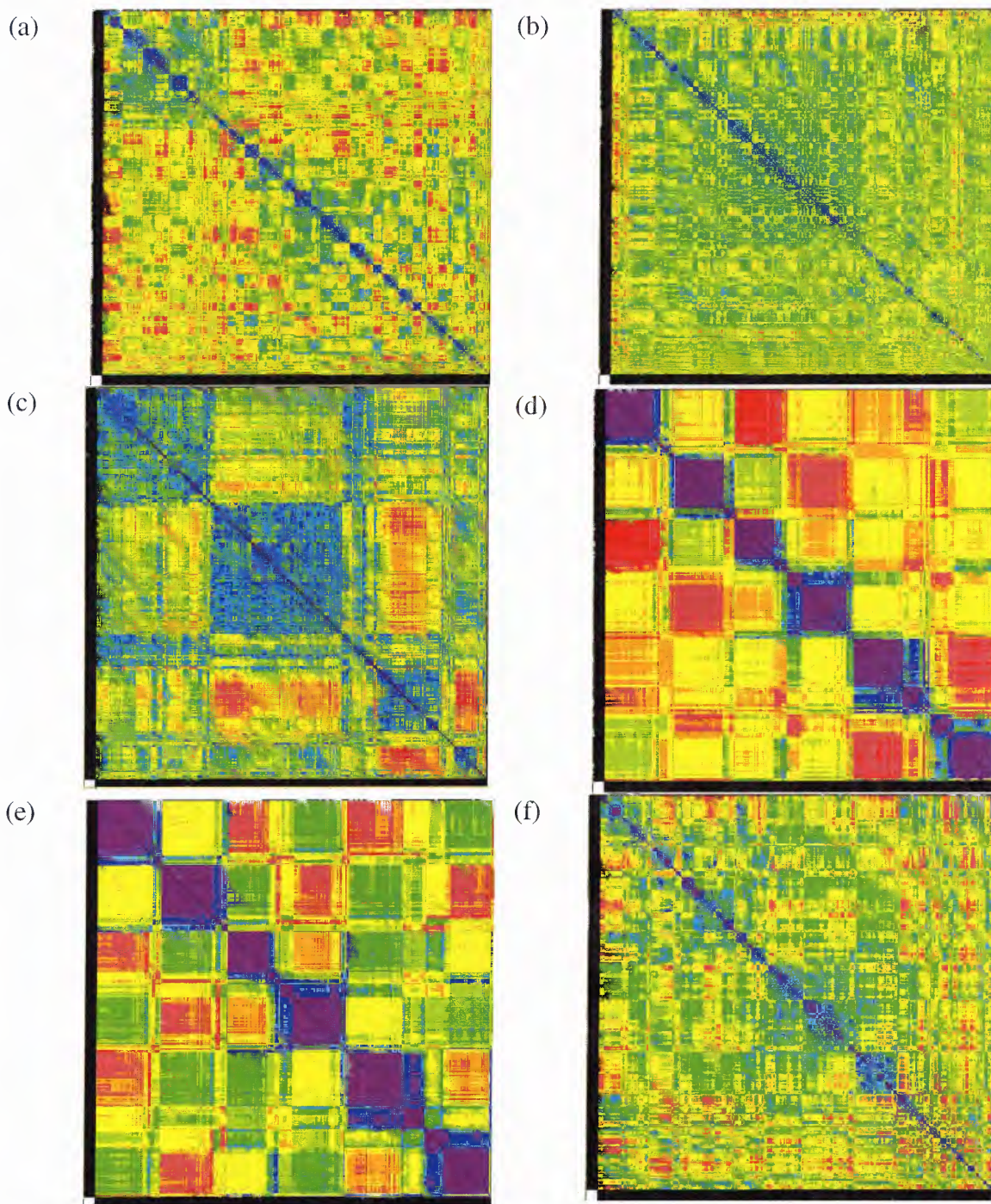
Figure 4.3 Distance maps for analog **3** clustering studies. Black borders on left and bottom sides represent generic-ordered conformers. Distance maps are color-coded by interconformational distances, from black through blue to red. Blue boxes along the diagonal represent clusters of conformers; large boxes contain many conformers, and small boxes contain only a few conformers. Distance maps are lettered by the feature set/alignment options identified in Table 3.2.

Figure 4.3 Distance maps for analog **3** clustering studies. Black borders on left and bottom sides represent generic-ordered conformers. Distance maps are color-coded by interconformational distances, from black through blue to red. Blue boxes along the diagonal represent clusters of conformers; large boxes contain many conformers, and small boxes contain only a few conformers. Distance maps are lettered by the feature set/alignment options identified in Table 3.2. (Continued)

show numerous small squares on the diagonal, indicating that major clusters were not found. The large amount of small squares on the diagonal is indicative of the diversity of the conformations when the analogs are viewed as a whole. These maps and their associated clustering studies were not analyzed further. Instead, attention was focused on clusters defined by A-side and B-side feature set/alignment options.

4.2.1.2 A-Side Feature Sets. The distance maps for the A-side analyses are shown in Figures 4.2(c), 4.2(d), 4.2(e) and 4.3(c), 4.3(d), 4.3(e), and correspond to feature set/alignment options *c*, *d*, and *e* in Table 3.2. Five of the six distance maps show relatively well-defined major clusters on the diagonal, and therefore require further analysis to select the most promising feature set/alignment option. Only the distance map in Figure 4.3(c) shows many small clusters on the diagonal, as well as off-diagonal blue areas representing similarity between large clusters, and was not used for further analysis. Comparison of the other two A-side feature set distance maps (options *d* and *e*) for both analogs shows a strikingly similar clustering pattern for the A-side clustering analyses.

The distance maps selected for analogs **2** (Figures 4.2(c), 4.2(d), and 4.2(e)) and **3** (Figures 4.3(d) and 4.3(e)) show six large clusters on the diagonal. For each map, the clustering level which gave an effective number of clusters closest to six was selected to calculate the percentage of conformers in each major cluster. These results are reported in Table 4.1. The table shows that for analog **2** feature set/alignment options *c* and *e* and analog **3** feature set/alignment option *d*, more than 99% of the conformers are included in the six major clusters. For analog **2** feature set/alignment option *d* and analog **3** feature set/alignment option *e*, these values are 95% and 97%, respectively. The A-side clustering studies result in evenly-populated clusters; the smallest of the major clusters for each feature set/alignment option contains 10% or more of the conformers in each data set. The largest minor clusters (selected from the clusters not included in the six major clusters) contain one-third or fewer conformers compared to the smallest major clusters.
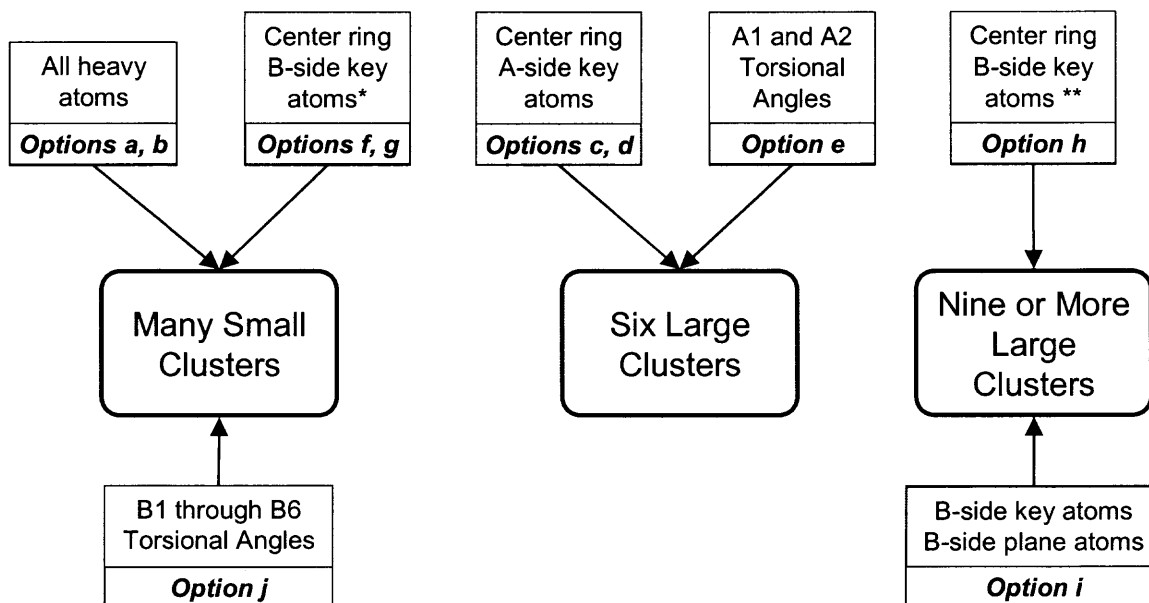
Figure 4.4 Distance map summary for feature set/alignment options found in Table 3.2. All feature set results are for all alignments, except for: * Center ring and center ring/B-side key atoms alignments only; for these alignment options, a few large clusters were present as well, and ** Oxygen and neighboring carbons alignment.

Table 4.1 Percentage of Conformers Included at Six Major Clusters, A-side Feature Sets[a]

Analog (Option)[b]	Clustering Level	Actual No. of Clusters	Effective No. of Clusters	Smallest Major Cluster	Largest Minor Cluster	Total No. NOT in Major Clusters
2(c)	720	9	6.02	10.4%	0.5%	0.8%
2(d)	717	12	6.81	10.4%	2.9%	4.7%
2(e)	719	10	6.02	11.0%	0.3%	0.7%
3(d)	731	9	6.23	13.4%	0.7%	0.9%
3(e)	729	11	6.62	13.3%	1.4%	3.0%

[a] Selected analog **2** and **3** A-side clustering levels.
[b] Feature set/alignment options listed in Table 3.2.

Plots of the effective number of clusters, k^*, versus the actual number of clusters, k, for the selected feature set/alignment options are given in Figures 4.5 (analog **2**) and 4.6 (analog **3**) for the 20 highest clustering levels. The straight line where $k = k^*$ is also plotted for comparison. Again, the results are similar for each of the five feature set/alignment options, although it is interesting to note where the large jumps in the

effective number of clusters occur. Options **2**(*c*), **2**(*e*) and **3**(*d*) are the only clustering studies that separate into six effective large clusters ($k^* = 6$) in the 10 highest clustering levels, with the effective number of clusters closer to the actual number of clusters than the other two A-side studies listed in the table.

These analyses led to the identification of options **2**(*c*), **2**(*e*) and **3**(*d*) as the most promising clustering studies for analogs **2** and **3**. However, because the goal is to use the clustering studies to select representative conformers of **2** and **3**, which are included in the same QSAR studies, selection of the same clustering feature set/alignment option is preferred. As noted above, option **3**(*c*) does not cluster well according to its distance map, so option *c* was not selected. Options **2**(*e*) and **3**(*e*) were selected for further analysis because the percentage of conformers in the six major clusters in option **2**(*e*) is slightly better than that in option **3**(*d*). Similarly, the percentage of conformers in the major clusters for option **3**(*e*) is better than option **2**(*d*) (3.0% not covered versus 4.7% not covered, respectively). Option *e* was used to confirm the optimal number of clusters for A-side clustering, as detailed in the section Review of Clustering Statistics.
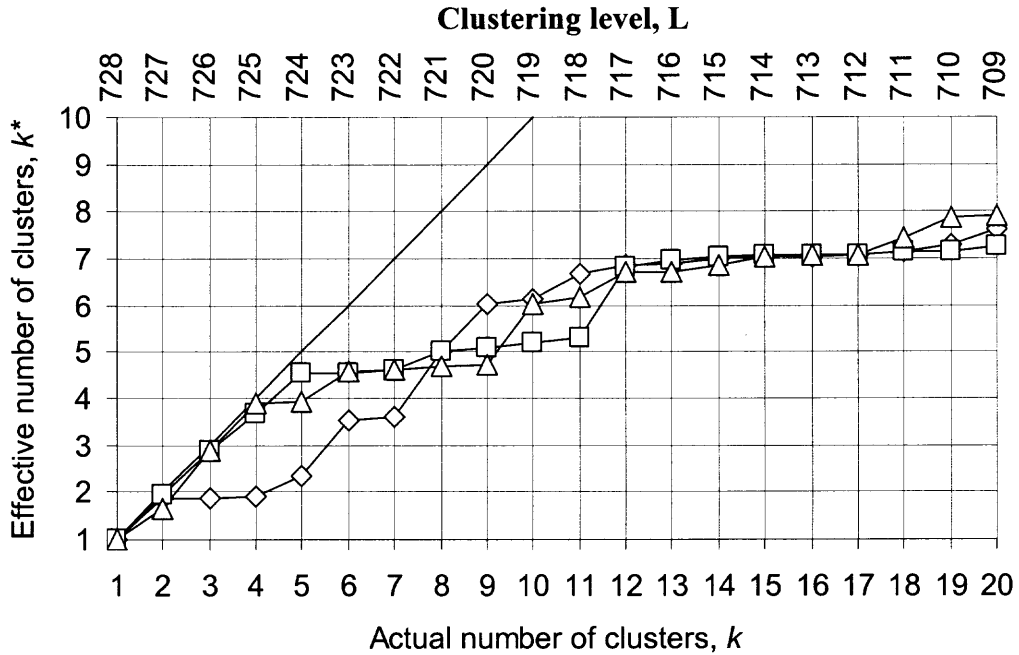
Figure 4.5 A-side effective vs. actual number of clusters for analog **2**. Feature set/alignment options **2**(*c*), **2**(*d*), and **2**(*e*), and theoretical maximum. Clustering level is noted on top x-axis for reference. Open diamond - Option **2**(*c*). Open square - Option **2**(*d*). Open triangle - Option **2**(*e*). Solid line - Theoretical maximum ($k = k^*$).
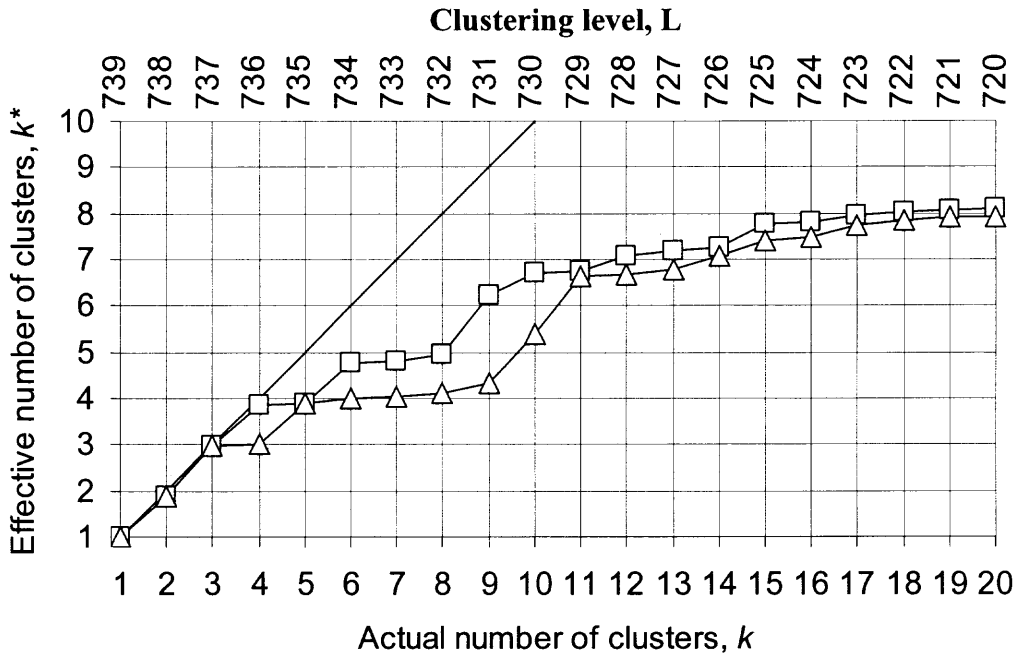
Figure 4.6 A-side effective vs. actual number of clusters for analog **3**. Feature set/alignment options **3**(*d*) and **3**(*e*), and theoretical maximum. Clustering level is noted on top x-axis for reference. Open square - Option **3**(*d*). Open triangle - Option **3**(*e*). Solid line - Theoretical maximum ($k = k^*$).

4.2.1.3 B-Side Feature Sets. The distance maps for the B-side analyses are shown in Figures 4.2(f) – 4.2(j) and 4.3(f) – 4.3(j), and correspond to feature set/alignment options $f - j$ in Table 3.2. For both analogs, options f and j result in many small clusters along the diagonal. Distance maps in Figures 4.2(f), 4.2(g) and 4.2(j) show many small clusters on the diagonal, and wide distribution of blue and green points, indicating distances between separate clusters are about the same as distances within clusters. Distance map 4.2(i) shows well-defined squares on the diagonal, but there are 15 or more squares present, not meeting the criteria of 10 or fewer major clusters. For analog **2**, only the distance map in Figure 4.2(h) shows less than 10 large clusters on the diagonal and was selected for further analysis. For analog **3**, Figures 4.3(h) and 4.3(i) show large clusters on the diagonal, although those in Figure 4.3(h) are less clearly defined. The distance maps for options f, g, and j for analog **3** show many small clusters on the diagonal.

The B-side distance maps show different sizes of large clusters along the diagonal as compared to the A-side distance maps. Nine large clusters can be qualitatively identified when looking for trends in the three distance maps shown in Figures 4.2(h), 4.3(h), and 4.3(i).

In contrast to the A-side distance maps that were comparable for analogs **2** and **3**, the B-side clustering studies for **2** and **3** do not match well for several feature set/alignment options. Of note is the comparison of Figure 4.2(i) and 4.3(i), where a distinctly different pattern is seen. Comparison of this same clustering option for the two analogs indicates that analog **2** may be able to access many more different types of B-side conformations than analog **3**.

Table 4.2 Percentage of Conformers Included at Nine Major Clusters, B-side Feature Sets[a]

Analog (Option)[b]	Clustering Level	Actual No. of Clusters	Effective No. of Clusters	Smallest Major Cluster Size	Largest Minor Cluster	Total No. NOT in Major Clusters
2(h)	719	10	5.88[b]	1.1%	0.1%	0.1%
3(h)	717	23	7.36[b]	2.6%	0.7%	2.7%
3(i)	725	16	7.83[b]	2.6%	0.8%	1.6%

[a] Selected analog **2** and **3** B-side clustering levels.
[b] Feature set/alignment options listed in Table 3.2.
[c] Effective number of clusters is not close to nine because sizes of major clusters are uneven.

The selected analog **2** (Figure 4.2(h)) and **3** (Figures 4.3(h) and 4.3(i)) distance maps all show approximately nine large clusters on the diagonal. The percentage of conformers left out of the nine major clusters is shown for the three B-side clusterings in Table 4.2 and indicates that the major clusters represent less of the entire data set than those of the selected A-side analyses. The selected B-side analysis option **2**(h) has a wider distribution of cluster sizes, as indicated by the lowest effective number of clusters (5.88) compared to the number of major clusters (nine) and by the small size of the smallest major cluster (1.1% of total set of conformations). Both of these attributes are due to a significant size range of the major clusters as seen in Figure 4.2(h). The B-side clustering studies produced a mixture of differently-sized major clusters for both **2** and **3**, although the analog **2** studies produced more variation in B-side cluster size. The smallest of the major clusters for each feature set/alignment option contains 3% or less of the conformers in each data set. The largest minor clusters (clusters not included in the nine major clusters) are not of significant size and follow a trend similar to the A-side studies.

The B-side effective number of clusters versus the actual number of clusters is plotted in Figures 4.7 (analog **2**) and 4.8 (analog **3**) for the 20 highest clustering levels of

the three clustering studies. It can be seen that option 3(*g*) has a minimal increase in effective number of clusters as the actual number of clusters increases, indicating splintering as opposed to formation of large clusters. The patterns for the other B-side clustering studies (option *h* for both analogs) are similar, and are not revealing in terms of where to determine the optimal number of major clusters. Although the review of the distance maps led to an estimate of nine major clusters, Figures 4.7 and 4.8 show that nine major clusters do not occur until *k*, the actual number of clusters, is more than 20. Based on these results, further analysis was not performed on the B-side clustering studies; only clustering option *e*, using A-side features, was selected for analysis of additional clustering statistics for each analog.
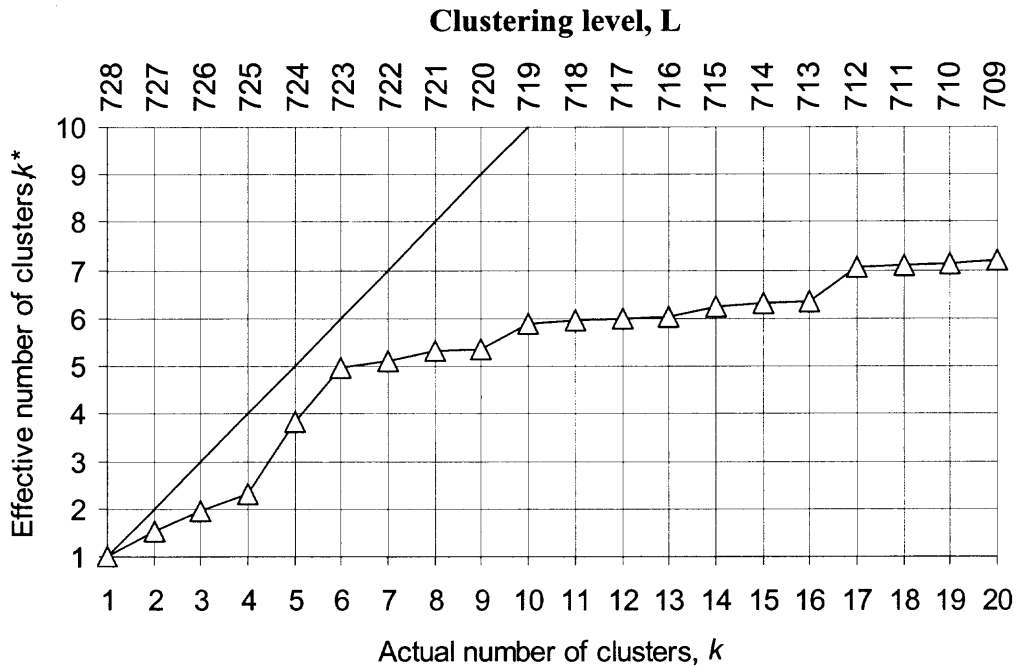
Figure 4.7 B-side effective vs. actual number of clusters for analog **2**. Feature set/alignment option **2**(*h*) and theoretical maximum. Clustering level is noted on top x-axis for reference. Open triangle - Option **2**(*h*). Solid line - Theoretical maximum (*k* = *k**).
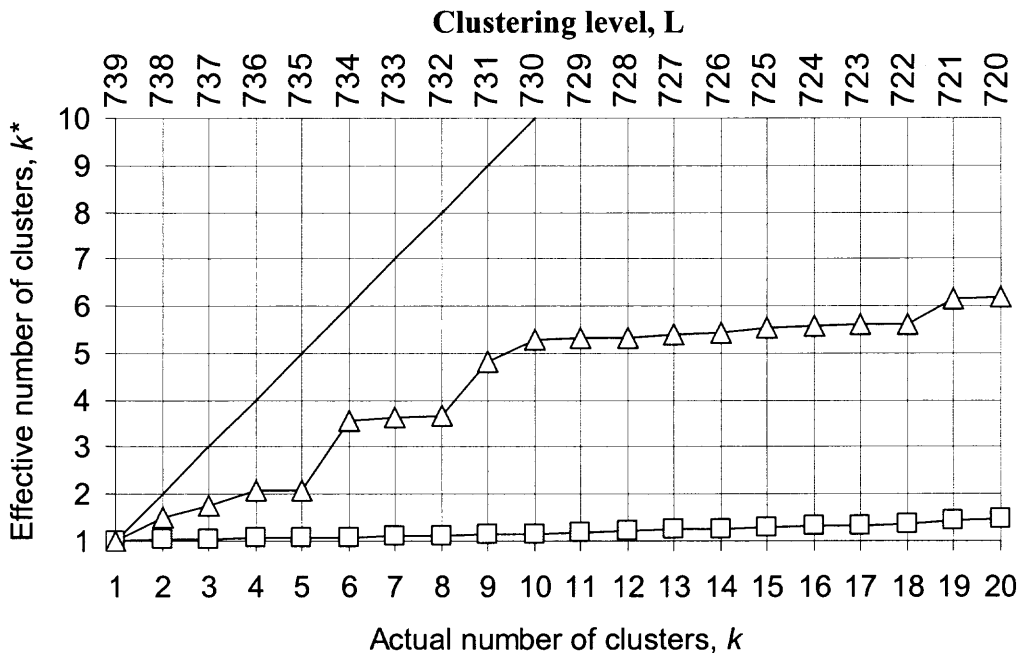
Figure 4.8 B-side effective vs. actual number of clusters for analog **3**. Feature set/alignment options **3**(*g*) and **3**(*h*), and theoretical maximum. Clustering level is noted on top x-axis for reference. Open square - Option **3**(*g*). Open triangle - Option **3**(*h*). Solid line - Theoretical maximum (*k* = *k**).

4.2.2 Review of Clustering Statistics

An analysis of the minimum separation ratio and the percentage change in the effective number of clusters based on Figures 3.5 and 3.6 was used to confirm the optimal number of clusters for options 2(*e*) and 3(*e*). Table 4.3 lists the 20 highest clustering levels in descending order, and their associated data: minimum separation ratio, critical distance, actual and effective number of clusters, and %ΔEff for the clustering levels of option 2(e). The highest minimum separation ratio is 1.23. Table 4.4 lists the same set of data for the 20 highest clustering levels of option 3(e). The analog **3** data shows an even smaller range of minimum separation ratio for high clustering levels, the highest value being 1.17. Both values of minimum separation ratio fall significantly below the recommended value of 2.0^{197} for identification of the optimal clustering level. Therefore, it appears that the minimum separation ratio is not a useful criterion for these clustering studies.

The %ΔEff values were compared for the 20 highest clustering levels as well. This value is a numerical representation of the trends seen in Figures 4.5 (analog **2**) and 4.6 (analog **3**), scaled by the number of actual clusters at the clustering level, as noted in equation 2.10. The %ΔEff is large for clustering levels where a large jump in the effective number of clusters occurs, as long as the actual number of clusters is low.

Following the procedure for selecting the optimal clustering level in Figure 3.6, review of Table 4.3 identified clustering level 719 (approximately six effective clusters, 10 actual clusters, 13% change, minimum separation ratio 1.12) for option 2(*e*). Similarly, review of Table 4.4 identified clustering level 729 (approximately six effective clusters, 11 actual clusters, 11% change, minimum separation ratio 1.06) for option 3(*e*).

Table 4.3 Analog **2**, Option *e* Clustering Statistics

Clustering Level	Minimum Separation Ratio[a]	Critical Distance[b]	Actual No. of Clusters	Effective No. of Clusters[c]	%ΔEff[d]
728	1.00	32.1	1	1.00	-
727	1.04	30.9	2	1.66	33%
726	1.18	26.3	3	2.87	40%
725	1.02	25.8	4	3.88	25%
724	1.05	24.6	5	3.93	1%
723	1.01	24.4	6	4.59	11%
722	1.23	19.8	7	4.63	1%
721	1.05	18.9	8	4.67	1%
720	1.02	18.6	9	4.71	0%
719	**1.12**	**16.7**	**10**	**6.02**	**13%**
718	1.05	15.9	11	6.18	1%
717	1.09	14.5	12	6.71	4%
716	1.01	14.4	13	6.72	0%
715	1.02	14.2	14	6.87	1%
714	1.03	13.7	15	7.05	1%
713	1.15	11.9	16	7.08	0%
712	1.10	10.9	17	7.09	0%
711	1.02	10.7	18	7.42	2%
710	1.01	10.6	19	7.86	2%
709	1.01	10.5	20	7.92	0%

[a] The distance between the most closely related clusters for a clustering level.

[b] The cutoff distance associated with a clustering level; corresponds to the distance between the clusters just formed.

[c] An estimate of how many large clusters are in the set of clusters.

[d] Percentage change in the effective number of clusters.

Table 4.4 Analog **3**, Option *e* Clustering Statistics

Clustering Level	Minimum Separation Ratio[a]	Critical Distance[b]	Actual No. of Clusters	Effective No. of Clusters[c]	%ΔEff[d]
739	1.00	29.0	1	1.00	-
738	1.07	27.3	2	1.89	44%
737	1.07	25.6	3	2.99	37%
736	1.09	23.5	4	3.02	1%
735	1.17	20.1	5	3.88	17%
734	1.02	19.8	6	4.01	2%
733	1.11	17.8	7	4.04	1%
732	1.00	17.8	8	4.11	1%
731	1.08	16.5	9	4.34	3%
730	1.09	15.1	10	5.40	11%
729	**1.06**	**14.2**	**11**	**6.62**	**11%**
728	1.01	14.1	12	6.68	0%
727	1.00	14.1	13	6.77	1%
726	1.04	13.5	14	7.08	2%
725	1.02	13.2	15	7.40	2%
724	1.00	13.2	16	7.46	0%
723	1.07	12.4	17	7.75	2%
722	1.01	12.2	18	7.85	1%
721	1.01	12.1	19	7.92	0%
720	1.01	12.1	20	7.93	0%

[a] The distance between the most closely related clusters for a clustering level.

[b] The cutoff distance associated with a clustering level; corresponds to the distance between the clusters just formed.

[c] An estimate of how many large clusters are in the set of clusters.

[d] Percentage change in the effective number of clusters.

4.2.3 Visualization of Clusters and Representative Conformers

Clustering levels representing six major clusters were chosen based on the selected A-side analyses (both option *e*, A1/A2 torsional angle feature set). For the purposes of the proposed QSAR work, the agreement of the A-side results for both analogs and the small number of similarly-sized A-side clusters led to a focus on the A-side for more detailed analysis. The representative conformers for each cluster were used to develop conformational families of analogs.

Two views of the conformers of **2** in the six major clusters at clustering level 719 of option *e* are shown in Figure 4.9. Structures in the figure are superimposed on the piperazine ring, with only the A-sides shown. The relative orientation of the naphthalene ring in the conformers is defined by torsional angles A1 and A2. The physical significance of torsional angles A1 and A2 is that they determine how the conformers form natural groups on the A-side of the molecule. The A1 torsional angle describes rotation about a $N(sp^3)$-$C(sp^3)$ bond and the A2 torsional angle describes rotation about a $C(sp^3)$-$C(sp^2)$ bond. Therefore, the A1 and A2 torsional angles of the conformers output by the random search technique should be close to the values of torsional angles found in the staggered low-energy conformations of compounds such as aminomethane and methylbenzene, which can be considered to be models for the A1 and A2 torsional angle rotational barriers, respectively. Figures 4.9(a) and 4.9(b) show that the six clusters form regularly around the A1 torsional angle, resulting in three pairs of clusters. The six clusters correspond to conformational energy minima for rotation around the A1 and A2 torsional angles. These patterns can be seen in Figure 4.10 which shows the potential energy surface in (A1, A2) space with conformers color-coded by cluster. The figure

shows the clusters are located about 120° apart on the A1 axis (at approximately -60°, +60°, and ±180°. The location of these clusters corresponds to rotational minima around the $N(sp^3)$-$C(sp^3)$ bond in A1, and is typical of the rotational energy minima in aminomethane. The figure also shows that most of the conformers are found clustered along the A2 axis at approximately A2 = -90° and +90°, while the remaining conformers are spread along the axis at intermediate values of A2. The location of the clusters corresponds to rotational minima around the $C(sp^3)$-$C(sp^2)$ bond in A2. This complex pattern of rotational minima is due to the effect of the large substituent groups on the carbons in the $C(sp^3)$-$C(sp^2)$ bond in A2. Similar results were obtained for option **3**(*e*).

Figure 4.11 shows the six pairs of representative conformers (representing six family templates) found in the clustering study that span the three-dimensional space occupied by analogs **2** and **3**. The attributes of the 12 representative conformers, six each for **2** and **3**, from the cluster studies with feature set/alignment option *e* are shown in Table 4.5. Clusters in Table 4.5 were ordered by increasing A1 and A2 values of the representative conformers, on a scale of -180° to +180°. The energy of each conformer is given relative to the global energy minimum (GEM) structure for each analog. The table shows that the A-sides of the representative conformers for the six clusters are very similar for the two analogs. The table also shows little difference between the relative energies of the representative conformers for analog **2** (7.54 to 9.50 kcal/mol) and analog **3** (3.53 to 5.45 kcal/mol). Neither analog has its GEM as a representative conformer.
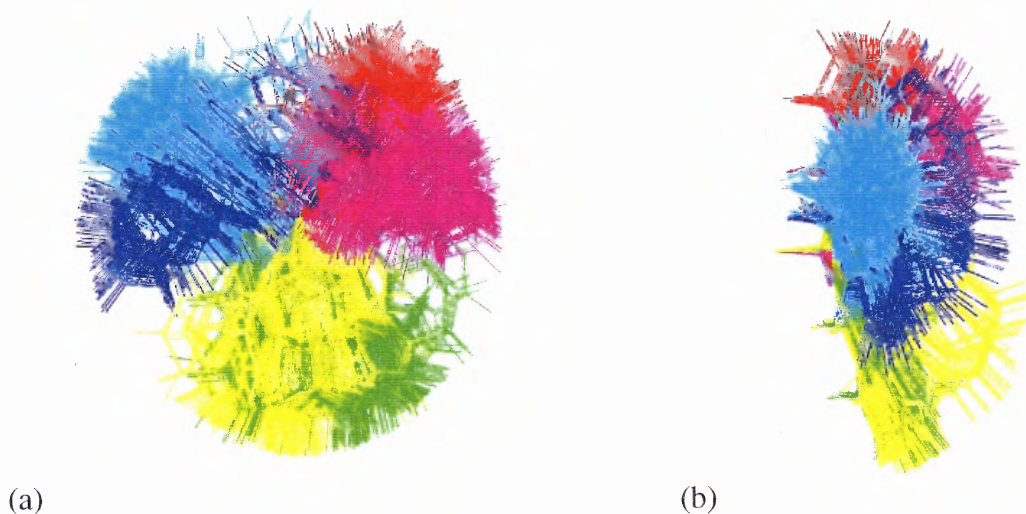
Figure 4.9 Two views of six major A-side clusters for analog **2**. Clusters are color-coded as per Table 4.5. Center ring and B-side are not shown for clarity. (a) Front view. (b) Side view.
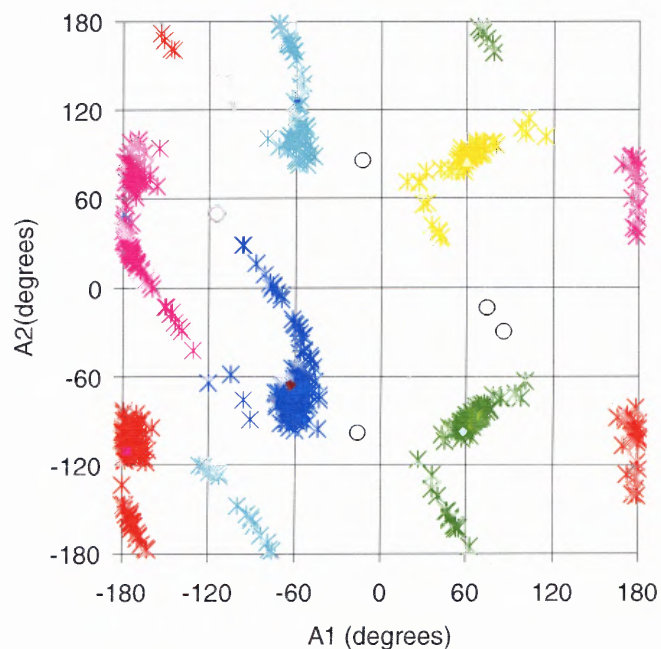
Figure 4.10 Cluster membership A1/A2 plot of the 728 conformers of analog **2**. Conformers are color-coded by their cluster, as noted in Table 4.5, with conformers not in a major cluster colored black.

(a)   (b)
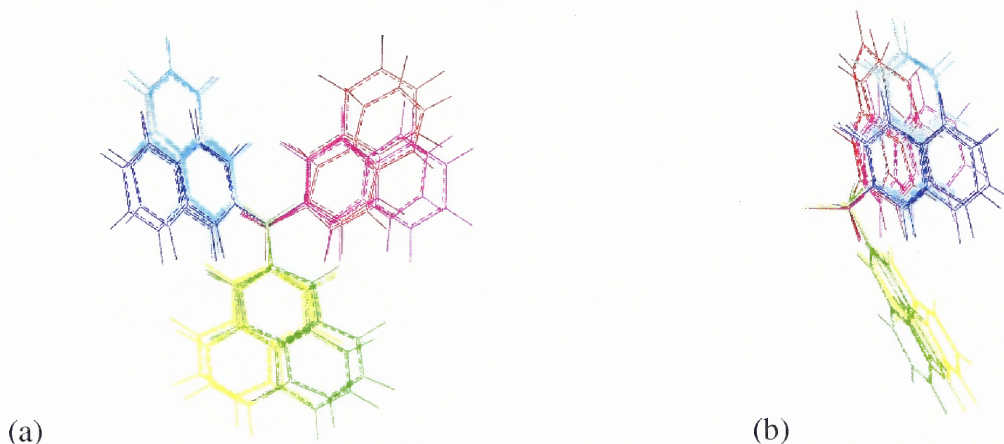
Figure 4.11   Two views of analog **2** and analog **3** representative conformers. Representative conformers are color-coded by their cluster, as noted in Table 4.5. Center ring and B-side are not shown for clarity. (a) Front view. (b) Side view.

Table 4.5 Representative Conformers for Analogs **2** and **3**[a]

Cluster	Analog_ Conformer Number[b]	A1 (degrees)	A2 (degrees)	Rel. Energy[c] (kcal/mol)	Cluster Color in Figures[d]
1	**2**_256	-179.2	-133.0	9.50	red
1	**3**_578	-174.6	-114.7	3.76	red
2	**2**_453	-176.2	72.4	8.16	magenta
2	**3**_653	-171.3	50.2	5.11	magenta
3	**2**_593	-60.7	-55.6	8.19	blue
3	**3**_246	-59.7	-64.8	3.53	blue
4	**2**_546	-62.5	114.8	9.50	cyan
4	**3**_258	-63.5	122.8	5.10	cyan
5	**2**_101	60.1	-96.0	7.54	green
5	**3**_489	65.2	-90.0	5.45	green
6	**2**_573	57.9	84.6	8.34	yellow
6	**3**_254	62.5	85.5	4.04	yellow

[a] Feature set/alignment option *e*.

[b] Conformer number in random search output.

[c] Relative energy is given compared to the global energy minimum of the entire set of conformers for each analog.

[d] Color of cluster, in Figure 4.9; cluster member, in Figure 4.10; and representative conformer pair, in Figure 4.11.

For each cluster, the feature set torsional angles A1 and A2 are similar for analogs **2** and **3**. These two angles control the relationship between the nitrogen atom in the center ring and the aromatic moiety (in these analogs, naphthalene) on the A-side, key characteristics of the majority of DA reuptake inhibitors. The template conformers can be broken down into three groups of two conformers each, based on their A1 values. Each group has A1 values offset by approximately 120°, indicating that they are truly representative of the three pairs of clusters shown in Figure 4.9. In each cluster, the A1 values are within 5° for each pair of representative conformers.

Analysis of the A2 values shows a difference of approximately 180° between the pairs of clusters with similar A1 values (clusters 1 and 2; clusters 3 and 4; clusters 5 and 6), having differences of A2 ranging from 165° to 205°. This indicates "flipping" of the naphthalene ring, by rotation around the $C(sp^3)$-$C(sp^2)$ bond in A2, by approximately 180°. This degree of variability from a perfect flip is attributed to the fact that the adjacent $N(sp^3)$-$C(sp^3)$ bond of the A1 torsional angle has only one large substituent, allowing a range of A2 angles around the $C(sp^3)$-$C(sp^2)$ bond.

The six representative conformers for the A-side clustering option **2**(e) were reviewed to determine the diversity of their B-side structures. The B-side option **2**(h), noted above as most suitable for further analysis, was used to determine the population of nine major clusters, and the locations of each of the six A-side representative conformers were identified. Three of the six representative conformers were found in one B-side cluster, and the other three were found in three different clusters. The three A-side representative conformers that were found in the same B-side cluster are clusters 1, 3, and 5, which all have negative A2 torsional angles.

4.2.4 Comparison of the RMSD of Representative Conformer Pairs

Figure 4.12 shows the 12 representative conformers for the selected clustering levels of the A-side option e superimposed on the center rings. Table 4.6 lists the clusters and the RMSD between the conformers of analogs **2** and **3** in each of the six major A-side clusters. The first RMSD column was calculated based on the same atoms used for feature set/alignment option e. The second RMSD column was calculated based on superposition of all heavy atoms in the molecules. The first RMSD values indicate that the five heavy atoms defining A1 and A2 have very similar locations in space for the representative conformer pairs. This can be seen in Figure 4.11. The increase in the RMSD for all heavy atoms is due to differences in the B-side atoms.

4.2.5 Comparison of the Conformational Space Coverage

Tables 4.7 and 4.8 show the radii of gyration of the conformational point cloud for each data set, along with the six major clusters selected. The minor clusters are included for comparison; their results are shown in italics. Torsional angle A1 and A2 values are only reported for the six major clusters.
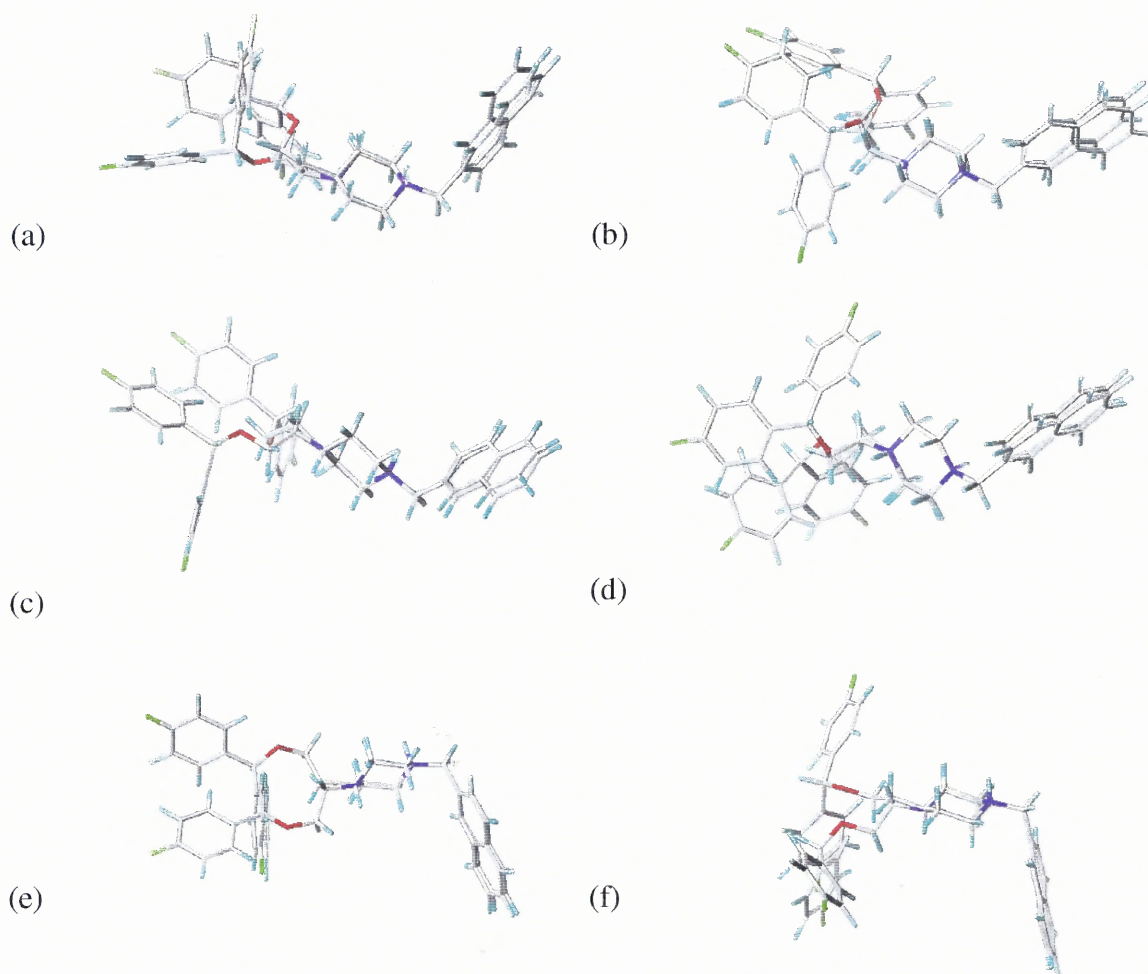
Figure 4.12 Six pairs of representative conformers from clustering study. Representative conformers for Cluster 1 (a) through Cluster 6 (f). Each pair contains a conformer of **2** and a conformer of **3**.

Table 4.6 Comparison of Representative Conformers of Analogs **2** and **3**

Cluster	RMSD of Atoms in Feature Set/Alignment Option e (Å)	RMSD of All Heavy Atoms (Å)
1	0.22	2.37
2	0.31	2.24
3	0.12	1.77
4	0.10	1.95
5	0.13	2.00
6	0.08	1.79

Table 4.7 Conformational Space Coverage for Analog **2**

Cluster	No. of Conformers[a]	Conf. No. of Rep.[b]	r_g[c]	A1 (degrees)	A2 (degrees)
1	111	256	25.6	-179.2	-133.0
2	152	453	20.3	-176.2	72.4
3	82	593	33.7	-60.7	-55.6
4	151	546	20.2	-62.5	114.8
5	80	101	16.0	60.1	-96.0
6	147	573	20.8	57.9	84.6
7	1	375	0.0	-	-
8	1	527	0.0	-	-
9	1	687	0.0	-	-
10	2	94	7.3	-	-
All	728	288	72.1	-	-

[a] Number of conformers in each cluster.
[b] Conformer number of representative conformer.
[c] Radius of gyration in Ångstroms.

Table 4.8 Conformational Space Coverage for Analog **3**

Cluster	No. of Conformers[a]	Conf. No. of Rep.[b]	r_g[c]	A1 (degrees)	A2 (degrees)
1	98	578	31.2	-174.6	-114.7
2	125	653	27.6	-171.3	50.2
3	143	246	22.2	-59.7	-64.8
4	122	258	24.4	-63.5	122.8
5	111	489	14.7	65.2	-90.0
6	121	254	21.7	62.5	85.5
7	10	557	3.7	-	-
8	2	268	2.5	-	-
9	1	589	0.0	-	-
10	5	427	7.3	-	-
11	1	241	0.0	-	-
All	739	522	74.0	-	-

[a] Number of conformers in each cluster.
[b] Conformer number of representative conformer.
[c] Radius of gyration in Ångstroms.

The radii of gyration for all conformers are nearly identical (less than 3% difference) for each analog, indicating that the piperazine and piperidine analogs cover a similar amount of conformational space. The amount of coverage for each cluster is not directly related to the number of conformers in each cluster. For example, analog **2** Cluster 3 has 82 conformers and reaches 33.7 Å, whereas analog **2** Cluster 4 has 151 conformers and reaches only 20.2 Å. This equates to almost twice as many conformers in Cluster 4 covering two-thirds of the conformational space covered by Cluster 3.

In addition to the absolute conformational space coverage, the relative conformational space coverage determined by review of the distance maps and related data is of interest. The A-side distance maps show clusters of approximately the same size, as viewed in the distance maps and by the effective number of clusters being near six when the number of major clusters is six.

4.3 Construction of Conformational Families

The representative conformers in Table 4.5 were used as templates to construct conformational families of analogs. For example, Family 1 consists of **2**, in conformation 256 (of the **2** random search results) and the other 22 piperazines in a minimized conformation that was derived from conformation 256, as well as **3**, in conformation number 578 (of the **3** random search results) and the other 20 piperidines in a minimized conformation that was derived from conformation 578. Members of each conformational family other than **2** and **3** themselves were minimized from the template conformations to avoid singularities and reach their local energy minima. Conformational families are therefore distinguished from clusters as being developed from the representative

conformer of each cluster; clusters contain only conformations of **2** or **3**, and the conformational families each contain all 45 analogs in the 3D-QSAR study.

4.4 QSAR Studies

Six analogs were selected as a test set, three piperazines (**5**, **8**, and **13**), and three piperidines (**29**, **32**, and **35**), the remaining 39 analogs forming the training set. The substituents varied; three substituted phenyl substituents, a benzothiophenyl substituent, a cyclopentanonyl substituent, and a quinolinyl substituent. These substituents are similar to those of the analogs in the training set. Figure 4.13 shows the 45 analogs, including both the training set and test set, created from the templates for Families 1 through 6.

4.4.1 QSAR on Training Set

The results of the preliminary set of QSAR studies are summarized in Table 4.9. Under "Best Results" for the LOO/CV models, the table reports the pair of steric and electrostatic cutoffs (of the nine combinations tested for each family) that gave the highest q^2. For comparison purposes, the "CoMFA Default" column shows that generally similar, but lower in all cases, values of q^2 were obtained by using SYBYL's steric/electrostatic default values of 30/30. The default results for Family 4 were most different from the highest q^2 results, with the CoMSIA steric/electrostatic field option yielding a q^2 value 0.078 higher than the q^2 value for the CoMFA 30/30 setting. Although all the "best" and "default" models found in the preliminary QSAR studies have values of q^2 below the generally-accepted criterion of 0.5 for a predictive CoMFA or CoMSIA model,[219] all are above the statistical 95% confidence limit of $q^2 = 0.3$.[220] SEP
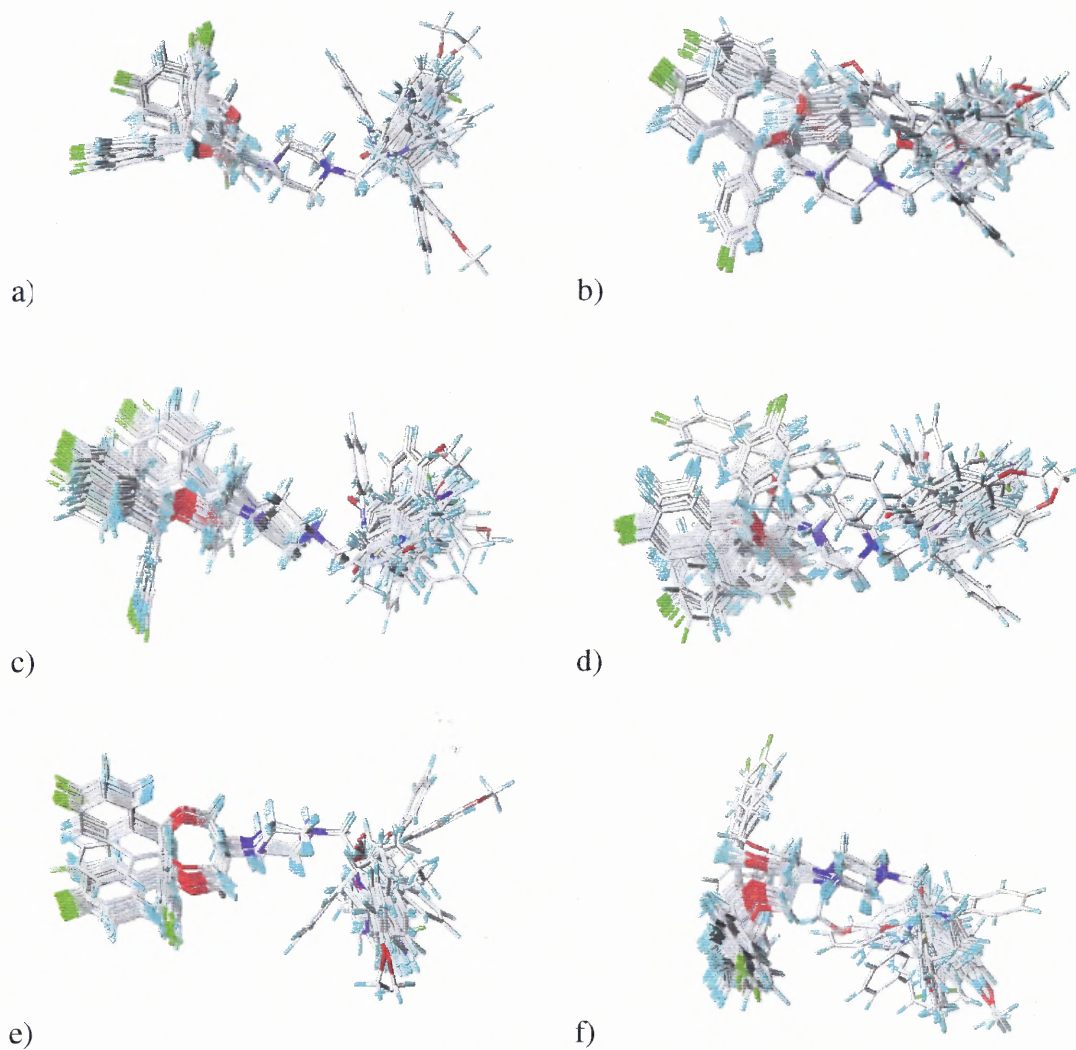
Figure 4.13 Conformational families, showing alignment of all analogs. (a) Family 1. (b) Family 2. (c) Family 3. (d) Family 4. (e) Family 5. (f) Family 6.

Table 4.9 GBR 12909 Analog QSAR Study - Preliminary Results

F^d	S/ES or COMSIA^e	LOO/CV PLS Best Results^a			LOO/CV PLS CoMFA Default^b			NCV PLS (Full Models)^c				
		q^2	C^f	SEP^g	q^2	C^f	SEP^g	r^2	SEE^h	F value^i	% S^j	% ES^k
1	50/10	0.323	2	0.705	0.287	1	0.714	0.680	0.485	38	74%	26%
2	HP	0.345	3	0.729	0.283	2	0.752	0.711	0.484	29	N/A	N/A
3	10/10	0.320	5	0.738	0.290	6	0.766	0.939	0.221	102	78%	22%
4	SE	0.344	5	0.725	0.266	5	0.767	0.901	0.282	60	45%	56%
5	10/10	0.374	2	0.678	0.324	2	0.705	0.739	0.438	51	82%	18%
6	30/10	0.388	2	0.671	0.368	2	0.682	0.686	0.480	39	79%	21%

^{a} Leave-one-out cross-validated models' highest q^2 found when changing steric and electrostatic cutoffs as noted in the Methods section.
^{b} Results for leave-one-out cross-validated model CoMFA run at default steric and electrostatic cutoffs of 30 kcal/mol each.
^{c} Non-cross-validated full models calculated using the optimal number of components noted for the "Best Results" section.
^{d} Family of analogs developed from representative conformers.
^{e} Steric/electrostatic cutoff parameters (each in kcal/mol) for highest q^2 value, or CoMSIA: abbreviation of CoMSIA field: HP = hydrophobic, SE = steric/electrostatic; other field did not produce any highest q^2 values.
^{f} Optimal number of components calculated from leave-one-out cross-validated model.
^{g} Standard error of prediction for best q^2 value.
^{h} Standard error of estimate for best r^2 value.
^{i} 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.
^{j} Percentage steric field contribution to full model. N/A = not applicable for hydrophobic field.
^{k} Percentage electrostatic field contribution to full model. N/A = not applicable for hydrophobic field.

values for the best models for each family are close (between 0.671 and 0.738). The NCV PLS (full model results based on the optimal number of components from the "Best Results") goodness-of-fit (r^2) values cover a wider range than the q^2 values, ranging from 0.680 to 0.939. The four families with CoMFA (rather than CoMSIA) QSAR models identified as the best show a predominance of the steric component, with 74% to 82% steric contribution. In contrast, Family 4, which has a steric/electrostatic (SE) CoMSIA model for the most predictive model, shows predominance of the electrostatic component

(55.5% electrostatic contribution versus 44.5% steric contribution). The hydrophobic field, which gave the best model for Family 2, does not provide a similar steric/electrostatic distribution.

The results of refining the best preliminary model for each family by Region Focusing are given in Table 4.10. For most conformational families, setting the column filtering (σ) value to 2 kcal/mol gave similar values of q^2 compared to the SAMPLS results ($\sigma=0$), with almost no difference (0.002) for Family 2 and 0.113 difference for Family 3, the filtered result having a lower q^2 value. Compared to the preliminary NCV PLS models of Table 4.9, 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 16% and 14% decreases in the steric contributions, respectively. All SAMPLS q^2 values increase in the focused models, with substantial increases seen for Families 1, 5 and 6. These three families meet the criteria of q^2 values more than 0.5. Acceptable r^2 values are seen for most families; however, the r^2 values for the Family 1 and Family 6 two-component models are just below 0.7. The r^2 values are approximately the same for the focused versus the preliminary models for all families. However, the r^2 values for the six-component models of Family 1 and Family 6 are significantly higher than those for the two-component models identified in the preliminary studies (two models are listed for Family 1 and Family 6 due to the instability of the six-component models as detailed below).

Table 4.10 GBR 12909 Analog QSAR Study - Focused Results

F^d	LOO/CV PLS Models, SAMPLS[a]			LOO/CV PLS Models, $\sigma = 2$[b]			NCV PLS (Full Models)[c]				
	q^2	C^e	SEP^f	q^2	C^e	SEP^f	r^2	SEE^g	F value[h]	% S^i	% ES^j
1	0.512	6	0.635	0.565	2	0.565	0.967	0.166	155	53%	47%
1 [k]	0.511	2	0.599	-	-	-	0.685	0.481	39	55%	45%
2	0.452	4	0.676	0.454	4	0.675	0.742	0.464	24	N/A	N/A
3	0.440	5	0.670	0.327	6	0.746	0.946	0.208	116	71%	29%
4	0.364	6	0.725	0.371	6	0.721	0.902	0.284	49	41%	59%
5	0.502	2	0.605	0.483	2	0.617	0.724	0.450	47	68%	32%
6	0.528	6	0.625	0.441	2	0.625	0.967	0.166	155	66%	34%
6 [k]	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] Column filtering (σ) set to 2 kcal/mol.

[c] No column filtering applied in full models.

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

[e] Optimal number of components calculated from focused leave-one-out cross-validated model.

[f] Standard error of prediction associated with the q^2 value.

[g] Standard error of estimate associated with the r^2 value.

[h] 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.

[i] Percentage steric field contribution to full model.

[j] Percentage electrostatic field contribution to full model.

[k] The six-component models were unstable for these families (Table 4.11). Therefore, the maximum number of components was reduced to five (from the default of six) and the QSAR studies were rerun, resulting in two-component models.

Table 4.11 Progressive Scrambling Results for GBR 12909 Analogs

F^a	C^b	Critical Point^c	$Q^{2\ d}$	cSDEP^e	$dq^{2\prime}/dr_{yy\prime}^2$ ^{f}
1	6	0.85	0.154	0.832	2.081
1	2	0.85	0.405	0.660	0.645
2	4	0.85	0.346	0.737	0.504
3	5	0.85	0.252	0.771	1.336
4	6	0.85	-1.454	1.555	37.007
5	2	0.85	0.386	0.671	0.730
6	6	0.85	0.266	0.774	1.628
6	2	0.85	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} Value of $r_{yy\prime}^2$ where the instantaneous slope $dq^{\prime 2}/dr_{yy\prime}^2$ is calculated.
^{d} Equals $(1 - (sSDEP)^2)$. Predictivity of the model using the scaled standard deviation of error of prediction (sSDEP) instead of the SDEP.
^{e} Calculated cross-validated standard deviation of error of prediction.
^{f} 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 $r_{yy\prime}^2$).

The stabilities of the QSAR models were estimated using SYBYL's Progressive Scrambling y-value scrambling procedure. Table 4.11 shows the results of 100 y-value scramblings on the eight different QSAR models listed in Table 4.10. The scrambling results for the six-component Family 1 and Family 6 models are poor because the values of $dq^{2\prime}/dr_{yy\prime}^2$ are greater than 1.2.[210] The maximum number of components was reduced to five (instead of the default six) and the QSAR studies were rerun. The LOO/CV SAMPLS runs for both families indicated that two components are optimal at this setting. The two-component Family 1 and Family 6 CoMFA models have q^2 values of 0.511 and 0.508, respectively (Table 4.10), and both have $dq^{2\prime}/dr_{yy\prime}^2$ values indicating stability.

Q^2 values were lower than the q^2 values in all cases, ranging from -1.454 to 0.405, but these are known to be more conservative than LOO/CV PLS q^2 values.[210] The results for Family 4 indicate a significant amount of instability. Calculated cross-validated standard deviation of error of prediction (cSDEP) values were similar to SEP values for

all but the six-component Family 1 model and the Family 4 model. Once again these values are lower than their LOO/CV counterparts. These results indicate that four of the eight listed models are expected to be stable and acceptably resistant to perturbation by random noise. The dual criteria of q^2 of 0.5 (or above) and stability to random noise were used to select three of these models (the two-component models for Families 1, 5, and 6) for detailed review, the fourth stable model (Family 2) having a q^2 value less than 0.5.

Training set DAT/SERT selectivities from the stable focused NCV PLS model results for Families 1, 5 and 6 are shown in Figure 4.14. The r^2 values of approximately 0.7 appear as deviations of data points from the ideal diagonal line in all three plots. Average residuals for the training sets are 0.00 for all three models; the averages of the absolute value of the errors are 0.35, 0.33, and 0.34 for the Family 1, 5, and 6 models, respectively. A list of the training set predictions for these models is included in Appendix C. Analog **21**, the cyclohexenophenyl piperazine analog, was predicted the poorest by Family 1 (residual = -1.13), and analog **20** was also predicted poorly (residual = 1.06). Analog **20** was predicted the poorest by both Family 5 and Family 6, with similar residuals (1.10 and 1.20, respectively). Both analogs **20** and **21** have fused rings, with a nonaromatic ring proximal to the A-side nitrogen, and an aromatic ring distal to the A-side nitrogen. Predictions for the similar piperazine analogs **16, 17, 18, 19,** and **22,** and the similar piperidine analogs **43** and **44,** were reviewed to determine if these compounds were predicted poorly as well. The range of experimental DAT/SERT selectivities for the seven piperazines is -0.87 to 1.61, or 2.48 log units. Analogs **19** (residuals = -0.76, -0.69, and -0.75, for Families 1, 5, and 6, respectively) and **22** (residuals = 0.11, 0.83, 1.02, for Families 1, 5, and 6, respectively) are also predicted

poorly for most families. However, analogs **16** and **18** are predicted well (residuals = -0.27, -0.45; -0.24, 0.08; and 0.40, and 0.11, for Families 1, 5, and 6, respectively), although one is a ketone and the other is an alcohol. Analog **17**, which is the same as analog **16** with a distal methoxy group, is predicted somewhat poorly (residuals = 0.61, 0.37, and 0.61, for Families 1, 5, and 6, respectively). The piperidine analog **43** is predicted somewhat poorly (residuals = 0.94, 0.45, and 0.56, for Families 1, 5, and 6, respectively), but its stereoisomer analog **44** is predicted well for two families (residuals = 0.56, 0.20, and 0.14, for Families 1, 5, and 6, respectively). Overall, these three CoMFA models appear to predict the nonaromatic/aromatic fused ring analogs relatively poorly. The trend of piperazines being more poorly predicted than the piperidines seems to be due to the higher percentage of these nonaromatic/aromatic fused ring substituents in the piperazine analog set.
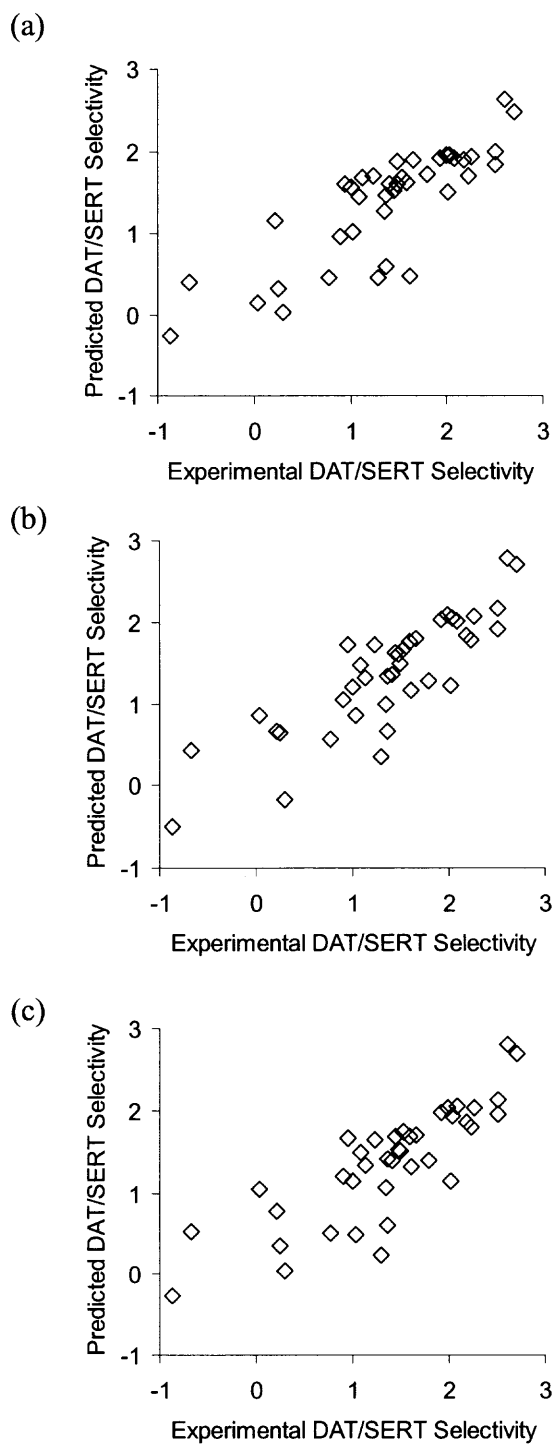
Figure 4.14 GBR 12909 analog QSAR study - training set predictions. (a) Family 1, two components. (b) Family 5, two components, (c) Family 6, two components.

Table 4.12 Original Test Set Predictions for GBR 12909 Analogs

F[a]	Molecule	Exp. Select.[b]	Pred. Select.[c]	Residual[d]
1	5	2.08	1.83	-0.25
	8	*-0.04*	*1.56*	*1.60*
	13	1.06	0.82	-0.24
	29	*0.60*	*1.72*	*1.13*
	32	1.62	1.68	0.06
	35	2.30	1.94	-0.35
5	5	2.08	1.59	-0.49
	8	*-0.04*	*1.78*	*1.82*
	13	1.06	0.66	-0.41
	29	*0.60*	*1.58*	*0.98*
	32	1.62	1.80	0.18
	35	2.30	2.05	-0.24
6	5	2.08	1.56	-0.52
	8	*-0.04*	*1.52*	*1.56*
	13	1.06	0.67	-0.39
	29	*0.60*	*1.57*	*0.98*
	32	1.62	1.73	0.11
	35	2.30	2.05	-0.25

[a] Conformational family of analogs developed from representative conformers.
[b] DAT/SERT selectivity defined in Equation (2.5).
[c] The DAT/SERT Selectivity predicted by the referenced family's best QSAR model.
[d] Equals (Pred. Select. - Exp. Select). Residual is calculated from the exact values, and therefore may vary slightly from the apparent difference of the rounded-off values in columns 3 and 4.

4.4.2 Prediction of Test Set DAT/SERT Selectivities

The three selected QSAR models were used to predict the DAT/SERT selectivities of six test set compounds (the piperazines **5**, **8**, and **13**, and the piperidines **29**, **32**, and **35**). Table 4.12 lists the predicted DAT/SERT selectivities and the associated residuals for these test set compounds. 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. However, the piperazine **8** is a significant outlier for all three families, predicted to have a much higher DAT/SERT selectivity than the experimental DAT/SERT selectivity for all three models. When this single compound is removed, reducing the test set from six to five compounds, for Family 1, the average residual drops

to 0.07; for Family 5, the average residual is reduced to 0.00; and for Family 6, the residual become -0.02. The reduction in error of prediction is less noticeable when considering the absolute values of the errors, with Family 1's average decreasing from 0.61 to 0.41; Family 5's average decreasing from 0.69 to 0.46; and Family 6's average decreasing from 0.63 to 0.45. Except for the one outlier, the piperazines were under-predicted by all three models. Two of the piperidines, analogs **32** and **35**, were predicted well for all three families. The third piperidine, analog **29**, was predicted poorly possibly because it is the only 2-substituted phenyl analog in the series, and is therefore somewhat dissimilar from the training set.

Interestingly, analog **8** was mentioned in a previous paper on **1** analogs because its experimental DAT/SERT selectivity (-0.04) was inconsistent with the DAT/SERT selectivities of similar compounds.[17] Analogs **9** and **10**, also piperazines, have quinolinyl A-side substituents with different connectivities. Experimental results show **9** has a better than average DAT/SERT selectivity (1.48) and **10** a somewhat lower but acceptable value (0.95). The DAT/SERT selectivity of **9**, the 3-quinolinyl analog, was predicted well, with residuals of 0.12, 0.15, and 0.09 for Families 1, 5, and 6, respectively. Analog **10**'s DAT/SERT selectivity had higher residuals of 0.66 for Family 1, 0.77 for Family 5 and 0.71 for Family 6.

The test set predictions were also used to validate the models, according the methods outlined by Golbraikh and Tropsha.[212] The outlying data for **8** was removed

Table 4.13 Test Set Correlation Validation for GBR 12909 Analogs[a]

F^b	n^c	R^d	$R^{2\,e}$	$R_0^{2\,f}$	k^g	$R'_0^{2\,h}$	$k'^{\,i}$	Max. of $(R_0^2$ or $R'_0{}^2)^j$
1	5	0.514	0.264	-0.848	0.939	0.256	0.950	0.256
1^k	7	0.680	0.462	-0.073	0.923	0.461	1.034	0.461
5	5	0.551	0.304	-0.224	0.917	0.269	0.975	0.269
5^k	7	0.819	0.671	0.620	0.895	0.655	1.085	0.655
6	5	0.550	0.302	-0.261	0.904	0.271	0.989	0.271
6^k	7	0.864	0.746	0.735	0.861	0.722	1.134	0.735

[a] Results for both reduced original (n=5, analogs **5**, **13**, **29**, **32**, and **35**) and extended (n=7, compound **1** and analogs **5**, **13**, **32**, **35**, **47**, and **48**) test sets are included. Validation method derived from Golbraikh and Tropsha.[212]

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

[c] Number of compounds in test set.

[d] Correlation coefficient for best-fit line of predicted DAT/SERT selectivity values for the reduced test set of five compounds.

[e] Coefficient of determination for the best-fit line of predicted DAT/SERT selectivity values for reduced test set of five compounds.

[f] Coefficient of determination for the zero-intercept line of predicted DAT/SERT selectivity values for test set, for the plot of Exp. DAT/SERT selectivity (y) vs. Pred. DAT/SERT selectivity (x).

[g] Slope of the zero-intercept line of predicted DAT/SERT selectivity values for test set, for the plot of Exp. DAT/SERT selectivity (y) vs. Pred. DAT/SERT selectivity (x).

[h] Coefficient of determination for the zero-intercept line of predicted DAT/SERT selectivity values for test set, for the plot of Pred. DAT/SERT selectivity (y) vs. Exp. DAT/SERT selectivity (x).

[i] Slope of the zero-intercept line of predicted DAT/SERT selectivity values for test set, for the plot of Pred. DAT/SERT selectivity (y) vs. Exp. DAT/SERT selectivity (x).

[j] Maximum of the coefficients of determination for the zero-intercept lines of the two plots.
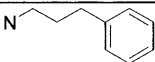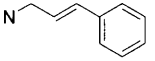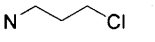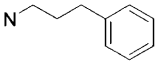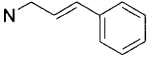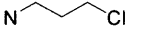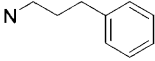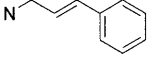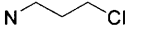
[k] Results for extended test set of seven compounds.

from the original test set for these calculations, since its residual is much larger than the residuals for the other compounds. Table 4.13 includes the test set correlation validation results for the reduced original test set of five compounds, and Figure 4.15 shows the plots of the predicted DAT/SERT selectivities versus the experimental DAT/SERT selectivities for the reduced original test set of five compounds.

(a) $y = 0.324x + 1.103$, $R^2 = 0.264$; $y = 0.939x$, $R^2 = -0.848$

(b) $y = 0.813x + 0.231$, $R^2 = 0.264$; $y = 0.950x$, $R^2 = 0.256$

(c) $y = 0.413x + 0.902$, $R^2 = 0.304$; $y = 0.917x$, $R^2 = -0.224$

(d) $y = 0.736x + 0.402$, $R^2 = 0.304$; $y = 0.975x$, $R^2 = 0.269$

(e) $y = 0.400x + 0.904$, $R^2 = 0.302$; $y = 0.904x$, $R^2 = -0.261$

(f) $y = 0.756x + 0.385$, $R^2 = 0.302$; $y = 0.989x$, $R^2 = 0.271$
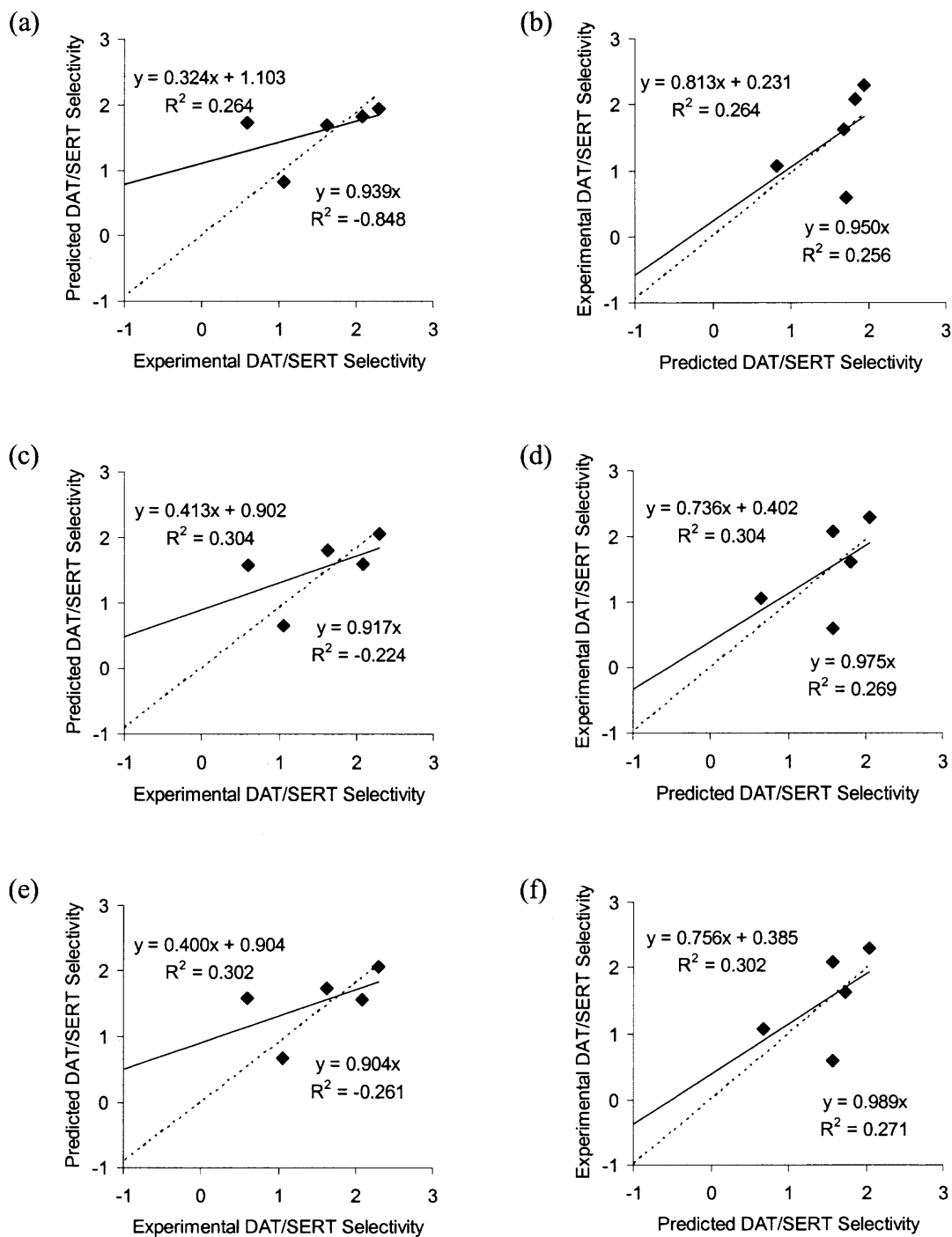
Figure 4.15 GBR 12909 analog QSAR study - reduced original test set predictions. Best-fit (yielding R^2) and zero-intercept (yielding R_0^2 and R'_0^2 in each pair of graphs, respectively) lines shown. (a) and (b) Family 1, two components. (c) and (d) Family 5, two components. (e) and (f) Family 6, two components.

Table 4.14 Predictions for Additional Extended Test Set Compounds

F^a	Molecule	R^b	Exp. Select.^{c}	Pred. Select.^{d}	Residual^e
1	1		1.54	1.70	0.16
	47		2.18	1.71	-0.47
	48		0.91	1.61	0.70
5	1		1.54	1.66	0.12
	47		2.18	1.77	-0.41
	48		0.91	1.14	0.23
6	1		1.54	1.61	0.07
	47		2.18	1.65	-0.53
	48		0.91	0.90	0.01

^{a} Conformational family of analogs developed from representative conformers.
^{b} A-side substituent, as noted in Figure 1.2. All three compounds are piperazines.
^{c} DAT/SERT selectivity defined in Equation (2.5).
^{d} The DAT/SERT Selectivity predicted by the referenced family's best QSAR model.
^{e} Equals (Pred. Select. - Exp. Select.) 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.

The reduced original test set coefficient of determination values for the best-fit lines (R^2) are similar. The R^2 values for the reduced original test set predictions are less than half the goodness-of-fit found for the training set (r^2), chiefly because of large residuals for analog **29**. The Golbraikh and Tropsha paper recommends having a test set of at least five compounds, so an extended test set of seven compounds was developed using three additional compounds (**1**, **47**, and **48**), leaving out analog **29** as an additional outlier. The structures and the predicted DAT/SERT selectivities for these three additional compounds, which are more flexible (have more rotatable bonds on the A-side) than the other compounds in the QSAR study, are shown in Table 4.14. The test set validation

(a) y = 0.455x + 0.854, R^2 = 0.462; y = 0.923x, R^2 = -0.073

(b) y = 1.015x + 0.032, R^2 = 0.462; y = 1.034x, R^2 = 0.461

(c) y = 0.708x + 0.341, R^2 = 0.671; y = 0.895x, R^2 = 0.620

(d) y = 0.947x + 0.228, R^2 = 0.671; y = 1.085x, R^2 = 0.655

(e) y = 0.772x + 0.163, R^2 = 0.746; y = 0.861x, R^2 = 0.735

(f) y = 0.967x + 0.266, R^2 = 0.746; y = 1.134x, R^2 = 0.722
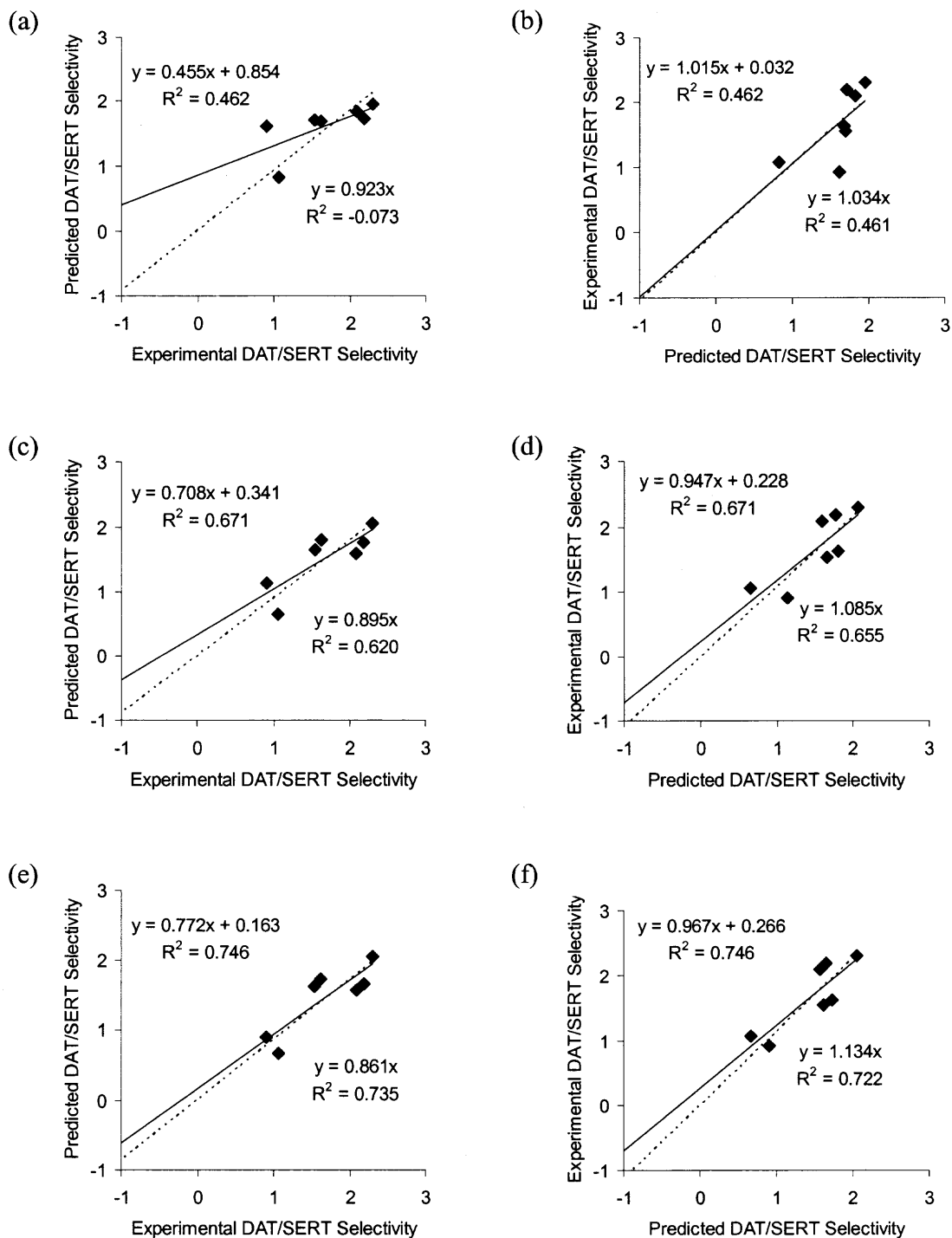
Figure 4.16 GBR 12909 analog QSAR study - extended test set predictions. Best-fit (yielding R^2) and zero-intercept (yielding R_0^2 and $R'_0{}^2$ in each pair of graphs, respectively) lines shown. (a) and (b) Family 5, two components. (c) and (d) Family 6, two components.

correlation values for the extended test set (n=7), listed in Table 4.13 and plotted in Figure 4.16, are higher than those for the reduced original test set (n=5). All three coefficients of determination for the Family 6 model are above 0.7 (0.746, 0.735, and 0.722), slightly more than the goodness-of-fit (r^2) for prediction of the training set compounds (0.685). The improved correlation of the extended test set results compared to the reduced original test set results can also be seen by comparing Figures 4.15 and 4.16. Details of the test set predictions are found in Appendix D. This method of CoMFA model validation indicates that the Family 6 model can predict test set compounds satisfactorily, although the presence of two outliers among nine candidates for the test set raises questions about the general applicability of the Family 6 model. The results of this study show that models with acceptable q^2 values of 0.5 or above do not necessarily predict activities well for all related compounds.

4.4.3 Prediction of Novel Compounds

Figure 4.17 shows the CoMFA steric/electrostatic contour maps for Family 6, with analog **3** shown as a reference molecule. Green areas indicate where an increase in bulk would lead to a higher DAT/SERT selectivity; yellow areas indicate where a decrease in bulk would lead to a higher DAT/SERT selectivity. Blue areas indicate where an increase in positive charge would lead to a higher DAT/SERT selectivity; red areas indicate where an increase in negative charge would lead to a higher DAT/SERT selectivity.
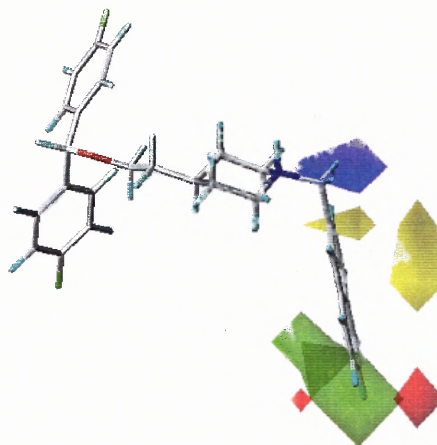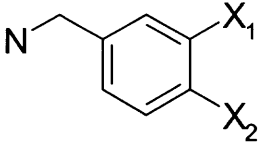
Figure 4.17 CoMFA steric/electrostatic contour map for Family 6 model.

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 DAT/SERT selectivity. However, the multiple small areas where less bulk leads to a higher DAT/SERT selectivity restrict opportunities for increasing DAT/SERT selectivity 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 DAT/SERT selectivities of analogs **30** and **36**.

Table 4.15 Novel Compounds with Predicted DAT/SERT Selectivities[a]

| -X$_1$ | -X$_2$ | Predicted DAT/SERT Selectivity | |
		Unminimized[b]	Minimized[c]
-OCN	-CF$_3$	2.90	2.81
-CH$_2$CN	-CF$_3$	2.87	2.79
-NHCN	-CN	2.83	2.50
-CH$_2$CN	-CN	2.77	2.72
-OCN	-CN	2.77	2.72
-NH$_2$	-CF$_3$	2.76	2.54
-F	-CF$_3$	2.76	2.74
-CH$_2$NH$_2$	-CF$_3$	2.75	2.76
-OH	-CF$_3$	2.72	2.59

[a] Compounds listed have predicted DAT/SERT selectivities of 2.70 or above, and were not already in the QSAR study.
[b] Structures were created and predictions made on unminimized conformations, for significant time savings allowing substitution of all 17 groups within a reasonable amount of time.
[c] Structures were created and then minimized. Predictions were made on minimized structures. Position 4 substituents were limited to the most active -CF$_3$ and -CN groups for efficiency.

The scaffold used for the novel compounds was the phenyl piperidine analog **26**, since the phenyl-substituted compounds **30** and **36** are the most active compounds in the QSAR series. An exhaustive search of replacing the position 3 and position 4 phenyl hydrogens (corresponding to positions 5, 6, and 7 of the 2-naphthyl analog) of analog **26** with one of 17 substituents similar to those found in the QSAR training set compounds was run and the DAT/SERT selectivities of the compounds (without minimization) were predicted. The complete list of substituents is included in Appendix E. Table 4.15 shows nine compounds that have predicted DAT/SERT selectivities higher than the most active analog, **36** (DAT/SERT selectivity of 2.70), for the Optimize QSAR run that did not perform minimization of each structure. The predicted DAT/SERT selectivities found in

additional runs, where minimization was performed, used only the two most active position 4 substituents and substituted each of the 17 groups at position 3.

Both the -CF$_3$ and -CN groups impart some bulk and negative electrostatics to position 4 of the phenyl ring. In addition, both position 4 substituents, -OCN and -CH$_2$CN are moderate electron withdrawing groups, causing the phenyl ring to be more positive. The -NH$_2$ group and -OH group are both strong electron donating groups, which lead to an increase in electron density at the ortho and para positions, and a decrease in electron density (a positive environment) near the N and O atoms of the substituents. Therefore, five of the substitutions at position 3 may address the blue area identified on the CoMFA contour map where a more positive environment is expected to yield a higher activity. Although these compounds do not show significant predicted DAT/SERT selectivity increases, their high activities indicate the potential for exploration of 3,4-disubstituted phenyls.

CHAPTER 5

DISCUSSION

5.1 Comparison of Piperazine and Piperidine Clustering Results

The results of the clustering study indicate that there is no major difference in the rotation of the A-side substituent when comparing the piperazine analog **2** to its piperidine counterpart **3**. Minor differences in the orientation of the A-side exist when comparing **2** and **3**. The results of the B-side analysis show less agreement between the two analogs; the presence of the methyne substituent instead of the nitrogen appears to lead to a wider distribution of cluster sizes. However, this could be an artefact of the calculation due to incomplete searching of conformational space or characteristics of the Tripos force field.

These results show that when independent clustering studies are performed on these piperazine and piperidine analogs of **1**, the results are strikingly similar for the A-side analysis. This clustering study found consistent results for the A-side of the piperazine analog, **2**, and the piperidine analog, **3**. The highly flexible B-side was more difficult to analyze, and may require additional analysis to accurately represent the possible conformations. The size of the clusters seen in the B-side clusterings varies to a greater extent, which may indicate that the random search did not fully populate each cluster or the smaller clusters represent a smaller range of low-energy conformers.

5.2 CoMFA and CoMSIA Studies

5.2.1 Using Multiple Fields and Field Settings for QSAR Development

A range of electrostatic and steric cutoffs for CoMFA columns and different CoMSIA columns were used in this study. A review of Table 4.9 shows that 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 previous work on methylphenidate showed from 0 to 30% increase in q^2 values between CoMFA runs at the default steric and electrostatic cutoffs of 30 kcal/mol each, with average q^2 increases of approximately 10% predictivity of data; however, a larger range of steric and electrostatic cutoffs was used.[44] The amount of time to perform eight extra CoMFA calculations, and three CoMSIA calculations is minimal compared to the time 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 QSAR studies. After focusing the models, Family 4 (static/electrostatic field CoMSIA model) had only a small increase in q^2 after focusing, making it by far the lowest q^2 value of the six (eight with the reduced models included) focused models. Family 2 (hydrophobic field CoMSIA model) dropped from third to fourth 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 Family2. This contrasts the other four families; the four CoMFA models were improved to a greater extent by

focusing (+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 CoMSIA models is somehow related to CoMSIA models themselves, or an artefact of these particular models. Larger studies of CoMSIA models would be necessary to determine if this is a general effect.

5.2.2 Use of Same Template for Piperazine and Piperidine Analogs.

Separate random searches were performed on the piperazine **2** and the piperidine **3**. QSAR studies using analog **2** as the only template (building the piperidine analogs of **1** by substituting a methyne substituent for the B-side nitrogen) were initially tested because the representative conformers clustered well for both analogs **2** and **3**. Predictivities were poor, therefore the present study was done using both **2** and **3** as templates. This brought the 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 (Figure 4.17) 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 templates to affect the CoMFA contour map significantly.

5.2.3 "Near-minimization" and Minimization

The Venanzi group's previous methylphenidate CoMFA study used a methylphenidate template with atom or group substitution for phenyl hydrogens, and no minimization for single atom substituents.[44] This method allows perfect alignment, which may be expected for single and few atom substituents. This series of **1** analogs is more diverse than the methylphenidate analogs, therefore minimization of each analog after changing a

substituent is much more important. Casual molecular modelers are cautioned to ensure complete minimization of their molecules; for a molecule of this size, the iterations were increased from the default of 100 iterations to 1,000 iterations. Any molecule that still had a high energy (more than 10 kcal/mol from other analogs in the series) was further checked by applying extra minimization steps manually to confirm the final value. In this study, 1,000 iterations were sufficient for proper minimization of each analog. Proper minimization of substituents was supported by the use of SYBYL library fragments for the A-side substituents.

For screening of novel potentially active compounds using Optimize QSAR, the initial exhaustive runs were performed on unminimized structures. The best results of this preliminary exhaustive run were used to develop a smaller Optimize QSAR run that only had two possible substituents in position 4, since these two substituents appeared in every compound predicted to be more active than analog **36**. The second run allowed the full set of the original 17 substituents for position 3, and the structures were minimized before activities were predicted. This combination of unminimized and minimized structures is an efficient way to target the most likely beneficial substituents, without the significant time effort of an exhaustive and fully minimized structure run.

5.2.4 Inclusion of Racemic Compounds in the QSAR Studies

Forty-three of the forty-eight compounds studied either have no chiral centers or were isolated as a specific stereoisomer. The inclusion of five compounds synthesized as racemic mixtures in this 3D-QSAR study was based on the data available at the time of the research. All five compounds (**13**, **14**, **15**, **16**, and **17**) are published in one paper,[20] and are cyclic ketones with one chiral center; three are single ring and two have two

fused rings. These compounds were modeled with the chirality that best aligned with the 2-naphthyl substituent in analog **2**, with the ketone group extending into the same area covered by the naphthyl substituent. Research into the use of racemic compounds in QSAR studies from the Journal of Medicinal Chemistry indicates that multiple QSAR studies based on data from racemic mixtures have been published in the first eight issues of 2005. For example, Budriesi et al.'s paper on selective myocardial calcium channel modulators[221] uses racemic data for all but one of their compounds, and makes chirality assumptions for both one and two chiral center compounds; this data is used as an input for a 3D-QSAR study. Other QSAR studies, such as Gallardo-Godoy et al.'s work,[222] do not explicitly mention racemic mixtures, but the structures of certain compounds (e.g., **29** and **30** in the referenced paper) implicitly indicate chiral centers.

A QSAR study on Family 6 performed with a training set of 35 compounds, excluding the four racemic compounds in the present study's training set (one racemic compound, **13**, being in the present study's test set), resulted in decreased q^2 values for all CoMFA and CoMSIA parameter/field combinations. The results are listed in Appendix F. The results of the present QSAR studies with and without the five racemic compounds indicates that the data based on racemic mixtures provides valuable information for this QSAR series. Colleagues at the NIH are presently working on expanding the GBR 12909 series of compounds, including additional stereoisomerically pure compounds.

5.2.5 LOO/CV with Column Filtering versus SAMPLS

Screening was performed using the SAMPLS method.[218] This method consistently gives the same results as a PLS run with σ, the column filtering cutoff value, set to zero. However, a "full" PLS run using a σ equal to 2 kcal/mol has been in the past

recommended to calculate q^2, to expedite calculations and to remove values that may not be as predictive as values that vary more widely. In this study, there were sometimes differences between the $\sigma = 0$ (SAMPLS) and $\sigma = 2$ results, from 0 to 11% explanation of activities in either direction. Performing SAMPLS calculations is expeditious for multiple studies, allowing performance of many field and parameter settings in a short amount of time. Performing at least both SAMPLS and $\sigma = 2$ cross-validated PLS runs for the best models from preliminary studies would be recommended to find the best σ setting for a particular data set and methodology, although discussions with Tripos personnel indicate that column filtering is becoming obsolete. Here the SAMPLS results are presented for the Preliminary CoMFA results, and both the $\sigma = 0$ and $\sigma = 2$ kcal/mol results for the Focused CoMFA runs because the models' q^2 values varied depending on the σ setting. For consistency, the SAMPLS results were used to determine the optimal number of components for the NCV runs.

CHAPTER 6

CONCLUSIONS

This work illustrates a QSAR study of a highly flexible family of drug-like molecules. Representative low-energy conformers were gleaned from large sets of random search conformers using hierarchical clustering. QSAR runs were performed on six conformational families, which were developed by modifying representative low-energy conformers of two template molecules. The process of developing a QSAR becomes parallel; considering different steric and electrostatic cutoffs for CoMFA and different fields for CoMSIA yields models of varying predictivity. Comparison of the q^2 values after focusing the QSAR models is a first way to determine potentially acceptable models. An acceptable q^2 must be backed up by internal validation, such as y-value scrambling,[210,211] and external validation, such as test set correlation validation,[212] to further ensure the validity and applicability of high-q^2 models. Validity of the QSAR model can be supported by predicting the activities of novel compounds. The results of this work will be provided to Drs. Rice and Rothman at the NIH to support their ongoing SAR work on the GBR 12909 family of analogs. Future collaborations are planned to continue 3D-QSAR model and pharmacophore development.

APPENDIX A

SOFTWARE SPECIFICATIONS AND INPUT PARAMETERS

This appendix contains the software specifications and parameter settings for the

SYBYL® molecular modeling program used to complete the QSAR studies of GBR

12909 analogs.

Software: SYBYL/Base Version 6.9.1

Manufacturer: Tripos, Inc., St. Louis, Missouri, USA

Platform: SGI Irix 6.5

SYBYL CoMFA Parameters:

All settings left as default (below), except for changes in electrostatic and steric cutoffs.

CoMFA Field Class: Tripos Standard

   Field Values: Type(s): Both

                  Dielectric: Distance

                  Smoothing: None

   Drop Electrostatics: Within Steric Cutoff for Each Row

   Transition: Smooth

   Region: Create Automatically

SYBYL CoMSIA Parameters:

All settings left as default (below), except for changes in selected field.

CoMSIA Field Parameters: Attenuation Factor: 0.3

   Region: Create Automatically

SYBYL PLS Parameters:

PLS: For all runs:

Scaling: CoMFA Standard

For cross-validated (leave-one-out) runs:

"Leave-One-Out" box checked

"Use SAMPLS" box: checked for Leave-One-Out SAMPLS runs

(Maximum Number of) Components: 6*

* Except when number of components needed to be limited in response to poor stability

of higher component model.

For full PLS runs:

"No Validation" box checked

Components: (as found in best cross-validated run)

APPENDIX B

SYBYL PROGRAMMING LANGUAGE SCRIPT EXAMPLE

In this appendix, a portion of a SPL script is included as an example of the scripts used to create and minimize the GBR 12909 analogs.

buildAll1.spl

Build base structures - one for C.3 atoms and one for C.ar atoms.

First, load in Analog 2 template for conformational family.

MOL MULT_IN M1

/afs/cad.njit.edu/research/chem/venanzi/4/Project/analog_creation/new_reps.mdb/c11REP

pDM324_00256.mol2

ORIENT BEST_VIEW M1

Remove the centroids, and the naphthalene aggregate (for future minimization

simplification).

REMOVE CENTROID (*) *

REMOVE AGGREGATE NAPHTHALENE_AGGREGATE

Add it to the conformational family database.

DATABASE OPEN

/afs/cad.njit.edu/research/chem/venanzi/4/Project/QSAR_II/1pALL.mdb UPDATE

MODIFY MOLECULE NAME M1 "1pDM-DM324-2-naphthylTEMPLATE"

DATABASE ADD "M1"

Remove some atoms, and save as Carbase (C.ar base structure)

REMOVE ATOM

M1(((((((((((((33)+66)+32)+31)+65)+30)+64)+29)+63)+62)+28)+27)+26)+61)

112

MODIFY MOLECULE NAME M1 "1pDM-Carbase"

Put in conformational family database, will delete at end.

DATABASE ADD "M1"

Change atoms 24 and 25 to C.3 atoms, then rename and put in database.

MODIFY ATOM TYPE M1((25)+24) C.3 C.3

MODIFY MOLECULE NAME M1 "1pDM-C3base"

DATABASE ADD "M1"

Make C.3 C.2 base.

MODIFY ATOM TYPE M1(25) C.2

MODIFY MOLECULE NAME M1 "1pDM-C32base"

DATABASE ADD "M1"

ZAP M1

PART II - Create 22 analogs from the base structures for the conformational family, either using

a database of side groups or directly using atom modification and adding hydrogens.

Build DM-325 (1-naphthyl)

Put C.ar base for conformational family in M1 area, works for first set of Analog 2 analogs. This will be repeated as analogs are built.

DATABASE GET "1pDM-Carbase" M1

Open database of A-side groups. Because it is last opened, it becomes default.

```
DATABASE OPEN

/afs/cad.njit.edu/research/chem/venanzi/4/Project/analog_creation/DMAsides.mdb

UPDATE


# Get the A-side group from the DMAsides database.

DATABASE GET "naphthalene" M2

FUSE M1(23) M2(14) M1(24) M2(4) M1(25) M2(10) |

# Put it in conformational family database by switching it to the default database.

DATABASE DEFAULT

/afs/cad.njit.edu/research/chem/venanzi/4/Project/QSAR_II/1pALL.mdb

MODIFY MOLECULE NAME M1 "1pDM-DM325-1-naphthyl"

DATABASE ADD "M1"

# Get rid of molecules.

ZAP M1

ZAP M2
```

(program continues in a similar manner for each analog)

APPENDIX C

CoMFA AND CoMSIA STUDY RESULTS

Detailed CoMFA and CoMSIA preliminary study results are contained herein.

Table C.1 Leave-one-out Cross-Validated CoMFA and CoMSIA Results

	Family 1		Family 2		Family 3		Family 4		Family 5		Family 6	
Field[a]	$q^{2\,b}$	C^c	$q^{2\,b}$	C^c	$q^{2\,b}$	C^c	$q^{2\,b}$	C^c	$q^{2\,b}$	C^c	$q^{2\,b}$	C^c
DACoMSIA	0.197	1	0.070	1	0.196	1	0.126	1	0.185	1	0.167	1
HPCoMSIA	0.233	1	**0.345**	**3**	0.288	5	0.141	1	0.349	3	0.178	1
SEComSIA	0.266	6	0.224	2	0.291	5	**0.344**	**5**	0.305	4	0.262	2
S10E10	0.296	1	0.315	2	**0.320**	**5**	0.235	6	**0.374**	**2**	0.383	2
S10E30	0.293	1	0.294	2	0.314	5	0.221	5	0.355	2	0.365	2
S10E50	0.293	1	0.294	2	0.314	5	0.223	5	0.355	2	0.365	2
S30E10	0.303	2	0.304	2	0.299	5	0.271	5	0.354	2	**0.388**	**2**
S30E30	0.287	1	0.283	2	0.290	6	0.266	5	0.324	2	0.368	2
S30E50	0.287	1	0.283	2	0.285	6	0.266	5	0.324	2	0.368	2
S50E10	**0.323**	**2**	0.294	2	0.320	6	0.231	4	0.368	2	0.380	2
S50E30	0.287	1	0.271	2	0.302	6	0.219	5	0.340	2	0.357	2
S50E50	0.286	1	0.271	2	0.298	6	0.219	5	0.339	2	0.356	2
Maximum q^2	0.323	6	0.345	3	0.320	5	0.344	5	0.374	2	0.388	2

[a] CoMFA or CoMSIA field settings. DACoMSIA: Hydrogen bond donor/hydrogen bond acceptor CoMSIA field. HPCoMSIA: Hydrophobic/hydrophilic CoMSIA field. SECoMSIA: Steric/electrostatic CoMSIA field. For CoMFA field settings, S represents steric cutoff value, and E represents electrostatic cutoff value. Thus S10E10 equals steric cutoff of 10 kcal/mol and electrostatic cutoff of 10 kcal/mol.
[b] Leave-one-out cross-validated predictivity, using SAMPLS.
[c] Optimal number of components determined by leave-one-out cross-validation method.
[d] Maximum q^2 of all of the field settings for each conformational family.

Table C.2 Training Set Predictions - Focused CoMFA Results

		DAT/SERT Selectivity[a]						
		Family 1			Family 5			Family 6		
	Exp.[b]	Pred.[c]	Res.[d]		Pred.[c]	Res.[d]		Pred.[c]	Res.[d]
2	1.59	1.63	0.04		1.76	0.17		1.69	0.10
4	0.89	0.97	0.08		1.05	0.16		1.21	0.32
6	1.65	1.89	0.24		1.81	0.16		1.71	0.06
7	2.23	1.70	-0.53		1.79	-0.44		1.81	-0.42
9	1.48	1.60	0.12		1.60	0.12		1.54	0.06
10	0.95	1.61	0.66		1.72	0.77		1.66	0.71
11	1.12	1.67	0.55		1.32	0.20		1.33	0.21
12	0.78	0.45	-0.33		0.57	-0.21		0.51	-0.27
14	1.29	0.46	-0.83		0.35	-0.94		0.24	-1.05
15	1.03	1.01	-0.02		0.85	-0.18		0.49	-0.54
16	0.29	0.02	-0.27		-0.16	-0.45		0.05	-0.24
17	-0.87	-0.26	0.61		-0.50	0.37		-0.26	0.61
18	0.25	0.33	0.08		0.65	0.40		0.36	0.11
19	1.36	0.60	-0.76		0.67	-0.69		0.61	-0.75
20	-0.67	0.39	1.06		0.43	1.10		0.53	1.20
21	1.61	0.48	-1.13		1.17	-0.44		1.31	-0.30
22	0.03	0.14	0.11		0.86	0.83		1.05	1.02
23	1.41	1.61	0.20		1.36	-0.05		1.40	-0.01
24	1.79	1.72	-0.07		1.28	-0.51		1.39	-0.40
25	1.35	1.28	-0.07		1.00	-0.35		1.07	-0.28
3	2.51	1.83	-0.68		1.92	-0.59		1.95	-0.56
26	1.24	1.70	0.46		1.72	0.48		1.64	0.40
27	1.53	1.68	0.15		1.67	0.14		1.74	0.21
28	1.08	1.44	0.36		1.48	0.40		1.50	0.42
30	2.60	2.63	0.03		2.79	0.19		2.81	0.21
31	2.51	2.00	-0.51		2.17	-0.34		2.13	-0.38
33	2.26	1.94	-0.32		2.08	-0.18		2.04	-0.22
34	2.03	1.95	-0.08		2.05	0.02		1.94	-0.09
36	2.70	2.48	-0.22		2.71	0.01		2.70	0.00
37	1.99	1.95	-0.04		2.10	0.11		2.04	0.05
38	2.08	1.91	-0.17		2.02	-0.06		2.05	-0.03
39	1.36	1.47	0.11		1.34	-0.02		1.42	0.06
40	1.93	1.92	-0.01		2.04	0.11		1.98	0.05
41	2.18	1.89	-0.29		1.85	-0.33		1.86	-0.32
42	1.49	1.87	0.38		1.49	0.00		1.51	0.02
43	0.21	1.15	0.94		0.66	0.45		0.77	0.56
44	1.00	1.56	0.56		1.20	0.20		1.14	0.14
45	1.44	1.53	0.09		1.63	0.19		1.68	0.24
46	2.02	1.51	-0.51		1.23	-0.79		1.15	-0.87

[a] DAT/SERT selectivity defined in Equation (2.5).

[b] Experimental DAT/SERT selectivity as listed in Table 3.1.

[c] The DAT/SERT Selectivity predicted by the referenced family's best QSAR model.

[d] Equals (Pred. Select. - Exp. Select) Residual is calculated from the exact values, and therefore may vary slightly from the apparent difference of the rounded-off values.

Table C.3 Statistics for Experimental Data and Training Set Predictions

	Exp.	Family 1	Family 5	Family 6
Minimum	-0.87	-0.26	-0.50	-0.26
Average	1.38	1.38	1.38	1.38
Median	1.44	1.61	1.48	1.50
Maximum	2.70	2.63	2.79	2.81
St. Dev.	0.83	0.69	0.71	0.69

APPENDIX D

PREDICTIONS FOR TEST SET COMPOUNDS

The SYBYL output for the predictions of all of the test set compounds are contained in this appendix.

D.1   Original Test Set Predictions

D.1.1   Family 1 Results

Predicting 1PDM-DL670-CYCLOPENTANONYL (M1)

Row 46 1PDM-DL670-CYCLOPENTANONYLX Column #17 (FOCUSS50E10) :

COMFA

... prediction of LSDR for 1PDM-DL670-CYCLOPENTANONYLX : 0.823415

^^^^^^^ Extrapolated 234 of 3202 terms (data outside range); sum is 0.00984208

^^^^^^^ Ignored 38 of 3202 terms (missing data, used column mean)

Predicting 1PDM-DM333-2-BENZOTHIOPHENYL (M2)

Row 46 1PDM-DM333-2-BENZOTHIOPHENYLX Column #17 (FOCUSS50E10) :

COMFA

... prediction of LSDR for 1PDM-DM333-2-BENZOTHIOPHENYLX : 1.82723

^^^^^^^ Extrapolated 6 of 3202 terms (data outside range); sum is 0.0231725

^^^^^^^ Ignored 50 of 3202 terms (missing data, used column mean)

Predicting 1PDM-DM337-2-QUINOLINYL (M3)

Row 46 1PDM-DM337-2-QUINOLINYLX Column #17 (FOCUSS50E10) : COMFA

... prediction of LSDR for 1PDM-DM337-2-QUINOLINYLX : 1.56428

^^^^^^^ Extrapolated 387 of 3202 terms (data outside range); sum is -0.00553617

^^^^^^^ Ignored 47 of 3202 terms (missing data, used column mean)

Predicting 1PTP-TP1134-2-METHOXYPHENYL (M4)

Row 46 1PTP-TP1134-2-METHOXYPHENYLX Column #17 (FOCUSS50E10) :

COMFA

... prediction of LSDR for 1PTP-TP1134-2-METHOXYPHENYLX : 1.72194

^^^^^^^ Extrapolated 175 of 3202 terms (data outside range); sum is 0.0722921

^^^^^^^ Ignored 42 of 3202 terms (missing data, used column mean)

Predicting 1PTP-TP227-4-TOLYL (M5)

Row 46 1PTP-TP227-4-TOLYLX Column #17 (FOCUSS50E10) : COMFA

... prediction of LSDR for 1PTP-TP227-4-TOLYLX : 1.67781

^^^^^^^ Extrapolated 1 of 3202 terms (data outside range); sum is -0.0180285

^^^^^^^ Ignored 44 of 3202 terms (missing data, used column mean)

Predicting 1PTP-TP231-4-CHLOROPHENYL (M6)

Row 46 1PTP-TP231-4-CHLOROPHENYLX Column #17 (FOCUSS50E10) :

COMFA

... prediction of LSDR for 1PTP-TP231-4-CHLOROPHENYLX : 1.94493

^^^^^^^ Ignored 41 of 3202 terms (missing data, used column mean)

D.1.2 Family 5 Results

Predicting 5PDM-DL670-CYCLOPENTANONYL (M1)

Row 46 5PDM-DL670-CYCLOPENTANONYLX Column #10 (FOCUSS10E10) :

COMFA

... prediction of LSDR for 5PDM-DL670-CYCLOPENTANONYLX : 0.65538

^^^^^^^ Extrapolated 371 of 3053 terms (data outside range); sum is -0.0861013

^^^^^^^ Ignored 38 of 3053 terms (missing data, used column mean)

Predicting 5PDM-DM333-2-BENZOTHIOPHENYL (M2)

Row 46 5PDM-DM333-2-BENZOTHIOPHENYLX Column #10 (FOCUSS10E10) :

COMFA

... prediction of LSDR for 5PDM-DM333-2-BENZOTHIOPHENYLX : 1.58704

^^^^^^^ Extrapolated 5 of 3053 terms (data outside range); sum is 4.8557e-09

^^^^^^^ Ignored 42 of 3053 terms (missing data, used column mean)

Predicting 5PDM-DM337-2-QUINOLINYL (M3)

Row 46 5PDM-DM337-2-QUINOLINYLX Column #10 (FOCUSS10E10) : COMFA

... prediction of LSDR for 5PDM-DM337-2-QUINOLINYLX : 1.77997

^^^^^^^ Extrapolated 375 of 3053 terms (data outside range); sum is 0.0699088

^^^^^^^ Ignored 51 of 3053 terms (missing data, used column mean)

Predicting 5PTP-TP1134-2-METHOXYPHENYL (M4)

Row 46 5PTP-TP1134-2-METHOXYPHENYLX Column #10 (FOCUSS10E10) :

COMFA

... prediction of LSDR for 5PTP-TP1134-2-METHOXYPHENYLX : 1.576

^^^^^^^ Extrapolated 178 of 3053 terms (data outside range); sum is -0.00353263

^^^^^^^ Ignored 44 of 3053 terms (missing data, used column mean)

Predicting 5PTP-TP227-4-TOLYL (M5)

Row 46 5PTP-TP227-4-TOLYLX Column #10 (FOCUSS10E10) : COMFA

... prediction of LSDR for 5PTP-TP227-4-TOLYLX : 1.80499

^^^^^^^ Extrapolated 1 of 3053 terms (data outside range); sum is -0.00926648

^^^^^^^ Ignored 46 of 3053 terms (missing data, used column mean)

Predicting 5PTP-TP231-4-CHLOROPHENYL (M6)

Row 46 5PTP-TP231-4-CHLOROPHENYLX Column #10 (FOCUSS10E10) :

COMFA

... prediction of LSDR for 5PTP-TP231-4-CHLOROPHENYLX : 2.05317

^^^^^^^ Ignored 43 of 3053 terms (missing data, used column mean)

D.1.3   Family 6 Results

Predicting 6PDM-DL670-CYCLOPENTANONYL (M1)

Row 46 6PDM-DL670-CYCLOPENTANONYLX Column #13 (FOCUSS30E10) :

COMFA

... prediction of LSDR for 6PDM-DL670-CYCLOPENTANONYLX : 0.667849

^^^^^^^ Extrapolated 66 of 3009 terms (data outside range); sum is -0.0094032

^^^^^^^ Ignored 45 of 3009 terms (missing data, used column mean)

Predicting 6PDM-DM333-2-BENZOTHIOPHENYL (M2)

Row 46 6PDM-DM333-2-BENZOTHIOPHENYLX Column #13 (FOCUSS30E10) :

COMFA

... prediction of LSDR for 6PDM-DM333-2-BENZOTHIOPHENYLX : 1.56269

^^^^^^^ Extrapolated 2 of 3009 terms (data outside range); sum is -0.00177293

^^^^^^^ Ignored 47 of 3009 terms (missing data, used column mean)

Predicting 6PDM-DM337-2-QUINOLINYL (M3)

Row 46 6PDM-DM337-2-QUINOLINYLX Column #13 (FOCUSS30E10) : COMFA

... prediction of LSDR for 6PDM-DM337-2-QUINOLINYLX : 1.51835

^^^^^^^ Extrapolated 152 of 3009 terms (data outside range); sum is -0.00616532

^^^^^^^ Ignored 52 of 3009 terms (missing data, used column mean)

Predicting 6PTP-TP1134-2-METHOXYPHENYL (M4)

Row 46 6PTP-TP1134-2-METHOXYPHENYLX Column #13 (FOCUSS30E10) :

COMFA

... prediction of LSDR for 6PTP-TP1134-2-METHOXYPHENYLX : 1.57328

^^^^^^^ Extrapolated 212 of 3009 terms (data outside range); sum is 7.01147e-05

^^^^^^^ Ignored 49 of 3009 terms (missing data, used column mean)

Predicting 6PTP-TP227-4-TOLYL (M5)

Row 46 6PTP-TP227-4-TOLYLX Column #13 (FOCUSS30E10) : COMFA

... prediction of LSDR for 6PTP-TP227-4-TOLYLX : 1.73129

^^^^^^^ Extrapolated 5 of 3009 terms (data outside range); sum is -0.0447089

^^^^^^^ Ignored 46 of 3009 terms (missing data, used column mean)

Predicting 6PTP-TP231-4-CHLOROPHENYL (M6)

Row 46 6PTP-TP231-4-CHLOROPHENYLX Column #13 (FOCUSS30E10) :

COMFA

... prediction of LSDR for 6PTP-TP231-4-CHLOROPHENYLX : 2.04843

^^^^^^^ Ignored 46 of 3009 terms (missing data, used column mean)

D.2 EXTENDED TEST SET PREDICTIONS

D.2.1 Family 1 Results

Predicting 1pDM-DM240 (M1)

Row 46 1PDM-DM240 Column #17 (FOCUSS50E10) : COMFA

... prediction of LSDR for 1PDM-DM240 : 1.71296

^^^^^^^ Extrapolated 49 of 3202 terms (data outside range); sum is 0.0305

^^^^^^^ Ignored 48 of 3202 terms (missing data, used column mean)

Predicting 1pDM-GBR12909 (M2)

Row 46 1PDM-GBR12909 Column #17 (FOCUSS50E10) : COMFA

... prediction of LSDR for 1PDM-GBR12909 : 1.70143

^^^^^^^ Extrapolated 59 of 3202 terms (data outside range); sum is 0.0339788

^^^^^^^ Ignored 47 of 3202 terms (missing data, used column mean)

Predicting 1pDM-YZ0902-CH2Cl (M3)

Row 46 1PDM-YZ0902-CH2CL Column #17 (FOCUSS50E10) : COMFA

... prediction of LSDR for 1PDM-YZ0902-CH2CL : 1.61162

^^^^^^^ Extrapolated 232 of 3202 terms (data outside range); sum is 0.0347802

^^^^^^^ Ignored 36 of 3202 terms (missing data, used column mean)

D.2.2 Family 5 Results

Predicting 5pDM-GBR12909 (M1)

Row 46 5PDM-GBR12909 Column #10 (FOCUSS10E10) : COMFA

... prediction of LSDR for 5PDM-GBR12909 : 1.65773

^^^^^^^ Extrapolated 71 of 3053 terms (data outside range); sum is 0.01869

^^^^^^^ Ignored 53 of 3053 terms (missing data, used column mean)

Predicting 5pDM-DM240 (M2)

Row 46 5PDM-DM240 Column #10 (FOCUSS10E10) : COMFA

... prediction of LSDR for 5PDM-DM240 : 1.76558

^^^^^^^ Extrapolated 88 of 3053 terms (data outside range); sum is 0.0164219

^^^^^^^ Ignored 52 of 3053 terms (missing data, used column mean)

Predicting 5pDM-YZ0902-ethylCl (M3)

Row 46 5PDM-YZ0902-ETHYLCL Column #10 (FOCUSS10E10) : COMFA

... prediction of LSDR for 5PDM-YZ0902-ETHYLCL : 1.14389

^^^^^^^ Extrapolated 24 of 3053 terms (data outside range); sum is 5.45094e-05

^^^^^^^ Ignored 32 of 3053 terms (missing data, used column mean)

D.2.3 Family 6 Results

Predicting 6pDM-GBR12909 (M1)

Row 46 6PDM-GBR12909 Column #13 (FOCUSS30E10) : COMFA

... prediction of LSDR for 6PDM-GBR12909 : 1.61151

^^^^^^^ Extrapolated 54 of 3009 terms (data outside range); sum is 0.00440965

^^^^^^^ Ignored 51 of 3009 terms (missing data, used column mean)

Predicting 6pDM-DM240 (M2)

Row 46 6PDM-DM240 Column #13 (FOCUSS30E10) : COMFA

... prediction of LSDR for 6PDM-DM240 : 1.64792

^^^^^^^ Extrapolated 96 of 3009 terms (data outside range); sum is 0.00417796

^^^^^^^ Ignored 49 of 3009 terms (missing data, used column mean)

Predicting 6pDM-YZ-0902-ethylCl (M3)

Row 46 6PDM-YZ-0902-ETHYLCL Column #13 (FOCUSS30E10) : COMFA

... prediction of LSDR for 6PDM-YZ-0902-ETHYLCL : 0.900819

^^^^^^^ Extrapolated 66 of 3009 terms (data outside range); sum is -0.0138962

^^^^^^^ Ignored 41 of 3009 terms (missing data, used column mean)

APPENDIX E

LIST OF SUBSTITUENTS FOR NOVEL COMPOUNDS

This appendix contains a table showing the 16 substituents, excluding the default of a

hydrogen atom, that were used in the Optimize QSAR studies to predict novel

compounds.

Table E.1 Substituents for Optimize QSAR Studies[a]

-Br	-CH$_2$CN	-NH$_2$	-CF$_3$
-Cl	-CH$_2$NH$_2$	-NHCHO	-CHO
-F	-CH$_3$	-NHCH$_3$	-CN
-OH	-OCN	-NHCN	-CONH$_2$

[a] Seventeenth substituent is the default of a hydrogen atom.

APPENDIX F

CoMFA AND CoMSIA STUDY RESULTS FOR FAMILY 6 WITH

NON-RACEMIC TRAINING SET

This appendix contains the results of a QSAR study on Family 6 using a reduced training set created by removing the four racemic compounds (**14**, **15**, **16**, and **17**) from the training set that was used in the bulk of the study.

TABLE F.1 CoMFA and CoMSIA Results for Family 6 with Non-Racemic Training Set[a]

Field[b]	q^{2} [c]	C[d]	Δq^{2} [e]
DACOMSIA	0.047	1	0.120
HPCOMSIA	0.088	1	0.090
SECOMSIA	0.127	1	0.135
S10E10	0.258	2	0.125
S10E30	0.247	2	0.118
S10E50	0.247	2	0.118
S30E10	0.275	2	0.113
S30E30	0.257	2	0.111
S30E50	0.257	2	0.111
S50E10	0.261	2	0.119
S50E30	0.239	2	0.118
S50E50	0.238	2	0.118
FOCUSED[f]	0.437	2	0.071

[a] Preliminary and focused QSAR study results for Family 6 conformational family when omitting racemic compounds from the training set.
[b] CoMFA or CoMSIA field settings. DACoMSIA: Hydrogen bond donor/hydrogen bond acceptor CoMSIA field. HPCoMSIA: Hydrophobic/hydrophilic CoMSIA field. SECoMSIA: Steric/electrostatic CoMSIA field. For CoMFA field settings, S represents steric cutoff value, and E represents electrostatic cutoff value. Thus S10E10 equals steric cutoff of 10 kcal/mol and electrostatic cutoff of 10 kcal/mol.
[c] Leave-one-out cross-validated predictivity, using SAMPLS.
[d] Optimal number of components determined by leave-one-out cross-validation method.
[e] Difference between q^2 with four racemic compounds included in QSAR training set (standard study; fifth racemic compound is in QSAR test set) and the non-racemic q^2 result shown here, for the indicated field setting.
[f] Focused CoMFA results for S30E10 field setting, compared models each have two components.

REFERENCES

1.   Gorelick, D. A., Gardner, E.L., Xi, Z.X. Agents in Development for the Management of Cocaine Abuse. *Drugs* **2004,** *64,* 1547-1573.

2.   Carroll, F. I. 2002 Medicinal Chemistry Division Award Address: Monoamine Transporters and Opioid Receptors. Targets for Addiction Therapy. *J. Med. Chem.* **2003,** *46,* 1775-1794.

3.   Kuhar, M. J.; Ritz, M. C.; Boja, J. W. The Dopamine Hypothesis of the Reinforcing Properties of Cocaine. *TINS* **1991,** *14,* 299-302.

4.   Kuhar, M. J. Molecular Pharmacology of Cocaine: A Dopamine Hypothesis and its Implications. *Ciba Found. Symp.* **1992,** *166,* 81-89.

5.   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.

6.   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.

7.   Singh, S. Chemistry, Design, and Structure-Activity Relationship of Cocaine Antagonists. *Chem. Rev.* **2000,** *100,* 925-1024.

8.   Dutta, A. K.; Zhang, S.; Kolhatkar, R.; Reith, M. E. A. Dopamine Transporter As Target for Drug Development of Cocaine Dependence Medications. *Eur. J. Pharmacol.* **2003,** *479,* 93-106.

9.   Glowa, J. R.; Wojnicki, F. H. E.; Matecka, D.; Bacher, J.; Mansbach, R. S.; Balster, R. L.; Rice, K. C. Effects of Dopamine Reuptake Inhibitors on Food- and Cocaine-Maintained Responding. I: Dependence on Unit Dose of Cocaine. *Exp. Clin. Psychopharmacol.* **1995,** *3,* 219-231.

10.   Mojsiak, J. Unpublished results. **2003,** National Institutes of Health.

11.   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,* 911-918.

12.   Prisinzano, T.; Rice, K. C.; Baumann, M. H.; Rothman, R. B. Development of Neurochemical Normalization ("Agonist Substitution") Therapeutics for

Stimulant Abuse: Focus on the Dopamine Uptake Inhibitor, GBR 12909. *Curr. Med. Chem. - Central Nervous Syst. Agents* **2004**, *4*, 47-59.

13. Vaughan, R. A.; Gaffaney, J. D.; Lever, J. R.; Reith, M. E. A.; Dutta, A. K. Dual Incorporation of Photoaffinity Ligands on Dopamine Transporters Implicates Proximity of Labeled Domains. *Mol. Pharmacol.* **2001**, *59*, 1157-1164.

14. Prisinzano, T.; Greiner, E.; Johnson II, E. M.; Dersch, C. M.; Marcus, J.; Partilla, J. S.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. Piperidine Analogues of GBR 12909: High Affinity Ligands for the Dopamine Transporter. *J. Med. Chem.* **2002**, *45*, 4371-4374.

15. Dutta, A. K.; Davis, M. C.; Reith, M. E. A. Rational Design and Synthesis of Novel 2,5-Disubstituted cis- and trans- Piperidine Derivatives Exhibiting Differential Activity for the Dopamine Transporter. *Bioorg. Med. Chem. Let.* **2001**, *11*, 2337 - 2340.

16. Matecka, D.; Rothman, R. B.; Radesca, L.; de Costa, B. R.; Dersch, C. M.; Partilla, J. S.; Pert, A.; Glowa, J. R.; Wojnicki, F. H. E.; Rice, K. C. Development of Novel, Potent, and Selective Dopamine Reuptake Inhibitors Through Alteration of the Piperazine Ring of 1-[2-(Diphenylmethoxy)ethyl]- and 1-[2-[Bis(4-fluorophenyl)methoxy]ethyl]-4-(3-phenylpropyl)piperazines (GBR 12935 and GBR 12909). *J. Med. Chem.* **1996**, *39*, 4704-4716.

17. 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. Heteroaromatic Analogs of 1-[2-(Diphenylmethoxy)ethyl]- and 1-[2-[Bis(4-fluorophenyl)methoxy]ethyl]-4-3-phenylpropyl)piperazines (GBR 12935 and GBR 12909) as High-Affinity Dopamine Reuptake Inhibitors. *J. Med. Chem.* **1997**, *40*, 705-716.

18. Lewis, D. B.; Matecka, D.; Zhang, Y.; Hsin, L.-W.; Dersch, C. M.; Stafford, D.; Glowa, J. R.; Rothman, R. B.; Rice, K. C. Oxygenated Analogues of 1-[2-(Diphenylmethoxy)ethyl]- and 1-[2-(Bis(4-fluorophenyl)methoxy]ethyl]-4-[3-phenylpropyl]piperazines (GBR 12935 and GBR 12909) as Potential Extended-Action Cocaine-Abuse Therapeutic Agents. *J. Med. Chem.* **1999**, *42*, 5029-5042.

19. Hsin, L.-W.; Dersch, C. M.; Baumann, M. H.; Stafford, D.; Glowa, J. R.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. Development of Long-Acting Dopamine Transporter Ligands as Potential Cocaine-Abuse Therapeutic Agents: Chiral Hydroxyl-Containing Derivatives of 1-[2-[Bis(4-fluorophenyl)methoxy]ethyl]-4-(3-phenylpropyl)piperazine and 1-[2-(Diphenylmethoxy)ethyl]-4-(3-phenylpropyl)piperazine. *J. Med. Chem.* **2002**, *45*, 1321-1329.

20. Lewis, D. B.; Zhang, Y.; Prisinzano, T.; Dersch, C. M.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. Further Exploration of 1-{2-[*bis*-4-Fluorophenyl)methoxy]ethyl}piperazine (GBR 12909): Role of N-Aromatic, N-Heteroaromatic, and 3-Oxygenated N-Phenylpropyl Substituents on Affinity for

the Dopamine and Serotonin Transporter. *Bioorg. Med. Chem. Let.* **2003**, *13*, 1385-1389.

21. Niznik, H. B., Tyndale, R.F., Sallee, F.R., Gonzalez, F.J., Hardwick, J.P., Inaba, T., Kalow, W. The Dopamine Transporter and Cytochrome P450IID1 (debrisoquine 4-hydroxylase) in Brain: Resolution and Identification of Two Distinct [^{3}H]GBR-12935 Binding Proteins. *Arch. Biochem. Biophys.* **1990**, *276*, 424-432.

22. Cramer III, R. D.; 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.

23. Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to Study Hydrogen-Bonding Properties and to Score Combinatorial Libraries. *J. Comput. Aided Mol. Des.* **1999**, *13*, 1-10.

24. Carroll, F. I.; Gao, Y.; Rahman, M. A.; Abrams, P.; Parham, K.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. Synthesis, Ligand Binding, QSAR, and CoMFA Study of 3β-(*p*-Substituted phenyl)tropane-2β–carboxylic Acid Methyl Esters. *J. Med. Chem.* **1991**, *34*, 2719-2725.

25. Froimowitz, M. Conformational Analysis of Cocaine, the Potent Analog 2ß-Carbomethoxy-3ß-(4-Fluorophenyl)Tropane (CFT), and Other Dopamine Reuptake Blockers. *J. Comput. Chem.* **1993**, *14*, 934-943.

26. Carroll, F. I.; Mascarella, S. W.; Kuzemko, M. A.; Gao, Y.; Abraham, P.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. Synthesis, Ligand Binding, and QSAR (CoMFA and Classical) Study of 3β-(3'-Substituted phenyl)-, 3β-(4'-Substituted phenyl)-, and 3β-(3',4'-Disubstituted phenyl)tropane-2β–carboxylic Acid Methyl Esters. *J. Med. Chem.* **1994**, *37*, 2865-2873.

27. Yang, B.; Wright, J.; Eldefrawi, M. E.; Pou, S.; MacKerell Jr., A. D. Conformational, Aqueous Solvation, and pK_a Contributions to the Binding and Activity of Cocaine, WIN 32065-2, and the WIN Vinyl Analog. *J. Am. Chem. Soc.* **1994**, *116*, 8722-8732.

28. Lieske, S. F.; Yang, B.; Eldefrawi, M. E.; MacKerell Jr., A. D.; Wright, J. (-)-3ß-Substituted Ecgonine Methyl Esters as Inhibitors for Cocaine Binding and Dopamine Uptake. *J. Med. Chem.* **1998**, *41*, 864-876.

29. Zhu, N.; Harrison, A.; Trudell, M. L.; Klein-Stevens, C. L. QSAR and CoMFA Study of Cocaine Analogs: Crystal and Molecular Structure of (-)-Cocaine Hydrochloride and N-Methyl-3β-(*p*-Fluorophenyl)Tropane-2β–Carboxylic Acid Methyl Ester. *Struct. Chem.* **1999**, *10*, 91-103.

30. Muszynski, I. C.; Scapozza, L.; Kovar, K.-A.; Folkers, G. Quantitative Structure-Activity Relationships of Phenyltropanes As Inhibitors of Three Monoamine

Transporters: Classical and CoMFA Studies. *Quant. Struct.-Act. Relat.* **1999**, *18*, 342-353.

31.   Hoffman, B. T.; Kopajtic, T.; Katz, J. L.; Newman, A. H. 2D QSAR Modeling and Preliminary Database Searching for Dopamine Transporter Inhibitors Using Genetic Algorithm Variable Selection of Molconn-Z Descriptors. *J. Med. Chem.* **2000**, *43*, 4151-4159.

32.   Davies, H. M. L.; Gilliatt, V.; Kuhn, L. A.; Saikali, E.; Ren, P.; Hammond, P. S.; Sexton, G. J.; Childers, S. R. Synthesis of 2β-Acyl-3β-(Substituted naphthyl)-8-azabicyclo[3.2.1] octanes and Their Binding Affinities at Dopamine and Serotonin Transport Sites. *J. Med. Chem.* **2001**, *44*, 1509-1515.

33.   Zhan, C. G.; Zheng, F.; Landry, D. W. Fundamental Reaction Mechanism for Cocaine Hydrolysis in Human Butylcholinesterase. *J. Am. Chem. Soc.* **2003**, *125*, 2462-2474.

34.   Paula, S.; Tabet, M. R.; Keenan, S. M.; Welsh, W. J.; Ball Jr., W. J. Three-Dimensional Structure-Activity Relationship Modeling of Cocaine Binding to Two Monoclonal Antibodies by Comparative Molecular Field Analysis. *J. Mol. Biol.* **2003**, *325*, 515-530.

35.   Paula, S.; Tabet, M. R.; Farr, C. D.; Norman, A. D.; Ball Jr., W. J. Three-Dimensional Quantitative Structure-Activity Relationship Modeling of Cocaine Binding by a Novel Human Monoclonal Antibody. *J. Med. Chem.* **2004**, *47*, 133-142.

36.   Kulkarni, S. S.; Grundt, P.; Kopajtic, T.; Katz, J. L.; Newman, A. H. Structure-Activity Relationships at Monoamine Transporters for a Series of *N*-Substituted 3α-(bis[4-fluorophenyl]methoxy)tropanes: Comparative Molecular Field Analysis, Synthesis, and Pharmacological Evaluation. *J. Med. Chem.* **2004**, *47*, 3388-3398.

37.   Yuan, H.; Kozikowski, A. P.; Petukhov, P. A. CoMFA Study of Piperidine Analogues of Cocaine at the Dopamine Transporter: Exploring the Binding Mode of the 3α-Substituent of the Piperidine Ring Using Pharmacophore-Based Flexible Alignment. *J. Med. Chem.* **2004**, *47*, 6137-6143.

38.   Robarge, M. J.; Agoston, G. E.; Izenwasser, S.; Kopajtic, T.; George, C.; Katz, J. L.; Newman, A. H. Highly Selective Chiral *N*-Substituted 3α-[Bis(4'-fluorophenyl)methoxy]tropane Analogues for the Dopamine Transporter: Synthesis and Comparative Molecular Field Analysis. *J. Med. Chem.* **2000**, *43*, 1085-1093.

39.   Newman, A. H.; Izenwasser, S.; Robarge, M. J.; Kline, R. H. CoMFA Study of Novel Phenyl Ring-Substituted 3-α-(Diphenylmethoxy)tropane Analogues at the Dopamine Transporter. *J. Med. Chem.* **1999**, *42*, 3502-3509.

40. Froimowitz, M.; Wu, K.-M.; Rodrigo, J.; George, C. Conformational Preferences of the Potent Dopamine Reuptake Blocker BTCP and Its Analogs and Their Incorporation into a Pharmacophore Model. *J. Comput. Aided Mol. Des.* **2000,** *14,* 135-146.

41. Froimowitz, M.; George, C. Conformational Analysis and a Crystal Structure of Bupropion, an Antidepressant with Dopamine Reuptake Blocking Activity. *J. Chem. Inf. Comp. Sci.* **1998,** *38,* 506-510.

42. Kulkarni, S. S.; Newman, A. H.; Houlihan, W. J. Three-Dimensional Quantitative Structure-Activity Relationships of Mazindol Analogues at the Dopamine Transporter. *J. Med. Chem.* **2002,** *45,* 4119-4127.

43. Froimowitz, M.; Patrick, K. S.; Cody, V. Conformational Analysis of Methylphenidate and Its Structural Relationship to Other Dopamine Blockers Such as CFT. *Pharm. Res.* **1995,** *12,* 1430-1434.

44. Venanzi, C. A.; Misra, M.; Gilbert, K. M.; Schweri, M. M.; Deutsch, H. M. Unpublished results. **2004,** New Jersey Institute of Technology.

45. Wang, S.; Sakamuri, S.; Enyedy, I. J.; Kozikowski, A. P.; Zaman, W. A.; Johnson, K. M. Molecular Modeling, Structure-Activity Relationships and Functional Antagonism Studies of 4-Hydroxy-1-methyl-4-(4-methylphenyl)-3-piperidyl 4-Methylphenyl Ketones as a Novel Class of Dopamine Transporter Inhibitors. *Bioorg. Med. Chem. Let.* **2001,** *9,* 1753 - 1764.

46. Kolhatkar, R.; Cook, C. D.; Ghorai, S. K.; Deschamps, J.; Beardsley, P. M.; Reith, M. E. A.; Dutta, A. K. Further Structurally Constrained Analogues of *cis-*(6-Benzhydrylpiperidin-3-yl)benzylamine with Elucidation of Bioactive Conformations: Discovery of 1,4-Diazabicyclo[3.3.1]nonane Derivatives and Evaluation of Their Biological Properties for the Monoamine Transporters. *J. Med. Chem.* **2004,** *47,* 5101-5113.

47. Nicklaus, M. C.; Wang, S.; Driscoll, J.; Milne, G. W. A. Conformational Changes of Small Molecules Binding to Proteins. *Bioorg. Med. Chem.* **1995,** *3,* 411-428.

48. Veith, M.; Hirst, J. D.; Brooks III, C. L. Do Active Site Conformations of Small Ligands Correspond to Low Free-Energy Solution Structures? *J. Comput. Aided Mol. Des.* **1998,** *12,* 563 - 572.

49. Boström, J.; Norrby, P.-O.; Liljefors, T. Conformational Energy Penalties of Protein-Bound Ligands. *J. Comput. Aided Mol. Des.* **1998,** *12,* 383-396.

50. Debnath, A. K. Comparative Molecular Field Analysis (CoMFA) of a Series of Symmetrical Bis-Benzamide Cyclic Urea Derivatives as HIV-1 Protease Inhibitors. *J. Chem. Inf. Comp. Sci.* **1999,** *38,* 761-767.

51. Debnath, A. K. Three-Dimensional Quantitative Structure-Activity Relationship Study on Cyclic Urea Derivatives as HIV-1 Protease Inhibitors: Application of Comparative Molecular Field Analysis. *J. Med. Chem.* **1999,** *42,* 249-259.

52. Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization Upon Binding. *J. Med. Chem.* **2004,** *47,* 2499-2510.

53. Benedetti, P.; Mannhold, R.; Cruciani, G.; Pastor, M. GBR Compounds and Mepyramines as Cocaine Abuse Therapeutics: Chemometric Studies on Selectivity using Grid Independent Descriptions. *J. Med. Chem.* **2002,** *45,* 1577-1584.

54. Guarnieri, F.; Weinstein, H. Conformational Memories and the Exploration of Biologically Relevant Peptide Conformations: An Illustration for the Gonadotropin-Releasing Hormone. *J. Am. Chem. Soc.* **1996,** *118,* 5580-5589.

55. 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.

56. Barnett-Norris, J.; Guarnieri, F.; Hurst, D. P.; Reggio, P. H. Exploration of Biologically Relevant Conformations of Anandamide, 2-Arachinodylglycerol, and Their Analogues Using Conformational Memories. *J. Med. Chem.* **1998,** *41,* 4861-4872.

57. Barnett-Norris, J.; Hurst, D. P.; Lynch, D. L.; Guarnieri, F.; Makriyannis, A.; Reggio, P. H. Conformational Memories and the Endocannabinoid Binding Site at the Cannabinoid CB1 Receptor. *J. Med. Chem.* **2002,** *45,* 3649-3659.

58. Greenidge, P. A.; Merette, S. A. M.; Beck, R.; Dodson, G.; Goodwin, C. A.; Scully, M. F.; Spencer, J.; Weiser, J.; Deadman, J. J. Generation of Ligand Conformations in Continuum Solvent Consistent with Protein Active Site Topology: Application to Thrombin. *J. Med. Chem.* **2003,** *46,* 1293-1305.

59. Bernard, D.; Coop, A.; MacKerell Jr., A. D. 2D Conformationally Sampled Pharmacophore: A Ligand-Based Pharmacophore to Differentiate δ Opioid Agonists from Antagonists. *J. Am. Chem. Soc.* **2003,** *125,* 3101-3107.

60. Gilbert, K. M.; Skawinski, W. J.; Misra, M.; Paris, K. A.; Naik, N. H.; Deutsch, H. M.; Venanzi, C. A. Conformational Analysis of Methylphenidate: Comparison of Molecular Orbital and Molecular Mechanics Methods. *J. Comput. Aided Mol. Des.* **2004,** *18,* 719-738.

61. Pandit, D.; Misra, M.; Gilbert, K. M.; Matecka, D.; Prisinzano, T.; Rice, K. C.; Venanzi, C. A. Unpublished results. **2004,** New Jersey Institute of Technology.

62.   Gilbert, K. M.; Venanzi, C. A. Unpublished results. **2004**, New Jersey Institute of Technology.

63.   Buolamwini, J. K.; Assefa, H. CoMFA and CoMSIA 3D QSAR and Docking Studies on Conformationally-Restrained Cinnamoyl HIV-1 Integrase Inhibitors: Exploration of a Binding Mode at the Active Site. *J. Med. Chem.* **2002**, *45*, 841-852.

64.   Vaughan, R. A.; Kuhar, M. J. Dopamine Transporter Ligand Binding Domains. Structural and Functional Properties Revealed by Limited Proteolysis. *J. Biol. Chem.* **1996**, *271*, 21672-21680.

65.   Chen, N.; Vaughan, R. A.; Reith, M. E. A. The Role of Conserved Tryptophan and Acidic Residues in the Human Dopamine Transporter as Characterized by Site-Directed Mutagenesis. *J. Neurochem.* **2001**, *77*, 1116-1127.

66.   Chen, N., Zhen, J., Reith, M.E.A. Mutation of Trp84 and Asp313 of the Dopamine Transporter Reveals Similar Mode of Binding Interaction for GBR 12909 and Benztropine as Opposed to Cocaine. *J. Neurochem.* **2004**, *89*, 853-864.

67.   Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.

68.   Pastor, M., Cruciani, G., McLay, I., Pickett, S., Clementi, S. Grid Independent Descriptors (GRIND). A Novel Class of Alignment-Independent Three-Dimensional Descriptors. *J. Med. Chem.* **2000**, *43*, 3233-3242.

69.   Volkow, N. D.; Wang, G.-J.; Fowler, J. S.; Fischman, M.; Foltin, R.; Abumrad, N. N.; Gatley, S. J.; Logan, E.; Wong, C.; Gifford, A.; Ding, Y.-S.; Hintzemann, R.; Pappas, N. Methylphenidate and Cocaine Have Similar *In Vivo* Potency to Block Dopamine Transporters in the Human Brain. *Life Sci.* **1999**, *65*, PL7-PL12.

70.   Carroll, F. I.; Howell, L. L.; Kuhar, M. J. Pharmacotherapies for Treatment of Cocaine Abuse: Preclinical Aspects. *J. Med. Chem.* **1999**, *42*, 2721-2736.

71.   Rothman, R. B.; Mele, A.; Reid, A. A.; Akunne, H.; Greig, N.; Thurkauf, A.; Rice, K. C.; Pert, A. Tight Binding Dopamine Reuptake Inhibitors as Cocaine Antagonists: A Strategy for Drug Development. *FEBS Letters* **1989**, *257*, 341-344.

72.   Cheng, Y.; Prusoff, W. H. Relationship Between the Inhibition Constant (K1) and the Concentration of Inhibitor Which Causes 50 Per Cent Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099-3108.

73.   Greiner, E.; Prisinzano, T.; Johnson II, E. M.; Dersch, C. M.; Marcus, J.; Partilla, J. S.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. Structure-Activity Relationship Studies of Highly Selective Inhibitors of the Dopamine Transporter:

N-Benzylpiperidine Analogues of [1-[2-[Bis(4-fluorophenyl)methoxy]ethyl]-4-(3-phenylpropyl)piperazine. *J. Med. Chem.* **2003**, *46*, 1465-1469.

74.   Dutta, A. K.; Coffey, L. L.; Reith, M. E. A. Highly Selective, Novel Analogs of 4-[2-(Diphenylmethoxy)ethyl]-1-benzylpiperidine for the Dopamine Transporter: Effect of Different Aromatic Substitutions on Their Affinity and Selectivity. *J. Med. Chem.* **1997**, *40*, 35-43.

75.   Ritz, M. C.; Boja, J. W.; Grigoriadis, D. E.; Zacek, R.; Carroll, F. I.; Lewin, A. H.; Kuhar, M. J. [³H]WIN 35,065-2: A Ligand for Cocaine Receptors in Striatum. *J. Neurochem.* **1990**, *55*, 1556-1562.

76.   Scheffel, U.; Boja, J. W.; Kuhar, M. J. Cocaine receptors: *In Vivo* Labeling with [³H]-(-)cocaine, [³H] WIN 35,065-2 and [³H]WIN 35,428. *Synapse* **1989**, *4*, 390-392.

77.   Madras, B. K.; Spealman, R. D.; Fahey, M. A.; Neumeyer, J. L.; Saha, J. K.; Milius, R. A. Cocaine Receptors Labeled by [³H]2β–Carbomethoxy-3β–(4–fluorophenyl)tropane. *Mol. Pharmacol.* **1989**, *36*, 518-524.

78.   Boja, J. W.; Patel, A.; Carroll, F. I.; Rahman, M. A.; Phillip, A.; Lewin, A. H.; Kopajtic, T. A.; Kuhar, M. J. [¹²⁵I]RTI-55: A Potent Ligand for Dopamine Transporters. *Eur. J. Pharmacol.* **1991**, *194*, 133-134.

79.   Singh, S., Basmadjian, G.P., Avor, K.S., Pouw, B., Seale, T.W. Synthesis and Ligand Binding Studies of 4'-Iodobenzoyl Esters of Tropanes and Piperidines at the Dopamine Transporter. *J. Med. Chem.* **1997**, *40*, 2474-2481.

80.   Silverthorn, M. L., Dersch, C.M., Baumann, M.H., Cadet, J.L., Partilla, J.S., Rice, K.C., Carroll, F.I., Becketts, K.M., Brockington, A., Rothman, R.B. Studies of the Biogenic Amine Transporters. V. Demonstration of Two Binding Sites for the Cocaine Analog [¹²⁵I]RTI-55 Associated with the 5-HT Transporter in Rat Brain Membranes. *J. Pharm. Expt. Ther.* **1995**, *273*, 213-222.

81.   Clarke, R. L.; Daum, S. J.; Gambino, A. J.; Aceto, M. D.; Pearl, J.; Levitt, M.; Kumisky, W. R.; Bogado, E. F. Compounds Affecting the Central Nervous System. 4. 3ß-Phenyltropane-2-carboxylic Esters and Analogs. *J. Med. Chem.* **1973**, *16*, 1260-1267.

82.   Madras, B. K.; Kamien, J. B.; Fahey, M. A.; Canfield, D. R.; Milius, R. A.; Saha, J. K.; Neumeyer, J. L.; Spealman, R. D. N-Modified Fluorophenyltropane Analogs of Cocaine with High Affinity for Cocaine Receptors. *Pharmacol. Biochem. Behav.* **1990**, *35*, 949-953.

83.   Balster, R. L.; Carroll, F. I.; Graham, J. H.; Mansbach, R. S.; Rahman, M. A.; Phillip, A.; Lewin, A. H.; Showalter, V. M. Potent Substituted-3β-phenyltropane Analogs of Cocaine Have Cocaine-Like Discriminative Stimulus Effects. *Drug Alcohol Dep.* **1991**, *29*, 145-151.

84. Milius, R. A.; Saha, J. K.; Madras, B. K.; Neumeyer, J. L. Synthesis and Receptor Binding of N-Substituted Tropane Derivatives. High-Affinity Ligands for the Cocaine Receptor. *J. Med. Chem.* **1991**, *34*, 1728-1731.

85. Carroll, F. I.; Abraham, P.; Lewin, A.; Parham, K. A.; Boja, J. W.; Kuhar, M. J. Isopropyl and Phenyl Esters of 3β-(4-Substituted phenyl)tropane-2β–carboxylic acids. Potent and Selective Compounds for the Dopamine Transporter. *J. Med. Chem.* **1992**, *35*, 2497-2500.

86. Carroll, F. I.; Gray, J. L.; Abraham, P.; Kuzemko, M. A.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. 3-Aryl-2-(3'-substiuted-1',2',4'-oxadiazol-5'-yl)tropane Analogues of Cocaine: Affinities at the Cocaine Binding Site at the Dopamine, Serotonin, and Norepinephrine Transporters. *J. Med. Chem.* **1993**, *36*, 2886-2890.

87. Meltzer, P. C.; Liang, A. Y.; Brownell, A.-L.; Elmaleh, D. R.; Madras, B. K. Substituted 3-Phenyl Tropane Analogs of Cocaine: Synthesis, Inhibition, of Binding at Cocaine Recognition Sites, and Positron Emission Tomography of Imaging. *J. Med. Chem.* **1993**, *36*, 855-862.

88. Wall, S. C.; Innis, R. B.; Rudnick, G. Binding of the Cocaine Analog 2 ß-Carbomethoxy-3 ß-(4-[125I]iodophenyl)tropane to Serotonin and Dopamine Transporters: Different Ionic Requirements for Substrate and 2 ß-Carbomethoxy-3 ß-(4-[125I]iodophenyl)tropane Binding. *Mol. Pharmacol.* **1993**, *43*, 264-270.

89. Boja, J. W.; Kuhar, M. J.; Kopajtic, T.; Yang, E.; Abraham, P.; Lewin, A. H.; Carroll, F. I. Secondary Amine Analogues of 3β-(4'-Substituted phenyl)tropane-2β-carboxylic Acid Esters and N-Norcaine Exhibit Enhanced Affinity for Serotonin and Norepinephrine Transporters. *J. Med. Chem.* **1994**, *37*, 1220-1223.

90. 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 2ß-Substituted 3ß-Phenyltropanes. *J. Med. Chem.* **1994**, *37*, 3875-3877.

91. Neumeyer, J. L.; Wang, S.; Gao, Y.; Milius, R. A.; Kula, N. S.; Campbell, A.; Baldessarini, R. J.; Zea-Ponce, Y.; Baldwin, R. M.; Innis, R. B. N-ω–Fluoroalkyl Analogs of (1R)-2ß-Carbomethoxy-3ß-(4-iodophenyl)-tropane (ßCIT): Radiotracers for Positron Emission Tomography and Single Photon Emission Computed Tomography of Dopamine Transporters. *J. Med. Chem.* **1994**, *37*, 1558-1561.

92. Boja, J. W.; Cadet, J. L.; Kopajtic, T. A.; Lever, J.; Seltzman, H. H.; Wyrick, C. D.; Lewin, A. H.; Abraham, P.; Carroll, F. I. Selective Labeling of the Dopamine Transporter by the High Affinity Ligand 3ß-(4-[125I]Iodophenyl)tropane-2ß-carboxylic Acid Isopropyl Ester. *Mol. Pharmacol.* **1995**, *47*, 779-786.

93. Carroll, F. I.; Kotian, P.; Dehghani, A.; Gray, J. L.; Kuzemko, M. A.; Parham, K. A.; Abraham, P.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. Cocaine and 3ß-(4'-Substituted Phenyl)tropane-2ß-Carboxylic Acid Ester and Amide Analogues. New

High-Affinity and Selective Compounds for the Dopamine Transporter. *J. Med. Chem.* **1995**, *38*, 379-388.

94. Kotian, P.; Abraham, P.; Lewin, A. H.; Mascarella, S. W.; Boja, J. W.; Kuhar, M. J.; Carroll, F. I. Synthesis and Ligand Binding Study of 3ß-(4'-Substituted phenyl)-2ß-(heterocyclic)tropanes. *J. Med. Chem.* **1995,** *38*, 3451-3453.

95. Xu, C.; Coffey, L. L.; Reith, M. E. A. Translocation of Dopamine and Binding of 2ß-carbomethoxy-3ß-(4-fluorophenyl)tropane (WIN 35,428) Measured Under Identical Conditions in Rat Striatal Synaptosomal Preparations. Inhibition by Various Blockers. *Biochem. Pharmacol.* **1995,** *49*, 339-350.

96. Holmquist, C. R.; Keverline-Frantz, K. I.; Abraham, P.; Boja, J. W.; Kuhar, M. J.; Carroll, F. I. 3α-(4'-Substituted phenyl)tropane-2β-carboxylic Acid Methyl Esters: Novel Ligands with High Affinity and Selectivity at the Dopamine Transporter. *J. Med. Chem.* **1996,** *39*, 4139-4141.

97. Kotian, P.; Mascarella, S. W.; Abraham, P.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J.; Carroll, F. I. Synthesis, Ligand Binding, and Quantitative Structure-Activity Relationship Study of 3β-(4'-Substituted phenyl)-2β-heterocyclic Tropanes: Evidence for an Electrostatic Interaction at the 2β-Position. *J. Med. Chem.* **1996,** *39*, 2753-2763.

98. Neumeyer, J. L.; Tamagnan, G.; Wang, S.; Gao, Y.; Milius, R. A.; Kula, N. S.; Baldessarini, R. J. *N*-Substituted Analogs of 2β–carbomethoxy-3β-(4'-iodophenyl)tropane (β-CIT) with Selective Affinity to Dopamine or Serotonin Transporters in Rat Forebrain. *J. Med. Chem.* **1996,** *39*, 543-548.

99. Reith, M. E. A.; Xu, C.; Coffey, L. L. Binding Domains for Blockers and Substrates on the Cloned Human Dopamine Transporter Studied by Protection Against *N*-Ethylmaleimide-Induced Reduction of 2β-Carbomethoxy-3β–(4–fluorophenyl)[³H]tropane ([³H]WIN 35,428) Binding. *Biochem. Pharmacol.* **1996,** *52*, 1435-1446.

100. Blough, B. E.; Abraham, P.; Mills, A. C.; Lewin, A. H.; Boja, J. W.; Scheffel, U.; Kuhar, M. J.; Carroll, F. I. 3β-(4-Ethyl-3-iodophenyl)nortropane-2β-carboxylic Acid Methyl Ester as a High-Affinity Selective Ligand for the Serotonin Transporter. *J. Med. Chem.* **1997,** *40*, 3861-3864.

101. Chang, A.-C.; Burgess, J. P.; Mascarella, S. W.; Abraham, P.; Kuhar, M. J.; Carroll, F. I. Synthesis and Transporter Binding Properties of 2,3-Diphenyltropane Stereoisomers. Comparison to 3β-Phenyltropane-2β-carboxylic Acid Esters. *J. Med. Chem.* **1997,** *40*, 1247-1251.

102. Emond, P.; Garreau, L.; Chalon, S.; Boazi, M.; Caillet, M.; Bricard, J.; Frangin, Y.; Mauclaire, L.; Besnard, J.-C.; Guilloteau, D. Synthesis and Ligand Binding of Nortropane Derivatives: *N*-Substituted 2β-Carbomethoxy-3β-(4'-

iodophenyl)nortropane and *N*-(3-Iodoprop-(2*E*-enyl)-2βcarbomethoxy-3β-(3',4'-disubstituted phenyl)nortropane. New High-Affinity and Selective Compounds for the Dopamine Transporter. *J. Med. Chem.* **1997,** *40,* 1366-1372.

103.   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.

104.   Nader, M. A.; Grant, K. A.; Davies, H. M. L.; Mach, R. H.; Childers, S. R. The Reinforcing and Discriminative Stimulus Effects of the Novel Cocaine Analog 2β-Propanoyl-3β-(4-Tolyl)-Tropane in Rhesus Monkeys. *J. Pharm. Expt. Ther.* **1997,** *280,* 541-550.

105.   Thiruvazhi, M.; Abraham, P.; Kuhar, M. J.; Carroll, F. I. Synthesis of the Isomers of (1*R*)-3-(phenylthio)tropane-2-carboxylic Acid Methyl Ester. A New Class of Ligands for the Dopamine Transporter. *Chem. Commun.* **1997,** 555-556.

106.   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 3β-Phenyltropane Derivatives. *J. Med. Chem.* **1997,** *40,* 858-863.

107.   Bennett, B. A.; Hollingsworth, C. K.; Martin, R. S.; Childers, S. R.; Ehrenkaufer, R. E.; Porrino, L. J.; Davies, H. M. L. Prolonged Dopamine and Serotonin Transporter Inhibition After Exposure to Tropanes. *Neuropharmacol.* **1998,** *37,* 123-130.

108.   Birmingham, A. M.; Nader, S. H.; Grant, K. A.; Davies, H. M. L.; Nader, M. A. Further Evaluation of the Reinforcing Effects of the Novel Cocaine Analog 2β-Propanoyl-3β-(4-tolyl)-tropane (PTT) in Rhesus Monkeys. *Psychopharmacol.* **1998,** *136,* 139-147.

109.   Carroll, F. I.; Lewin, A. H.; Kuhar, M. J. 3β-Phenyl-2β-Substituted Tropanes - An SAR Analysis. *Med. Chem. Res.* **1998,** *8,* 59-65.

110.   Daunais, J. B.; Hart, S. L.; Smith, H. R.; Letchworth, S. R.; Davies, H. M. L.; Sexton, T.; Bennett, B. A.; Childers, S. R.; Porrino, L. J. Long-Acting Blockade of Biogenic Amine Transporters in Rat Brain by Administration of the Potent Novel Tropane 2β-Propanoyl-3β–(2–Napthyl)-Tropane. *J. Pharm. Expt. Ther.* **1998,** *285,* 1246-1254.

111.   Jiang, S.; Chang, A.-C.; Abraham, P.; Kuhar, M. J.; Carroll, F. I. Synthesis and Transporter Binding Properties of (*R*)-2β,3β- and (*R*)-2α,3α-Diaryltropanes. *Bioorg. Med. Chem. Let.* **1998,** *8,* 3689-3692.

112. Kozikowski, A. P.; Araldi, G. L.; Prakash, K. R. C.; Zhang, M.; Johnson, K. M. Synthesis and Biological Properties of New 2β-Alkyl- and 2β-Aryl-3-(Substituted phenyl)tropane Derivatives: Stereochemical Effect of C-3 on Affinity and Selectivity for Neuronal Dopamine and Serotonin Transporters. *J. Med. Chem.* **1998,** *41,* 4973-4982.

113. Neumeyer, J. L.; Baldessarini, R. J.; Kula, N. S.; Tamagnan, G. Novel Phenyltropanes: Affinity to Monoamine Transporters in Rat Forebrain. In *Progress in Alzheimer's and Parkinson's Diseases*; Fisher, A., Hanin, I., Yoshida, M. (ed.); Plenum Press: New York, 1998; pp. 739-745.

114. Prakash, K. R. C.; Araldi, G. L.; Smith, M. P.; Zhang, M.; Johnson, K. M.; Kozikowski, A. P. Synthesis and Biological Activity of New 6- and 7-Substituted 2β-Butyl-3-Phenyltropanes as Ligands for the Dopamine Transporter. *Med. Chem. Res.* **1998,** *8,* 43-58.

115. Chalon, S.; Garreau, L.; Emond, P.; Zimmer, L.; Vilar, M.-P.; Besnard, L.-C.; Guilloteau, D. Pharmacological Characterization of (*E*)-*N*-(3-iodoprop-2-enyl)-2β-Carbomethoxy-3β-(4'-methylphenyl)nortropane as a Selective and Potent Inhibitor of Neuronal Dopamine Transporter. *J. Pharm. Expt. Ther.* **1999,** *291,* 648-654.

116. Kuhar, M. J.; McGirr, K. M.; Hunter, R. G.; Lambert, P. D.; Garrett, B. E.; Carroll, F. I. Studies of Selected Phenyltropanes at Monoamine Transporters. *Drug Alcohol Dep.* **1999,** *56,* 9-15.

117. Lewin, A. H.; Sorensen, J. B.; Dustman, J. A.; Boweb, J. P. Computational Methods for Conformational Analysis of Unsymmetrical 1,3-Diamines: 3-Aminotropanes. *J. Comput. Chem.* **1999,** *20,* 1371-1378.

118. Freedland, C. S.; Smith, H. R.; Hart, S. L.; Daunais, J. B.; Davies, H. M. L.; Porrino, L. J. A Comparison of the Behavioral Effects of the Repeated Administration of PTT, 2-Propanoyl-3β-(4-tolyl)tropane and Cocaine. *Brain Res.* **2000,** *869,* 98-104.

119. Zhao, L.; Johnson, K. M.; Zhang, M.; Flippen-Anderson, J. L.; Kozikowski, A. P. Chemical Synthesis and Pharmacology of 6- and 7-Hydroxylated 2-Carbomethoxy-3-(*p*-tolyl)tropanes: Antagonism of Cocaine's Locomotor Stimulant Effects. *J. Med. Chem.* **2000,** *43,* 3283-3294.

120. Blough, B. E.; Keverline, K. I.; Nie, Z.; Navarro, H.; Kuhar, M. J.; Carroll, F. I. Synthesis and Transporter Binding Properties of 3β-[4'-(Phenylalkyl, -phenylalkenyl, and -phenylalkynl)phenyl] tropane-2 β-carboxylic Acid Methyl Esters: Evidence of a Remote Phenyl Binding Domain on the Dopamine Transporter. *J. Med. Chem.* **2002,** *45,* 4029-4037.

121. Xu, L.; Izenwasser, S.; Katz, J. L.; Kopajtic, T.; Klein-Stevens, C.; Zhu, N.; Lomenzo, S. A.; Winfield, L.; Trudell, M. L. Synthesis and Biological Evaluation of 2-Substituted 3β-Tolytropane Derivatives at Dopamine, Serotonin and Norepinephrine Transporters. *J. Med. Chem.* **2002,** *45,* 1203-1210.

122. Kim, D. I.; Schweri, M. M.; Deutsch, H. M. Synthesis and Pharmacology of Site Specific Cocaine Abuse Treatment Agents: 8-Substituted Isotropane (3-Azabicyclo[3.2.1]octane) Dopamine Uptake Inhibitors. *J. Med. Chem.* **2003,** *46,* 1456-1464.

123. Carroll, F. I.; Runyon, S. P.; Abraham, P.; Navarro, H.; Kuhar, M. J.; Pollard, G. T.; J.L., H. Monoamine Transporter Binding, Locomotor Activity, and Drug Discrimination Properties of 3-(4-Substituted-phenyl)tropane-2-carboxylic Acid Methyl Ester Isomers. *J. Med. Chem.* **2004,** *47,* 6401-6409.

124. Kozikowski, A. P.; Araldi, G. L.; Boja, J.; Meil, W. M.; Johnson, K. M.; Flippen-Anderson, J. L.; George, C.; Saiah, E. Chemistry and Pharmacology of Piperidine-Based Analogues of Cocaine. Identification of Potent DAT Inhibitors Lacking the Tropane Skeleton. *J. Med. Chem.* **1998,** *41,* 1962-1969.

125. Newman, A. H.; Allen, A. C.; Izenwasser, S.; Katz, J. L. Novel 3α-(Diphenylmethoxy)tropane Analogs: Potent Dopamine Uptake Inhibitors Without Cocaine-Like Behavioral Profiles. *J. Med. Chem.* **1994,** *37,* 2258-2261.

126. Newman, A. H.; Kline, R. H.; Allen, A. C.; Izenwasser, S.; George, C.; Katz, J. L. Novel 4'-Substituted and 4',4"-Disubstituted 3α-(Diphenylmethoxy)tropane Analogs as Potent and Selective Dopamine Uptake Inhibitors. *J. Med. Chem.* **1995,** *38,* 3933-3940.

127. Meltzer, P. C.; Liang, A. Y.; Madras, B. K. 2-Carbomethoxy-3-(diarylmethoxy)-1αH, 5αH-tropane Analogs: Synthesis and Inhibition of Binding at the Dopamine Transporter and Comparison with Piperazines of the GBR Series. *J. Med. Chem.* **1996,** *39,* 371-379.

128. Agoston, G. E.; Wu, J. H.; Izenwasser, S.; George, C.; Katz, J.; Kline, R. H.; Newman, A. H. Novel *N*–Substituted 3α–[Bis(4'-fluorophenyl)methoxy]tropane Analogues: Selective Ligands for the Dopamine Transporter. *J. Med. Chem.* **1997,** *40,* 4329-4339.

129. Kline, R. H.; Izenwasser, S.; Katz, J. L.; Joseph, D. B.; Bowen, W. D.; Newman, A. H. 3'-Chloro-3α-(diphenylmethoxy)tropane But Not 4'-Chloro-3α–(diphenylmethoxy)tropane Produces a Cocaine-Like Behavioral Profile. *J. Med. Chem.* **1997,** *40,* 851-857.

130. Newman, A. H.; Agoston, G. E. Novel Benztropine [3α-(diphenylmethoxy)tropane] Analogs as Probes for the Dopamine Transporter. *Curr. Med. Chem.* **1998,** *5,* 305-319.

131. Newman, A. H.; Robarge, M. J.; Howard, I. M.; Wittkopp, S. L.; George, C.; Kopajtic, T.; Izenwasser, S.; Katz, J. L. Structure-Activity Relationships at Monoamine Transporters and Muscarinic Receptors for *N*-Substituted-3α-(3'-chloro-,4'-chloro-, and 4',4"-dichloro-Substituted-diphenyl)methoxytropanes. *J. Med. Chem.* **2001**, *44*, 633-640.

132. Zou, M.; Kopajtic, T.; Katz, J. L.; Wirtz, S.; Justice, J., J.B.; Newman, A. H. Novel Tropane-Based Irreversible Ligands for the Dopamine Transporter. *J. Med. Chem.* **2001**, *44*, 4453 - 4461.

133. Zou, M.-F.; Kopajtic, T.; Katz, J. L.; Newman, A. H. Structure-Activity Relationship Comparison of (S)-2β-Substituted 3α-(Bis[4-fluorophenyl]methoxy)tropanes and (R)-2β-Substituted 3β-(3,4-Dichlorophenyl)tropanes at the Dopamine Transporter. *J. Med. Chem.* **2003**, *46*, 2908-2916.

134. Heikkila, R. E.; Babington, R. G.; Houlihan, W. J. Pharmacological Studies with Several Analogs of Mazindol: Correlation Between Effects on Dopamine Uptake and Various In Vivo Responses. *Eur. J. Pharmacol.* **1981**, *71*, 277-286.

135. Javitch, J. A.; Blaustein, R. O.; Snyder, S. H. [³H]Mazindol Binding Associated With Neuronal Dopamine and Norepinephrine Uptake Sites. *Mol. Pharmacol.* **1984**, *26*, 33-44.

136. Zimanyi, I.; Lajtha, A.; Reith, M. E. A. Comparison of Characteristics of Dopamine Uptake and Mazindol Binding in Mouse Striatum. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1989**, *340*, 626-632.

137. Kozikowski, A. P.; Xiang, L.; Tanaka, J.; Bergmann, J. S.; Johnson, K. M. Use of Nitrile Oxide Cycloaddition (NOC) Chemistry in the Synthesis of Cocaine Analogues: Mazindol Binding and Dopamine Uptake Studies. *Med. Chem. Res.* **1991**, *1*, 312-321.

138. Houlihan, W. J.; Boja, J. W.; Parrino, V. A.; Kopajtic, T. A.; Kuhar, M. J. Halogenated Mazindol Analogs as Potential Inhibitors of the Cocaine Binding Site at the Dopamine Transporter. *J. Med. Chem.* **1996**, *39*, 4935-4941.

139. Xu, C.; Reith, M. E. A. WIN 35,428 and Mazindol Are Mutually Exclusive in Binding to the Cloned Human Dopamine Transporter. *J. Pharm. Expt. Ther.* **1997**, *282*, 920-927.

140. Houlihan, W. J.; Boja, J. W.; Kopajtic, T. A.; Kuhar, M. J.; Degrado, S. J.; Toledo, L. Positional Isomers and Analogs of Mazindol as Potential Inhibitors of the Cocaine Binding Site on the Dopamine Transporter Site. *Med. Chem. Res.* **1998**, *8*, 77-90.

141. Houlihan, W. J.; Kelly, L.; Pankuch, J.; Koletar, J.; Brand, L.; Janowsky, A.; Kopajtic, T. Mazindol Analogues as Potential Inhibitors of the Cocaine Binding Site at the Dopamine Transporter. *J. Med. Chem.* **2002,** *45*, 4097-4109.

142. Houlihan, W. J.; Ahmad, U. F.; Koletar, J.; Kelly, L.; Brand, L.; Kopajtic, T. A. Benzo- and Cyclohexanomazindol Analogues as Potential Inhibitors of the Cocaine Binding Site at the Dopamine Transporter. *J. Med. Chem.* **2002,** *45*, 4110-4118.

143. Patrick, K. S.; Caldwell, R. W.; Ferris, R. M.; Breese, G. R. Pharmacology of the Enantiomers of *threo*-Methylphenidate. *J. Pharm. Expt. Ther.* **1987,** *241*, 152-158.

144. Ding, Y.-S.; Gatley, S. J.; Thanos, P. K.; Shea, C.; Garza, V.; Xu, Y.; Carter, P.; King, P.; Warner, D.; Taintor, N. B.; Park, D. J.; Pyatt, B.; Fowler, J. S.; Volkow, N. D. Brain Kinetics of Methylphenidate (Ritalin) Enantiomers After Oral Administration. *Synapse* **2004,** *53*, 168-175.

145. Deutsch, H. M.; Shi, Q.; Gruszecka-Kowalik, E.; Schweri, M. Synthesis and Pharmacology of Potential Cocaine Antagonists. 2. Structure-Activity Relationship Studies of Aromatic Ring-Substituted Methylphenidate Analogs. *J. Med. Chem.* **1996,** *39*, 1201-1209.

146. Froimowitz, M.; Deutsch, H. M.; Shi, Q.; Wu, K.-M.; Glaser, R.; Adin, I.; George, C.; Schweri, M. M. Further Evidence for a Dopamine Reuptake Pharmacophore. The Effect of N-methylation on *Threo*-Methylphenidate and Its Analogs. *Bioorg. Med. Chem. Let.* **1997,** *7*, 1213-1218.

147. Deutsch, H. M. Structure-Activity Relationships for Methylphenidate Analogs and Comparisons to Cocaine and Tropanes. *Med. Chem. Res.* **1998,** *8*, 91-99.

148. Froimowitz, M.; Wu, K.-M.; George, C.; VanDerveer, D.; Shi., Q.; Deutsch, H. M. Crystal Structures of Analogs of *Threo*-Methylphenidate. *Struct. Chem.* **1998,** *4*, 295-303.

149. Glaser, R.; Adin, I.; Shiftan, D.; Shi, Q.; Deutsch, H. M.; George, G.; Wu, K.-M.; Froimowitz, M. Solution and Solid-State Conformational and Structural Analysis of the *N*-Methyl Derivatives of (±)-*Threo*-Methylphenidate, (±)-*Erythro*-Methylphenidate and (±)-*Threo-p*-methyl-methylphenidate HCl Salts. *J. Org. Chem.* **1998,** *63*, 1785-1794.

150. Thai, D. L.; Sapko, M. T.; Reiter, C. T.; Bierer, D. E.; Perel, J. M. Asymmetric Synthesis and Pharmacology of Methylphenidate and Its Para-Substituted Derivatives. *J. Med. Chem.* **1998,** *41*, 591-601.

151. Deutsch, H. M.; Dunn, T.; Ye, X.; Schweri, M. M. Synthesis and Pharmacology of Site-Specific Cocaine Abuse Treatment Agents: The Role of the Phenyl Group

in Highly Modified Methylphenidate Analogs as Dopamine Reuptake Inhibitors. *Med. Chem. Res.* **1999,** *9,* 213-222.

152.   Deutsch, H. M.; Ye, X.; Shi, Q.; Liu, Z.; Schweri, M. M. Synthesis and Pharmacology of Site Specific Cocaine Abuse Treatment Agents: A New Synthetic Methodology for Methylphenidate Analogs Based on the Blaise Reaction. *Eur. J. Med. Chem.* **2001,** *36,* 303-311.

153.   Schweri, M. M.; Deutsch, H. M.; Massey, A. T.; Holtzman, S. G. Biochemical and Behavioral Characterization of Novel Methylphenidate Analogs. *J. Pharm. Expt. Ther.* **2002,** *301,* 527-535.

154.   Meltzer, P. C.; Wang, P.; Blundell, P.; Madras, B. K. Synthesis and Evaluation of Dopamine and Serotonin Transporter Inhibition by Oxacyclic and Carbacyclic Analogues of Methylphenidate. *J. Med. Chem.* **2003,** *46,* 1538-1545.

155.   Misra, M. Comparative Molecular Field Analysis (CoMFA) of Phenyl Ring Substituted Methylphenidates. Master's Thesis, New Jersey Institute of Technology, Newark, NJ, 1999.

156.   Gilbert, K. M. Comparative Molecular Field Analysis (CoMFA) of Protonated Methylphenidate Phenyl-Substituted Analogs. Master's Thesis, New Jersey Institute of Technology, Newark, NJ, 2002.

157.   Deutsch, H. M.; Kim, D. I.; Holtzman, S. G.; Schweri, M. M.; Spealman, R. D. The Synthesis and Evaluation of New Methylphenidates: Restricted Rotation Analogs, Preliminary Results, *College on the Problems of Drug Dependence, 64th Annual Meeting,* Quebec City, Canada, June 9-13, 2002.

158.   Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972,** *56,* 2257-2261.

159.   Stephens, P. J., Devlin, F.J., Chabalowski, C.F., Frisch, M.J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994,** *98,* 11623-11627.

160.   Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. E.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985,** *107,* 3902-3909.

161.   Dewar, M. J. S.; Zoebisch, E. G. Extension of AM1 to the Halogens. *J. Molec. Struct. (Theochem)* **1988,** *49,* 1-21.

162.   Dewar, M. J. S.; Yate-Ching, Y. AM1 Parameters for Sulfur. *Inorg. Chem.* **1990,** *29,* 3881-3890.

163. Clark, M.; Cramer III, R. D.; Van Opdenbosch, N. Validation of the General Purpose TRIPOS 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982-1012.

164. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J., J.A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Menucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T. A.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.11.3*, Gaussian Inc.: Wallingford, CT, 2002.

165. van der Zee, P.; Koger, H. S.; Gootjes, J.; Hespe, W. Aryl 1,4-dialk(en)ylpiperizines as Selective and Very Potent Inhibitors of Dopamine Uptake. *Eur. J. Med. Chem.* **1980**, *15*, 363-370.

166. Matecka, D.; Rice, K. C.; Rothman, R. B.; de Costa, B. R.; Glowa, J. R.; Wojnicki, F. H.; Pert, A.; George, C.; Carroll, F. I.; Silverthorn, M. L.; Dersch, C. M.; Becketts, K. M.; Partilla, J. S. Synthesis and Absolute Configuration of Chiral Piperazines Related to GBR 12909 as Dopamine Reuptake Inhibitors. *Med. Chem. Res.* **1994**, *5*, 43-53.

167. Matecka, D.; Radesca, L.; de Costa, B.; Rothman, R. B.; Dersch, C.; Akunne, H.; Lewis, B.; Partilla, J.; Xu, H.; Pert, A.; Rice, K. C. Synthesis, Receptor Binding and Behavioral Studies of N-(2-Diphenylmethoxyethyl)-N'-(3-phenylpropyl)homopiperazine, A Novel GBR 12935 Analog. *NIDA Res. Monogr.* **1993**, *132*.

168. McKinney, J. D.; Richard, A.; Waller, C.; Newman, M. C.; Gerberick, F. The Practice of Structure Activity Relationships (SAR) in Toxicology. *Toxicol. Sci.* **2000**, *56*, 8-17.

169. Deutsch, H. M.; Schweri, M. M.; Culbertson, C. T.; Zalkow, L. H. Synthesis and Pharmacology of Irreversible Affinity Labels as Potential Cocaine Antagonists: Aryl 1,4-dialkylpiperizines Related to GBR-12783. *Eur. J. Pharmacol.* **1992**, *220*, 173-180.

170. Zhang, Y.; Joseph, D. B.; Bowen, W. D.; Flippen-Anderson, J. L.; Dersch, C. M.; Rothman, R. B.; Jacobson, A. E.; Rice, K. C. Synthesis and Biological Evaluation of Tropane-like 1-{2-[Bis(4-fluorophenyl)methoxy]ethyl}-4-(3-phenylpropyl)piperazine (GBR 12909) Analogues. *J. Med. Chem.* **2001**, *44*, 3937 - 3945.

171. Zhang, S.; Reith, M. E. A.; Dutta, A. K. Design, Synthesis, and Activity of Novel cis- and trans-3,6-Disubstituted Pyran Biomimetics of 3,6-Disubstituted Piperidine as Potential Ligands for the Dopamine Transporter. *Bioorg. Med. Chem. Let.* **2003**, *13*, 1591-1595.

172. Dutta, A. K.; Meltzer, P. C.; Madras, B. K. Positional Importance of the Nitrogen Atom in Novel Piperidine Analogs of GBR 12909: Affinity and Selectivity for the Dopamine Transporter. *Med. Chem. Res.* **1993**, *3*, 209-222.

173. Akunne, H. C.; Dersch, C. M.; Cadet, J. L.; Baumann, M. H.; Char, G. U.; Partilla, J. S.; de Costa, B. R.; Rice, K. C.; Carroll, F. I.; Rothman, R. B. Studies of the Biogenic Amine Transporters. III. Demonstration of Two Binding Sites for [^{3}H]GBR 12935 and [^{3}H]BTCP in Rat Caudate Membranes. *J. Pharm. Expt. Ther.* **1994**, *268*, 1462-1475.

174. Baumann, M. H.; Phillips, J. M.; Ayestas, M. A.; Ali, S. F.; Rice, K. C.; Rothman, R. B. Preclinical Evaluation of GBR 12909 Decanoate as a Long-Acting Medication for Methamphetamine Dependence. *Ann. N. Y. Acad. Sci.* **2002**, *965*, 92-108.

175. Baumann, M. H.; Char, G. U.; de Costa, B. R.; Rice, K. C.; Rothman, R. B. GBR 12909 Attenuates Cocaine-Induced Activation of Mesolimbic Dopamine Neurons in the Rat. *J. Pharm. Expt. Ther.* **1994**, *271*, 1216-1222.

176. Glowa, J. R.; Fantegrossi, W. E.; Lewis, D. B.; Matecka, D.; Rice, K. C.; Rothman, R. B. Sustained Decrease in Cocaine-Maintained Responding in Rhesus Monkeys with 1-[2-[Bis(4-flurophenyl)methoxy]ethyl]-4-(3-hydroxy-3-phenylpropyl)piperazinyl Decanoate, a Long-Acting Ester Derivative of GBR 12909. *J. Med. Chem.* **1996**, *39*, 4689-4691.

177. Glowa, J. R.; Rice, K. C.; Matecka, D.; Rothman, R. B. Phentermine/Fenfluramine Decreases Cocaine Self-Administration in Rhesus Monkeys. *NeuroReport* **1997**, *8*, 1347-1351.

178. Glowa, J. R.; Wojnicki, F. H. E.; Matecka, D.; Rice, K. C.; Rothman, R. B. Effects of Dopamine Reuptake Inhibitors on Food- and Cocaine-Maintained Responding: II. Comparisons With Other Drugs and Repeated Administrations. *Exp. Clin. Psychopharmacol.* **1995**, *3*, 232-239.

179. Gu, X.-H.; Yu, H.; Jacobson, A. E.; Rothman, R. B.; Dersch, C. M.; George, C.; Flippen-Anderson, J. L.; Rice, K. C. Design, Synthesis, and Monoamine Transporter Binding Site Affinities of Methoxy Derivatives of Indatraline. *J. Med. Chem.* **2000**, *43*, 4868-4876.

180. Rothman, R. B. High Affinity Dopamine Reuptake Inhibitors as Potential Cocaine Antagonists: A Strategy for Drug Development. *Life Sci.* **1990**, *46*, PL17-PL21.

181. Rothman, R. B.; Baumann, M. H. Monoamine Transporters and Psychostimulant Drugs. *Eur. J. Pharmacol.* **2003,** *479,* 23-40.

182. Rothman, R. B.; Baumann, M. H.; Dersch, C. M.; Appel, J.; Houghten, R. A. Discovery of Novel Peptidic Dopamine Transporter Ligands by Screening a Positional Scanning Combinatorial Hexapeptide Library. *Synapse* **1999,** *33,* 239-246.

183. Rothman, R. B.; Becketts, K. M.; Radesca, L. R.; de Costa, B. R.; Rice, K. C.; Carroll, F. I.; Dersch, C. M. Studies of the Biogenic Amine Transporters. II. A Brief Study on the Use of [^{3}H]DA-Uptake-Inhibition to Transporter-Binding Inhibition for the *in vitro* Evaluation of Putative Cocaine Antagonists. *Life Sci.* **1993,** *53,* PL267-PL272.

184. Rothman, R. B.; Cadet, J. L.; Akunne, H. C.; Silverthorn, M. L.; Baumann, M. H.; Carroll, F. I.; Rice, K. C.; de Costa, B. R.; Partilla, J. S.; Wang, J.-B.; Uhl, G.; Glowa, J. R.; Dersch, C. M. Studies of the Biogenic Amine Transporters. IV. Demonstration of a Multiplicity of Binding Sites in Rat Caudate Membranes for the Cocaine Analog [^{125}]RTI-55. *J. Pharm. Expt. Ther.* **1994,** *270,* 296-309.

185. Rothman, R. B.; Grieg, N.; Kim, A.; de Costa, B. R.; Rice, K. C.; Carroll, F. I.; Pert, A. Cocaine and GBR 12909 Produce Equivalent Motoric Responses at Different Occupancy of the Dopamine Transporter. *Pharmacol. Biochem. Behav.* **1992,** *43,* 1135-1142.

186. Rothman, R. B.; Lewis, B.; Dersch, C. M.; Xu, H.; Radesca, L.; de Costa, B. R.; Rice, K. C.; Kilburn, R. B.; Akunne, H. C.; Pert, A. Identification of a GBR 12935 Homolog, LR 1111, Which Is Over 4,000-fold Selective for the Dopamine Transporter, Relative to Serotonin and Norepinephrine Transporters. *Synapse* **1993,** *14,* 34-39.

187. Rothman, R. B.; Mele, A.; Reid, A. A.; Akunne, H.; Grieg, N.; Thurkauf, A.; de Costa, B. R.; Rice, K. C.; Pert, A. Preliminary Evidence that the High Affinity Dopamine Reuptake Inhibitor, GBR 12909, Antagonizes the Ability of Cocaine to Elevate Extracellular Levels of Dopamine. *NIDA Res. Monogr.* **1991,** *105,* 359-360.

188. Rothman, R. B.; Mele, A.; Reid, A. A.; Akunne, H. C.; Greig, N.; Thurkauf, A.; de Costa, B. R.; Rice, K. C.; Pert, A. GBR 12909 Antagonizes the Ability of Cocaine to Elevate Extracellular Levels of Dopamine. *Pharmacol. Biochem. Behav.* **1991,** *40,* 387-397.

189. Rothman, R. B.; Partilla, J. S.; Baumann, M. H.; Dersch, C. M.; Carroll, F. I.; Rice, K. C. Neurochemical Neutralization of Methamphetamine With High-Affinity Nonselective Inhibitors of Biogenic Amine Transporters: A Pharmacological Strategy for Treating Stimulant Abuse. *Synapse* **2000,** *35,* 222-227.

190. Rothman, R. B.; Silverthorn, M. L.; Glowa, J. R.; Matecka, D.; Rice, K. C.; Carroll, F. I.; Partilla, J. S.; Uhl, G. R.; Vandenbergh, D. J.; Dersch, C. M. Studies of the Biogenic Amine Transporters. VII. Characterization of a Novel Cocaine Binding Site Identified with [^{125}]RTI-55 in Membranes Prepared from Human, Monkey and Guinea Pig Caudate. *Synapse* **1998**, *28*, 322-338.

191. Villemagne, V. L.; Rothman, R. B.; Yokoi, F.; Rice, K. C.; Matecka, D.; Dannals, R. F.; Wong, D. F. Doses of GBR 12909 That Suppress Cocaine Self-Administration in Non-Human Primates Substantially Occupy Dopamine Transporters as Measured by [^{11}C] WIN 35,428 PET Scans. *Synapse* **1999**, *32*, 44-50.

192. Wojnicki, F. H. E.; Rothman, R. B.; Rice, K. C.; Glowa, J. R. Effects of Phentermine on Responding Maintained under Multiple Fixed-Ratio Schedules of Food and Cocaine Presentation in the Rhesus Monkey. *J. Pharm. Expt. Ther.* **1999**, *288*, 550-560.

193. Zhang, Y.; Rothman, R. B.; Dersch, C. M.; de Costa, B. R.; Jacobson, A. E.; Rice, K. C. Synthesis and Transporter Binding Properties of Bridged Piperazine Analogues of 1-{2-[Bis(4-fluorophenyl)methoxy]ethyl}-4-(3-phenylpropyl)piperazine (GBR 12909). *J. Med. Chem.* **2000**, *43*, 4840-4849.

194. Saunders, M. Stochastic Exploration of Molecular Mechanics Energy Surfaces. Hunting for the Global Minimum. *J. Am. Chem. Soc.* **1987**, *109*, 3150-3152.

195. Perkins, T. D. J.; Dean, P. M. An Exploration of a Novel Strategy for Superposing Several Flexible Molecules. *J. Comput. Aided Mol. Des.* **1993**, *7*, 155 - 172.

196. Kelly, K. Exhaustive and Iterative Clustering of the Protein Databank. *J. Chem. Comp. Group* **1998**.

197. Downs, G. M.; Barnard, J. M. Clustering Methods and Their Uses in Computational Chemistry. In *Rev. Comput. Chem.*; Lipkowitz, K. B.; Boyd, D. B. (ed.); John Wiley & Sons: New York, 2002; pp. 1-40.

198. Shenkin, P. S.; McDonald, D. Q. Cluster Analysis of Molecular Conformations. *J. Comput. Chem.* **1994**, *15*, 899-916.

199. Theiler, J.; Gisler, G. A Contiguity-Enhanced k-Means Clustering Algorithm for Unsupervised Multispectral Image Segmentation. *Proc. SPIE* **1997**, *3159*, 108-118.

200. Swanson, R. M. Entropy Measures Amount of Choice. *J. Chem. Educ.* **1990**, *67*, 206-208.

201. Shenkin, P. S.; Erman, B.; Mastrandrea, L. D. Information-Theory Entropy as a Measure of Sequence Variability. *Proteins: Struct. Funct. Gen.* **1991**, *11*, 297-313.

202. Klebe, G. Comparative Similarity Indices Analysis: CoMSIA. In *3D QSAR In Drug Design*; Kubinyi, H. (ed.); Kluwer Academic: Dordrecht, The Netherlands, 1998; pp. 87-104.

203. Kearsley, S. K., Smith, G.M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Method.* **1990**, *3*, 615-633.

204. Wold, S.; Johansson, E.; Cocchi, M. PLS - Partial Least Squares Projections to Latent Structures. In *3D QSAR in Drug Design: Theory, Methods, and Applications*; Kubinyi, H. (ed.); ESCOM: Leiden, 1993; pp. 523-550.

205. Kim, K. H.; Greco, G.; Novellino, E. A Critical Review of Recent CoMFA Applications. In *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H.; Folkers, G.; Martin, Y. C. (ed.); Kluwer Academic: Dordrecht, 1998; pp. 257-315.

206. Lindgren, F., Geladi, P., Rannar, S., Wold, S. Interactive Variable Selection (IVS) for PLS. Part 1: Theory and Algorithms. *J. Chemometrics* **1994**, *8*, 349-363.

207. So, S. S.; Karplus, M. Three-Dimensional Quantitative Structure-Activity Relationships from Molecular Similarity Matrixes and Genetic Neural Networks. 1. Methods and Validations. *J. Med. Chem.* **1997**, *40*, 4337-4359.

208. Hoffman, B.; Cho, S. J.; Zheng, W.; Wyrick, S.; Nichols, D. E.; Mailman, R. B.; Tropsha, A. Quantitative Structure-Activity Relationship Modeling of Dopamine D(1) Antagonists Using Comparative Molecular Field Analysis, Genetic Algorithms-Partial Least-Squares, and K Nearest Neighbor Methods. *J. Med. Chem.* **1999**, *42*, 3217-3226.

209. Tetko, I. V.; Kovalishyn, V. V.; Livingstone, D. J. Volume Learning Algorithm Artificial Neural Networks for 3D QSAR Studies. *J. Med. Chem.* **2001**, *44*, 2411-2420.

210. 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. (ed.); Prous Science, S.A.: Barcelona, Spain, 2001; pp. 475-485.

211. Clark, R. D.; Fox, P. C. Statistical Variation in Progressive Scrambling. *J. Comput. Aided Mol. Des.* **2004**, *18*, 563-576.

212. Golbraikh, A., Tropsha, A. Beware of q^2! *J. Mol. Graph. Mod.* **2002**, *20*, 269-276.

213. Berfield, J. L.; Wang, L. C.; Reith, M. E. A. Which Form of Dopamine is the Substrate for the Human Dopamine Transporter: the Cationic or the Uncharged Species? *J. Biol. Chem.* **1999**, *274*, 4876-4882.

214. Skawinski, W. J. Unpublished results. **2004**, New Jersey Institute of Technology.

215. Misra, M.; Banerjee, A.; Davé, R. N.; Venanzi, C. A. Novel Feature Extraction Technique for Fuzzy Relational Clustering of a Flexible Dopamine Reuptake Inhibitor. *J. Chem. Inf. Model.* **2005,** *In press.*

216. *XCluster (part of the Macromodel software suite),* Version 8.5; Schrödinger, L.L.C.: 2003.

217. Gasteiger, J., Marsili, M. Iterative Partial Equalization of Orbital Electronegativity: A Rapid Access to Atomic Charges. *Tetrahedron* **1980,** *36,* 3219-3228.

218. Bush, B. L.; Nachbar Jr., R. B. Sample-Distance Partial Least Squares: PLS Optimized for Many Variables, with Application to CoMFA. *J. Comput. Aided Mol. Des.* **1993,** *7,* 587-619.

219. *Tripos Bookshelf, Sybyl Version 6.9.1, [QSAR Manual],* Tripos, Inc.: St. Louis, MO, 2003.

220. Clark, M.; Cramer III, R. D. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993,** *12,* 137-145.

221. Budriesi, R., Carosati, E., Chiarini, A., Cosimelli, B., Cruciani, G., Ioan, P., Spinelli, D., Spisani, R. A New Class of Selective Myocardial Calcium Channel Modulators. 2. Role of the Acetal Chain in Oxadiazol-3-one Derivatives. *J. Med. Chem.* **2005,** *48,* 2445-2456.

222. Gallardo-Godoy, A.; Fierro, A.; McLean, T. H.; Castillo, M.; Cassels, B. K.; Reyes-Parada, M.; Nichols, D. E. Sulfur-Substituted a-Alkyl Phenethylamine as Selective and Reversible MAO-A Inhibitors: Biological Activities, CoMFA Analysis, and Active Site Modeling. *J. Med. Chem.* **2005,** *48,* 2407-2419.