ABSTRACT

PHARMACOPHORE DERIVATION USING DISCOTECH AND COMPARISON OF SEMI-EMPIRICAL, AB INITIO AND DENSITY FUNCTIONAL COMFA STUDIES FOR SIGMA 1 AND SIGMA 2 RECEPTOR-LIGANDS

by
Dawoon Jung

This study describes the development of pharmacophore and CoMFA models for sigma receptor ligands. CoMFA studies were performed for 48 bioactive sigma 1 receptor-ligands using [H^3](+) pentazocine as the radioligand, for 30 PCP derivatives for sigma 1 receptor-ligands using [^{3}H](+)SKF10047 as the radioligand and for 24 bioactive sigma 2 receptor-ligands using the radioligand [H^3](+)DTG in the presence of pentazocine. Distance Comparisons (DISCOtech) was used as the starting point for CoMFA studies. The conformers, derived by DISCOtech were optimized using AM1, or HF/3-21G* in Gaussian 98. The optimized geometries were aligned with the pharmacophore, derived using DISCOtech. Atomic charges were calculated using AM1, HF/3-21G*, B3LYP/3-21G*, MP2/3-21G* methods in Gaussian 98. The CoMFA Maps that were developed using Sybyl 6.9 were compared on steric and electrostatic field differences. With leave-one-out cross validation the numbers of optimal components were decided. Using these numbers of optimal components no cross validation was performed in a training set. After a test set, it was known that CoMFA models derived from HF/3-21G* optimized geometries were more reliable in predicting bioactivities than CoMFA models derived from AM1 optimized geometries.

PHARMACOPHORE DERIVATION USING DISCOTECH AND COMPARISON OF SEMI-EMPIRICAL, AB INITIO AND DENSITY FUNCTIONAL COMFA STUDY FOR SIGMA 1 AND SIGMA 2 RECEPTOR-LIGANDS

by
Dawoon Jung

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 2003

APPROVAL PAGE

PHARMACOPHORE DERIVATION USING DISCOTECH AND COMPARISON OF SEMI-EMPIRICAL, AB INITIO AND DENSITY FUNCTIONAL COMFA STUDIES FOR SIGMA 1 AND SIGMA 2 RECEPTOR-LIGANDS

Dawoon Jung

Dr. Tamara M. Gund, Advisor
Professor of Chemistry, NJIT

Date

Dr. Carol A. Venanzi, Committee Member
Distinguished Professor of Chemistry, NJIT

Date

Dr. Barbara B. Kebbekus, Committee Member
Professor of Chemistry, NJIT

Date

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

Date

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

Date

BIOGRAPHICAL SKETCH

Author: Dawoon Jung

Degree: Doctor of Philosophy

Date: May 2003

Undergraduate and Graduate Education:

- Doctor of Philosophy in Chemistry,
  New Jersey Institute of Technology, Newark, NJ, 2003
- Master of Science in Chemistry,
  New Jersey Institute of Technology, Newark, NJ, 2000
- Master of Science in Chemistry,
  Yonsei University, Seoul, Korea, 1997
- Bachelor of Science in Chemistry,
  Yonsei University, Seoul, Korea, 1995

Major: Chemistry

Presentations and Publications:

Jung, D.; Floyd, J.; Gund, T. M. A CoMFA Study using semi-empirical, density functional, ab initio methods and Pharmacophore Derivation using DISCOtech on Sigma 1 Ligands, submitted to J. Comput. Chem. April in 2003.

Jung, D.; Gund, T. M. A CoMFA Study using semi-empirical, density functional, ab initio methods and Pharmacophore Derivation using DISCOtech on Sigma 2 Ligands, submitted to J. Computer-Aid. Des. April in 2003.

Gund, T. M.; Jung, D., Pharmacophore Derivation using DISCOtech on PCP derivatives for sigma 1 receptor-ligands and CoMFA Study using semi-empirical, density functional, ab initio methods submitted J. Mol. Graphics Modell. April in 2003.

Bozzelli, J. W.; Jung, D. Theoretical Investigation on Stability of the C· H2OCl Radical, *J. Phys. Chem. A.* **2001**, *105(16)*, 3941-3946.

Jung, D.; Chen, C.-J.; Bozzelli, J. W. Structures, Rotation Barrier, and Thermodynamic Properties H_{f298}, S_{298}, and $C_{p(T)}$ of Chloromethyl Hypochlorites CH_3OCl, CH_2ClOCl, $CHCl_2OCl$, and CCl_3OCl, *J. Phys. Chem. A.* **2000**, *104(42)*, 9581-9590.

Jung, D.; Bozzelli, J. W. Structure, Torsional Potentials, and Thermodynamic Properties H_{f298}, S_{298}, and $C_{p(T)}$ of Chloro-Dimethyl Ethers: CH_2ClOCH_3, $CHCl_2OCH_3$, and CCl_3OCH_3. Density Functional and ab Initio Calculations. *J. Phys. Chem. A.* **2001**, *105(22)*, 5420-5430.

S. H. Lee, D. W. Jung, J. B. Kim and Y. -R. Kim Effect of altervalent cation-doping on catalytic activity of neodymium sesquioxide for oxidative coupling of methane, *Applied Catalysis A:* **1997**, *164 (1-2)*, 159-169.

To my beloved family

ACKNOWLEDGEMENT

Drug design, Molecular modeling, Semi-emprical, Ab initio, Density functional calculation, CoMFA, QSAR, DISCOtech pharmacophore, and sigma receptor, all of them are my old friends. I could meet them in the field of drug design. I really appreciate my advisor, Professor, Tamara M. Gund to introduced me great friends. I could make them good buddies under her kind guidence.

Dr. Chistopher C. Van Dyke is my committee member to teach me how to interact other friends; Docking, Unity, Virtual screen, and Biopolymer; I thank him with all my heart.

Distinguished Professor, Carol A. Venanzi displayed me the right road to take care of friends; Professor, Barbara B. Kebbekus encouraged me to make a good friendship and Professor, Sanjay V. Malhotra introduced me new friends (Transition states in organic reactions, TS1-4, TR1-4). I appreciate their kindness help and happy to have great committee members.

My wife, Boki and my daughter, Seulgi are the sources of my happiness and endleess love from my parents is my driving force to finish this dissertation and to be Dr. Jung. I appreciate and love them.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

GENERAL INTRODUCTION

Molecular modeling which includes CoMFA has been an important tool in many areas of research for determining molecular structure, function and for drug design. Although there have been some molecular modeling studies on the sigma receptor, few have been done to differentiate between sigma 1 and sigma 2 subtypes, and only one has been performed in our group on sigma subtype 2.

The research marks a beginning in combining molecular modeling studies and CoMFA studies of sigma subtypes to understand each subtype's specific pharmacophore and its structural activity relationship using different calculational methods.

The objectives of this research were:

1) to derive pharmacophores for sigma receptor subtypes using highly selective and potent ligands of sigma subtypes.
2) To understand the differences between the sigma 1 receptor and sigma 2 receptor site binding requirements by studying three-dimensional quantitative structure activity relationships (3D-QSAR) applying the CoMFA method using the pharmacophore results from the first step as alignment rules
3) To compare semi-empirical, density functional, and ab initio calculations to CoMFA studies on sigma receptor ligands.
4) To design new ligands for each subtype using CoMFA results.

Chapter 1 gives outlines, and objectives of this research. Introduction of sigma receptor-ligands is present in Section 1.1.

Capter 2 is a description of various quantum mechanics methods for ab initio, density functional, and semi-empirical methods used in this study. This part consists of nine main Sections. Section 2.1 contains the Schrödinger Equation. Section 2.2 explains Hartree-Fock Self-Consistent Field theory used in this study. Section 2.3 suggests limitations of the HF method in electron correlations. Section 2.4 describes the Møller-Plesset Method used in this study. Section 2.5 gives multiconfiguration SCF methods. Section 2.6 describes density functional theory that has been used in this research. Section 2.7 contains properties derived from the wavefunction as electrical properties and atomic charges used in this study. Section 2.8 explains basis set effects by minimal basis sets, split valence basis sets, polarized basis sets, diffuse functions, and high angular momentum basis sets. Section 2.9 describes semi-empirical methods.

Chapter 3 contains QSAR methodology. Section 3.1 describes statistical concepts. Section 3.2 gives approaches to developing a QSAR.

Chapter 4 contains comparative molecular field analysis (CoMFA) studies using semi-empirical, density functional, ab initio methods and pharmacophore derivation using DISCOtech on sigma 1 receptor-ligands. Section 4.1 includes an introduction of chapter 4. Section 4.2 describes materials and methods. It gives 48 bioactive compounds from literature data, and computational methods used in this study. Section 4.3 shows results in chapter 4 in comparative molecular field analysis. Section 4.4 validates this CoMFA models and designs of new Ligands using the CoMFA models. Section 4.5 concludes in CoMFA studies of sigma 1 receptor-ligands using radioligand, [^{3}H](+) pentazocine.

Chapter 5 contains pharmacophore derivation using DISCOtech on PCP derivatives for sigma 1 receptor-ligands and CoMFA studies using semi-empirical, density functional, ab initio methods. Section 5.1 gives an introduction of PCP Derivatives for sigma 1 receptor-ligands. Section 5.2 includes biological data of 30 phenyl cyclohexyl piperidine derivatives and computational methods used in chapter 5. It covers choice of initial conformations, pharmacophore information from DISCOtech. geometry optimization, atomic charge calculations, alignments, and CoMFA models. Section 5.3 describes results and discussions in chapter 5. It shows results of comparative molecular field analysis and their validations of CoMFA models. Using these CoMFA models, new ligands are designed. Section 5.4 shows conclusions in CoMFA studies of PCP derivatives for sigma 1 receptor-ligands.

Chapter 6 contains CoMFA studies using semi-empirical, density functional, ab initio methods and pharmacophore derivation using DISCOtech on sigma 2 receptor-ligands. Section 6.1 introduces sigma 2 receptor ligands. Section 6.2 shows a selection of ligands, choice of initial conformations, pharmacophore information using DISCOtech, geometry optimization and atomic charge calculations, alignments of optimized molecules, and their CoMFA models. Section 6.3 contains results and discussions of comparative molecular field analysis, validation of the CoMFA models, and design of new ligands. Section 6.4 gives conclusions of CoMFA studies of Sigma 2 receptor-ligands.

Chapter 7 displays general conclusions on this study. Suggestions for further work is also given in Section 7.1.

1.1 Sigma Receptor Ligands

It is now well established that sigma (σ) receptors represent a unique binding site in the brain and peripheral organs, distinct from any other known proteins. However, when they were initially proposed by Martin and colleagues [1] to account for the psychotomimetic effects of *N*-allylnormetazocine ((±)-SKF-10,047) in the morphine-dependent chronic spinal dog, they were initially classifed as `opiate/σ' sites. It was rapidly evident that most of the behaviors elicited by the drug were resistant to blockade by classical opiate receptor antagonists naloxone or naltrexone [2]. The sigma (σ) receptors were thus distinguished from other classical *μ*-, *κ*-, and *δ*-opiate receptors [3]. The sigma (σ) receptors were then confounded with the high affinity phencyclidine (PCP) binding sites, located within the ion channel associated with the NMDA-type of glutamate receptor, because of similar affinities of these sites for several compounds, including PCP and (+)-SKF-10,047 [3]. The confusion was cleared up by the availability of more selective drugs, including dizocilpine or thienylcyclidine for the PCP site; 1,3-di-*o*-tolylguanidine (DTG), (+)-pentazocine, (+)-3-(3-hydroxyphenyl)-*N*-(1-propyl)-piperidine ((+)-3-PPP), igmesine (JO-1784), (+)-*cis*-*N*-methyl-*N*-[2-(3,4-dichlorophenyl)ethyl]-2-(1-pyrrolidinyl) cyclohexylamine (BD737), among others, for the *σ* site. The pharmacological identification of sigma (σ) sites was characterized by their ability to bind several chemically unrelated drugs with high affinity, including psychotomimetic benzomorphans, PCP and derivatives, cocaine and derivatives, amphetamines, certain neuroleptics, many new `atypical' antipsychotic agents, anticonvulsants, cytochrome P450 inhibitors, monoamine oxidase inhibitors, histaminergic receptor ligands, peptides

from the neuropeptide Y (NPY) and calcitonin gene-related peptide (CGRP) families, and several steroids [4-7].

The pharmacological identification and localization of σbinding sites was achieved using various radioligands, including [3H](+)-SKF-10,047, [3H](+)-3-PPP, [3H]haloperidol, [3H]DTG, [3H](+)-pentazocine [8-12]. Biochemical studies allowed the distinction of two classes of sigma (σ) sites, termed σ 1 and σ 2 [13]. The two sites can be distinguished based on their different drug selectivity patterns and molecular weights. The σ 1 site is a 25-30 kDa single polypeptide, and the σ 2 site is an 18– 21 kDa protein that has not yet been cloned [14-16]. The σ 1 site presents a high affinity and stereoselectivity for the (+)-isomers of SKF-10,047, pentazocine and cyclazocine, whereas σ 2 sites have lower affinity and show the reverse stereoselectivity [15]. DTG, (+)-3-PPP and haloperidol are non-discriminating ligands with high affinity on both subtypes. In addition, σ 1 sites are allosterically modulated by phenytoin [17] and sensitive to pertussis toxin and to the modulatory effects of guanosine triphosphate [18-20]. It also has been shown that several drugs, such as haloperidol, reduced haloperidol, α-(4-fluorophenyl)-4-(5-fluoro-2-pyrimidinyl)-1-piperazine butanol (BMY-14,802), rimcazole, or *N,N*-dipropyl-2-(4-methoxy-3-(2-phenylethoxy)phenyl) ethylamine (NE-100) act as antagonists in several physiological and behavioral tests relevant to the σ 1 pharmacology [21-25]. However, most of them are non-selective and also bind to other pharmacological targets.

The σ 1 receptor cDNA has been cloned from guinea-pig liver [26], human placental cell line, T leukemia Ichikawa cell line and human brain [27-29], mouse kidney and brain

[30,31], and rat brain [32,33]. The amino acid sequences of the purified proteins are highly similar, with a 87–92% identity and 90–93% homology between species. The protein sequence also shared a similarity, 33% identity and 66% homology, with a fungal sterol C8–C7 isomerase [26]. However, it shares no homology to the related mammalian enzyme or any other mammalian protein, indicating that the σ1 receptor is a distinct entity from any other known receptors and that an identical σ1 receptor is expressed in peripheral tissues and brain. The promoter region sequence of the σ1 receptor contains consensus sequences for the liver-specific transcription factors nuclear factor (NF)-1/L, activator protein (AP)-1, AP-2, IL-6RE, NF-GMa, NF-GMb, NF-κB, steroid response element, GATA-1, Zeste, for the xenobiotic responsive factor called the arylhydrocarbon receptor, and for a putative signal for retention in the endoplasmic reticulum [29], suggesting that the receptor transcription could be related by immediate early genes.

In this research, there are three different studies for sigma receptors; (1) 43 molecules as a training set and 5 molecules as a test set for the sigma 1 receptor-ligands, using the radioligand [^{3}H](+) pentazocine, (2) 24 PCP derivatives as a training set and three molecules as a test set for the sigma 1 receptor-ligands, using the radioligand [^{3}H](+)SKF10047, and (3) 21 molecules as a training set and three molecules as a test set for sigma 2 receptor-ligands, using [^{3}H]DTG in the presence of (+) pentazocine.

CHAPTER 2

QUANTUM CHEMICAL METHODS FOR MOLECULAR MODELING

Many aspects of molecular structure and dynamics can be modeled using classical methods in the form of molecular mechanics and dynamics. The classical force field is based on empirical results, averaged over a large number of molecules. Because of this extensive averaging, the results can be good for standard systems, but there are many important questions in chemistry that cannot be addressed by means of this empirical approach. If one wants to know more than just structure or other properties that are derived only from the potential energy surface, in particular properties that depend directly on the electron density distribution, one has to resort to a more fundamental and general approach: quantum chemistry. The same holds for all non-standard cases for which molecular mechanics is not applicable.

Quantum chemistry is based on the postulates of Quantum Mechanics. In this chapter, some basic aspects of the theory of quantum chemistry are recalled with an emphasis on their practical implications for the molecular modeler. In quantum chemistry, the system is described by a wavefunction which can be found by solving the Schrödinger equation. This equation relates the stationary states of the system and their energies to the Hamiltonian operator, which can be viewed as the recipe for obtaining the energy associated with a wavefunction describing the positions of the nuclei and electrons in the system. In practice the Schrödinger equation cannot be solved exactly and approximations have to be made. The approach is called "ab initio" when it makes no use

of empirical information, except for the fundamental constants of nature such as the mass of the electron, Planck's constant etc., that are required to arrive at numerical predictions. In spite of the necessary approximations, ab initio theory has the conceptual advantage of generality, and the practical advantage that (with experience) its successes and failures are more or less predictable.

The major disadvantage of ab initio quantum chemistry are the heavy demands on computer power. Therefore, further approximations have been applied for a long time which go together with the introduction of empirical parameters into the theoretical model. This has led to a number of semi-empirical quantum chemical methods, which can be applied to larger systems, and give reasonable electronic wavefunctions so that electronic properties can be predicted. Compared with ab initio calculations their reliability is less and their applicability is limited by the requirement for parameters, just like in molecular mechanics.

In general, one should apply quantum chemistry for "small" systems, which can be treated at a very high level, when electronic properties are sought (electric moments, polarizabilities, shielding constants in NMR and ESR, etc.) and for "non-standard" structures, for which no valid molecular mechanics parameters are available. Examples are conjugated pi systems, organometallic compounds and other systems with unusual bond or atom types, excited states, reactive intermediates, and generally structures with unusual electronic effects [34].

2.1 The Schrödinger Equation

The energies and wavefunctions of stationary states of a system are given by the solutions of the Schrödinger Equation: $\hat{H}\Psi = E\Psi$. In this equation $\hat{H}$ is the Hamiltonian operator which in this case gives the kinetic and potential energies of a system of atomic nuclei and electrons. It is analogous to the classical kinetic energy of the particles and the Coulomb electrostatic interactions between the nuclei and electrons. Ψ is a wavefunction, one of the solutions of the eigenvalue equation. This wavefunction depends on the coordinates of the electrons and the nuclei. The Hamiltonian is composed of three parts: the kinetic energy of the nuclei, the kinetic energy of the electrons, and the potential energy of nuclei and electrons.[35-39]

Schrödinger equation: $\hat{H}\Psi_{e,n} = E\Psi_{e,n}$, Hamiltonian: $\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{e,n}$

Four approximations are commonly (but not necessarily) made:

- time independence; looking at states that are stationary in time.
- neglect of relativistic effects; this is warranted unless the velocity of the electrons approaches the speed of light, which is the case only in heavy atoms with very high nuclear charge.
- Born-Oppenheimer approximation; separation of the motion of nuclei and electrons.
- orbital approximation; the electrons are confined to certain regions of space.

The Born-Oppenheimer approximation implies the separation of nuclear and electronic wavefunctions, the total wavefunction being a product of the two:

Born-Oppenheimer: $\Psi_{e,n} = \chi_n \psi_e$

The motivation behind this is that the electrons are so much lighter than the nuclei that their motion can easily follow the nuclear motion. In practice, this approximation is usually valid. From this point, the electronic wavefunction ψ_e is investigated and ψ_e is obtained by solving the electronic Schrödinger equation:

$$\hat{H}_e(\mathrm{R_n})\psi(r_e) = E_e(\mathrm{R_n})\psi(r_e)$$

This equation still contains the positions of the nuclei, however not as variables but as parameters.

The electronic Hamiltonian contains three terms: kinetic energy, electrostatic interaction between electrons and nuclei, and electrostatic repulsion between electrons. In order to simplify expressions and to make the theory independent of the experimental values of physical constants, atomic units are introduced:

e = 1 charge of electron

m = 1 mass of the electron

$\hbar$= 1 Planck's constant divided by 2 pi

Derived atomic units of length and energy are:

1 bohr = $a_0 = \dfrac{\hbar^2}{me^2} = 0.529$ Å

1 hartree = $\dfrac{e^2}{a_0} = 4.3598 \times 10^{-18}$ J = 627.51 kcal/mol

With these units the electronic Hamiltonian is:

$$\hat{H}_e = -\tfrac{1}{2}\sum_{i=1}^{n}\Delta_i - \sum_{i=1}^{n}\sum_{A=1}^{N}\frac{Z_A}{|R_A - r_i|} + \sum_{i<j}^{n}\frac{1}{r_{ij}}$$

The Δsymbol is the Laplace operator $\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$ (also called "del-squared"). The total energy in the Born-Oppenheimer model is obtained by adding the nuclear repulsion energy to the electronic energy:

$$E_{tot} = E_e + E_n$$

$$E_n = \sum_{A<B}^{N}\frac{Z_A Z_B}{|R_A - R_B|}$$

The total energy defines a potential energy hypersurface E=f(Q) which can be used to subsequently solve a Schrödinger equation for the nuclear motion:

$$\left[\hat{T}_n + E(\mathrm{R_n})\right]\Phi(\mathrm{R_n}) = \varepsilon_i \Phi(\mathrm{R_n})$$

In the following Section, the important problem of solving the electronic Schrödinger equation will be studied.

2.2 Hartree-Fock Self-Consistent Field Theory

The electronic Hamiltonian contains two terms that act on one electron at a time, the kinetic energy and the electron-nucleus attraction, and a term that describes the pairwise repulsion of electrons. The latter depends on the coordinates of two electrons at the same time, and has turned out to be a practical computational bottleneck, which can be passed only for very small systems:

$$\hat{H}^1 = \sum_i^n \hat{H}_i^1 = -\tfrac{1}{2}\sum_{i=1}^n \Delta_i - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{|R_A - r_i|}$$

$$\hat{H}^2 = \sum_{i<j}^n \hat{H}_{ij}^2 = \sum_{i<j}^n \frac{1}{r_{ij}}$$

To avoid this problem the independent particle approximation is introduced: the interaction of each electron with all the others is treated in an average way. Suppose:

$$\hat{H}^2 = \sum_{i<j}^n \hat{H}_{ij}^2 = \sum_i^n \hat{V}_i^{av}$$

Then the Schrödinger equation which initially depended on the coordinates x (representing spatial and spin coordinates) of all electrons can be reduced to a set of equations:

$$\sum_{i=1}^n \left(\hat{H}_i^1 + \hat{V}_i^{av}\right)\psi(x_1, x_2, .., x_n) = E\psi(x_1, x_2, .., x_n)$$

$$\left(\hat{H}_i^1 + \hat{V}_i^{av}\right)\phi_i(x_1) = \hat{F}_i\phi_i(x_1) = \varepsilon_i\phi_i(x_1)$$

The wavefunctions $\phi_i(x_1)$ are called one-electron spin-orbitals.

The obvious problem is that for each electron the potential due to all other electrons has to be known, but initially none of these is known. In practice trial orbitals are used which are iteratively modified until a self-consistent solution (a "Self-Consistent Field") is obtained, which can be expressed as a solution to the Hartree-Fock equations:

$$\left(\hat{H}_i^1 + \hat{V}_i^{av}\right)\phi_i = \hat{F}\phi_i = \varepsilon_i\phi_i$$

It is important to realize that convergence of the SCF procedure is by no means guaranteed. Many techniques have been developed over the years to speed up convergence, and to solve even difficult cases. In practice, difficulties often occur with systems with an unusual structure, where the electrons "do not know where to go". The eigenvalues are interpreted as orbital energies. The orbital energies have an attractively simple physical interpretation: they give the amount of energy necessary to take the electron out of the molecular orbital, which corresponds to the negative of the experimentally observable ionization potential (Koopmans' Theorem) [35-39]:

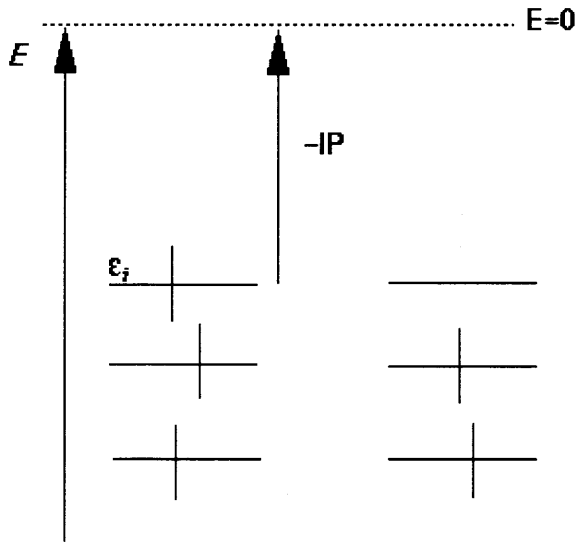


Figure 2.1 The eigenvalues, interpreted as orbital energies.

In addition to being a solution of the electronic Schrödinger equation the wavefunction must be normalized and satisfy the Pauli principle. The normalization condition is connected with the interpretation of the wavefunction as a distribution function which when integrated over entire space should give a value of one:

$$\int \psi^* \psi dx = 1$$

in "bra-ket" notation:

$$\langle \psi | \psi \rangle = 1$$

The Pauli principle states that the wavefunction must change sign when two independent electronic coordinates are interchanged:

$$\psi_n(x_1, x_2, .., x_i, .., x_k, .., x_m) = -\psi_n(x_1, x_2, .., x_k, .., x_i, .., x_m)$$

For a two-electron system the spin-orbitals $\varphi_1(x_1, y_1, z_1, \sigma_1)$ and $\varphi_2(x_2, y_2, z_2, \sigma_2)$ (in which sigma is either alpha or beta spin state) can be combined as follows:

$$\psi_n = \frac{1}{\sqrt{2}}\{\varphi_1(x_1, .., \sigma_1)\varphi_2(x_2, .., \sigma_2) - \varphi_2(x_1, .., \sigma_1)\varphi_1(x_2, .., \sigma_2)\}$$

According to the definition of a determinant this antisymmetrized product is equal to:

$$\psi_n = \frac{1}{\sqrt{2}}\begin{vmatrix} \varphi_1(x_1, .., \sigma_1) & \varphi_1(x_2, .., \sigma_2) \\ \varphi_2(x_1, .., \sigma_1) & \varphi_2(x_2, .., \sigma_2) \end{vmatrix}$$

This type of wavefunction is known as a Slater determinant, commonly abbreviated as:

$$\psi_n = n^{-1/2}|\varphi_1\varphi_2 \ldots \varphi_n|$$

An important property of the SCF method is that its solutions satisfy the Variation Principle, which states that the expectation value of the energy evaluated with an inexact wavefunction is always higher than the exact energy:

$$E_\psi = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_{\psi_{exact}}$$

As a consequence the lowest energy is associated with the best approximate wavefunction and energy minimization is equivalent with wavefunction optimization. The energies of Slater determinants from a Hartree-Fock calculation are readily expressed in one- and two-electron integrals. For the ground state it is:

$$E = \sum_{i}^{occ} H_{ii}^{1} + \sum_{i<j}^{occ} \left[(ii|jj) - (ij|ij) \right]$$

Here, the following abbreviations have been used:

$$H_{ii}^{1} = \left\langle i \middle| \hat{H}_{i}^{1} | j \right\rangle = \int \varphi_{i}^{*}(x_i) \hat{H}_{i}^{1} \varphi_{j}(x_i) dx_i$$

$$(ij|kl) = \iint \varphi_{i}^{*}(x_1) \varphi_{j}^{*}(x_1) \frac{1}{r_{12}} \varphi_{k}(x_2) \varphi_{l}(x_2) dx_1 dx_2$$

The two-electron integral $(ii|jj)$ which describes the repulsion between two electrons each localized in one orbital is called a Coulomb integral, $(ij|ij)$ for which a classical picture cannot be drawn so easily is called the Exchange integral.

In many cases it is advantageous to apply the restriction that electrons with opposite spin pairwise occupy the same spatial orbital. This leads to the Restricted Hartree Fock method (RHF), as opposed to the Unrestricted version (UHF). An important advantage of the RHF method is that the magnetic moments associated with the electron spin cancel exactly for the pair of electrons in the same spatial orbital, so that the SCF wavefunction is an eigenfunction of the spin operators and . Note that the UHF wavefunction is more flexible than the RHF wavefunction, thus can approximate the exact solution better and give a lower energy. In practice RHF is mostly used for closed shell systems, UHF for open shell species. RHF models for open shell systems and more advanced models can used when necessary. The total energy for a closed shell ground state RHF model can be written as:

$$E_{tot} = E_n + 2\sum_{i}^{occ} H_{ii} + \sum_{i,j}^{occ} (ii|jj) - \sum_{i,j}^{occ} (ij|ij)$$

The orbital energy in this case is:

$$\varepsilon_i = H_{ii} + \sum_{i,j}^{occ} \left[2(ii|jj) - (ij|ij) \right]$$

2.3 Limitations of the HF Method; Electron Correlation

Restricted Hartree-Fock SCF theory has some painful shortcomings. Consider for example the dissociation of the H_2 molecule:

H+ + H- <------ H-H --------> H. + H.

A "dissociation catastrophe" occurs because the separated hydrogen atoms cannot be described using doubly occupied orbitals, so that H2 tends to dissociate in H+ and H-, which can be described with a doubly occupied orbital on H-. This problem does not occur in the UHF method, but this method has the disadvantage that it does not give pure spin states.

An additional limitation of the HF method in general is that due to the use of the independent particle approximation the instantaneous correlation of the motions of electrons is neglected, even in the Hartree-Fock limit. The difference between the exact energy (determined by the Hamiltonian) and the HF energy is known as the correlation energy: Ecorrelation = Eexact - EHF < 0

Even though EHF is approximately 99% of Esub>exact the difference may be

chemically important. Several approaches are known that try to calculate the correlation energy after Hartree-Fock calculations (post-HF methods).

- Configuration Interaction (CI),
- Møller-Plesset Perturbation Theory and
- Multi-Configuration SCF (MCSCF or CASSCF).

HF theory gives a wavefunction which is represented as a Slater determinant. In the conceptually simple Configuration Interaction (CI) method, a linear combination of Slater determinants is constructed, using the unoccupied "virtual" orbitals from the SCF-calculation:

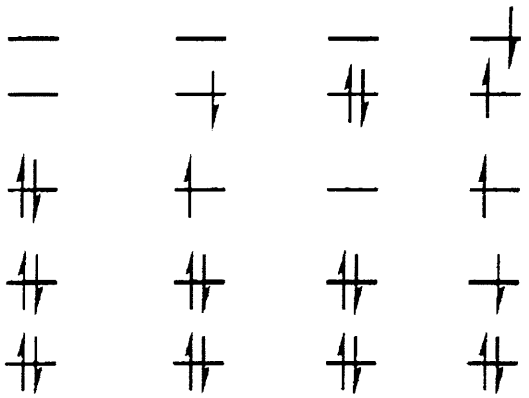


Figure 2.2 Configuration Interaction (CI) method, a linear combination of slater determinants

The total wavefunction is written as:

$$\Psi = \phi_{HF} + \sum_{ij^*} c^S_{ij^*} \phi_i^{j^*} + \sum_{ijk^*l^*} c^D_{ijk^*l^*} \phi_{ij}^{k^*l^*} + \ldots$$

In principle, the exact correlation energy can be obtained from a full CI calculation in which all configurations are taken into consideration. Unfortunately this is not possible for all but the smallest systems. Moreover, the problem is aggrevated when the size of the basis set is increased, on the way towards the Hartree-Fock limit. Thus, the theoretical limit of the exact (time-independent, non-relativistic) Schrödinger equation cannot be reached.

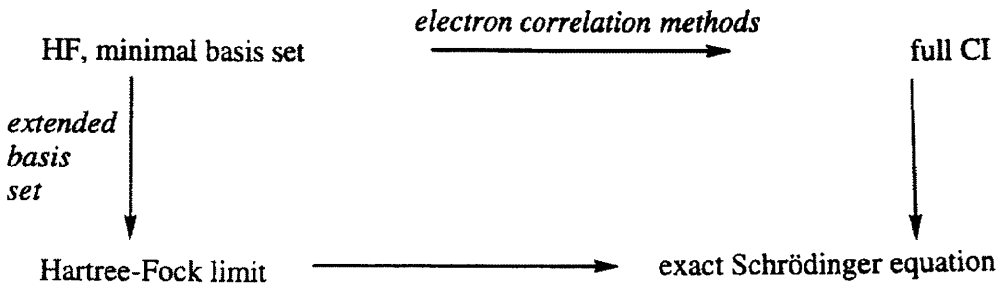


Figure 2.3 The Hartree-Fock limit as the size of the basis set is increased.

Even for small systems the number of excited configurations is enormously large. A popular way to truncate the CI expansion is to consider only singly and doubly excited configurations (CI-SD). The energy, calculated as the expectation value of the Hamiltonian for CISD is:

$$E = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} = E_{HF} + \sum_{i<j} \sum_{k^* < l^*} c^{D}_{ijk^*l^*} \left[\left(ij | k^* l^* \right) - \left(ik^* | jl^* \right) \right]$$

To perform the calculation one needs the two-electron integrals over Molecular Orbitals. The computation of these is very time-consuming, even when the integrals over AO's are available:

$$\left(ij | kl \right) = \sum_{\mu\nu\lambda\sigma} c_{i\mu} c_{j\nu} c_{k\lambda} c_{l\sigma} \left(\mu\nu | \lambda\sigma \right)$$

In general, CI is not the practical method of choice for the calculation of correlation energy because full CI is not possible, convergence of the CI expansion is slow, and the integral transformation time-consuming. Moreover truncated CI is not size-consistent, which means that the calculation of two species at large separation does not give the same energy as the sum of the calculations on separate species. This is because a different selection of excited configurations is made in the two calculations. An advantage of the CI method is that it is variational, so the calculated energy is always greater than the exact energy. Although CI is not recommendable as a method for ground states CI-singles (CIS) has been advocated as an approach to computation of excited state potential energy surfaces [40].

2.4 Møller-Plesset Method

A different approach to electron correlation has become very popular in recent years: Møller-Plesset perturbation theory. The basic idea is that the difference between the Fock operator and the exact Hamiltonian can be considered as a perturbation:

$$\hat{H} = \hat{F} + \hat{V}$$

Corrections can be made to any order of the energy and the wavefunction:

$$E = E_{HF} + E^{(1)} + E^{(2)} + E^{(3)} + \ldots$$

$$\Psi = \Psi_{HF} + \Psi^{(1)} + \Psi^{(2)} + \Psi^{(3)} + \ldots$$

The most popular method is the lowest level of correction, MP2.

$$E_{corr}^{MP2} = \sum_{ijk^*l^*} \frac{2\left(ik^*|jl^*\right) - \left(il^*|jk^*\right)}{\varepsilon_i + \varepsilon_j - \varepsilon_{k^*} - \varepsilon_{l^*}} \left(ik^*|jl^*\right)$$

An enormous practical advantage is that MP2 is fast (of the same order of magnitude as SCF), while it is rather reliable in its behavior, and size consistent. A disadvantage is that it is not variational, so the estimate of the correlation energy can be too large. Subsequent MP-levels MP3, MP4 (usually MP4 SDQ) are more complicated and much more time-consuming. For example, for pentane (C5H12) with the 6-31G(d) basis set (99 basis functions) an MP2 energy calculation took about 4 times the amount of time needed for SCF, while MP4 took almost 90 times that time [39].

2.5 Multiconfiguration SCF Method

Multiconfiguration SCF (MCSCF) or Complete Active Space SCF (CASSCF) is a special method in which HF-orbitals are optimized simultaneously with a "small" CI. This can be used to study problems where the Hartree-Fock method is inappropriate (e.g. when there

are low-lying excited states), or to generate a good starting wavefunction for a subsequent CI calculation.

$$\Psi_{MCSCF} = \sum_{k=1}^{n} c_k \left| \phi_{k1} \phi_{k2} \phi_{k3} \cdots \phi_{kn} \right|$$

The MCSCF method requires considerable care in the selection of the basis set and especially the active space, and should not be considered for routine use. In contrast to the HF, MPn and CI methods, MCSCF does not provide a "model chemistry" because each problem requires different choices. MCSCF methods are essential for the study of processes in which transitions between potential energy surfaces occur, such as in photochemical reactions [41, 42]. A combination of MP2 with MCSCF has recently been explored by Roos et al. [43]. This seems to be a very promising method for excited states.

Other methods to determine the correlation energy are under development. At this point it is useful to note another promising development, that of *density functional theory*. This is a method in which the two-electron integrals are not computed in the conventional way. Application of this approach to molecular systems is still in its infancy, but rapid developments are to be expected in the next few years, in particular driven by the desire to be able to compute larger systems, e.g. metal complexes and organometallic compounds.

2.6 Density Functional Theory [44]

The existence of correlations between the particles, the main formal difficulty encountered in treating a materials problem in quantum mechanics, is a familiar one in many contexts. The positions and motions of the particles that make up a molecule or material are correlated because the particles interact with each other and exert forces upon each other as they move. In quantum mechanics, the situation is further compounded by the mysterious forces that devolve from the Pauli exclusion principle governing electrons. This causes correlations to appear even between (fictitious) noninteracting particles that have no direct interaction with each other. Such forces are referred to as exchange forces because they have to do with the set of rules in quantum mechanics that govern what happens when the labels characterizing indistinguishable particles are exchanged.

Whether due to interactions (e.g., the Coulomb force) or exchange, correlations can be characterized as either long- or short-range. The former can be dealt with by averaging techniques and a mean-field or a self-consistent field (meaning that the field experienced by an atom depends on the global distribution of atoms). Short-range correlations involve the local environment around a particular atom, i.e., deviations of the local environment from average behavior, and are much more difficult to treat. In large part, the central problem of quantum methods in chemistry and condensed matter physics has been the search for more and more accurate ways of incorporating short-range correlations into mean-field theory. The massive cpu requirement of codes that employ modern methods such as coupled clusters or Quantum Monte Carlo bear witness to the degree of difficulty of the problem. These methods are applicable only to relatively small

molecules or very simple crystalline solids and their scaling properties as the system size increases are very unfavorable.

Fortunately, the fine details of short range correlations are often of only minor importance so that a theory based on the concept of a mean or self-consistent field is sufficiently accurate for many purposes. Where this is not the case, as in the high temperature ceramic superconductors, or valence-mixed solids, one refers to strongly correlated systems, implying that the shortrange correlations between electrons due to exchange and their mutual Coulomb repulsions must be accounted for very accurately if even the qualitative features of observed behavior are to be reproduced.

Several promising methods of dealing with the problem of strong correlations have been developed in recent years but this is still at the cutting edge of research in condensed matter physics and none of these methods is quite ripe for inclusion in a suite a general software tools. An important advance in the calculation of the energy of collections of atoms and the forces on each atom was made by Kohn and Sham (1965), who showed how a mean-field theory could be applied to this problem. In their method, the electron density plays a crucial role so that, although the term has more general applicability, the Kohn-Sham method is commonly referred to as density functional theory. This has since advanced to become a very important method for determining the energy of many-electron, and therefore many-atom systems. In addition, Kohn-Sham density functional theory is equally applicable to molecules (bounded collections of atoms) and crystalline materials (where a specific unit cell is repeated throughout space).

In density functional theory, the energy is not written in terms of the many-electron wavefunction as is conventional in quantum chemistry, but as a functional of the electron density. Kohn and Sham proposed that the functional for a system of electrons with external field Vext(x) be written in the form

$$E_{KS}[\rho(x)] = T_s[\rho(x)] + E_{es}[\rho(x)] + E_{xc}[\rho(x)] + E_{ext}[\rho(x)] \tag{2.1}$$

where the terms refer to the kinetic energy of non-interacting electrons having density ρ (x), the electrostatic energy, the so-called exchange-correlation energy, and the potential energy of non-interacting electrons having density ρ(x) in the external field Vext(x). The important advance of Kohn and Sham was the correction of a defect of earlier forms for the density functional (such as the Thomas-Fermi-Dirac functional) with regard to reproducing the shell structure of atoms. This is achieved in the Kohn-Sham functional via the kinetic energy term which is expressed by a set of orbitals, Φn, emanating from a one-particle Schrodinger equation;

$$T_s[\rho(x)] = \sum_n a_n \int dx\, \phi_n(x) \left\{ -\frac{\hbar^2 \nabla^2}{2m} \right\} \phi_n(x) \tag{2.2}$$

$$\rho(x) = \sum_n a_n \left|\phi_n(x)\right|^2 \tag{2.3}$$

The link between Ts and ρ(x) is then indirect, via the orbitals, Φn, in terms of which

Here the an are occupation numbers that determine the electron configuration. Ts[ρ(x)] and ρ(x), as given by Equation 2.2 and Equation 2.3, provide the required link between a density and the kinetic energy with which it is associated.

For purposes of practical calculation, the Kohn-Sham functional must be supplemented by an approximation for the exchange and correlation term. The traditional approximation, proposed by Kohn and Sham, is referred to as the "local density approximation" (LDA) and takes the form

$$E_{xc}^{h}[\rho(\mathbf{x})] = \int d\mathbf{x}\, \rho(\mathbf{x})\ \varepsilon_{xc}^{h}(\rho(\mathbf{x})) \tag{2.4}$$

where is the exchange correlation energy of a homogeneous electron gas having density r. Although this form of the exchange correlation energy appears to be valid only in the limit that the electron density is slowly varying (in which case Equation 2.4 is the first term in a gradient expansion), a posteriori calculation showed that the expression remains relatively accurate in general, even when the density is so rapidly varying that a gradient expansion of it does not exist. Arguments of a dimensional nature having nothing to do with gradient expansions help to explain the general accuracy of Equation 2.4 and suggest why this expression gives a reasonable estimate of the exchange-correlation energy irrespective of the nature of the density distribution. The quantity εhxc(ρ) has been calculated in several ways by different groups. The calculations give similar, but not identical results. The differences to be expected on switching from one LDA functional to another are, in general, only marginal.

The LDA remained the approximation of choice for Exc for many years (and is still for some applications, particularly in extended systems). In applications to molecules, however, it was found that the LDA tends to overbinding (too large values of molecular binding energies). This can be understood as a consequence of a known defect of Equation 4. In regions of low electron density. Here, the exact form of ε_{hxc} is known (it is some kind of electrostatic interaction having the functional form of a power law) and deviates greatly from the LDA which falls off exponentially with the electron density. This means that the exchange correlation contribution emanating from regions of low electron density is underestimated, which, in turn, implies that the difference in energy between two systems whose electron distributions have different "surface areas" will be in error. This is the case when two atoms combine to form a molecule and the sign of the effect is consistent with overbinding of the molecule.

Over the past decade, a class of corrections to the LDA has been developed that correct this deficit to a large extent by going over explicitly to the power law form in regions of low density. This is usually done by introducing a dependence on the gradient of the density and the new class of corrected exchange-correlation functionals is referred to as gradient corrected or Generalized Gradient Approximations (GGA). The use of gradient corrections has little influence on local properties such as bond lengths or vibration frequencies, but does lead usually to a significant improvement in global changes in the energy such as those that result when two atoms form to make a molecule, or a molecule binds on a surface. The hunt for yet further improvement in exchange-correlation functionals continues, though this is unsystematic and there is no guarantee

that higher accuracy can be attained than is already exhibited by the functionals commonly in use today.

The energy of a system of electrons in an external field (such as that due to a collection of nuclei) is given by minimizing the density functional Equation 2.1. This is equivalent to solving a set of Kohn- Sham equations comprising a one-particle Schrodinger equation together with a so-called self-consistency condition. The Schrodinger equation links the input potential to the output density of the Schrodinger equation:

$$\{-\frac{\nabla^2}{2m} + V_{eff}(\mathbf{x}) - \varepsilon_n\}\phi_n(\mathbf{x}) = 0 \tag{2.5a}$$

$$V_{eff}(\mathbf{x}) = V_{ext}(\mathbf{x}) + \Phi(\rho(\mathbf{x})) + \mu_{xc}(\rho(\mathbf{x})) \tag{2.5b}$$

where F is the Coulomb potential corresponding to $\rho(x)$,

$$\mu_{xc}(\rho) \equiv \frac{d}{d\rho}\left[\rho\,\varepsilon_{xc}^{h}(\rho)\right] \tag{2.6}$$

and the output density, $\rho(x)$, is given in terms of the orbitals by Equation 2.3, Equation 2.5a, and Equation, 2.5b are usually solved by iteration. Beginning with a start potential, Equation 2.5a is solved and its out density calculated from the orbitals via Equation 2.3. Then this density is used to form a new potential for Equation 2.5a. The self- consistency cycle is then continued until the in potential and the out-density satisfy Equation 5b to some desired accuracy. This often involves many iterations because the self-consistency

procedure is inherently unstable. Sophisticated "feedback" techniques are necessary to prevent oscillations.

Once self-consistency is achieved, the calculational output includes the energy, Equation 2.1, it's derivatives with respect to the nuclear coordinates (i.e., the atomic forces), the eigenvalues of Equation 2.5a, (which in extended systems give the energy bands), and the one electron orbitals $\rho(x)$. According to formal density functional theory, only the energy and its derivatives (the forces on the ions) have physical significance. However, practical calculation over many decades has shown that many other quantities, calculated approximately in a "one-electron picture" using the eigenvalues (energy bands) and orbitals or Equation 2.5a, are given with equal accuracy. These include (in many cases) the optical absorption, which is treated by assuming the electrons of the system to be excited from occupied to unoccupied levels as the result of photon absorption, and the magnetic structure of materials. This is calculated using a spin-polarized version of the theory in which the electrons of up-spin and down-spin may experience different potentials. It is then possible for the system to adopt a symmetry broken configuration wherein there is a preponderance of one kind of spin and therefore a magnetic state. The use of the local spin-density approximation for the exchange correlation energy, which is analogous to Equation 2.4 but with allowance for different densities for up- and down-spins, gives surprisingly accurate data for the magnetic structure of metals and alloys. Spin-polarized calculations are also important in dealing with open-shell atoms and molecules.

In short, the solution of the Kohn-Sham equations, Equation 2.5a and Equation 2.5b, for a collection of atoms, whether in a molecule, cluster or extended solid provides

a wealth of information about the system. This includes structural information, such as the equilibrium geometry, and a wide variety of important electronic properties. In addition, dynamical and thermal behavior can be studied using forces generated by the solution of the Kohn-Sham equation in, e.g., molecular dynamics calculations (so-called ab-initio molecular dynamics). Although Eqs. 5 are very much simpler than standard quantum mechanics - because the Coulomb interaction is treated via a mean-field - this does not mean that they can be easily solved. The functional dependence of the exchange correlation energy density on the electron density is non analytic, so exact, analytic solutions are not possible even for the hydrogen atom. Methods yielding numerically exact solutions are possible, but only for very small systems (atoms and small, light molecules). In general, approximate methods must be used. Over the years, a number of standard methods have been applied with varying degrees of success. Each has strengths and weaknesses in terms of the systems and/or properties for which it is most accurate.

2.7 Properties Derived from the Wavefunction

The electronic wavefunction which is computed in ab initio as well as semi-empirical quantum chemical methods can be used to derive observable quantities of a molecule, but it can also be analyzed and used to rationalize certain chemical phenomena.

2.7.1 Electrical Properties

The electric dipole moment μ of a molecule can be calculated directly from the positions of the nuclei and the electronic wavefunction [45]:

$$\mu = 2.5416\left(\sum_{A}^{nuclei} Z_A r_A - \sum_{\mu}^{N}\sum_{\nu}^{N} P_{\mu\nu}\langle\mu|r|\nu\rangle\right)$$

The dipole moment can be viewed as the first term of an expansion of the electric field due to the molecule, the next higher term being the quadrupole moment. It is also possible to obtain the dipole moment and polarizabilities directly as derivatives of the energy with respect to a uniform electric field [35]. The electrostatic potential of the molecule represents the interaction between the charge distribution of the molecule and a unit point charge located at some position p:

$$\varepsilon_p = \sum_{A}^{nuclei} \frac{Z_A}{R_{Ap}} - \sum_{\mu}^{N}\sum_{\nu}^{N} P_{\mu\nu} \frac{\langle\mu|\nu\rangle}{r_{1p}}$$

Calculation of the molecular electrostatic potential at the surface of the molecule (described by the total electron density) can indicate how the molecule will interact with polar molecules or charged species. Visualization of this can be nicely accomplished using color coding [45].

2.7.2 Atomic Charges

Although concepts like atomic point charges or bond dipoles are widely used in molecular mechanics, there is no unique definition of atomic charge in a molecule. All ways to attribute a part of the electron density to individual atoms are to a certain extent arbitrary. As a first analysis, or as a way to compare related systems, Mulliken Population Analysis can be applied. The electron density distribution (the probability of finding an electron in a volume element d**r**) is:

$$\rho(r) = \sum_{\mu}^{N} \sum_{\nu}^{N} P_{\mu\nu} \varphi_{\mu} \varphi_{\nu}$$

Integrated over entire space this gives the total number of electrons ($S_{\mu\nu}$ is the overlap):

$$\int \rho(r) dr = \sum_{\mu}^{N} \sum_{\nu}^{N} P_{\mu\nu} S_{\mu\nu} = n$$

This can be separated into diagonal and off-diagonal terms, where the former represent the net population of the basis orbitals and the latter are make up the overlap population.

$$\sum_{\mu}^{N} P_{\mu\mu} + 2 \sum_{\mu}^{N} \sum_{<\nu}^{N} P_{\mu\nu} S_{\mu\nu} = n$$

$P_{\mu\mu}$ net population of φ_{μ}

$Q_{\mu\nu} = 2P_{\mu\nu}S_{\mu\nu} =$ overlap population of φ_{μ} and φ_{ν}

In the Mulliken scheme the overlap population is simply shared between the contributing atoms, which leads to the following charge for each basis orbital:

$$q_\mu = P_{\mu\mu} + \sum_{\mu \neq \nu} P_{\mu\nu} S_{\mu\nu}$$

Summing of the charges in the orbitals associated with each atom gives the atomic charge. An important disadvantage of the Mulliken population analysis is that extended basis sets can lead to unphysical results, e.g. charges of more than 2e, which result from the fact that the basis orbitals centered at one atom actually describe electron density close to another nucleus. Population Analysis based on *Natural Atomic Orbitals* does not have this problem.An approach which may be physically more relevant is to fit charges at the atomic positions to the *molecular electrostatic potential* measured at a grid of points. This still leaves some arbitrariness in the choice of the grid, and the procedure is computationally much more demanding than the other types of population analysis.

2.8 Basis Set Effects

A basis set is the mathematical description of the orbitals within a system(which in turn combine to approximate the total electronic wavefunction) used to perform the theoretical calculation. Larger basis sets more accurately approximate the orbitals by imposing fewer restrictions on the locations of the electrons in space. In the true quantum mechanical picture, electrons have a finite probability of existing anywhere in space. The standard basis sets use linear combinations of gaussian functions to form the orbitals. Gaussian offers a range of pre-defined basis sets, which may be classified by the number and types of functions that they contain. Basis sets assign a group of basis functions to

each atom within a molecule to approximate its orbitals. These basis functions themselves are composed of gaussian functions; the former are then referred to as contracted gaussians (or contracted functiosn), and the latter are referred to as primitives.

2.8.1 Minimal Basis Sets

Minimal basis sets contain the minimum number of basis functions needed for each atom, as in these examples:

H: 1s

C: 1s, 2s, 2px, 2py, 2pz

Minimal basis sets used fixed-size atomic-type orbitals. The STO-3G [46] basis set is a minimal basis set (although it is not the smallest possible basis set). It uses three gaussian primitives per basis function, which accounts for the "3G" in its name. "STO" stands for "Slater-type orbitals", and the STO-3G basis set approximates Slater orbitals with gaussian functions.

2.8.2 Split Valence Basis Sets

The first way that a basis set can be made larger is to increase the number of basis function per atom. Split valence basis sets, such as 3-21G [47] and 6-31G [48], have two (or more) sizes of basis function for each valence orbital. For example, hydrogen and carbon are represented as:

H: 1s, 1s′

C: 1s, 1s′, 2s, 2s′, 2px, 2px′, 2py, 2py′, 2pz, 2pz′

Where the primed and unprimed orbitals differ in size.

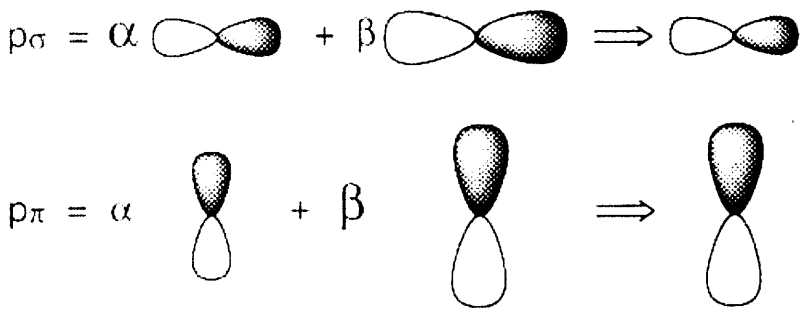


Figure 2.4 The Diagram of split valence basis sets.

The double zeta basis sets, such as the Dunning-Huzinaga basis set (D95), form all molecular orbitals from linear combinations of two sizes of functions for each atomic orbitals. Similarly, triple zeta basis sets, like 6-311G [49], use three sizes of contracted functions for each orbital-type.

2.8.3 Polarized Basis Sets

Split valence basis sets allow orbitals to change size, but not to change shape. Polarized basis sets remove this limitation by adding orbitals with angular momentum beyond what is required for the ground state to the description of each atom. For example, polarized basis sets add d functions to carbon atoms and f functions to transition metals, and some of them add p functions to hydrogen atoms.

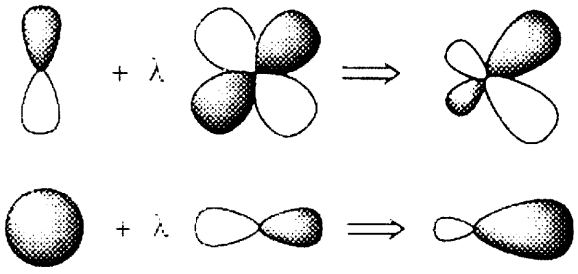


Figure2.5 The diagram of polarized basis sets.

Experience suggests that d-type functions are required on second row and heavier main-group elements even though they are not occupied in the free atoms. (This situation is very much like that found for alkali and alkaline earth elements where p-type functions, while not occupied in the ground-state atoms, are required for proper description of bonding in molecules.) This applies not only to molecules with expanded valence octets (so-called "hypervalent molecules") but also to normal-valent systems. The polarized basis set 6-31G(d), indicates that it is the 6-31G basis set with d functions added to heavy atoms. This basis set is becoming the standard basis set for calculations involving up to medium-sized systems. This basis set is also known as 6-31G*. The popular polarized basis set is 6-31G(d,p), also known as 6-31G**, which adds p functions to hydrogen atoms in addition to the d functions on heavy atoms [50]. Other basis set which has proven to be quite successful for molecules incorporating heavy main-group elements is 3-21G*, constructed from 3-21G basis sets by the addition of a set of d-type functions on second-row and heavier by main-group elements only [51].

2.8.4 Diffuse Functions

Basis sets with diffuse functions are important for systems where electrons are relatively far from the nucleus: molecules with lone pairs, anions and other systems with significant negative charge, systems in their excited states, systems with low ionization potentials, descriptions of absolute acidities. Diffuse functions are large-size versions of s- and p-type functions (as opposed to the normal, contracted functions). They allow orbitals to occupy a larger region of space.

The 6-31+G(d) basis set is the 6-31G(d) basis set with diffuse functions adde3d to heavy atoms. The double plus version, 6-31++G(d), adds diffuse functions to the hydrogen atoms as well. This addition is usually relatively inexpensive, but seldom makes a tremendous difference in accuracy. Diffuse functions have no significant effect on the optimized structure of methanol but do significantly affect the bond angles in methoxide anion. They are required to produce an accurate structure for the anion [39].

2.8.5 High Angular Momentum Basis Sets

Even larger basis sets are now practical for many systems. Such basis sets add multiple polarization functions per atom to the triple zeta basis set. For example, the 6-311G(2d,p) basis set adds two d functions per heavy atom instead of just one, while the 6-311++G(3df, 3pd) basis set contains three sets of valence region functions, diffuse functions on both heavy atoms and hydrogens, and multiple polarization functions: 3 d and 1 f function on heavy atoms and 3 p and 1 d function on hydrogen atoms. These basis sets are useful for describing the interactions between electrons in electron correlation methods; they are not generally needed for Hartree-Fock calculations [39].

2.9 Semi-empirical Method

Ab initio quantum chemical methods are limited in their practical applicability because of their heavy demands of cpu-time and storage space on disk or in the computer memory. At the Hartree-Fock level the problem is seen to be in the large number of two-electron integrals that need to be evaluated. Without special tricks this is proportional to the fourth power of the number of basis functions. In practice this can be reduced to something close to the third power for larger molecules, e.g. because use is made of the fact that integrals between orbitals centered on distant atoms need not be calculated because they will be zero anyway.

Still, the size of systems that can be treated is limited, and this holds much more strongly for correlated treatments. MP2 for example formally scales with the fifth power of the number of basis functions. Therefore there is a place for more approximate methods that retain characteristics of the quantum-chemical approach, in particular the calculation of a wavefunction from which electronic properties can be derived. In this Section, There is overview of commonly used semi-empirical methods [45].

The semi-empirical methods are based on the Hartree-Fock approach. A Fock-matrix is constructed and the Hartree-Fock equations are iteratively solved. The approximations are in the construction of the Fock matrix, in other words in the energy expressions. Recall how the Fock matrix elements are expressed as integrals over atomic basis functions:

$$F_{\mu\nu} = \langle \mu | \hat{F} | \nu \rangle = \langle \mu | \hat{H}_1 | \nu \rangle + \sum_{\lambda,\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \tfrac{1}{2}(\mu\lambda|\nu\sigma) \right]$$

in which P is the density matrix:

$$P_{\lambda\sigma} = 2\sum_{i}^{occ} c_{i\lambda} c_{i\sigma}$$

To simplify matters drastically, the Zero Differential Overlap (ZDO) approximation assumes:

$$\varphi_{\mu}(r)\varphi_{\nu}(r) = 0 \qquad \text{for } \mu \neq \nu$$

which implies that

$$S_{\mu\nu} = \langle \mu | \nu \rangle = \delta_{\mu\nu} \text{ in which } \delta_{\mu\nu} = 0 \text{ if } \mu \neq \nu,\ \delta_{\mu\nu} = 1$$

This can be justified when the atomic basis orbitals are orthogonalized (Löwdin orthogonalization).

As a result of the ZDO approximation many two-electron integrals vanish:

$$(\mu\nu|\lambda\sigma) = \delta_{\mu\nu}\delta_{\lambda\sigma}(\mu\mu|\lambda\lambda)$$

Another common feature of semi-empirical methods is that they only consider the valence electrons.The core electrons are accounted for in a core-core repulsion function, together with the nuclear repulsion energy. In the most popular semi-empirical methods used today (MNDO, AM1 and PM3) the ZDO approximation is only applied to basis functions on different atoms. This is called the NDDO approximation (Neglect of

integrals by parameters, which can either have fixed values, or depend on the distance between the atoms on which the basis functions are located. At this stage empirical parameters can be introduced, which can be derived from measured properties of atoms or diatomic olecules. In the modern semi-empirical methods the parameters are however mostly devoid of this physical significance: they are just optimized to give the best fit of the computed molecular properties to experimental data. Different semi-empirical methods differ in the details of the approximations (e.g. the core-core repulsion functions) and in particular in the values of the parameters. Note that in contrast to molecular mechanics, only parameters for single atoms and for atom pairs are needed. The number of published parameters increases steadily. The semi-empirical methods can be optimized for different purposes. The MNDO, AM1 and PM3 methods were designed to reproduce heats of formation and structures of a large number of organic molecules. Other semi-empirical methods are specifically optimized for spectroscopy, e.g. INDO/S or CNDO/S, which involve CI calculations and are quite good at prediction of electronic transitions in the UV/VIS spectral region.

Some even more approximate methods are still quite useful. In the Hückel and Extended Hückel methods the whole sum over two-electron integrals is replaced by a single diatomic parameter (the resonance integral), so that no search for a self-consistent field is necessary (nor possible). These methods have proven extremely valuable in qualitative and semi-quantitative MO theories of pi-electron systems and of organometallic systems [36]. For pi-electron systems ZDO treatments have been developed that take only pi-centers (p-atomic orbitals) into account, but do perform the SCF calculation. In the MM2 and MM3 programs pi-electron calculations are used to

adjust the force constants and equilibrium values of bond lengths to the prevailing bond order. The pi-bond order between two atoms is simply the sum over Mos of the product of the coefficients of the basis functions on the atoms in the MO, multiplied by the occupation number of the MO:

$$P_{rs}^{\pi} = \sum_{i} n_i c_{ri} c_{si}$$

For a given geometry the pi-electron calculation is done, and the bond-orders computed. Then the force field is adjusted: the force constants for stretching and torsion are scaled and the equilibrium bond length for the bonds between the pi-centers are calculated.

$$k_{s_{ij}} = k_s(0) + c_s p_{ij}^{\pi}$$

$$r_{0_{ij}} = r_0(0) + d_s p_{ij}^{\pi}$$

When the geometry changes too much, the pi-electron treatment is repeated to adjust the force field to the new situation. For the pi-electron calculation, the pi-system is treated as if it is planar. Otherwise the bond order for a twisted semi-single bond would become smaller as the bond is twisted more, and the " restoring force" towards planarity (conjugation) would vanish.

CHAPTER 3

QSAR METHODOLOGY

Drug design is an iterative process which begins with a compound that displays an interesting biological profile and ends with optimizing both the activity profile for the molecule and its chemical synthesis. The process is initiated when the chemist conceives a hypothesis which relates the chemical features of the molecule (or series of molecules) to the biological activity. Without a detailed understanding of the biochemical process(es) responsible for activity, the hypothesis generally is refined by examining structural similarities and differences for active and inactive molecules. Compounds are selected for synthesis which maximize the presence of functional groups or features believed to be responsible for activity.

The combinatorial possibilities of this strategy for even simple systems can be explosive. As an example, the number of compounds required for synthesis in order to place 10 substituents on the four open positions of an asymmetrically disubstituted benzene ring system is approximately 10,000. The alternative to this labor intensive approach to compound optimization is to develop a theory that quantitatively relates variations in biological activity to changes in molecular descriptors which can easily be obtained for each compound. A Quantitative Structure Activity Relationship (QSAR) can then be utilized to help guide chemical synthesis. This chapter develops the concepts used to derive a QSAR and reviews the application of these techniques to medicinal research.

3.1 Statistical Concepts

Computational chemistry represents molecular structures as numerical models and simulates their behavior with the equations of quantum and classical physics. Available programs enable scientists to easily generate and present molecular data including geometries, energies and associated properties (electronic, spectroscopic and bulk). The usual paradigm for displaying and manipulating these data is a table in which compounds are defined by individual rows and molecular properties (or descriptors) are defined by the associated columns. A QSAR attempts to find consistent relationships between the variations in the values of molecular properties and the biological activity for a series of compounds so that these "rules" can be used to evaluate new chemical entities.

A QSAR generally takes the form of a linear equation

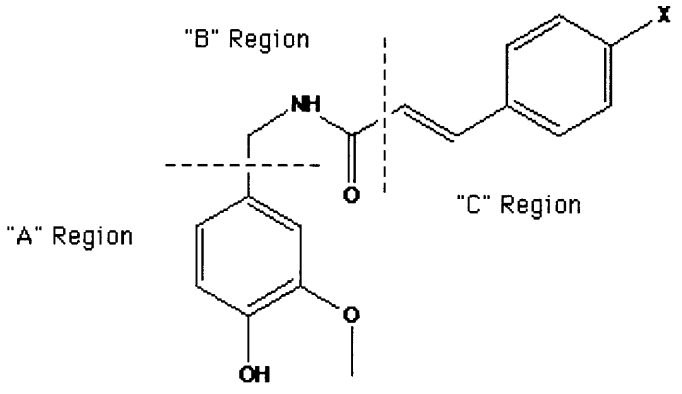
$$\text{Biological Activity} = \text{Const} + (C1 \bullet P1) + (C2 \bullet P2) + (C3 \bullet P3) + \ldots$$

where the parameters P1 through Pn are computed for each molecule in the series and the coefficients C1 through Cn are calculated by fitting variations in the parameters and the biological activity. Since these relationships are generally discovered through the application of statistical techniques, a brief introduction to the principles behind the derivation of a QSAR follows.

The work reported from The Sandoz Institute for Medical Research on the development of novel analgesic agents [53] can be used as an example of a simple QSAR. In this study, vanillylamides and vanillylthioureas related to capsaicin were prepared and their activity was tested in an in vitro assay which measured [54] Ca^{2+}

influx into dorsal root ganglia neurons. The data, which was reported as the EC50 (μM), is shown in Table 3.1 (note that compound 6f is the most active of the series).

Table 3.1 Capsaicin Analogs Activity Data



Compound Number	Name	X	EC50(μM)
1	6a	H	11.80 ± 1.90
2	6b	Cl	1.24 ± 0.11
3	6d	NO2	4.58 ± 0.29
4	6e	CN	26.50 ± 5.87
5	6f	C6H5	0.24 ± 0.30
6	6g	N(CH3)2	4.39 ± 0.67
7	6h	I	0.35 ± 0.05
8	6i	NHCHO	???

In the absence of additional information, the only way to derive a best "guess" for the activity of 6i is to calculate the average of the values for the current compounds in the series. The average, 7.24, provides a guess for the value of compound 8 but, how good is this guess? The graphical presentation of the data points is shown in Figure 3.1.

The standard deviation of the data, **s**, shows how far the activity values are spread about their average. This value provides an indication of the quality of the guess by showing the amount of variability inherent in the data. The standard deviation is calculated as shown below.

$$s = \sqrt{\frac{(11.8 - 7.24)^2 + (1.24 - 7.24)^2 + (...)^2}{7 - 1}}$$

$$s = \sqrt{\frac{539.41}{6}} = 9.48$$

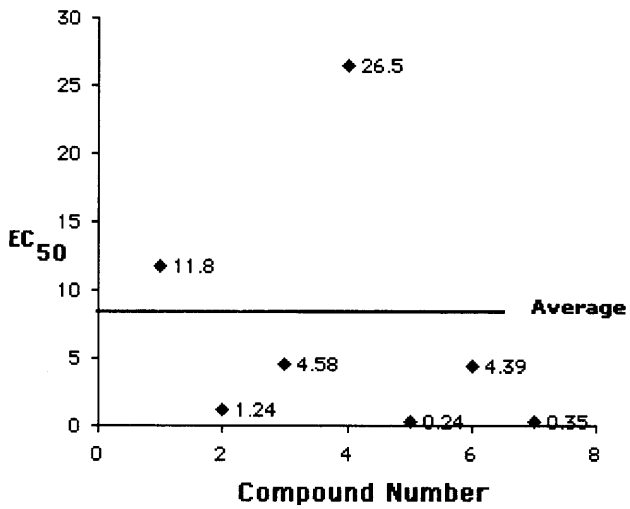


Figure 3.1 Capsaicin analogs activity data.

Rather than relying on this limited analysis, one would like to develop an understanding of the factors that influence activity within this series and use this understanding to predict activity for new compounds. In order to accomplish this objective, one needs:

- binding data measured with sufficient precision to distinguish between compounds;
- a set of parameters which can be easily obtained and which are likely to be related to receptor affinity;
- a method for detecting a relationship between the parameters and binding data (the QSAR) and
- a method for validating the QSAR.

The QSAR equation is a linear model which relates variations in biological activity to variations in the values of computed (or measured) properties for a series of molecules. For the method to work efficiently, the compounds selected to describe the "chemical space" of the experiments (the training set) should be diverse. In many synthesis campaigns, compounds are prepared which are structurally similar to the lead structure. Not surprisingly, the activity values for this series of compounds will frequently span a limited range as well. In these cases, additional compounds must be made and tested to fill out the training set.

The quality of any QSAR will only be as good as the quality of the data which is used to derive the model. Dose-response curves need to be smooth, contain enough points to assure accuracy and should span two or more orders of magnitude. Multiple readings for a given observation should be reproducible and have relatively smaller errors. The issue being addressed is the signal-to-noise ratio. The variation of the readings obtained by repeatedly testing the same compound should be much smaller than the variation over the series. In cases where the data collected from biological experiments do not follow

these guidelines, other methods of data analysis should be utilized since the QSAR models derived from the data will be questionable.

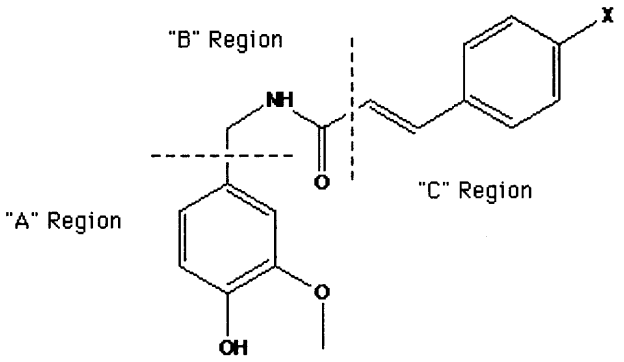
Once biological data has been collected, it is often found that the data is expressed in terms which cannot be used in a QSAR analysis. Since QSAR is based on the relationship of free energy to equilibrium constants, the data for a QSAR study must be expressed in terms of the free energy changes that occur during the biological response. When examining the potency of a drug (the dosage required to produce a biological effect), the change in free energy can be calculated to be proportional to the inverse logarithm of the concentration of the compound.

$\Delta G0 = - 2.3RTlogK = \log 1/[S]$

Further, since biological data are generally found to be skewed, the log transformation moves the data to a nearly normal distribution. Thus, when measuring responses under equilibrium conditions, the most frequent transformation used is to express concentration values (such as IC50, EC50, etc.) as log[C] or log 1/[C]. The transformed data for the capsaicin agonists are shown in Table 3.2.

The effect of this transformation on the spread of the data relative to the average is shown in Figure 3.2. Note that the data points, projected onto the Y-axis, have become more uniformly distributed.

Table 3.2 Capsaicin Analogs Transformed Data



Compound Number	Name	X	EC50(mM)	Log EC50	Log 1/EC50
1	6a	H	11.80 ± 1.90	1.07	-1.07
2	6b	Cl	1.24 ± 0.11	0.09	-0.09
3	6d	NO_2	4.58 ± 0.29	0.66	-0.66
4	6e	CN	26.50 ± 5.87	1.42	-1.42
5	6f	C_6H_5	0.24 ± 0.30	-0.62	0.62
6	6g	$N(CH_3)_2$	4.39 ± 0.67	0.64	-0.64
7	6h	I	0.35 ± 0.05	-0.46	0.46
8	6i	NHCHO	?? ± ??	??	??

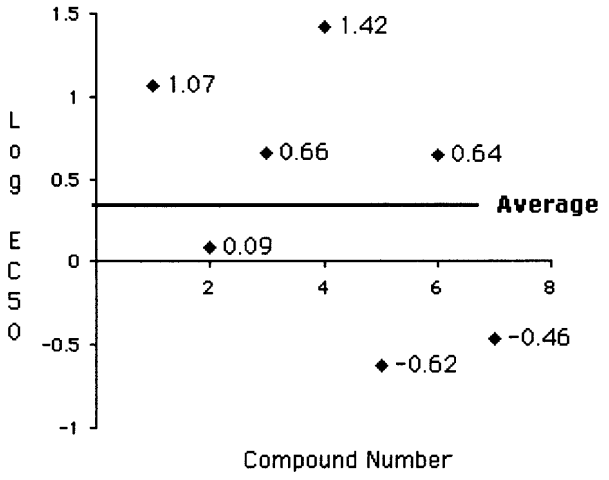


Figure 3.2 Capsaicin analogs transformed data.

Given the transformed data, the best guess for the activity of 6i is still the average of the data set (or 0.40). As before, the error associated with this guess is calculated as the square root of the average of the squares of the deviations from the average.

$$s = \sqrt{\frac{(1.07 - 0.40)^2 + (0.09 - 0.40)^2 + (...)^2}{7-1}}$$

$$s = \sqrt{\frac{3.4906}{6}} = 0.76$$

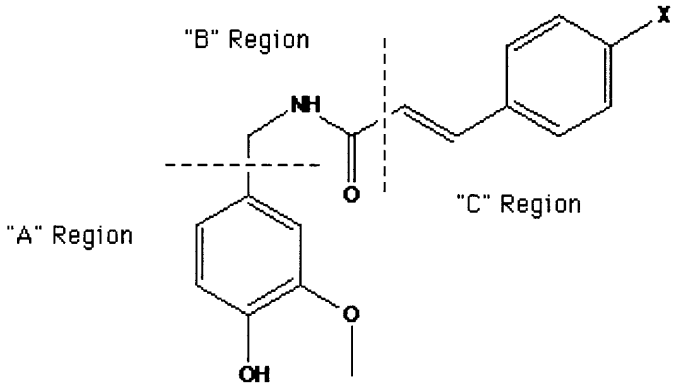
This is an example data set intended to show the general approach; real data sets would have many more compounds and descriptors. The purpose of a QSAR is to highlight relationships between activity and structural features.

There are several potential classes of parameters used in QSAR studies. Substituent constants and other physico-chemical parameters (such as Hammett sigma constants) measure the electronic effects of a group on the molecule. Fragment counts are used to enumerate the presence of specific substructures. Other parameters can include topological descriptors and values derived from quantum chemical calculations.

The selection of parameters is an important first step in any QSAR study. If the association between the parameter(s) selected and activity is strong, then activity predictions will be possible. If there is only weak association, knowing the value of the parameter(s) will not help in predicting activity. Thus, for a given study, parameters should be selected which are relevant to the activity for the series of molecules under investigation and these parameters should have values which are obtained in a consistent manner.

The Sandoz group divided their analysis of capsaicin analogs into three regions: the A-region which was occupied by an aromatic ring; the B-region which was defined by an amide bond; and the C-region which was occupied by a hydrophobic side-chain (See figure in Table 3.1). The hypothesis for the C-region assumed that a small, hydrophobic substituent would increase activity. Given this assumption, the parameters selected to best define this characteristic were molar refractivity (size) and π, the hydrophobic substituent constant. These values are given in Table 3.3.

Table 3.3 Capsaicin Analogs Parameter Values



Compound Number	Name	X	LogEC50	π	MR
1	6a	H	1.07	0	1.03
2	6b	Cl	0.09	0.71	6.03
3	6d	NO_2	0.66	-0.28	7.36
4	6e	CN	1.42	-0.57	6.33
5	6f	C_6H_5	-0.62	1.96	25.36
6	6g	$N(CH_3)_2$	0.64	0.18	15.55
7	6h	I	-0.46	1.12	13.94
8	6i	NHCHO	??	??	??

The data above can be analyzed for relationships by two means: graphically and statistically. The most visual approach to a problem with a limited number of variables is graphical. In this case, a plot of activity versus either molar refractivity or hydrophobicity

gives some insight into the relationship between the parameters and activity. The plots derived by the Sandoz group are reproduced in Figure 3.3.

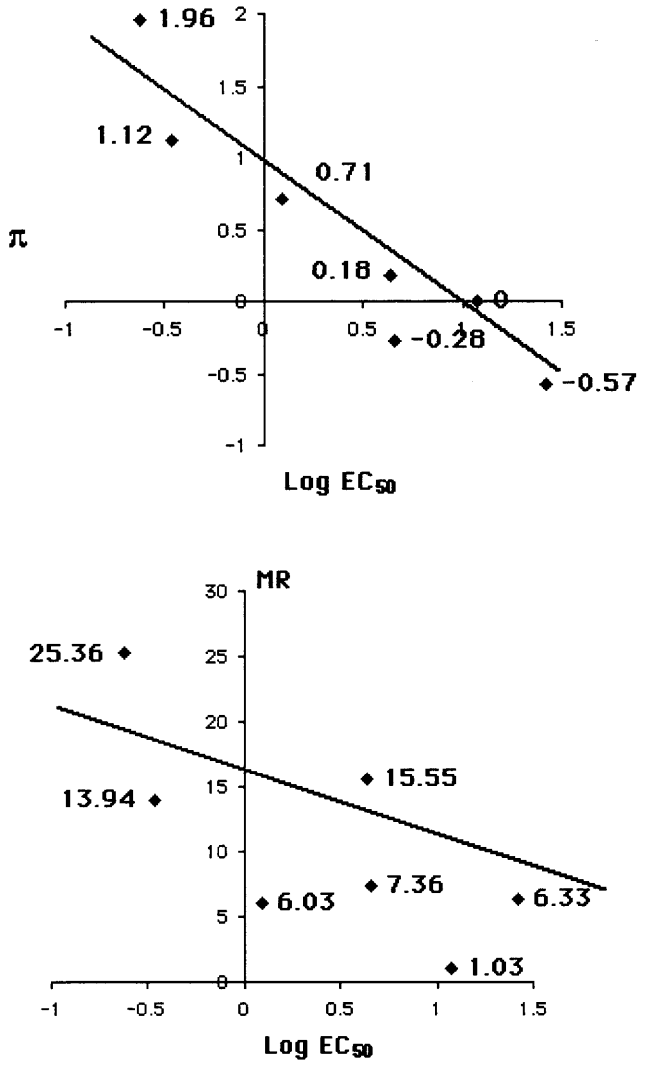


Figure 3.3 Capsaicin analogs parameter values.

The graph provide insight into the the activity for compound 6i. The values for either the hydrophobicity or molar refractivity parameters for this compound provide a

good estimate for activity, since this is a simple example where only two values are examined. In more complex situations however, where multiple parameters are correlated to activity, statistics is used to derive an equation which relates activity to the parameter set. The linear equation which defines the best model for this set of data is

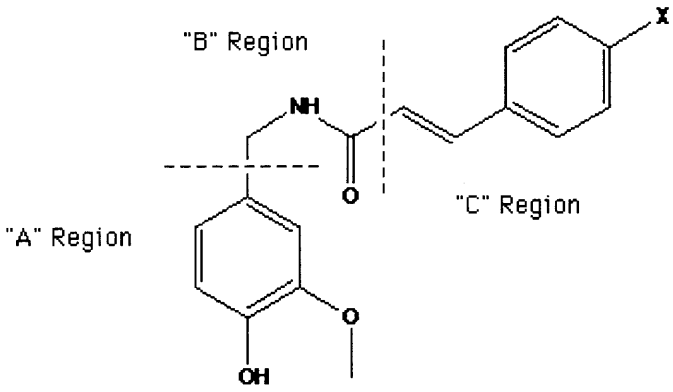
$$\text{Log EC50} = 0.764 - (0.817)\pi$$

How much confidence should be placed in this model? The first step to answering this question is to determine how well the equation predicts activities for known compounds in the series. The equation above estimates the average value for the EC50 based on the value for π; because assays vary, it is not surprising that individual values will differ from the regression estimate. The difference between the calculated values and the actual (or measured) values for each compound is termed the residual from the model. The calculated values for activity and their residuals (or the errors of the estimate for individual values) are shown in Table 3.4.

The residuals are one way to quantify the error in the estimate for individual values calculated by the regression equation for this data set. The standard error for the residuals is calculated by taking the root-mean-square of the residuals (in this calculation, the denominator shown as decremented by two to reflect the estimation of two parameters).

$$S = \sqrt{\frac{0.28^2 + (-0.12)^2 + (-0.36)^2 + \ldots + (-0.34)^2}{(7-2)}} = 0.28$$

Table 3.4 Capsaicin Analogs Calculated Values



Compound Number	Compound Name	X	Log EC50	π	Calculated Log EC50	Residual
1	6a	H	1.07	0	0.79	0.28
2	6b	Cl	0.09	0.71	0.21	-0.12
3	6d	NO_2	0.66	-0.28	1.02	-0.36
4	6e	CN	1.42	-0.57	1.26	0.16
5	6f	C_6H_5	-0.62	1.96	-0.81	0.19
6	6g	$N(CH_3)_2$	0.64	0.18	0.65	-0.01
7	6h	I	-0.46	1.12	-0.12	-0.34
8	6i	NHCHO	??	-0.98	1.6	??

In order to be an improved model, the standard deviation of the residuals calculated from the model should be smaller than the standard deviation of the original data. The standard error about the mean was previously calculated to be 0.76 whereas the standard error from the QSAR model is 0.28. Clearly, the the use of linear regression has improved the accuracy of the analysis. The plot of measured values versus calculated is shown in Figure 3.4 with a 45° line.

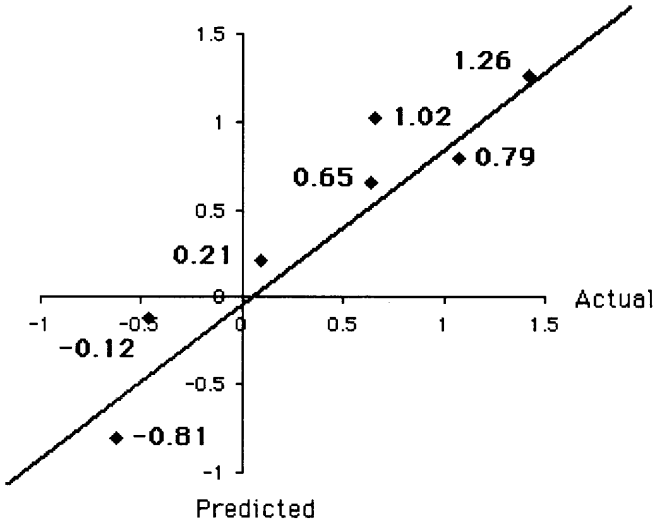


Figure 3.4 Capsaicin analogs predicted versus actual EC50 values.

There are several assumptions inherent in deriving a QSAR model for a series of compounds. First, it is assumed that parameters can be calculated (or measured in some cases) more accurately and cheaply than activity can be measured. Second, it is assumed that deviations from the best fit line follow a normal (Gaussian) distribution. Finally, it is assumed that any variation in the line described by the QSAR equation is independent of the magnitude of both the activity and the parameters. Given these assumptions, the quality of the model can be gauged using a variety of techniques.

Variation in the data is quantified by the correlation coefficient, r, which measures how closely the observed data tracks the fitted regression line. Errors in either the model or in the data will lead to a bad fit. This indicator of fit to the regression line is calculated as:

$$r^2 = \frac{\text{Sum-of-Squares of the deviations from the regression line}}{\text{Sum-of-Squares of the deviations from the mean}}$$

$$r^2 = \frac{\text{Regression Variance}}{\text{Original Variance}}$$

where the Regression Variance is defined as the Original Variance minus the Variance around the regression line. The Original Variance is the sum-of-the-squares distances of the original data from the mean. This can be viewed graphically as shown in Figure 3.5.

The calculation is carried out as follows:

Original Variance = (1.07 - 0.40)2 + (0.09 - 0.40)2 + ...

Original Variance = 3.49

Variance around the line = (0.28)2 + (- 0.12)2 + (- 0.36)2 + ...

Variance around the line = 0.40

Regression Variance = Original Variance - Variance around the line

Regression Variance = 3.49 - 0.40 = 3.09

r^2 = Regression Variance/Original Variance

$r^2 = 3.09/3.49$

$r^2 = 0.89$

Possible values reported for r^2 fall between 0 and 1. An r^2 of 0 means that there is no relationship between activity and the parameter(s) selected for the study. An r^2 of 1 means there is perfect correlation. The interpretation of the r^2 value for the capsaicin analogs is that 89% of the variation in the value of the Log EC50 is explained by variation in the value of π, the hydrophobicity parameter.

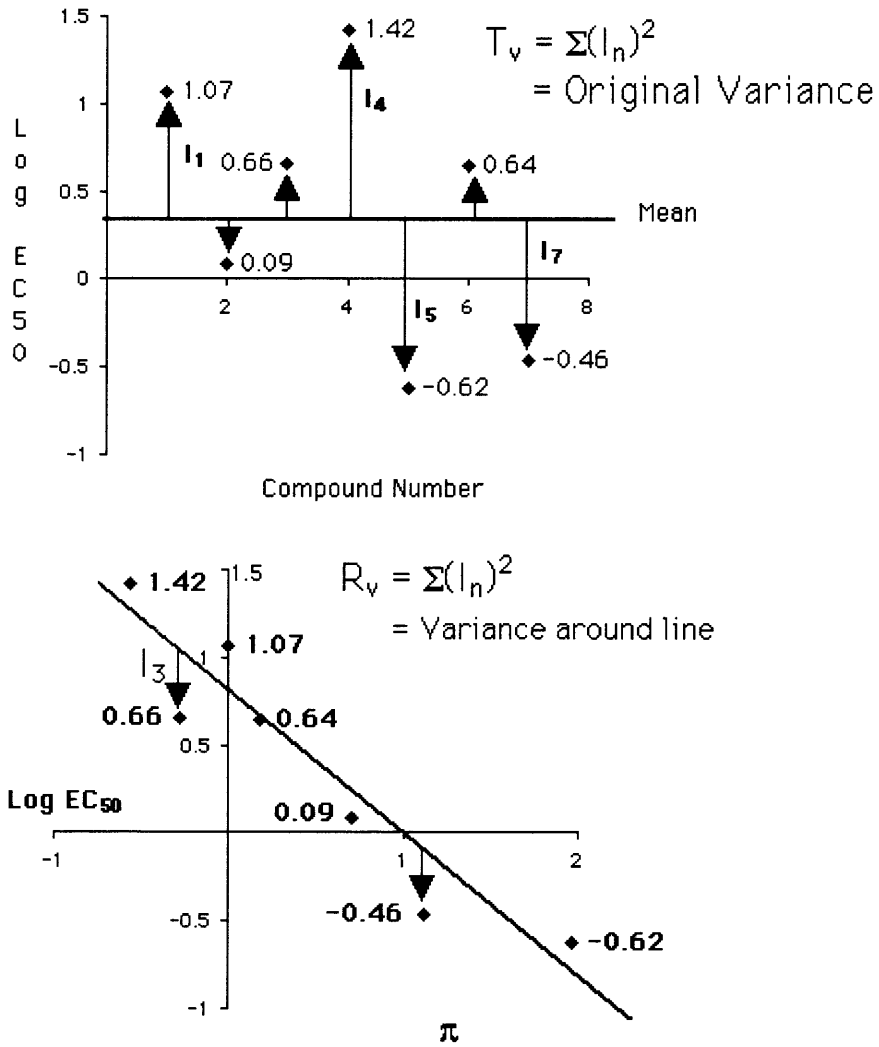


Figure 3.5 Capsaicin analogs derivation of r^2 values.

While the fit of the data to the regression line is excellent, how can one decide if this correlation is based purely on chance? The higher the value for r^2 the less likely that the relationship is due to chance. If many explanatory variables are used in a regression equation, it is possible to get a good fit to the data due to the flexibility of the fitting process; a line will fit two points perfectly, a quadratic curve will fit three, multiple linear regression will fit the observed data if there are enough explanatory variables2. Given the

assumption that the data has a Gaussian distribution, the F statistic below assesses the statistical significance of the regression equation. The F statistic is calculated from r^2 and the number of data points (or degrees of freedom) in the data set. The F ratio for the capsaicin analogs is calculated as:

$$F_{1,n} = (n-2)\frac{r^2}{1 - r^2} = (7-2)\frac{0.89}{1 - 0.89} = 40.46$$

This value often appears as standard output from statistical programs or it can be checked in statistical tables to determine the significance of the regression equation. In this case, the probability that there is no relationship between activity and the π value is less than 1% (p=0.01). It is found that hydrophobicity values correlate well with biological activity. Does the addition of a size parameter (MR) improve the model? In order to analyze a relationship which is possibly influenced by several variables (or properties), it is useful to assess the contribution of each variable. π and MR appear to be somewhat correlated in this data set so the order of fitting can influence how much the second variable helps the first. Multiple linear regression is used to determine the relative importance of multiple variables to the overall fit of the data.

Multiple linear regression attempts to maximize the fit of the data to a regression equation (minimize the squared deviations from the regression equation) for the biological activity (maximize the r^2 value) by adjusting each of the available parameters up or down. Regression programs often approach this task in a stepwise fashion. That is, successive regression equations will be derived in which parameters will be either added or removed until the r^2 and s values are optimized. The magnitude of the coefficients

derived in this manner indicate the relative contribution of the associated parameter to biological activity.

There are two important caveats in applying multiple regression analysis. The first is based on the fact that, given enough parameters any data set can be fitted to a regression line. The consequence of this is that regression analysis generally requires significantly more compounds than parameters; a useful rule of thumb is three to six times the number of parameters under consideration. The difficulty is that regression analysis is most effective for interpolation and it is extrapolation that is most useful in a synthesis campaign (i.e., the region of experimental space described by the regression analysis has been explained, but projecting to a new, unanalyzed region can be problematic).

Using multiple regression for the capsaicin analogs, one can derive the following equation which relates hydrophobicity and molar refractivity to biological activity.

$$\text{Log EC50} = 0.762 - (0.819)\pi + (0.011)\text{MR}$$

$$s = 0.313,\ r^2 = 0.888$$

To judge the importance of a regression term, three items need to be considered.

1. Statistical significance of the regression coefficient.

2. The magnitude of the typical effect bixi (in this case, 0.011 •25.36).

3. Any cross-correlation with other terms.

As more terms are added to multiple linear regression, r^2 always gets larger. The previous calculations (r^2 = 0.89) were recomputed carrying three significant figures so that rounding does not lead to confusion. These results of this analysis indicate that,

within this series, steric bulk is not an important factor in activity. The influence of the hydrophobicity constant confirms the presence of a hydrophobic binding site. Given the limited number of substituents in this analysis, it is unlikely that more can be learned from further analysis.

This Section has developed the fundamental mathematics of QSAR studies. Several authors have published reviews of QSAR and have discussed various aspects of the methods3-8. Each of the examples to follow uses these techniques to derive information about the chemical factors which are important for activity.

3.2 Approaches to Developing a QSAR

Drugs exert their biological effects by participating in a series of events which include transport, binding with the receptor and metabolism to an inactive species. Since the interaction mechanisms between the molecule and the putative receptor are unknown in most cases (i.e., no bound crystal structures), one is reduced to making inferences from properties which can easily be obtained (molecular properties and descriptors) to explain these interactions for known molecules. Once the relationship is defined, it can be used to aid in the prediction of new or unknown molecules.

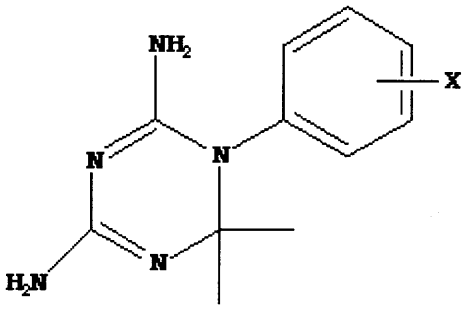
The first approach to developing quantitative relationships which described activity as a function of chemical structure relied on the principles of thermodynamics. The free-energy terms ΔE, ΔH and ΔS were represented by a series of parameters which could be derived for a given molecule. Electronic effects such as electron donating and withdrawing tendencies, partial atomic charges and electrostatic field densities were defined by Hammett sigma (σ) values, resonance parameters (R values), inductive

parameters (F values) and Taft substituent values (ρ^*, σ^*, Es). Steric effects such as molecular volume and surface area were represented by values calculated for Molar Refractivity and the Taft steric parameter. Enthalpic effects were calculated using partition coefficients (LogP) or the hydrophobic parameter, π, which was derived from the partition coefficient. In addition, an assortment of structural indices were used to describe the presence of specific functional groups at positions within the molecule. The linear equation which described the relationship between activity and this parameter set was the Hansch equation

$$\log 1/[C] = A(\log P) - B(\log P)2 + C(Es) + D(\rho\sigma) + E + \ldots$$

Multiple linear regression analysis was used to derive the values of the coefficients. In general, Hansch type studies were performed on compounds which contained a common template (usually a rigid one such as an aromatic ring) with structural variation limited to functional group changes at specific sites.

Hansch utilized this approach in his analysis of 256 4,6-diamino-1,2-dihydro-2,3-dimethyl-1-(X-phenyl)-s-triazines which were active against tumor dihydrofolate reductase9. It was demonstrated that for 244 of the compounds, activity could be correlated to the presence of hydrophobic groups at the three and four positions of the N-phenyl ring. The parameters used to derive this correlation were the hydrophobic constant (π) and molar refractivity constant (MR) for meta and para substituents on the N-phenyl ring and six indicator variables I1-I6 which were used to indicate the presence (a value of 1) or absence (a value of 0) of specific structural features. The equation which was formulated from these data using the method of least squares is shown in Figure 3.6.



log 1/[C] = 0.680(π3) - 0.118(π3)2 + 0.230(MR4) - 0.024(MR4)2 + 0.238(I1) - 2.530(I2) - 1.991(I3) + 0.877(I4) + 0.686(I5) + 0.704(I6) + 6.489

n = 244, r = 0.923, s = 0.377

Figure 3.6 Analysis of the Baker Triazines.

The optimal values for MR4 (4.7) and π3 (2.9) were obtained from the partial derivatives of the equation. Note that the number of compounds in the data set was reduced to 244. Hansch and Silipo reported improvements in the value for r and s by removing 12 compounds which were incorrectly predicted by a factor of 10 or more. While there are limits to the Hansch approach, it permitted complex biological systems to be modeled successfully using simple parameters. The approach has been used successfully to predict substituent effects in a wide number of biological assays. The main problem with the approach was the large number of compounds which were required to adequately explore all structural combinations. Further, the analysis methods did not lend themselves to the consideration of conformational effects. Several authors

have published articles which provide additional background on the Hansch approach [55-57].

Alternative approaches to compound design have been suggested which avoid the combinatorial problem found in Hansch type analyses. Free and Wilson used a series of substituent constants which related biological activity to the presence of a specific functional group at a specific location on the parent molecule12. The relationship between biological activity and the presence or absence of a substituent was then expressed by the following equation:

$$\text{Activity} = A + \Sigma_i \Sigma_j G_{ij} X_{ij}$$

where A was defined as the average biological activity for the series, Gij the contribution to activity of a functional group **i** in the jth position and Xij the presence (1.0) or absence (0.0) of the functional group **i** in the jth position.

The procedure used the equation above to build a matrix for the series and represented this matrix as a series of equations. Substituent constants then were derived for every functional group at every position. Statistical tests were used to test the importance of the constants. If the models were shown to be valid, the model was used to predict activity values for compounds which had not been prepared. In general, while a large number of compounds are required to explore the effects of multiple substitution patterns, the Free-Wilson approach substantially reduces the number of analogs required. However, the method demands that the effects of substituents are additive.

In 1972, John Topliss published a paper which detailed methodology to automate the Hansch approach [58]. The method assumed that the lead compound of interest

contained at least one phenyl ring which could serve as the template for functional group modifications. The first modification to the template was preparation of the para-chloro derivative to examine lipophilicity. Additional substitution patterns were then made sequentially in an attempt to explore and optimize the relationship between activity and the hydrophobic and electronic character of the molecule. While the Topliss approach is easy to follow, it has several drawbacks. The primary problems are that the procedure is not applicable to all types of studies and that there is a high degree of risk associated with its use (it essentially ignores the possibility of interactions between substituents as it changes one substituent at a time).

The use of classical QSAR was expanded during the 1960's as a means of correlating observed activity to chemical properties. However, there are many areas where these techniques could not be used or where they failed to provide useful correlations. These included situations in which activity was found to be determined by 3-dimensional geometry, where poor training sets of compounds were used or the set of compounds were too small or insufficiently diverse and cases where biological activity could not be well quantified. Many of these problems were addressed by extensions to the Hansch method and the development of alternative approaches to QSAR.

There are cases where biological activity values cannot be determined accurately for a variety of reasons, e.g. lack of sensitivity of a particular test system. Alternative statistical techniques can be used in these cases; the problem is simplified to a classification scheme in which compounds are labeled as active, partially active, inactive, etc. The resulting data set is then searched for patterns which predict these categories. The methods which have been used for this type of analysis include SIMCA (Soft,

Independent Modeling of Class Analogy) [59], ADAPT (Automated Data Analysis by Pattern recognition Techniques) [60], CASE (Computer Automated Structure Evaluation) [61] and CSA (Cluster Significance Analysis) [62].

Pattern recognition methods [63] attempt to define the set of parameter values which will result in clustering compounds of similar activity into regions of n-dimensional space. The methods used to accomplish this goal can be parametric or nonparametric. Parametric methods search the n-dimensional space for clusters of compounds based on their calculated properties. These methods do not use derived values (e.g., mean vectors and covariance matrices), but instead use the original data to find clustering definitions and apply iterative procedures to find the linear set of parameters which best define the classification scheme.

Where the methods described above develop discriminant functions, SIMCA methods use Principal Component Analysis (PCA) to describe the data set. The objective of PCA is to create a reduced number variables which describe biological activity or chemical properties into a relatively few independent ones. This is accomplished through an analysis of the correlation matrix of biological or chemical properties.

Principle component analysis can be used to create derived variables for each class (e.g., active and inactive) separately by decomposing the correlation matrix; this method is useful to point out redundancies or interrelationships among the variables. PCA seeks to find simplified relationships in data by transforming the original parameters into a new set of uncorrelated variables which are termed principal components. The symmetric correlation matrix is decomposed by an eigenvalue decomposition. The largest eigenvalue and its eigenvector are used to form a linear combination of the original

variables with maximum variance. Successively smaller eigenvalues and vectors produce linear combinations of the original variables with diminishing variance. Successive eigenvectors are independent of one another. The simplification is derived by disregarding eigenvectors associated with small eigenvalues. In summary, the procedure finds the set of orthogonal axes for the data which decompose variance in the data.

Another approach to examining the effects of chemical structure on activity was developed by the Jurs' group. Rather than rely on multivariant statistics to highlight these relationships, Jurs used the combination of cluster analysis and pattern recognition techniques as a tool to develop these correlations. The ADAPT program generated a data set of molecular descriptors (topological, geometrical and physicochemical) derived from three dimensional model building, projected these data points onto an n-dimensional surface and analyzed them using pattern recognition methods. The goal of this analysis was to discriminate between active and inactive compounds in a series.

Jurs has reported several applications of the methodology contained in ADAPT. In one study of chemical carcinogens [64], a linear discriminant function was derived from a set of 28 calculated structure features including fragment descriptors, substructure descriptors, environment descriptors, molecular connectivity descriptors and geometric descriptors. Two hundred and nine compounds from twelve structural classes (130 carcinogens, 79 noncarcinogens) were selected for this study. The program was used to identify a training set of 192 compounds which was used to find the best set of descriptors and analyze the entire data set. A predictive success of 90% for carcinogenic compounds and 78% for noncarcinogenic compounds was obtained in randomized testing. The CASE program extended the techniques in ADAPT by using topological

methods to define substructural fragments which were essential for activity. CASE was able to differentiate between positional isomers. Both CASE and ADAPT are limited to analyzing structurally similar data sets.

The analysis methods described to this point have not explicitly incorporated the contribution of three dimensional shape in the analysis of the activity of a molecule. While the use of chemical graph indexes [65], intermolecular binding distances [66], molecular surface areas [67] and electrostatic potentials [68] contain some information about the 3-D shape of molecules, the Hopfinger [69] and Marshall [70] groups were the first to exhaustively analyze these effects.

In 1979, Marshall extended the 2-D approach to QSAR by explicitly considering the conformational flexibility of a series as reflected by their 3-D shape [70]. The first step of the Active Analog Approach was to exhaustively search the conformations of a compound which was highly active in a particular biological assay. The result of the search was a map of interatomic distances which was used to filter the conformational searches of subsequent molecules in the series. The implicit assumption of the method was that all compounds which display similar activity profiles were able to adopt similar conformations. Once the "active conformation" was determined, molecular volumes for each molecule were calculated and superimposed. Regression analysis of the volumes was used to establish a relationship to biological activity. Marshall and co-workers commercialized the Active Analog Approach and a suite of other drug design techniques in the SYBYL [71] molecular modeling program.

Hopfinger and co-workers also used 3-D shape in QSAR. In molecular shape analysis [72] of the Baker Triazines, the common space shared by all molecules of a series and the differences in their potential energy fields were computed. When these calculations were combined with a set of rules for overlapping the series, comparative indicies of the shape of different molecules were obtained. Inclusion of these shape descriptors in standard Hansch analysis schemes lead to improved descriptions relating computed parameters to biological activity such that no compounds in the original data set had to be eliminated from the calculations. The techniques developed by Hopfinger and co-workers were made available in the CAMSEQ, CAMSEQ-II, CHEMLAB and CAMSEQ-M computer programs.

In 1988, Richard Cramer proposed that biological activity could be analyzed by relating the shape-dependent steric and electrostatic fields for molecules to their biological activity [73]. Additionally, rather than limiting the analysis to fitting data to a regression line, CoMFA (Comparative Molecular Field Analysis) utilized new methods of data analysis, PLS (Partial Least Squares) and cross-validation, to develop models for activity predictions. The approach used in the CoMFA procedure requires that the scientist define alignment rules for the series which overlap the putative pharmacophore for each molecule; the active conformation and alignment rule must be specified. Once aligned, each molecule is fixed into a three-dimensional grid by the program and the electrostatic and steric components of the molecular mechanics force field, arising from interaction with a probe atom (e.g., an SP3 C atom), are calculated at intersecting lattice points within the 3-D grid. The equations which result from this exercise have the form

Act1 = Const1 + a1(stericxyz) + b1(stericxyz) + ... + a'1(estaticxyz) + b'1(estaticxyz) + ...

Act2 = Const2 + a2(stericxyz) + b2(stericxyz) + ... + a'2(estaticxyz) + b'2(estaticxyz) + ...

Actn = Constn + an(stericxyz) + bn(stericxyz) + ... + a'n(estaticxyz) + b'n(estaticxyz) + ...

Traditional regression methods require that the number of parameters must be considerably smaller than the number of compounds in the data set (or the number of degrees of freedom in the data). The data tables which result from CoMFA analysis have far more parameters than compounds. PLS, which removes this limitation, is used to derive the coefficients for all of the steric and electrostatic terms. PLS essentially relies upon the fact that the correlations among near parts of a molecule are similar so that the real dimensionality is smaller that the number of grid points. Since these coefficients are position dependant, substituent patterns for the series are elucidated which define regions of steric bulk and atomic charge associated with increased or decreased activity. The size of the model (the number of components27 needed for the best model) and the validity of the model as a predictive tool are assessed using cross-validation.

As opposed to traditional regression methods, cross-validation evaluates the validity of a model by how well it predicts data rather than how well it fits data. The analysis uses a "leave-one-out" scheme; a model is built with N-1 compounds and the Nth compound is predicted. Each compound is left out of the model derivation and predicted in turn. An indication of the performance of the model is obtained from the cross-validated (or predictive) r^2 which is defined as

r^2 (cross-validated) = (SD - Press)/SD

The SD is the Sum-of-Squares deviation for each activity from the mean. Press (or Predictive Sum of Squares) is the sum of the squared differences between the actual and that predicted when the compound is omitted from the fitting process.

As it was discussed, values for conventional r^2 range from 0 to 1. Values for the cross-validated r^2 are reported by the method to range from -1 to 1. Negative values indicate that biological activity values are estimated by the mean of the activity values better than they are by the model (i.e, the predictions derived from the model are worse than no model). Once a model is developed which has the highest cross-validated r^2, this model is used to derive the conventional QSAR equation and conventional r^2 and s values. The results of the final model are then visualized as contour maps of the coefficients.

The first CoMFA study reported analyzed the binding affinities of 21 steroid structures to human corticosteroid-binding globulins and testosterone-binding globulins. This class of compounds is rigid and was selected to eliminate conformationally dependant effects from the study. The models for each steroid were built from coordinates from the Cambridge Crystallographic Database which were minimized using the Tripos force field. Side chain positioning was accomplished using systematic conformational searching. The Field Fit algorithm was used to align each structure within the fixed lattice (the 3-D grid used to calculate the CoMFA field effects). The fit of the regression line for the predicted versus actual binding values for the corticosteroids showed a cross-validated r^2 of 0.65 (conventional r^2 = 0.897, s = 0.397). For the testosterone-binding steroids, the cross-validated r^2 was 0.555 (conventional r^2 = 0.873, s = 0.453).

As noted, CoMFA starts with defined pharmacophore and overlap rules and derives a 3-D model which can be used to predict activity for new chemical entities. The DISCOtech program is used to identify pharmacophores from databases of chemical structures and biological activity. DISCOtech [74-78] can find different conformers using the Tripos field in reasonable energy limits and suggest possible matcing pharmacophore.

Developing a quantitative structure activity relationship is difficult. Molecules are typically flexible and it is possible to compute many possibly useful properties that might relate to activity. Early in a research program there are typically few compounds to model. It is clear that many training compounds need to span through the space and model fitting techniques need to address not only deriving a fit, but the predictive quality of the fit. While these methods have not discovered a new compound, they have aided scientists in examining the volumes of data generated in a research program. As the methods evolve, they will find broader applications in areas such as combinatorial chemistry.

CHAPTER 4

COMPARATIVE MOLECULAR FIELD ANALYSIS (COMFA) STUDIES USING SEMI-EMPIRICAL, DENSITY FUNCTIONAL, AB INITIO METHODS AND PHARMACOPHORE DERIVATION USING DISCOTECH ON SIGMA 1 RECEPTRO LIGANDS

4.1 Introduction

Sigma receptors have been the focus of extensive studies because of their potentially important roles in biochemical, physiological, and behavioral processes [80-84]. Potential therapeutic usage has been foreseen in psychiatric diseases [85], in the treatment of cocaine abuse, in neuroprotection, in the treatment of schizophrenia, and in mediating antipsychotic effects of inhibiting neurotransmitter release [86-89]. Although the exact structure of the sigma receptor is unknown, it is possible to refer to the receptor properties by finding important commonalities of the most active sigma1 ligands. Comparative molecular field analysis (CoMFA) of the three-dimensional quantitative structure-activity relationship (3D-QSAR) proves to be appropriate for such problems. CoMFA is a widely used 3D-QSAR method for correlating biological activities with three-dimensional structural properties which may be described by a steric and an electrostatic molecular field [90-99]. The CoMFA method systematically samples the steric and electrostatic fields surrounding a set of ligands and constructs a 3D-QSAR model by correlating these 3D steric and electrostatic fields with the corresponding experimental binding affinities.

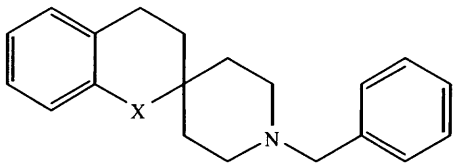
In my research, there are three different CoMFA studies; (1) 43 molecules as a training set and 5 molecules as a test set for the sigma 1 receptor-ligands, using the radioligand [^{3}H](+) pentazocine, (2) 24 PCP derivatives as a training set and three

molecules as a test set for the sigma 1 receptor-ligands, using the radioligand [^{3}H](+)SKF10047, and (3) 21 molecules as a training set and three molecules as a test set for sigma 2 receptor-ligands, using [^{3}H]DTG in the presence of (+) pentazocine.

In these studies, the pharmacophore of sigma receptor-ligands were developed by DISCOtech using Sybyl 6.9 and the suggested conformers by DISCOtech were optimized using AM1 or HF/3-21G* calculation by Gaussian 98 and atomic charges were calculated using AM1, HF/3-21G*, B3LYP/3-21G*, MP2/3-21G* methods. They were analyzed comparatively on steric and electrostatic field contour map to do CoMFA study.

Ab Initio and Density functional produces reliable structures and good result in QSAR compared to semi-empirical, and molecular mechanics but they are expensive to calculate bulky molecules like medicinal ligands or proteins. As CPU has been faster, it has been easier to calculate higher level methods. It is good time to apply quantum mechanics in molecular modeling of 3D QSAR.

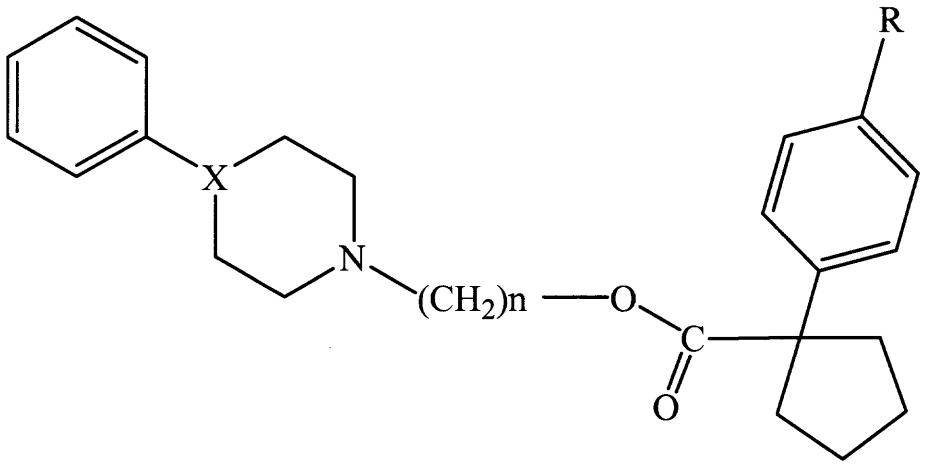
Table 4.1 Binding and Functional Data for Spipethiane and Its Analogs [104]



Compounds No.	X	K_i(nM) σ_1	K_i(nM) σ_2	σ_2/σ_1
1 [a]	CO	1.4± 0.08	10±3	7
2	S	0.5± 0.02	416± 43	832
3	SO	25±0.12	1700±166	68
4	SO_2	20± 1.2	1800± 230	90
5	O	0.62 ±0.07	22±2.4	35

[a] Compound of test sets.

Table 4.2 Binding and Functional Data for Piperidine and Piperazine Analogs [105]

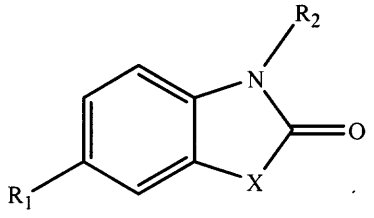


Compounds No.	R	n	X	K_i(nM) σ_1
6	H	2	CH	3.95± 1.14
7	NO_2	2	CH	0.05± 0.02
8	I	2	CH	1.44 ±0.31
9 [a]	CN	2	CH	1.30± 0.13
10	Cl	2	CH	1.34± 0.15
11	OCH_3	2	CH	1.02± 0.23
12	H	2	N	62.8 ±9.77
13	NO_2	2	N	2.78± 0.47
14	I	2	N	22.8± 4.99
15	Cl	2	N	76.9± 9.77
16	H	3	CH	0.50 ±0.11
17 [a]	NO_2	3	CH	0.27 ±0.08
18	Cl	3	CH	1.51± 0.12
19	I	3	CH	0.88 ±0.22
20	OCH_3	3	CH	0.65± 0.18
21	H	4	CH	0.51± 0.12
22	H	5	CH	0.61 ±0.02
23	H	6	CH	1.21 ±0.05

[a] Compounds of test sets.

Table 4.3 Binding and Functional Data for Benzoxazolone and Benzothiazolone Analogs

[106]



Compounds No.	R_1	R_2	X	K_i(nM) σ_1	K_i(nM) σ_2	σ_2/σ_1
24	C_3H_7	$(CH_2)_2$- $C_5H_{10}N$	S	0.6±0.3	18.1±6.2	29
25	C_2H_5CO	$(CH_2)_3$- $C_5H_{10}N$	S	2.3±0.3	202±18	87
26 [a]	C_2H_5CO	$(CH_2)_3$- $C_5H_{10}N$	O	8.5±0.5	496±50	58
27	C_4H_9CO	$(CH_2)_2$- $C_5H_{10}N$	O	23±3	84±9	4
28	C_4H_9CO	$(CH_2)_3$- $C_5H_{10}N$	O	40±3	155±13	4
29	C_2H_5CO	$(CH_2)_2$- $C_5H_{10}N$	S	47±5	481±21	10
30	C_4H_9CO	$(CH_2)_2$- $C_5H_{10}N$	S	49±4	232±35	5
31	C_6H_5CO	$(CH_2)_2$- C_4H_8N	S	54±3	225±16	4
32	C_2H_5CO	$(CH_2)_2$- $C_5H_{10}N$	O	72±6	900±42	13
33	C_7H_7	$(CH_2)_2$- C_4H_8NO	O	72±6	515±25	7
34	C_6H_5CO	$(CH_2)_2$- $C_5H_{10}N$	S	83±7	1906±257	23
35	C_6H_5CO	$(CH_2)_2$- C_2H_6N	S	105±6	>10000	>95
36	C_5H_{11}	$(CH_2)_2$- C_4H_8NO	O	177±21	966±93	5
37	C_6H_5CO	$(CH_2)_2$- C_4H_8N	O	185±10	125±11	0.7
38	C_6H_5CO	$(CH_2)_2$- $C_5H_{10}N$	O	236±25	780±69	3
39	C_4H_9CO	$(CH_2)_2$- C_4H_8N	O	323±20	685±50	2
40 [a]	C_6H_5CO	$(CH_2)_2$- C_2H_6N	O	365±15	>10000	>27
41	C_2H_5CO	$(CH_2)_2$- C_4H_8N	S	571±42	919±87	1.6
42	C_2H_5CO	$(CH_2)_2$- C_4H_8N	O	617±32	494±33	0.8
43	C_4H_9CO	$(CH_2)_2$- C_2H_6N	S	628±40	4612±525	7

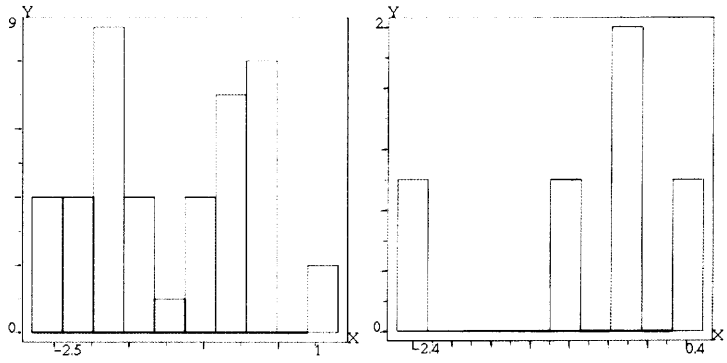
[a] Compounds of Test sets.

4.2 Materials and Methods

4.2.1 Selection of Ligands

The bioactivity of ligands for sigma receptor subtype 1 was critically evaluated for 48 compounds found in the literature. All values were obtained using the radioligand [^{3}H](+) pentazocine. The K_i values were converted to pK_i values ($pK_i = -\log K_i$). Most of the obtained compounds belonged to three structurally different families. These are spipethiane [104] molecules shown in Table 4.1, piperidine and piperazine analogues [105] shown in Table 4.2, and benzoxazolone and benzothiazolone compounds [106] shown in Table 4.3. The remaining compounds [107-109] are shown in Table 4.4. Test compounds were labeled in these tables.

The varied structural diversity and the homogeneous repatriation of the affinities are necessary to obtain meaningful results from a 3D-QSAR study using the CoMFA method. It is also important that the test set reflects the affinity range of the training set to assure a complete evaluation. A training set containing 43 compounds and a test set of five compounds were used to assess the predictive power of the model. The test set was created to contain compounds of all three families. Histogram pictures of train and test sets are shown in Figure 4.1. The range of binding affinities for the training set was –2.79 to 1.30 log units, while –2.56 to 0.57 log units for the test set.

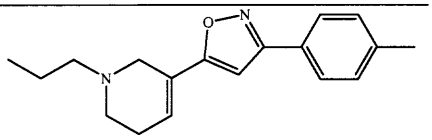
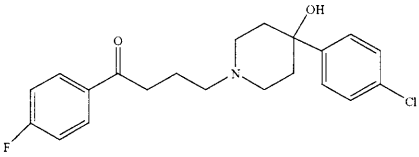
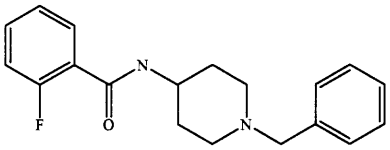
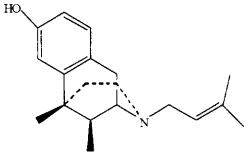
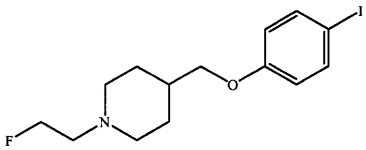


a b

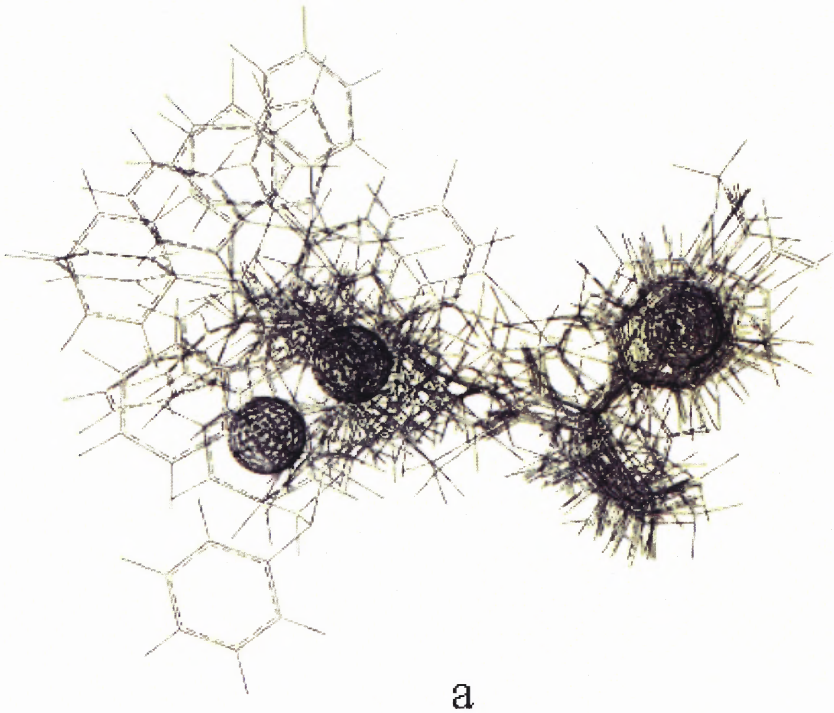
Figure 4.1 Histogram of pKi (abscissa) vs number of molecules (ordinate).

[a] training set, [b] test set.

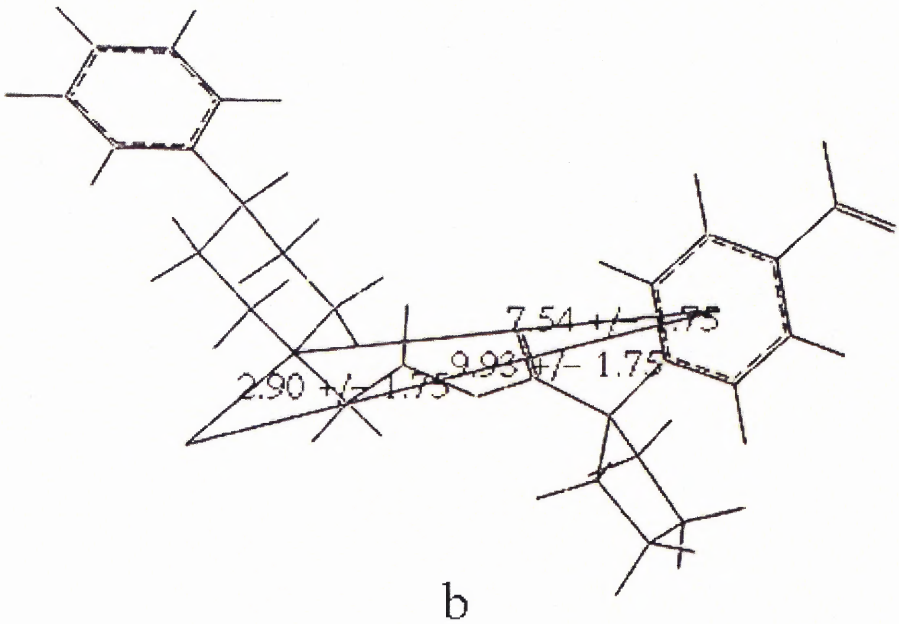
Table 4.4 Other Sigma 1 Ligands Examined in This Study

Number	K_i(nM) σ_1	K_i(nM) σ_2	σ_2/σ_1	Compounds
44 [107], PD144418	0.08±0.01	1377±179	17212.5	
45 [107], Haloperidol	1.2±0.20	26±5.4	21.66667	
46 [108], Flurobenzamide [a]	3.39	406	119.764	
47 [107], (+) –Pentazocine	5.8±1.0	1253± 519	216.0345	
48 [109], Piperidine [b]	0.84	102	121.4286	

[a] N-(N-benzylpiperidin-4-yl)-2-flurobenzamide [108]. [b] 1-(2-fluoroethyl)-4-[(4-iodophenoxy)methyl]piperidine from reference [109].



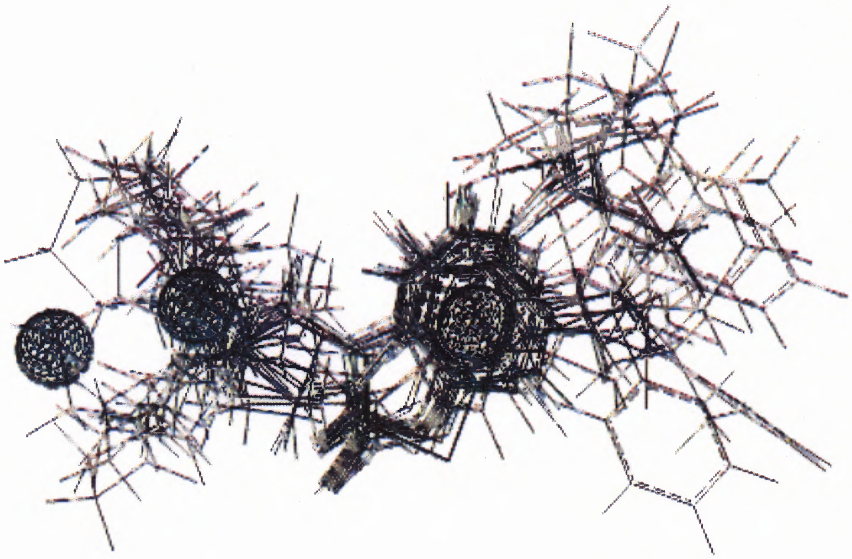
a



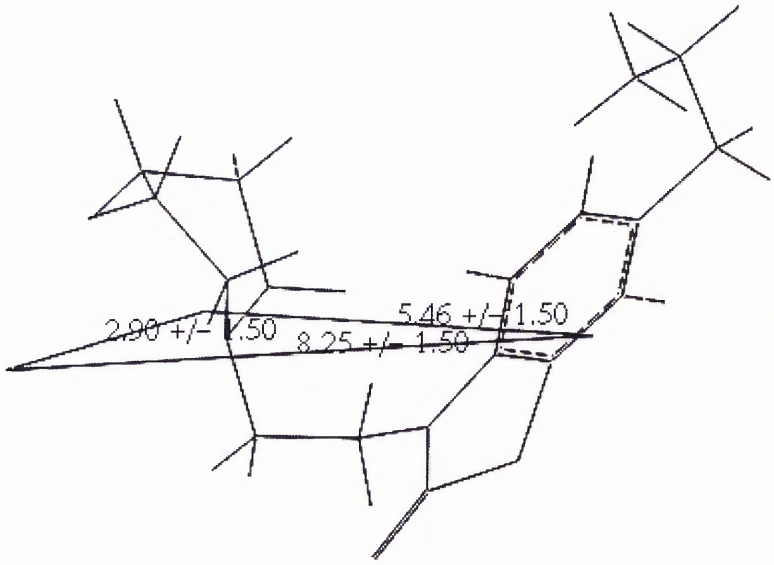
b

Figure 4.2 DISCOtech pharmacophore for Piperidines.

[a] all piperidines in Table 4.2, [b] DISCOtech model with compound number 7 in Table 4.2.



a



b

Figure 4.3 DISCOtech pharmacophore for Benzoxazolone and Benzothiazolones.

[a] all molecules in Table 4.3, [b]. DISCOtech model with compound number 24 in Table 4.3.

4.2.2 Choice of Initial Conformations

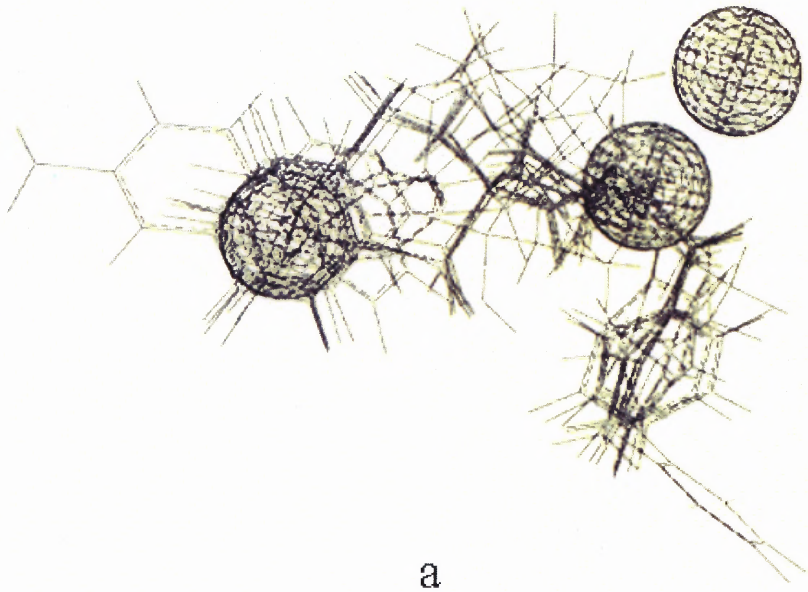
The CoMFA study began with the selection of the three-dimensional conformation for each compound. Initial structures were generated by building with SYBYL6.9 [71] default bond distances and angles, and minimized with the tools MAXIMIN2 in SYBYL6.9, in which the Tripos force field was applied with a distance-dependent dielectric function. Then, DISCOtech [74-78] in SYBYL6.9 [71] was used to search possible conformations and proper pharmacophores using these initial molecules.

4.2.3 Pharmacophore Information

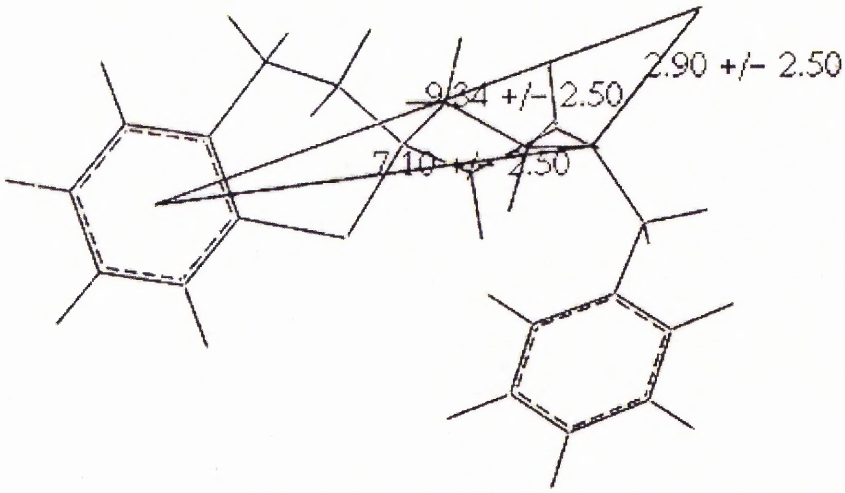
DISCOtech derived a pharmacophore model based on the following categories; piperidine and piperazine in Figure 4.2, benzoxazolone and benzothiazolone in Figure 4.3 and spipethiane and the other selective ligands in Figure 4.4. DISCOtech found possible conformations within reasonable energy (in this study, 25 kcal/mol) and suggested a proper pharmacophore model. The overall pharmacophores are the triangles including nitrogen, the center of phenyl ring and the lone pair of electrons. These are in agreement with the earlier σ1 receptor ligands pharmacophore [109] shown in Figure 4.5.

4.2.4 Geometry Optimzation and Atomic Charges

The conformers derived by DISCOtech were optimized with Gaussian 98 [79] in semi-empirical AM1 [99-102] calculation or ab initio HF/3-21G* [99,103] method. Theses geometries were single point calculated for atomic charges in semi-empirical AM1, density functional B3LYP/3-21G*, or ab initio HF/3-21G* and MP2/3-21G* levels according to Mulliken population.



a



b

Figure 4.4. DISCOtech pharmacophore for Spipethiane and other Ligands.

[a] all molecules in Table 4.1 and 4.4, [b] DISCOtech model with compound number 2 in Table 4.1.

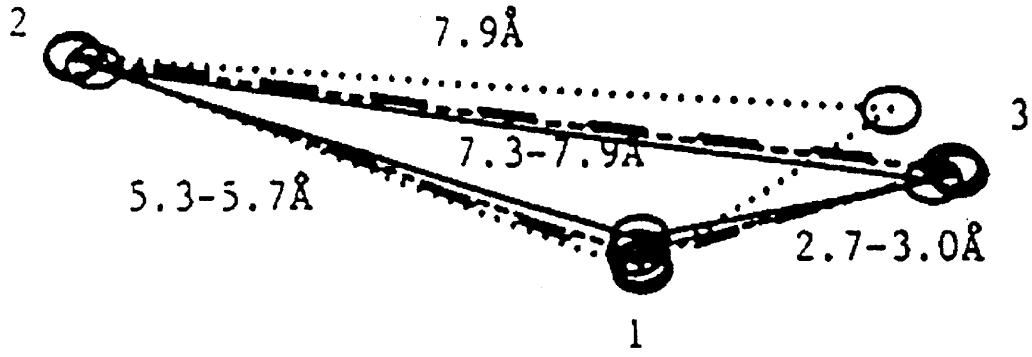


Figure 4.5 Earlier σ1 receptor ligands pharmacophore; 1 = Nitrogen, 2 = the center of phenyl ring, 3 = the lone pair of electrons; pharmacophore was fitted from Progesterone, PRE-084, PPP3, Haloperidol, and Pentazocine.

4.2.5 Alignment

Alignment of the presumed bound conformations of the training set compounds is also an essential prelude to the CoMFA study. The AM1 or HF/3-21G* optimized conformers were aligned by a match function in SYBYL6.9 using a template molecule in order a to c in Figure 4.6; the distance from N atom to a lone pair of electrons was scaled in 1.4Å because it was reported that 1.4Å performed in the best result on CoMFA study for the distance from N atom to a lone pair of electrons [109]. The aligned 48 molecules, used in training and test sets are shown by optimization methods in Figure 4.7 for AM1 and in Figure 4.8 for HF/3-21G*.

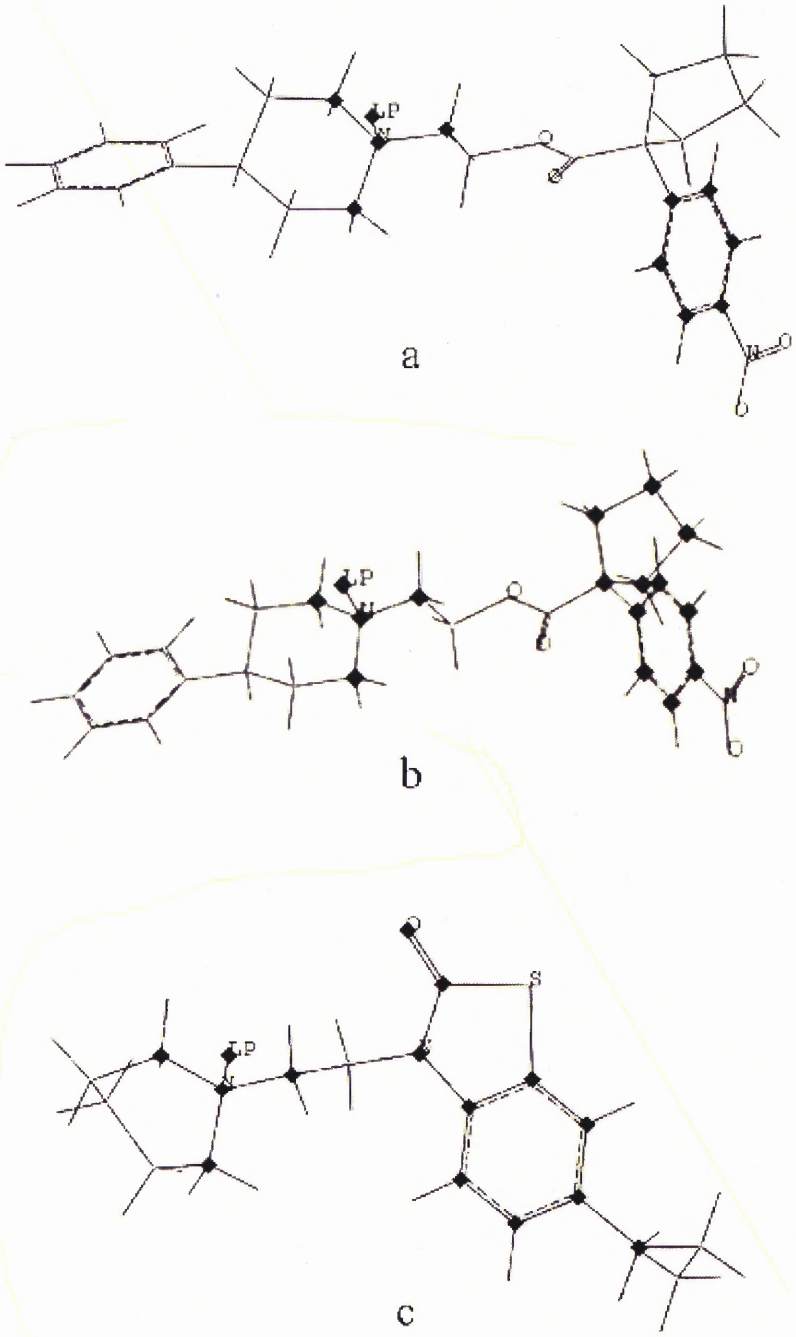


Figure 4.6 Alignment of atoms and groups for CoMFA Study.

Bold atoms were used as match points. [a] Molecule, number 7 was used as the template to match alignments of the most active compounds in Table 4.2 and 4.3 (number 7, 24), 5 spipethiane analogues listed in Table 4.1 and all compounds listed in Table 4.4. [b] Compound, number 7 is used as the template of 18 piperidine and piperazine analogs listed in Table 4.2. [c] Molecule, number 24 was used as the template of 20 benzoxazolone and benzothiazolone analogs listed in Table 4.3.

Table 4.5 The Number of Optimal components [i] and q^2 by "Leave-One-Out" Using SAMPLS [111] by the Training Set of 43 Molecules

Charge// Geometry	Term	C. 1 [j]	C. 2	C. 3	C. 4	C. 5	C. 6	C. 7	C. 8	C. 9	C. 10
AM1//AM1 [a]	S.E.E [k]	0.856	0.839	**0.798**	0.812	0.813	0.853	0.865	0.878	0.888	0.901
	q2 [l]	0.429	0.465	**0.528**	0.524	0.536	0.503	0.502	0.502	0.506	0.506
HF//AM1 [b]	S.E.E	0.859	0.846	0.831	0.828	**0.826**	0.863	0.892	0.903	0.916	0.93
	q2	0.425	0.455	0.488	0.505	**0.52**	0.491	0.471	0.473	0.474	0.474
B3//AM1 [c]	S.E.E	0.859	0.849	0.832	**0.824**	0.839	0.875	0.908	0.919	0.928	0.943
	q2	0.425	0.451	0.487	**0.509**	0.505	0.477	0.452	0.454	0.459	0.459
MP2//AM1 [d]	S.E.E	0.859	0.847	0.831	0.828	**0.826**	0.863	0.892	0.904	0.917	0.931
	q2	0.425	0.455	0.488	0.505	**0.52**	0.491	0.471	0.472	0.473	0.473
HF//HF [e]	S.E.E	0.957	1.016	0.947	0.911	0.848	**0.835**	0.84	0.85	0.867	0.907
	q2	0.286	0.215	0.335	0.4	0.495	**0.523**	0.531	0.533	0.529	0.499
B3//HF [f]	S.E.E	0.957	1.031	0.95	0.917	0.848	**0.839**	0.847	0.856	0.874	0.911
	q2	0.286	0.193	0.331	0.392	0.494	**0.519**	0.523	0.527	0.521	0.495
MP2//HF [g]	S.E.E	0.954	1.008	0.943	0.89	0.851	**0.836**	0.846	0.86	0.876	0.915
	q2	0.292	0.228	0.342	0.427	0.491	**0.521**	0.524	0.522	0.518	0.491
AM1//HF [h]	S.E.E	0.964	0.981	0.909	0.854	0.821	0.815	**0.805**	0.811	0.838	0.867
	q2	0.276	0.269	0.387	0.474	0.526	0.545	**0.569**	0.575	0.56	0.543

[a-d] Charges were calculated using AM1,[a] HF/3-21G*,[b] B3LYP/3-21G*,[c] and MP2/3-21G*,[d] methods and all geometries were optimized using AM1 method. [e-h] Geometries were optimized using HF/3-21G* and charges were calculated using HF/3-21G*,[e] B3LYP/3-21G*,[f] MP2/3-21G*,[g] and AM1[h] methods. [i] The number of optimal components is the bold one in Table in each level. [j] Component number, [k] standard error of estimate, and [l] leave-one-out cross-validated regression of co-efficient.

4.2.6 CoMFA Model

The CoMFA was carried out using the QSAR option of SYBYL6.9. The grid dimensions were running from 0 to 10Å along the X axis, from -6 to 6Å along the Y axis, and from -4 to 2Å along the Z axis, shown as Figure 4.9 to minimize randomness of statistics and focus on the pharmacophore in this study. This column shows the number of lattice interSections located "inside" that molecule, which is a very crude volume estimate. A sp^3-hybridized carbon atom was probed with a +1.0 unit charge, 2.0Å for grid spacing, and the default 30kcal/mol energy cutoff for steric and electrostatic fields.

Partial least squares analysis regresses a target property against predictors calculated as steric and electrostatic components of the intermolecular interaction field. Scaling was used as the CoMFA standard. The SAMPLS (SAMple-distance PLS) algorithm developed by Bruce Bush [111] is used to determine "leave-one-out" cross-validation q^2. The method for cross-validation serves two purposes to find out whether the CoMFA model was productively useful, and if useful, to decide how many components to use for the best model. This number of optical components was considered by the 5% rule; if the q^2 increases by at least 5% upon increasing the number of components by one, then it is justified to add an additional component. The PLS analysis was then repeated without cross-validation using the optimum number of components. This final analysis yielded a predictive model, and a CoMFA coefficient contour plot for the steric and electrostatic potentials.

4.3 Results

4.3.1 Comparative Molecular Field Analysis

The CoMFA model in this study, required three to seven optimal components in different calculations to explain the variance in binding affinity to sigma1 receptors in Table 4.5. All crossvalidated q^2 were more than 0.5 in Table 4.5, which surpassed the generally accepted criterion, 0.3 [112] for statistical validity. This number means that such a model to account for 50.9 to 56.9% of the actual variance in activity among additional similar sigma1 ligands in Table 4.5. Using these optimized geometries, the q^2 results derived from the HF/3-21G method are higher than those from AM1.

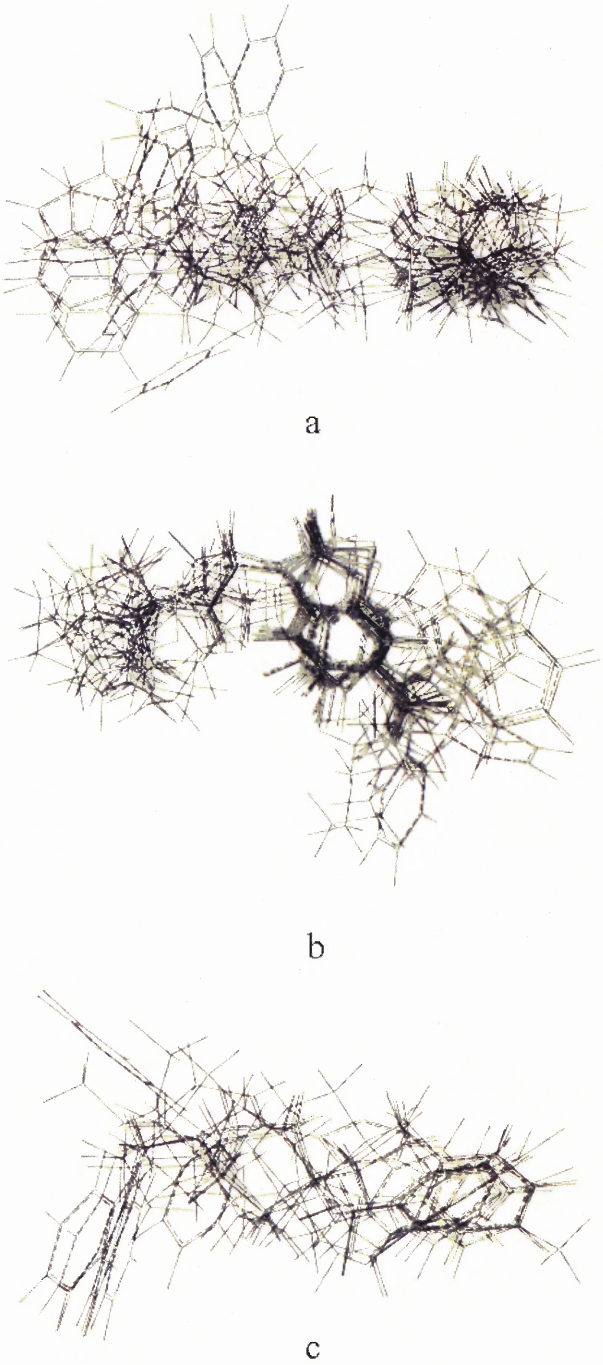
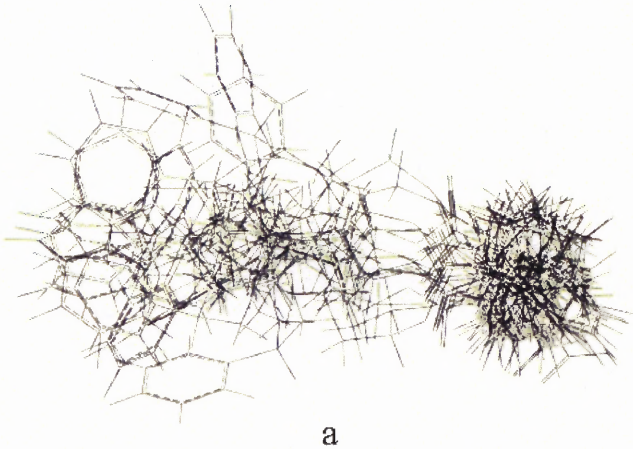
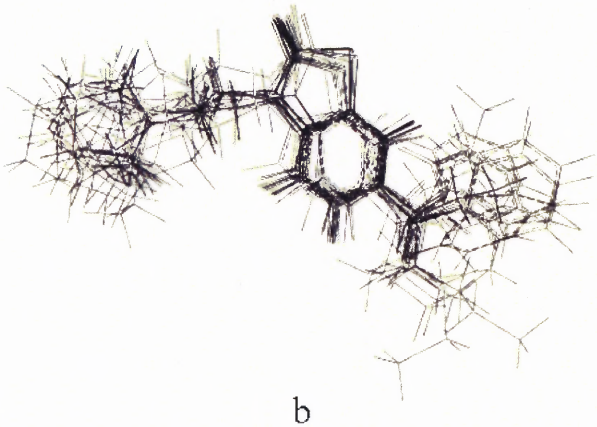


Figure 4.7 Alignments of all molecules, optimized in AM1 method.

[a] AM1 optimized 18 piperidine and piperazine analogs listed in Table 4.2. [b] AM1 optimized 20 benzoxazolone and benzothiazolone analogs listed in Table 4.3. [c] AM1 optimized 5 spipethiane analogues listed in Table 4.1 and the 5 other compounds listed in Table 4.4.



a



b

c

Figure 4.8 Alignments of all molecules, optimized in HF/3-21G* method.

[a] HF/3-21G* optimized 18 piperidine and piperazine analogs listed in Table 4.3. [b] HF/3-21G* optimized 20 benzoxazolone and benzothiazolone analogs listed in Table 4.3. [c] HF/3-21G* optimized 5 spipethiane analogues listed in Table 4.1 and the other compounds listed in Table 4.4.

Otherwise, AM1 optimized geometries suggest that AM1 atomic charges produce higher q^2 than others. The HF/3-21G* optimized geometries also suggest that AM1 atomic charges produce higher q^2 than others.The q^2 value and optimal component for HF//AM1 and MP2//AM1 are the same (q^2 =0.520, n = 5) in Table 4.5. Having the crossvalidation to confirm the predictive ability, a PLS analysis was performed without any validation to derive the best predictive model for use in graphics and in numerical prediction. R^2 measures of fit were 91.1 to 98.9%, and the standard errors of estimate were 0.130 to 0.346. The steric fields contributed 43.7 to 51.4% of the model's information, while the electrostatic fields represented the other 48.6 to 56.3% in Table 4.6. The relationship is shown between experimental and predicted pK_i values for the non-cross-validated analysis in Table 4.7 and Figure 4.10 to 4.11. The AM1//HF shows higher R^2 than any others but AM1//AM1 displays the lowest R^2 value in Table 4.6. This suggests that not only atomic charge, but also optimized geometry is very important in a CoMFA model. In the case of AM1 optimized geometry, HF/3-21G* or MP2/3-21G* charge calculations increase the predictive ability (R^2 = 0.966 for HF/3-21G* and MP2/3-21G*) from that of AM1 charge calculation (R^2 = 0.911). Ab initio calculations for atomic charge produce the same predictive ability for AM1 and HF/3-21G* optimized geometries. The R^2 values for HF//AM and MP2//AM equal 0.966 and those of HF//HF and MP2//HF are 0.977 in Table 4.6.

Table 4.6 QSAR Reports by Non-Crossvalidation Using SAMPLS [111] by the Training Set of 43 Molecules

Theory	S. E.[i]	R^2 [j]	F values [k]	Steric. [l]	Electro. [m]
AM1//AM1[a]	0.346	0.911	(n1=3,n2=39) 133.309	0.514	0.486
HF//AM1[b]	0.219	0.966	(n1=5,n2=37) 212.332	0.488	0.512
B3//AM1[c]	0.284	0.942	(n1=4,n2=38) 153.675	0.494	0.506
MP2//AM1[d]	0.219	0.966	(n1=5,n2=37) 212.036	0.489	0.511
HF//HF[e]	0.184	0.977	(n1=6,n2=36) 253.860	0.44	0.56
B3//HF[f]	0.184	0.977	(n1=6,n2=36) 254.315	0.437	0.563
MP2//HF[g]	0.182	0.977	(n1=6,n2=36) 259.344	0.44	0.56
AM1//HF[h]	0.13	0.989	(n1=7,n2=35) 437.947	0.481	0.519

[a-h] See in Table 4.5. [i] Standard error of estimation. [j] non-crossvalidated regression of co-efficient using training set of 43 molecules in Table 4.1-4.3. [k] F-statistic analysis [l] Steric contribution to this CoMFA field. [m] Electrostatic contribution to this CoMFA field.

4.3.2 Contour Map

The contour maps, obtained from the training set compounds are shown in Figure 4.12 and 4.13 by the type of calculational method. The most active compound (number 7) is shown in each contour map. The results are viewed as regions surrounding the sigma ligands where steric bulk or electrostatic potential most strongly affects the sigma1 activity.

Table 4.7 Experimental and Predicted Bioactivities (pK_i) by the Training Set of 43 Molecules using Various Calculation Methods

Number	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//HF[f]	MP2//[g]	AM1//[h]
	pK_i	AM1	AM1	AM1	Predicted pK_i		HF	HF	HF
2	0.3	0.31	0.33	0.24	0.33	0	0.03	0.02	0.22
3	-1.4	-1.48	-1.36	-1.48	-1.36	-0.16	-1.28	-0.13	-1.35
4	-1.3	-1.47	-1.22	-1.36	-1.22	-0.63	-1.41	-0.69	-1.33
5	0.21	0.27	0.39	0.08	0.39	-1.24	0.27	-1.26	0.34
6	-0.6	-0.8	-1.04	-0.93	-1.05	-0.57	-0.56	-0.56	-0.55
7	1.3	0.89	1.15	1.1	1.15	1.26	1.22	1.27	1.26
8	-0.16	-0.02	-0.22	-0.01	-0.23	-0.17	-0.19	-0.15	-0.1
10	-0.13	0.09	-0.09	0.1	-0.09	-0.18	-0.25	-0.14	-0.14
11	-0.01	-0.14	0.03	0.09	0.03	-0.06	-0.09	-0.05	0.08
12	-1.8	-1.15	-1.52	-1.32	-1.52	-1.72	-1.7	-1.74	-1.84
13	-0.44	-0.37	-0.23	-0.14	-0.23	-0.48	-0.46	-0.48	-0.39
14	-1.36	-0.57	-1.16	-0.84	-1.16	-1.48	-1.5	-1.49	-1.44
150	-1.89	-2.31	-2.16	-2.3	-2.16	-1.68	-1.7	-1.69	-1.73
16	0.3	0.14	0.45	0.39	0.45	0.37	0.37	0.39	0.26
18	-0.18	-0.16	-0.22	-0.07	-0.22	-0.08	-0.06	-0.08	-0.2
19	0.06	0.2	-0.03	0.17	-0.03	0.03	0.05	0.01	0.09
20	0.19	0.1	0.08	0.09	0.08	0.12	0.15	0.16	0.22
21	0.29	-0.4	0.11	-0.08	0.11	0.16	0.19	0.14	0.12
22	0.21	0.1	0.4	0.28	0.39	0.17	0.2	0.11	0.2
23	-0.08	0.08	-0.11	0.08	-0.11	-0.09	-0.07	-0.12	-0.16
24	0.22	-0.33	-0.17	-0.31	-0.17	-0.01	0.02	-0.02	0.07
25	-0.36	-0.34	-0.36	-0.35	-0.36	-0.19	-0.18	-0.23	-0.38
27	-1.36	-1.95	-1.69	-1.83	-1.69	-1.35	-1.35	-1.33	-1.38
28	-1.6	-1.14	-1.72	-1.46	-1.71	-1.78	-1.78	-1.73	-1.87
29	-1.67	-1.16	-1.43	-1.41	-1.43	-1.56	-1.55	-1.56	-1.5

[a-h] See in Table 4.5.

Table 4.7 Experimental and Predicted Bioactivities (pK_i) by the Training set of 43 Molecules Using Various Calculation Methods (Continued)

Number	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//HF[f]	MP2//[g]	AM1//[h]
	pK_i	AM1	AM1	AM1	Predicted pK_i		HF	HF	HF
30	-1.69	-1.89	-1.86	-1.84	-1.86	-1.79	-1.76	-1.78	-1.67
31	-1.73	-1.7	-1.89	-1.85	-1.89	-2.06	-2.06	-2.05	-1.95
32	-1.86	-1.66	-2.31	-2.1	-2.31	-1.68	-1.7	-1.69	-1.88
33	-1.86	-1.81	-1.49	-1.48	-1.49	-1.97	-1.96	-1.94	-1.89
34	-1.92	-1.83	-1.91	-1.86	-1.91	-1.78	-1.78	-1.76	-1.86
35	-2.02	-2.42	-2.28	-2.44	-2.28	-2.48	-2.48	-2.48	-2.33
36	-2.25	-2.31	-2.13	-2.11	-2.13	-2.34	-2.33	-2.33	-2.37
37	-2.27	-2.69	-2.1	-2.28	-2.1	-2.22	-2.22	-2.17	-2.16
38	-2.37	-2.03	-2.26	-2.13	-2.26	-2.07	-2.08	-2.1	-2.19
39	-2.51	-2.54	-2.53	-2.41	-2.53	-2.59	-2.58	-2.59	-2.59
41	-2.76	-2.11	-2.41	-2.24	-2.41	-2.78	-2.79	-2.6	-2.74
42	-2.79	-2.89	-2.64	-2.79	-2.64	-2.78	-2.79	-2.97	-2.74
43	-2.8	-2.35	-2.61	-2.57	-2.61	-2.28	-2.28	-2.34	-2.45
44	1.1	1.26	1.16	1.32	1.16	1.19	1.14	1.15	1.15
45	-0.08	-0.21	0.11	-0.33	0.11	-0.16	-0.17	-0.13	-0.06
46	-0.53	-0.48	-0.33	-0.51	-0.33	-0.51	-0.52	-0.48	-0.52
47	-0.76	-0.68	-0.82	-0.91	-0.82	-0.63	-0.63	-0.69	-0.7
48	0.08	-0.33	-0.16	-0.44	-0.16	0.3	0.35	0.37	0.19

[a-h] See in Table 4.5.

4.4 Discussion

4.4.1 Validation of the CoMFA Model

The test compounds selected were piperidine and piperazine analogs, benzoxazolone and benzothiazolone, spipethiane analogs from all three families in Table 4.8. The range of binding affinities for the test set was –2.51 to 0.57 log units. The predictive utility of the CoMFA model for five ligands in the test set were considered satisfactory and HF/3-21G* optimized geometry produced higher accuracy than AM1 optimized geometry. The

derived CoMFA model is shown in Table 4.8. The best CoMFA model is satisfactory in both statistical significance and predictive ability. AM1//HF (HF/3-21G* optimized geometry and AM1 charge) shows higher predictive ability ($R^2 = 0.989$) but HF//HF (HF/3-21G* optimized geometry and charge) processed the best CoMFA model. Although the predictive ability ($R^2 = 0.977$) is lower than that of AM1//HF, the test set is more accurate.

4.4.2 Design of New Ligands

CoMFA model displays the spatial distribution of important steric and electrostatic properties affecting the activity. Applying the quantitative model of the sigma 1 receptor, one can predict the activity of different sigma 1 ligands. A careful investigation of the CoMFA model with the most bioactive compound (number 7) revealed that attachment of electron withdrawing groups to cyclopentane could improve its bioactivity (pK_i) from the original predicted value of 1.30 to 1.79 in HF/3-21G* optimized geometry and atomic charges. These new molecules are shown in Table 4.9.

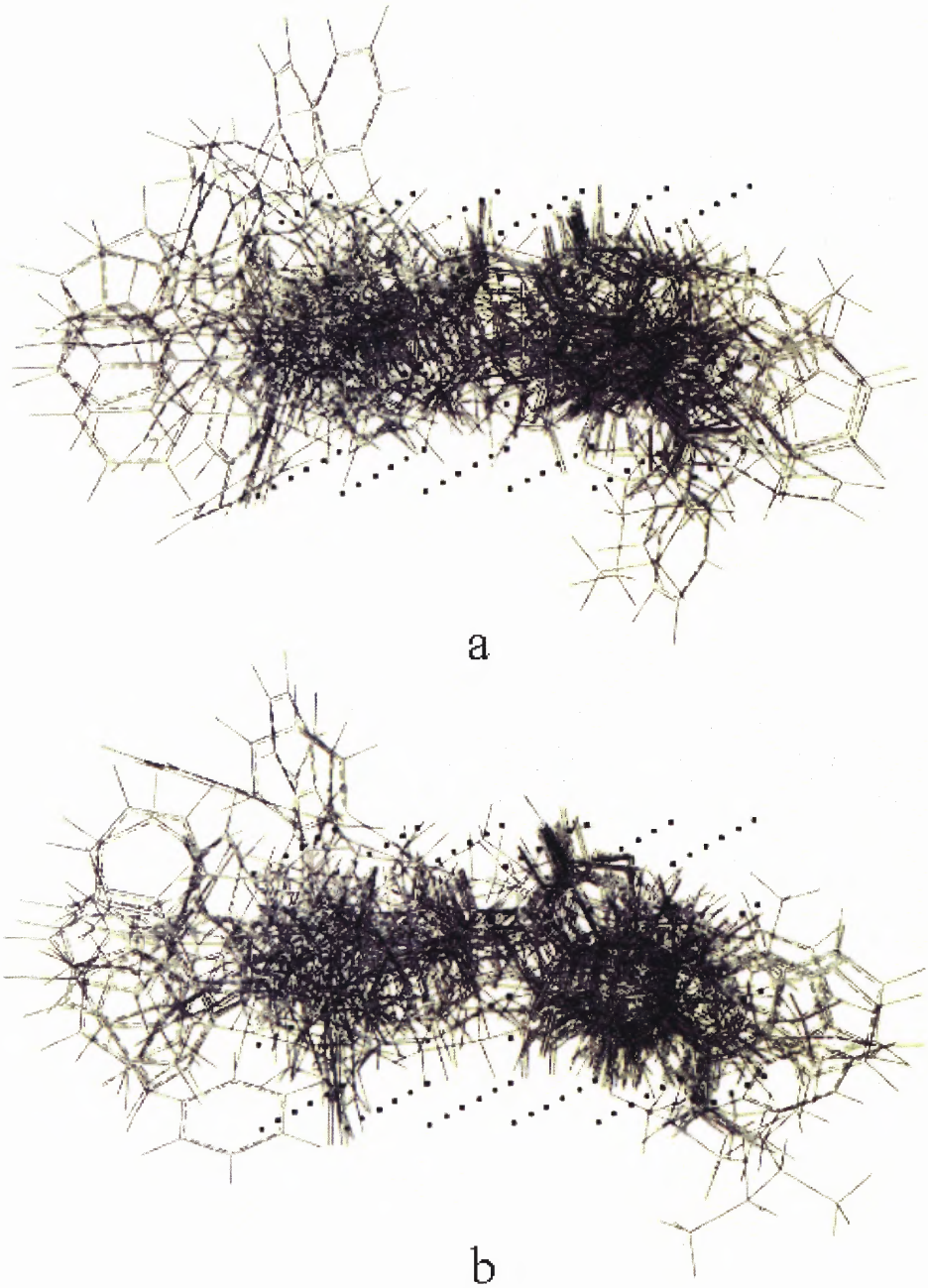


Figure 4.9 Alignments and grid spacing (2Å) of sp3 carbon (+1 charge) for all molecules by calculational methods.

[a] AM1 optimized molecules, [b] HF/3-21G* optimized molecules.

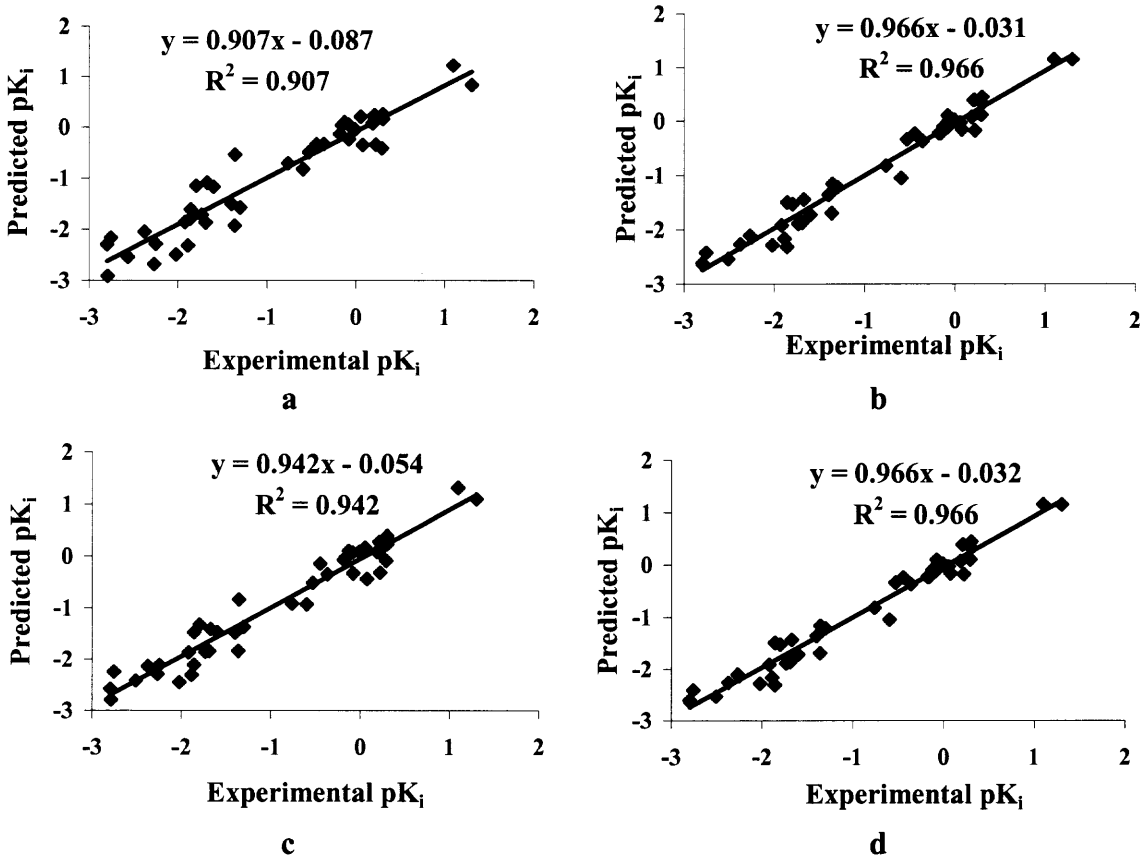


Figure 4.10 Graph of experimental pKi(-logKi) versus CoMFA predicted bioactivity

All geometries were optimized using AM1. Charges were calculated using [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, and [d] MP2/3-21G* methods.

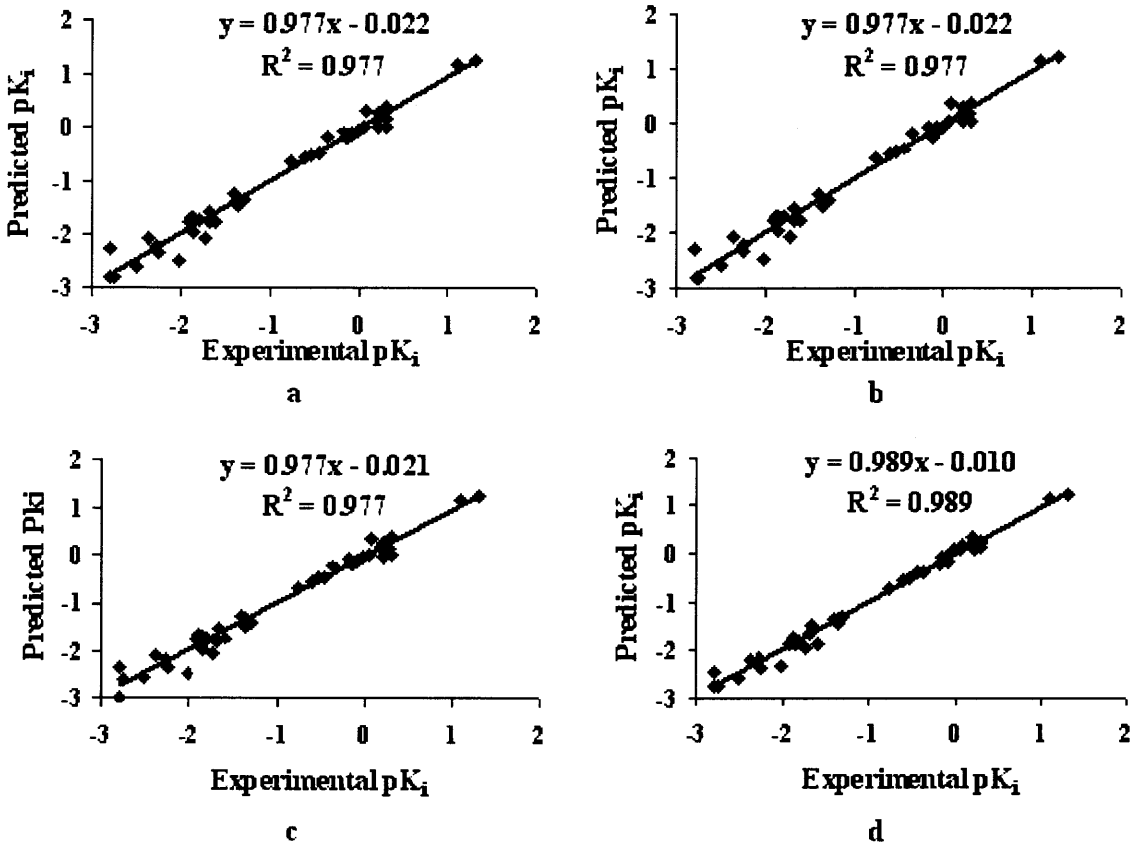


Figure 4.11. Graph of experimental pKi(-logKi) versus CoMFA predicted bioactivity

All geometries were optimized using HF/3-21G*. Charges were calculated using [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, and [d] AM1 methods.

4.5 Conclusions in CoMFA Studies of Sigma 1 Receptor-Ligands

This study proves the predictive abilities of the CoMFA model. The results suggest that ab initio HF/3-21G* optimized geometries show higher q^2 and R^2 than semi-empirical AM1 optimizations. Furthermore, it is in agreement with Tonmunphean's report. [100] This study also demonstrates that a single CoMFA model can be built from all piperidine and piperazine analogs, benzoxazolone and benzothiazolone analogs, and spipethiane analogues examined. Together, the present studies enhance the information available about ligand interactions with the sigma subtype 1 receptor. This CoMFA model may prove useful for designing new and more potent sigma1 ligands.

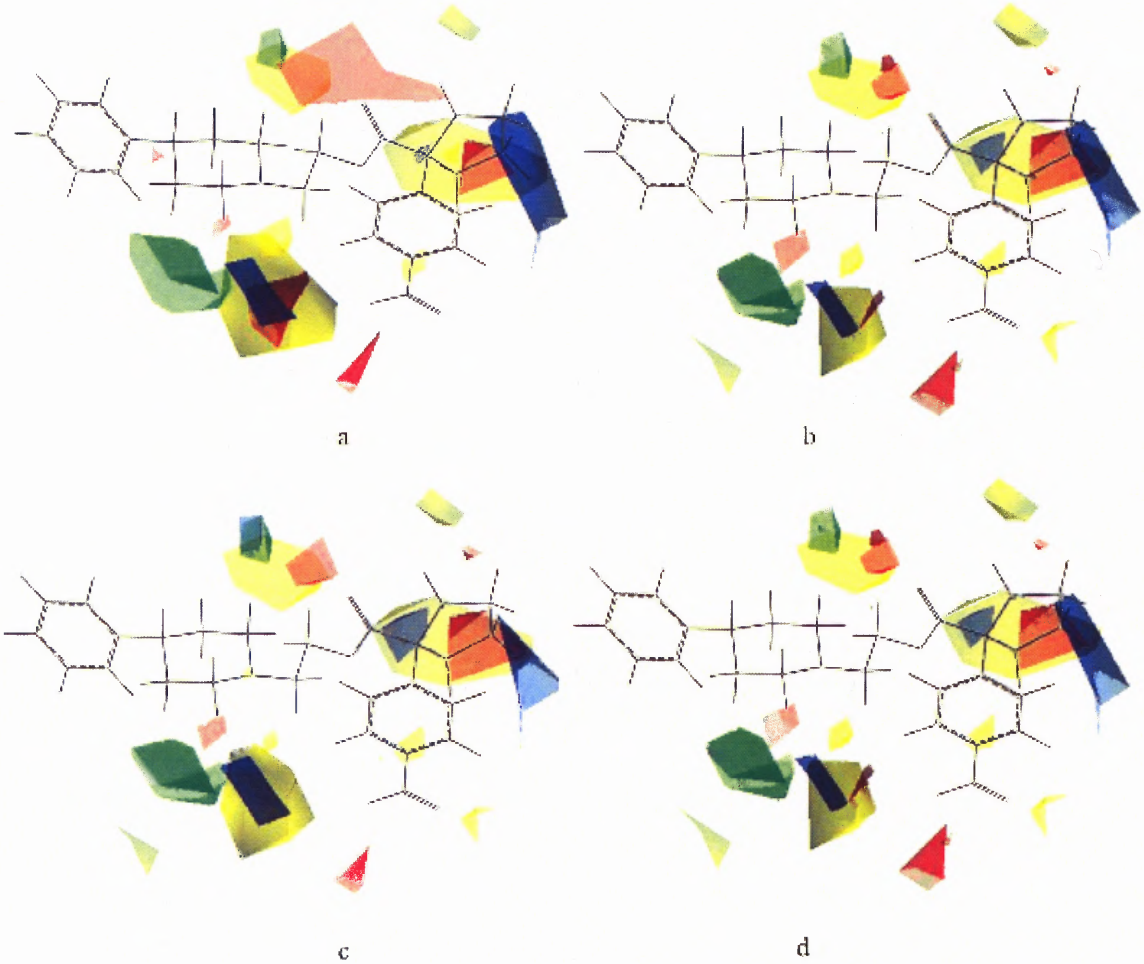


Figure 4.12 CoMFA contour map for compound (number 7) derived by various charge and AM1 optimized geometry methods [a-d] using steric [e] and electrostatic [f] fields.

All geometries were optimized using AM1. Charges were calculated using [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, and [d] MP2/3-21G* methods. [e] The contour maps of the steric field are shown in yellow and green; green areas are regions where more bulky substituents are desirable, yellow areas are regions where less bulk is favorable. [f] The contour maps of the electrostatic field are shown in red and blue. the red areas are the regions where more negative charge is favorable for the higher σ1 activity and blue areas are regions where more positive charge is favorable.

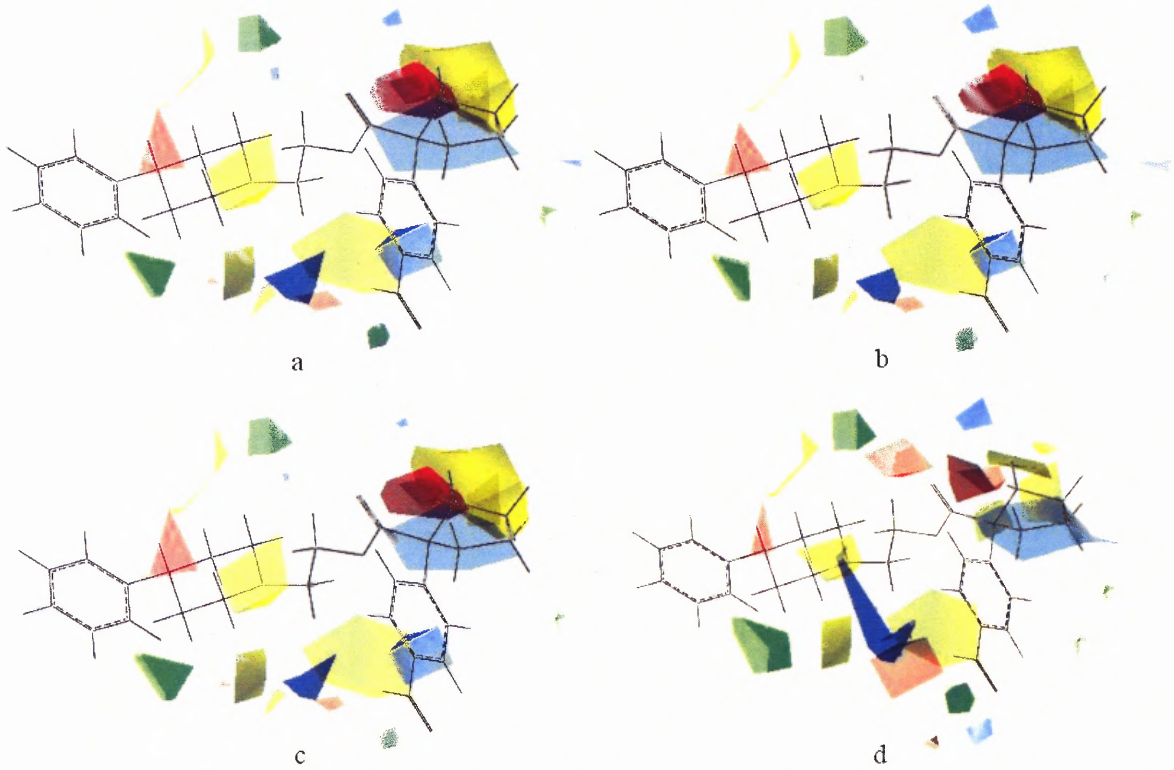


Figure 4.13. CoMFA contour map for compound (number 7) derived by various charge and HF/3-21G* optimized methods [a-d] using steric [e] and electrostatic [f] fields.

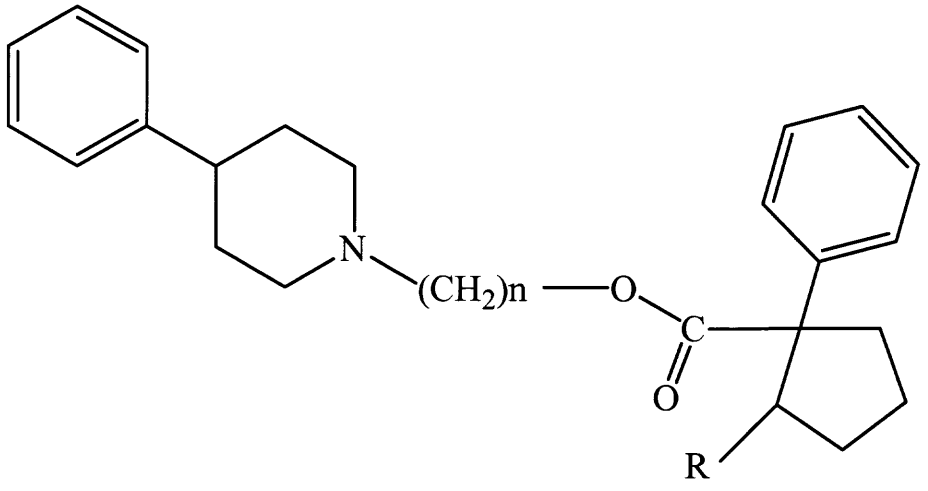
All geometries were optimized using HF/3-21G*. Charges were calculated using [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, and [d] AM1 methods. [e-f] See Figure 4.12.

Table 4.8 Experimental and Predicted Bioactivities (pK_i) by Test Set of 5 Molecules using Various Calculation Methods

Number	Experiment	AM1//AM1[a]	HF//AM1[b]	B3//AM1[c]	MP2//AM1[d]	HF//HF[e]	B3//HF[f]	MP2//HF[g]	AM1//HF[h]
	pK_i	Predicted pKi							
1	-0.15	0.21	0.5	0.27	0.5	-0.26	-0.36	-0.24	-0.01
9	-0.11	0.15	0.34	0.46	0.35	-0.08	-0.12	-0.05	-0.21
17	0.57	-0.24	0.24	0.11	0.26	0.21	0.16	0.22	0.2
26	-0.93	-0.76	-1.33	-1.26	-1.33	-0.9	-0.89	-0.97	-0.92
40	-2.56	-2.44	-2.14	-2.22	-2.14	-2.28	-2.18	-2.29	-2.19

[a-h] See in Table 4.5.

Table 4.9 Prediction of Bioactivity for New Ligands



R	Experimental	HF//HF [a]	B3//HF [b]	MP2//HF [c]	AM1//HF [d]
	pK_i	Predicted pKi			
H	1.3	1.26	1.22	1.27	1.26
NO_2	Unknown	1.79	1.74	1.80	1.48
$COCH_3$	Unknown	1.60	1.59	1.61	1.52
COH	Unknown	1.52	1.51	1.53	1.43
$COOCH_3$	Unknown	1.41	1.40	1.42	1.35
COOH	Unknown	1.47	1.44	1.48	1.39
SO_3H	Unknown	1.63	1.58	1.65	1.63
CH_3	Unknown	1.25	1.21	1.26	1.26
NH_2	Unknown	1.41	1.36	1.41	1.30
F	Unknown	1.49	1.42	1.50	1.34
Cl	Unknown	1.36	1.31	1.37	1.33
I	Unknown	1.31	1.25	1.31	1.26

All geometries were optimized using HF/3-21G*. Charges were calculated using [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, and [d] AM1 methods.

CHAPTER 5

PHARMACOPHORE DERIVATION USING DISCOTECH ON PCP DERIVATIVES FOR SIGMA 1 RECEPTOR-LIGANDS AND COMFA STUDIES USING SEMI-EMPIRICAL, DENSITY FUNCTIONAL, AB INITIO METHODS

5.1 PCP Derivatives for Sigma 1 Receptor-Ligands

Gund [113] reported bioactivity of PCP derivatives, PRE 084, 078, 079, and 082 for the sigma 1 receptor. The sigma 1 receptor represents a unique intracellular neuronal protein modulating several neurotransmitter responses with relevant effects on cognitive functions. There have been several research articles about PRE 084, the sigma 1 receptor-ligand; (1) antidepressant effects [114], (2) improving spatial memory capacities of aged rats [115-118] (3) Ca^{2+} signaling via $sigma_1$-receptors [119], and (4) a new strategy against cocaine addiction and toxicity [120-121]. The sigma 1 receptor was recently cloned in several animal species and in the human [29,31,32,34]. The protein obtained shows a 223-amino acid sequence and shares no homology with any known protein, in particular with classical ionotropic or metabotropic neurotransmitter receptors. However, the protein mediates a very efficient neuromodulatory action, affecting several neurotransmitters systems, including the acetylcholine and N-methyl aspartate (NMDA)-type of glutamatergic receptor [6].

In this study, the pharmacophore of PCP derivatives were developed by DISCOtech using Sybyl 6.9 and it is the first pharmacophore excluding the lone pair of nitrogen for bioactive sigma 1 receptor-ligands; when the nitrogen of PRE 079 was methylated, a ligand had no lone pair of electron from nitrogen but still showed sigma 1 bioactivity. The suggested conformers by DISCOtech were optimized using AM1 or

HF/3-21G* calculation by Gaussian 98 and atomic charges were calculated using AM1, HF/3-21G*, B3LYP/3-21G*, MP2/3-21G* methods. A CoMFA map of steric and electrostatic fields was analyzed comparatively. Using the CoMFA predicting property module, the effect of insertion of various groups such as methylene, ethylene, propane, and butane between the cyclohexyl and amine group of phencyclidine on the bioactivity of sigma 1 receptor-ligands was investigated.

5.2 Materials and Methods

5.2.1 Biological Data

Sigma 1 receptor affinity values for 30 PCP derivatives, found in the literature [113] were listed in Table 5.1. All values had been obtained using the radioligand, [^{3}H](+)SKF10047 for the sigma 1 receptor. Potencies at sigma 1 receptor was analyzed by various insertion of methylene, ethylene, or carboxyl ethylene between the cyclohexyl group and amine of phencyclidine, by several derivatives with various sizes of cycloalkyl group, and by the effects of phenyl ring substitutions. The IC50 values were converted to pIC50 values [pIC50 = -log(IC50)]. A training set containing 27 compounds and a test set of 3 compounds were used to assess the predictive power of the model. Histogram pictures of train and test sets are shown in Figure 5.1. The range of binding affinities for the training set was –3.75 to –0.71 log units, and –3.03 to –1.41 log units for the test set in Figure 5.1.

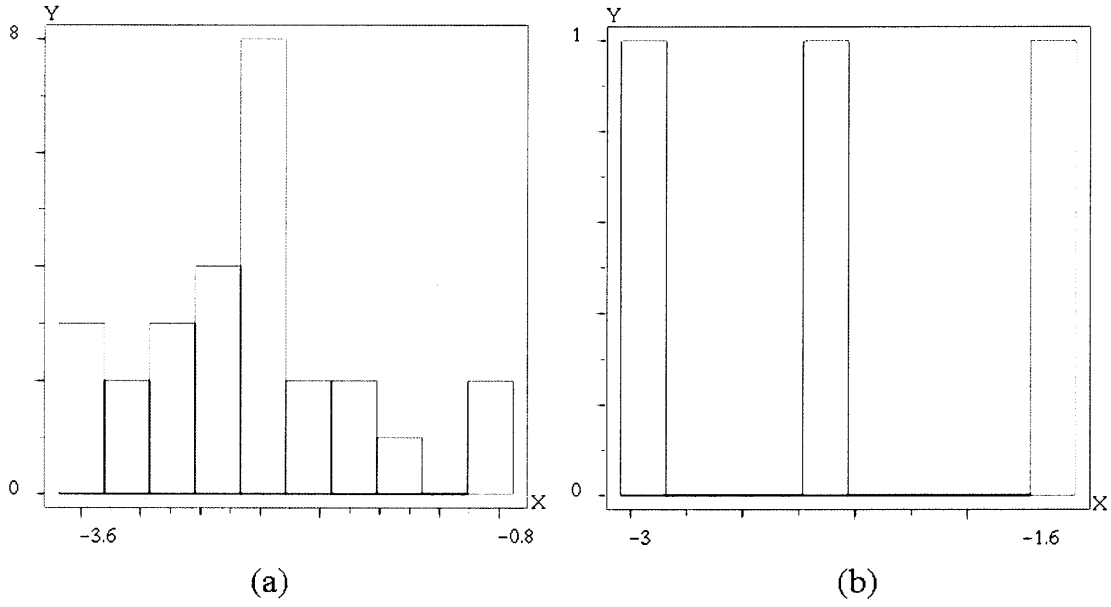


Figure 5.1 Histogram of pIC50 (abscissa) vs number of molecules (ordinate).

[a] Training set, [b] Test set

Table 5.1 PCP Derivatives and Their Bioactivities for Sigma 1 Receptor-Ligands Using Radioligand, [^{3}H]+SKF-10047 [113]

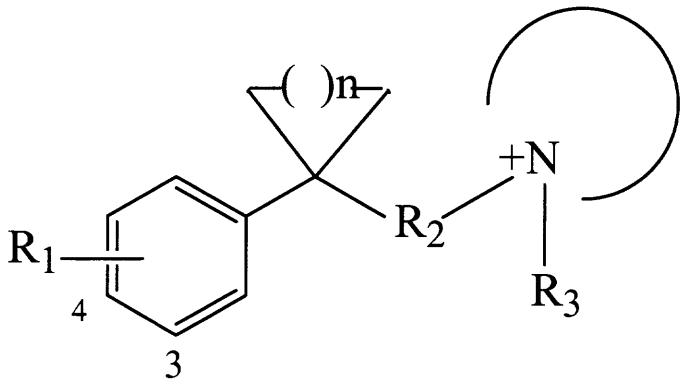


Table 5.1 PCP Derivatives and Their Bioactivities for Sigma 1 Receptor-Ligands Using Radioligand, [^{3}H]+SKF-10047 [113] (Continued)

Nunber	Amine	R1	R2	R3	n	sigma 1 receptor IC50 (nm)
1	Piperidine	H	CH_2	H	3	104 ± 24
2 [a]	Piperidine	4-CH_3	CH_2	H	3	223 ± 76
3	Piperidine	4-NO_2	CH_2	H	3	249 ± 108
4	Piperidine	3-Cl	CH_2	H	3	71 ± 13
5	Piperidine	3-OCH_3	CH_2	H	3	326 ± 117
6	Piperidine	4-OCH_3	CH_2	H	3	221 ± 50
7	Piperidine	3,4-OCH_3	CH_2	H	3	387 ± 18
8	Piperidine	H	CH_2	H	2	995 ± 214
9	Piperidine	H	CH_2	H	0	5589 ± 214
10 [a, b]	Piperidine	H	C(O)O-$(CH_2)_2$	H	3	26 ± 7
11 [a]	Pyrrolidine	3-OH	CH_2	H	3	1082 ± 263
12	Pyrrolidine	4-OH	CH_2	H	3	1735 ± 293
13	Pyrrolidine	4-NO_2	CH_2	H	3	282 ± 65
14	Pyrrolidine	3-Cl	CH_2	H	3	248 ± 39
15	Pyrrolidine	4-OCH_3	CH_2	H	3	546 ± 54
16	Pyrrolidine	3,4-OCH_3	CH_2	H	3	1317 ± 167
17	Pyrrolidine	H	C2H5	H	3	141 ± 42
18 [c]	Pyrrolidine	H	C(O)O-$(CH_2)_2$	H	3	5.1 ± 1.7
19	Pyrrolidine	H	C(O)O-$(CH_2)_2$	CH3	3	242 ± 123
20	Morpholine	H	CH_2	H	3	710 ± 68
21	Morpholine	H	C(O)O-$(CH_2)_2$	H	3	44 ± 7
22	Morpholine	H	C(O)O-$(CH_2)_2$	H	2	454 ± 80
23 [d]	Morpholine	H	C(O)O-$(CH_2)_2$	H	0	1463 ± 102
24	$N(C_2H_5)_2$	H	CH_2	H	3	208 ± 85
25	$N(C_2H_5)_2$	4-NO_2	CH_2	H	3	304 ± 140
26 [e]	$N(C_2H_5)_2$	H	C(O)O-$(CH_2)_2$	H	3	30 ± 5
27	$N(CH_3)_2$	H	CH_2	H	3	4130 ± 235
28	$N(CH_3)_2$	4-CH_3	CH_2	H	3	541 ± 70
29	$N(CH_3)_2$	3-OCH_3	CH_2	H	3	2851 ± 1034
30 [f]	$N(CH_3)_2$	H	C(O)O-$(CH_2)_2$	H	3	9.2 ± 0.8

[a] Compounds of a test set. [b] PRE-082, [c] PRE-079, [d] PRE-084, [e] PRE-083, and [f] PRE-078.

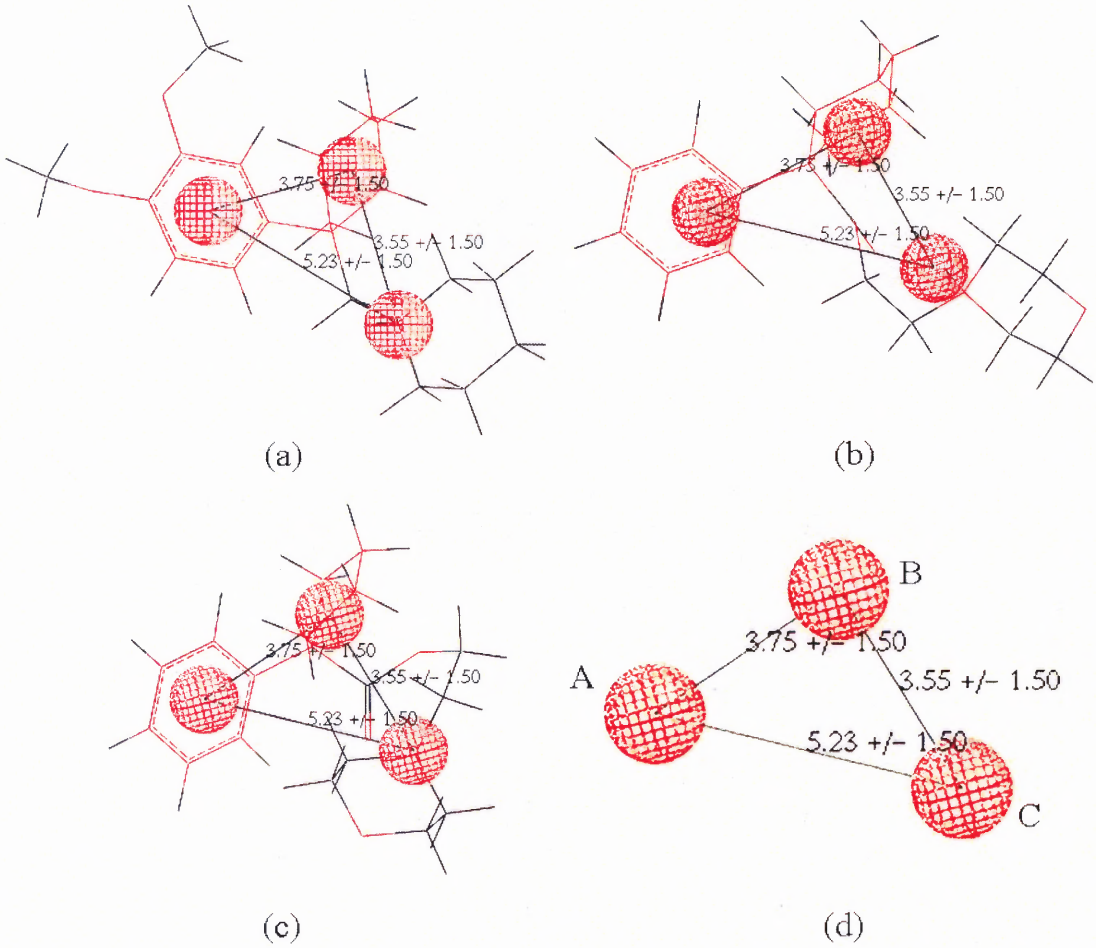


Figure 5.2 DISCOtech pharmacophore of PCP derivatives for sigma 1 receptor-ligands.

[a] compound number 7, [b] compound number 21, and [c] compound number 22 in Table 5.1. [d] DISCOtech model; A is the center of a phenyl ring. B is the center of a hydrophobic ring C is a nitrogen atom.

5.2.2 Computational Methods

The initial conformer searching and pharmacophore study were analyzed using DISCOtech [74-78] on SYBYL6.9 [71] and the optimization of geometry and calculation of atomic charge were performed using Gaussian 98 [79]. All CoMFA models were derived using SYBYL6.9.

5.2.3 Choice of Initial Conformations

The CoMFA study began with the selection of the three-dimensional conformation for each compound. Initial structures were generated by building with SYBYL6.9 [71] default bond distances and angles, and minimized with the tools MAXIMIN2 in SYBYL6.9 in which the Tripos force field was applied with a distance-dependent dielectric function. Then, DISCOtech [74-78] in SYBYL6.9 was used to search possible conformations and proper pharmacophores using these initial molecules. The energy limit was 35 kcal/mol. The reference structure in computing DISCOtech models was compound 18, because it is the most bioactive PCP derivative for sigma 1 receptor ligands, displayed Table 5.1.

5.2.4 Pharmacophore Information

DISCOtech [74-78] derived a pharmacophore model based on PCP derivatives for sigma 1 receptor ligands. DISCOtech found possible conformations within reasonable energy boundaries (in this study, 35 kcal/mol) and suggested a proper pharmacophore model. The overall pharmacophore is a triangle that includes two centers of a hydrophobic ring, and a nitrogen atom in Table 5.2. This is a new trial to derive a pharmacophore model for sigma 1 receptor-ligands without the lone pair of a nitrogen atom because compound 19 does not contain a lone pair of nitrogen but is still active at the sigma 1 receptor. There was a previous report about the fitting of PRE-084 and other sigma 1 receptor ligands using the lone pair as a pharmacophoric point [109].

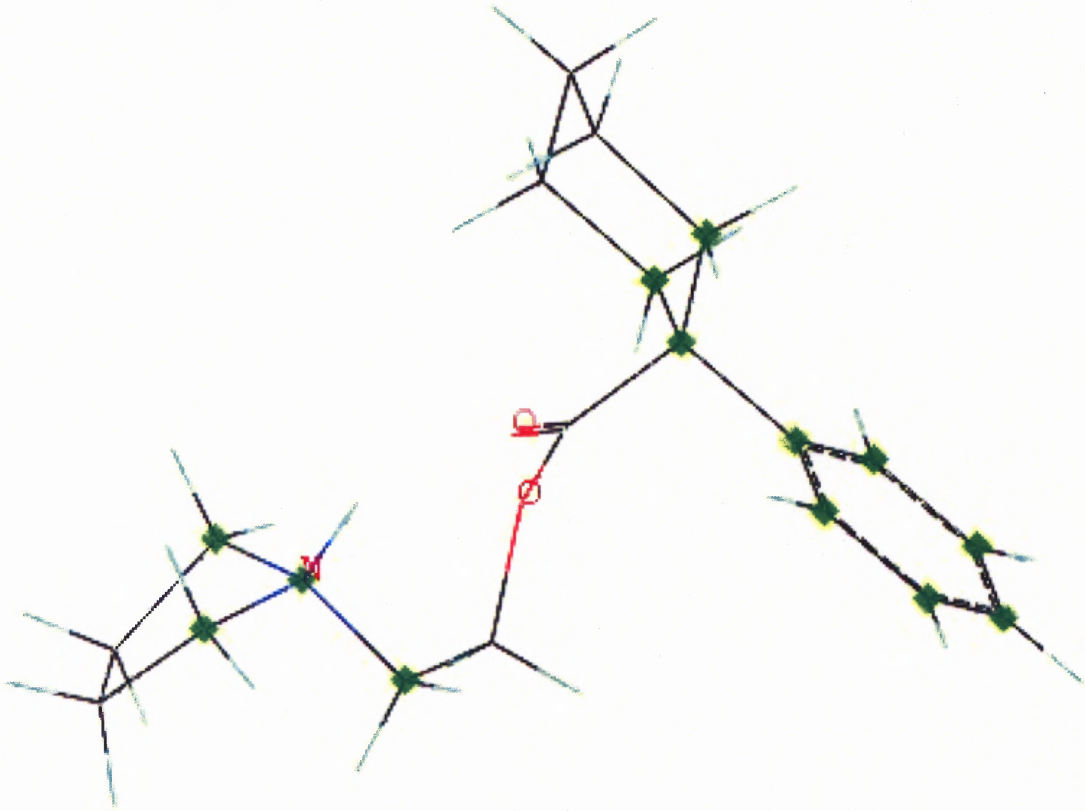
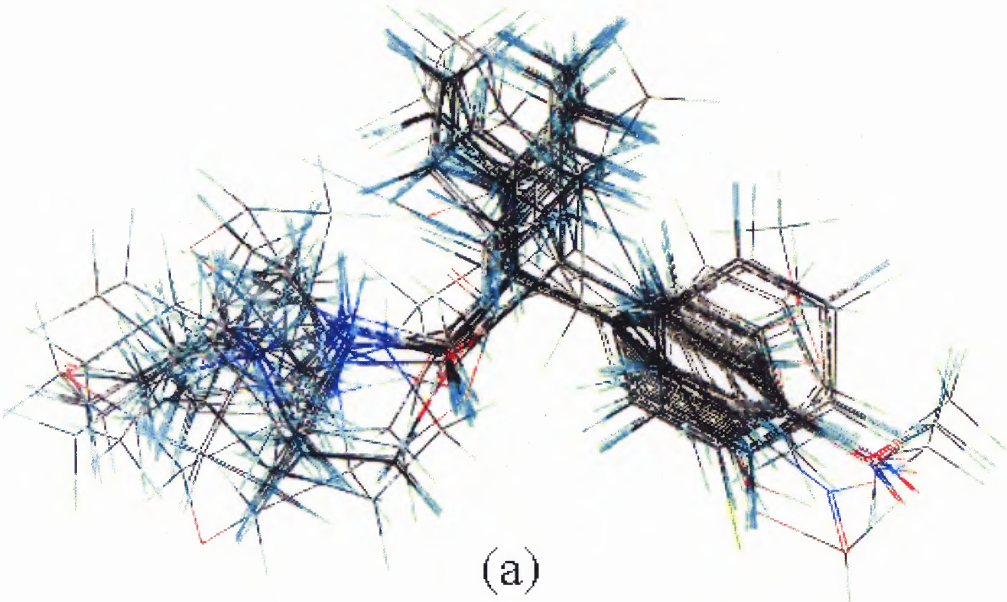
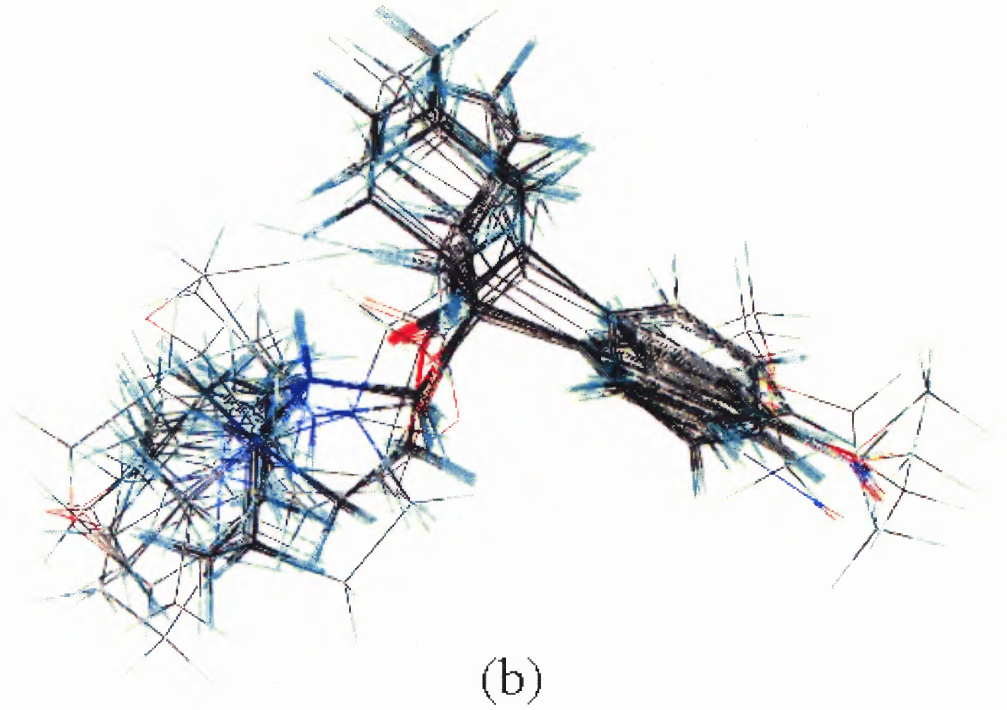


Figure 5.3 A Template molecule (compound 18 in Table 5.1) and the green bolded atoms were used for the alignment of all 30 Molecules.



(a)



(b)

Figure 5.4 Alignments of all molecules, optimized using (a) AM1 and (b) HF/3-21G* methods.

5.2.5 Geometry Optimzation and Atomic Charges

The conformers, derived by DISCOtech [74-78] were optimized with ab initio HF/3-21G* [99,103] method or with semi-empirical AM1 [99-102] calculation with Gaussian 98 [79]. Theses geometries were calculated for atomic charges in semi-empirical AM1, density functional B3LYP/3-21G*, ab initio HF/3-21G* and MP2/3-21G* levels according to Mulliken population using Gaussian 98.

Table 5.2 The Number of Optimal Components [i] and q^2 by "Leave-One-Out" Using SAMPLS [111] by the Training Set of 27 Molecules

Charge// Geometry	Term	C. 1 [j]	C. 2	C. 3	C. 4	C. 5	C. 6	C. 7	C. 8
AM1//AM1 [a]	S.E.E. [k]	**0.589**	0.667	0.669	0.702	0.775	0.803	0.836	0.851
	q2 [l]	**0.373**	0.226	0.256	0.215	0.086	0.068	0.039	0.056
HF//AM1 [b]	S.E.E.	**0.590**	0.672	0.667	0.728	0.774	0.833	0.856	0.887
	q2	**0.370**	0.215	0.260	0.157	0.089	-0.005	-0.008	-0.025
B3//AM1 [c]	S.E.E.	**0.591**	0.666	0.669	0.701	0.765	0.791	0.825	0.837
	q2	**0.368**	0.230	0.250	0.218	0.111	0.094	0.064	0.088
MP2//AM1 [d]	S.E.E.	**0.593**	0.664	0.674	0.699	0.741	0.786	0.801	0.824
	q2	**0.364**	0.233	0.244	0.223	0.166	0.106	0.117	0.115
HF//HF [e]	S.E.E.	0.538	0.471	**0.469**	0.547	0.555	0.613	0.647	0.690
	q2	0.477	0.614	**0.634**	0.524	0.533	0.456	0.424	0.381
B3//HF [f]	S.E.E.	0.538	0.468	**0.465**	0.516	0.548	0.579	0.638	0.674
	q2	0.477	0.620	**0.641**	0.577	0.543	0.515	0.440	0.409
MP2//HF [g]	S.E.E.	0.538	0.471	**0.468**	0.548	0.555	0.613	0.649	0.689
	q2	0.477	0.614	**0.635**	0.523	0.531	0.457	0.421	0.382
AM1//HF [h]	S.E.E.	0.539	**0.461**	0.488	0.517	0.600	0.625	0.681	0.726
	q2	0.474	**0.632**	0.603	0.575	0.454	0.435	0.363	0.313

[a-d] Charges were calculated in [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, [d] MP2/3-21G* and all geometries were optimized in AM1 method. [e-h] Charges were calculated in [e] HF/3-21G*, [f] B3LYP/3-21G*, [g] MP2/3-21G*, [h] AM1, and all geometries were optimized in HF/3-21G* methods. [i] The number of optimal components is the bold one in Table in each level. [j] Component number, [k] standard error of estimate, and [l] leave-one-out cross-validated regression of co-efficient.

5.2.6 Alignment

Alignment of the presumed bound conformations of the training set compounds is also an essential prelude to the CoMFA study. The AM1 or HF/3-21G* optimized conformers were aligned by "match" function in SYBYL6.9 using a template compound (number 18 in Table 5.1) using the green bolded atoms, shown Figure 5.3. The aligned 30 molecules, used in training and test sets are shown by optimization methods in Figure 5.4 for (a) AM1 and (b) for HF/3-21G*.

5.2.7 CoMFA Model

Auto CoMFA columns were calculated using the Tripos Standard CoMFA field class. It extended 4 Å beyond every molecule in all directions, and had a 2 Å spacing, and a probe atom of C.3 (sp^3 carbon) and a charge of +1 with a dielectric function of 1/r, a dielectric constant ε of 1 and the default of 30 kcal/mol energy cutoff for steric and electrostatic fields. Partial least squares analysis, regresses a target property against predictors calculated as steric and electrostatic components of the intermolecular interaction field. Scaling was used as the CoMFA standard. The SAMPLS (SAMple-distance PLS) algorithm developed by Bruce Bush [111] was used to determine "leave-one-out" cross-validation q^2. The method for cross-validation serves two purposes; (1) to find out whether the CoMFA model was productively useful, and (2) if useful, to decide how many components to use for the best model. The number of optical components was considered by the 5% rule; if the q^2 increases by at least 5% upon increasing the number of components by one, then it is justified to add an additional component. The Partial Least Squares (PLS) analysis was then repeated without cross-validation using the

optimum number of components. This final analysis yielded a predictive model, and a CoMFA coefficient contour plot for the steric and electrostatic potentials contributions.

5.3 Results and Discussion

5.3.1 Comparative Molecular Field Analysis

There were two types of CoMFA models by AM1 geometry optimization methods. The CoMFA model in this study required 2 or 3 optimal components using HF/3-21G* optimized geometries, but required 1 optimal component using AM1 optimized geometries by

Table 5.3 QSAR Reports by Non-Crossvalidation using SAMPLS [111] by the Training Set of 27 Molecules

Theory	S. E.[i]	R^2 [j]	F values [k]	Steric. [l]	Electro. [m]
AM1//AM1 [a]	0.477	0.589	(n1=1,n2=25) 35.764	0.812	0.188
HF//AM1 [b]	0.468	0.604	(n1=1,n2=25) 38.120	0.669	0.331
B3//AM1 [c]	0.478	0.587	(n1=1,n2=25) 35.578	0.801	0.199
MP2//AM1 [d]	0.478	0.586	(n1=1,n2=25) 35.453	0.802	0.198
HF//HF [e]	0.239	0.905	(n1=3,n2=23) 73.007	0.828	0.172
B3//HF [f]	0.237	0.906	(n1=3,n2=23) 74.247	0.855	0.145
MP2//HF [g]	0.238	0.906	(n1=3,n2=23) 73.875	0.828	0.172
AM1//HF [h]	0.308	0.835	(n1=2,n2=24) 60.886	0.811	0.189

[a-h] See in Table 5.2. [i] Standard error of estimation. [j] Leave-one-out non-cross-validated regression of co-efficient using training set of 21 molecules in Table 5.1. [k] F-statistic analysis [l] Steric contribution to this CoMFA field. [m] Electrostatic contribution to this CoMFA field.

different calculations to explain the variance in binding affinity to sigma 1 receptor-ligands. in Table 5.2. All crossvalidated q^2 were more than 0.6 by the CoMFA model using HF/3-21G* optimized geometries, but they were less than 0.4 using AM1

optimized geometries. Anyway it surpassed the generally accepted criterion, 0.3 [112] for statistical validity. The highest q^2 (0.641) was for HF/3-21G* optimized geometries and B3LYP/3-21G* atomic charge calculations. The CoMFA models of AM1 optimized geometries produced lower q^2 (0.364-0.373) than those of HF/3-21G* optimized geometries (0.632-0.641). It suggests that CoMFA models obtained from HF/3-21G* optimized geometries give more accurate predicition of activity among additional similar sigma 1 ligands than CoMFA models obtained from AM1 optimized geometries.

Having the crossvalidation to confirm the predictive ability, a PLS analysis was performed without any validation to derive the best predictive model for use in graphics and in numerical prediction. CoMFA models using AM1 optimized geometries gave lower R^2 (0.586-0.604) values and higher standard errors of estimate (0.468-0.478) than CoMFA models using HF/3-21G* optimized geometries (Table 5.3). R^2 measures of fit were 0.835 to 0.906 by CoMFA model using HF/3-21G* optimized geometries; only AM1 atomic charge displayed a little lower R^2 of 0.835 but other Gaussian type calculations (HF, B3LYP, MP2 methods with 3-21G* basis set) showed R^2 higher than 0.9 (0.905, 0.906, 0.906, respectively). On the other hand, the standard errors of estimation were 0.237 to 0.308 from CoMFA model using HF/3-21G* optimized geometries; only AM1 atomic charge displayed a little higher standard error of 0.308 but other Gaussian type calculations (HF, B3LYP, MP2 methods with 3-21G* basis set) showed lower standard error of estimation (0.239, 0.237, 0.238, respectively) (Table 5.3).

Table 5.4 Experimental and Predicted Bioactivities (pIC50) by the Training Set of 27 Molecules using Various Calculation Methods

Compounds	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//[f]	MP2//[g]	AM1//[h]
	pIC50	AM1	AM1	AM1	Predicted pIC50		HF	HF	HF
1	-2.02	-2.74	-2.67	-2.72	-2.73	-2.2	-2.19	-2.2	-2.39
3	-2.4	-2.77	-2.69	-2.77	-2.79	-2.31	-2.26	-2.31	-2.24
4	-1.85	-2.37	-2.37	-2.38	-2.36	-2.21	-2.2	-2.21	-2.44
5	-2.51	-2.42	-2.38	-2.4	-2.39	-2.19	-2.2	-2.21	-2.18
6	-2.34	-2.43	-2.42	-2.43	-2.42	-2.23	-2.2	-2.22	-2.35
7	-2.59	-2.58	-2.51	-2.57	-2.57	-2.65	-2.61	-2.65	-2.5
8	-3	-2.85	-2.79	-2.85	-2.84	-2.66	-2.66	-2.65	-2.78
9	-3.75	-2.84	-2.8	-2.84	-2.83	-3.67	-3.68	-3.67	-3.57
12	-3.24	-2.77	-2.74	-2.78	-2.8	-2.98	-2.97	-2.98	-2.93
13	-2.45	-2.53	-2.5	-2.53	-2.54	-2.61	-2.57	-2.61	-2.37
14	-2.39	-2.79	-2.75	-2.8	-2.8	-2.54	-2.52	-2.53	-2.85
15	-2.74	-2.68	-2.63	-2.68	-2.69	-2.93	-2.91	-2.93	-2.94
16	-3.12	-2.79	-2.74	-2.79	-2.8	-3.04	-3.03	-3.03	-3.2
17	-2.15	-2.6	-2.57	-2.59	-2.6	-1.93	-1.92	-1.93	-2.19
18	-0.71	-1.33	-1.42	-1.34	-1.34	-1.23	-1.25	-1.23	-1.28
19	-2.38	-1.47	-1.49	-1.47	-1.48	-2.29	-2.3	-2.29	-2.19
20	-2.85	-3	-2.97	-3.03	-2.99	-2.92	-2.96	-2.92	-2.98
21	-1.64	-1.4	-1.46	-1.42	-1.4	-1.34	-1.35	-1.34	-1.21
22	-2.66	-2.19	-2.21	-2.18	-2.19	-2.73	-2.72	-2.73	-2.65
23	-3.17	-2.58	-2.56	-2.56	-2.57	-3.49	-3.47	-3.49	-3.34
24	-2.32	-2.91	-2.85	-2.9	-2.9	-2.3	-2.31	-2.29	-2.61
25	-2.48	-2.88	-2.81	-2.89	-2.91	-2.61	-2.6	-2.62	-2.39
26	-1.48	-1.33	-1.39	-1.33	-1.33	-1.22	-1.25	-1.23	-1.1
27	-3.62	-3.14	-3.77	-3.15	-3.13	-3.34	-3.39	-3.34	-3.16
28	-2.73	-2.97	-2.91	-2.96	-2.96	-2.92	-2.97	-2.92	-2.9
29	-3.46	-3.07	-3	-3.07	-3.08	-3.41	-3.44	-3.41	-3.13
30	-0.96	-1.58	-1.61	-1.59	-1.59	-1.06	-1.08	-1.06	-1.13

[a-h] See in Table 5.2.

This study shows that CoMFA models from HF/3-21G* optimized geometries of 27 PCP derivatives are successfull in interpreting QSAR (Quantum Structure-Activity Relationships) through PLS studies but CoMFA models obtained from AM1 optimized geometries of the 27 compounds are not good predictors of activity for sigma 1 receptor-ligands. The experimental and predicted bioactivity using different CoMFA models are listed in Table 5.4 and their graphs are shown in Figures 5.5 and 5.6.

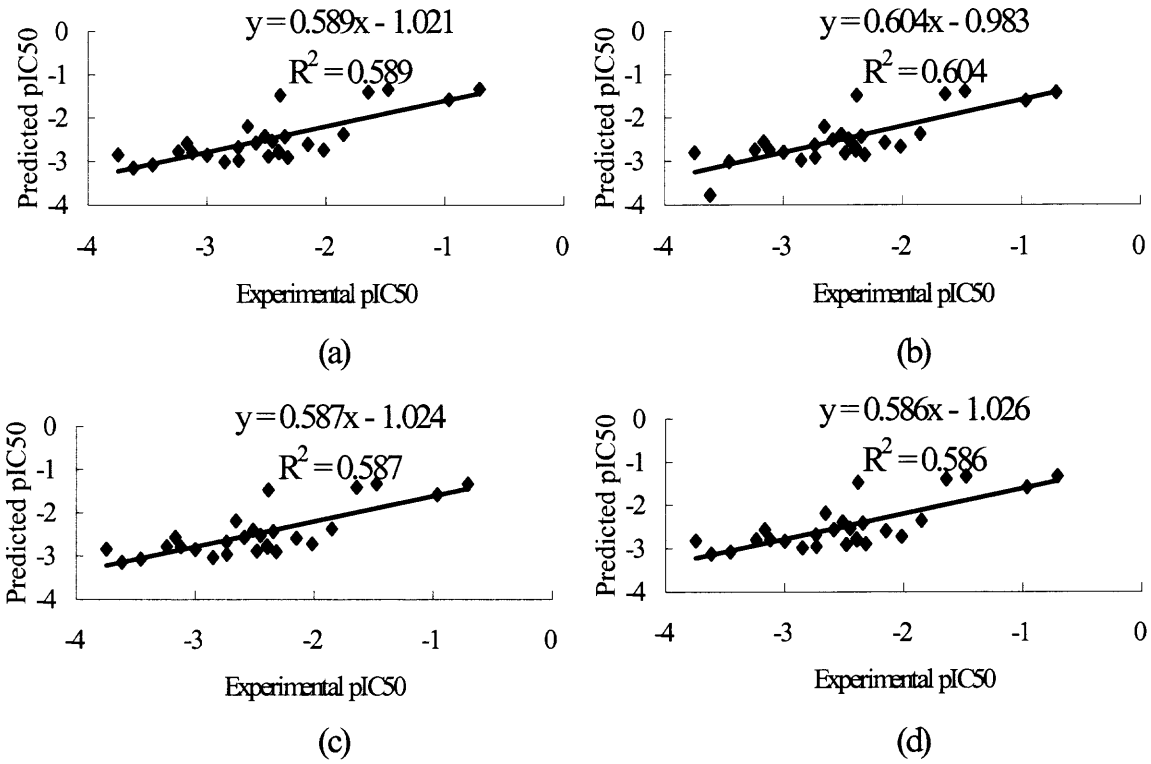


Figure 5.5 Graph of experimental pIC50 versus predicted bioactivity by the CoMFA model using different calculational methods.

[a-d] All geometries were optimized in AM1 and atomic charges were calculated in [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, [d] MP2/3-21G* methods.

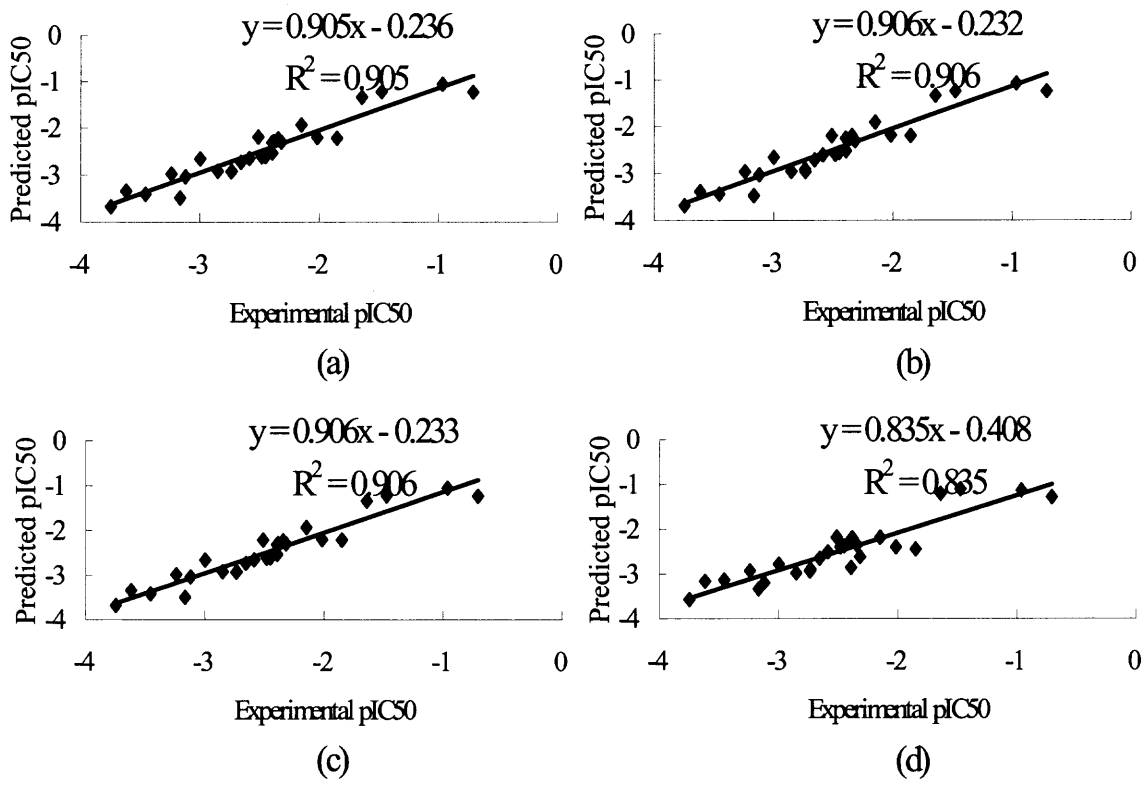


Figure 5.6 Graph of experimental pIC50 versus predicted bioactivity by the CoMFA model using different calculational methods.

[a-d] All geometries were optimized in HF/3-21G* and atomic charges were calculated in [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, [d] AM1, methods.

Table 5.5 Experimental and Predicted Bioactivities (pIC50) by Test Set of Three Molecules using Various Calculation Methods

Number	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//HF[f]	MP2//[g]	AM1//[h]
	pK_i	AM1	AM1	AM1	Predicted pK_i		HF	HF	HF
2	-2.35	-2.7	-2.64	-2.69	-2.7	-2.16	-2.15	-2.15	-2.34
10	-1.41	-1.33	-1.37	-1.32	-1.33	-1.28	-1.3	-1.28	-1.18
11	-3.03	-2.37	-2.37	-2.36	-2.35	-2.82	-2.84	-2.81	-3.02

[a-h] See in Table 5.2.

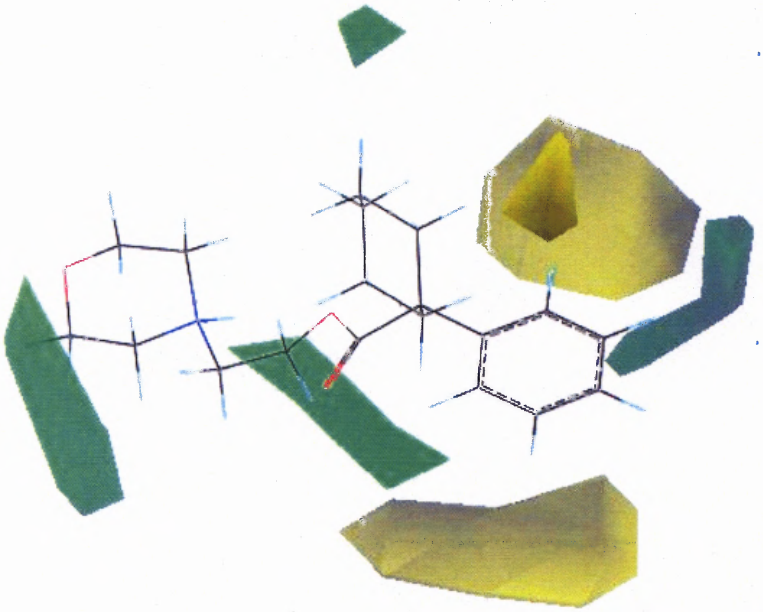


Figure 5.7 CoMFA steric contour map for compound number 23 (PRE-084), derived by 27 PCP derivatives for sigma 1 receptor-ligands; geometries were optimized in AM1 and atomic charges were calculated in AM1, HF/3-21G*, B3LYP/3-21G*, MP2/3-21G methods; they showed the same contour map for steric field.

5.3.2 Validation of CoMFA Models

Three PCP derivatives were selected for a test set to validate of CoMFA models. The range of binding affinities for the test set was –3.03 to –1.41 log units and the predicted range of pIC50 for a test set was –2.70 to –1.33 log units by CoMFA models using AM1 optimized geometries, and –3.02 to –1.30 log units by CoMFA models using HF/3-21G* optimized geometries. The predictive utility of the CoMFA model for three ligands in the test set was considered satisfactory only when using HF/3-21G* optimized geometries. In Table 5.5, CoMFA models using AM1 optimized geometries failed to predict the proper activity range for compound 11.

5.3.3 Design of New Ligands

CoMFA model displayed the spatial distribution of important steric and electrostatic properties affecting the activities in Figures 5.7 to 5.10. The contour maps of the steric field are shown in yellow and green; the green areas (80% contribution) are regions where more bulk is desirable, and yellow (20% contribution) areas are regions where less bulk is favorable for the higher sigma 1 activity in Figures 5.7 and 5.8. CoMFA models of AM1 optimized geometries, displayed a similar contour map for steric field analysis in Figure 5.7 and CoMFA models using HF/3-21G* optimized geometries showed a little different steric contour map by atomic charge calculations; AM1 atomic charges show different steric field analysis from atomic charge calculations of Gaussian type (HF, B3LYP, MP2 with 3-21G* basis set) in Figure 5.8. The contour maps of the electrostatic field are shown in red and blue. The red areas (80% contribution) are the regions where more negative charge is favorable and blue areas (20% contribution) are the regions where more negative charge is disfavorable for higher sigma 1 activity. The CoMFA electrostatic contour model of AM1 and HF/3-21G* optimized geometries, are displayed in Figures 5.9 and 5.10. A careful investigation of the CoMFA model with PCP derivatives revealed that insertion of various groups such as ethylene, propane, and butane between the cyclohexyl and amine group of phencyclidine or its derivatives could improve its bioactivity (pIC50) from the original PCP derivatives from methyl insertion between the cyclohexyl and amine group of phencyclidine or its derivatives (compounds 1, 7, 20, 24, 27 in Table 5.1). These new molecules are shown in Table 5.6. From methane to propane, as the number of carbon atoms of insertion was increased, the sigma

1 bioactivity was increased but when butane was inserted, the bioactivity was decreased (Table 5.6).

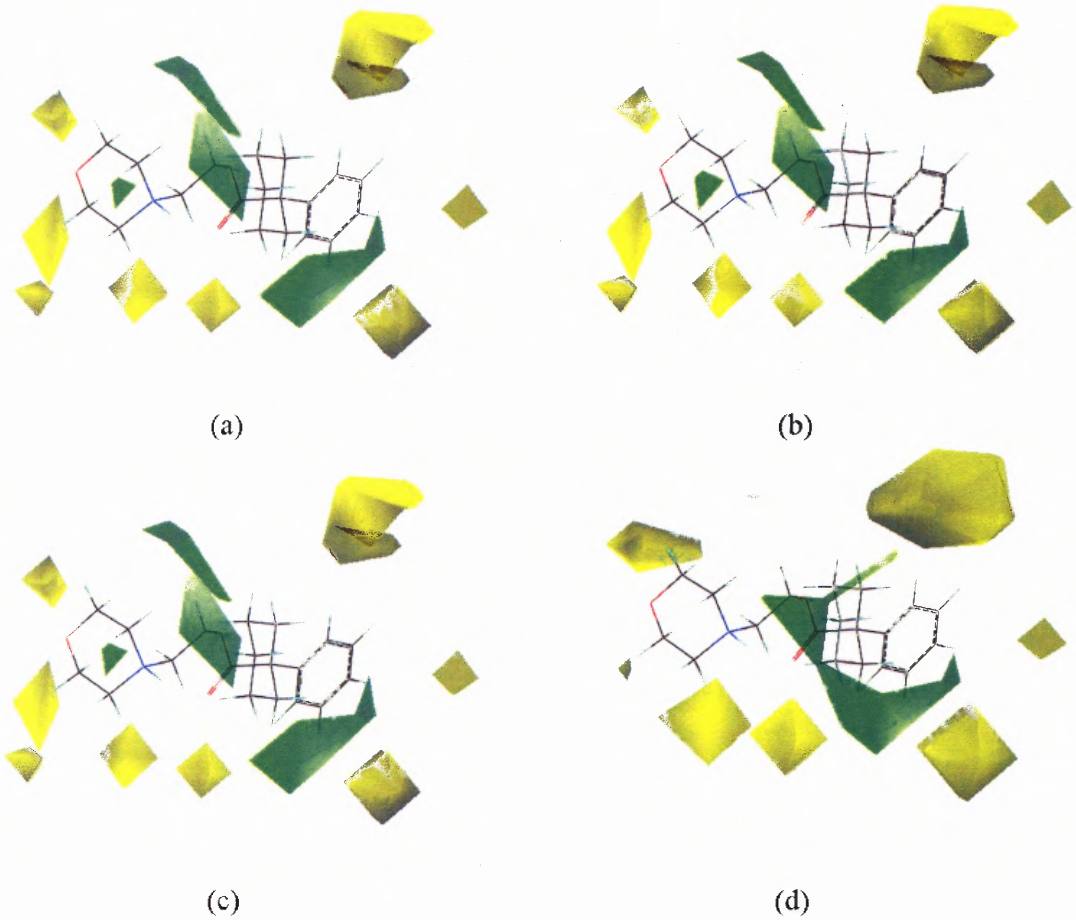


Figure 5.8 CoMFA steric contour map for compound number 23 (PRE-084), derived by 27 PCP derivatives for sigma 1 receptor-ligands using various charge and HF/3-21G* geometry optimization methods.

[a-d] All geometries were optimized in HF/3-21G* and atomic charges are calculated in [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, [d] AM1 methods.

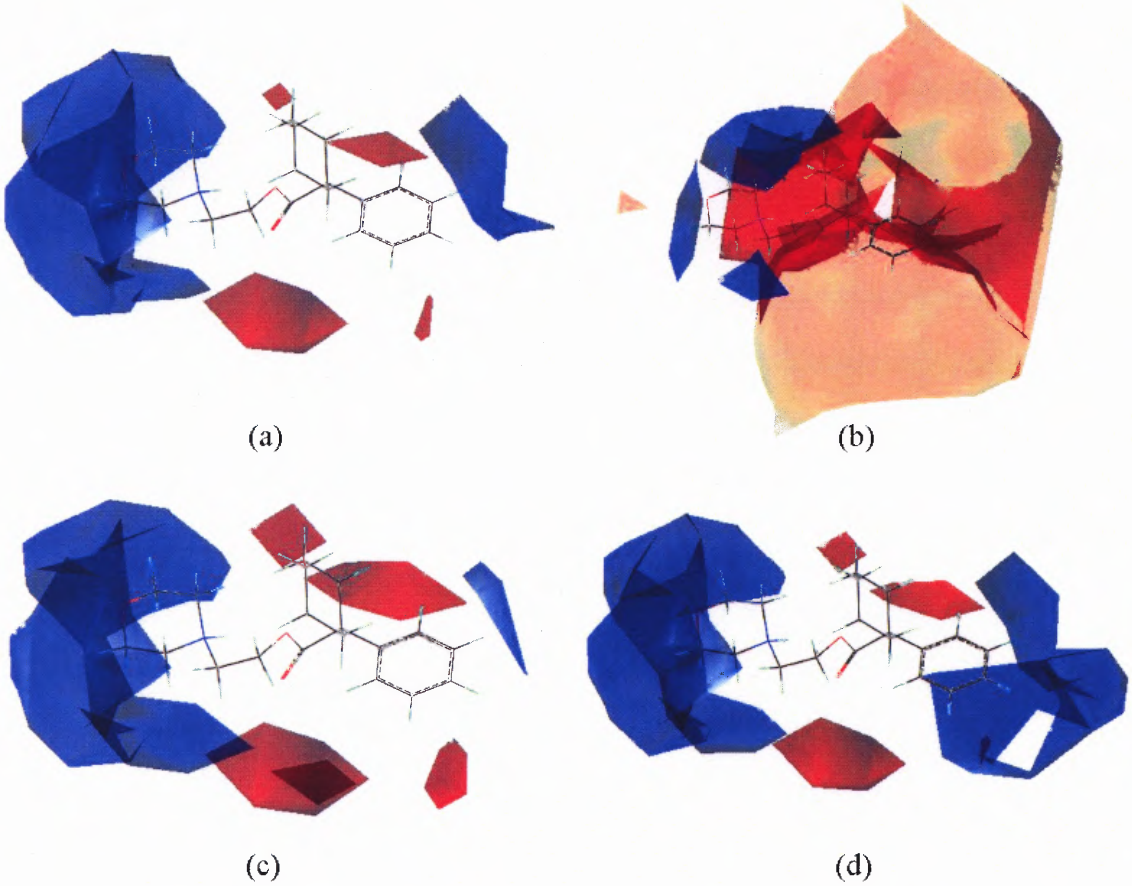


Figure 5.9 CoMFA electrostatic contour map for compound number 23 (PRE-084), derived by 27 PCP derivatives for sigma 1 receptor-ligands using various charge and AM1 geometry optimization methods.

[a-d] All geometries were optimized in AM1 and atomic charges were calculated in [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, [d] MP2/3-21G methods.

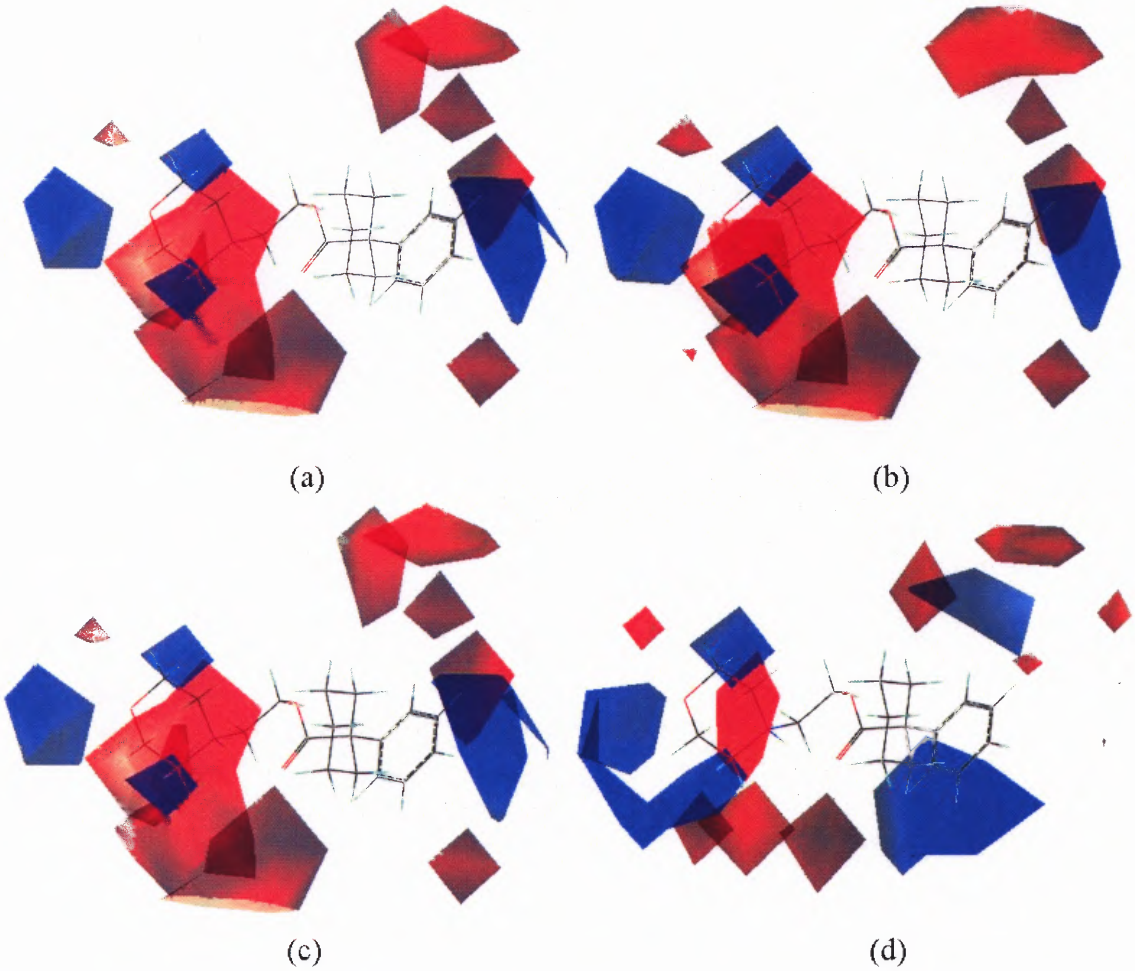


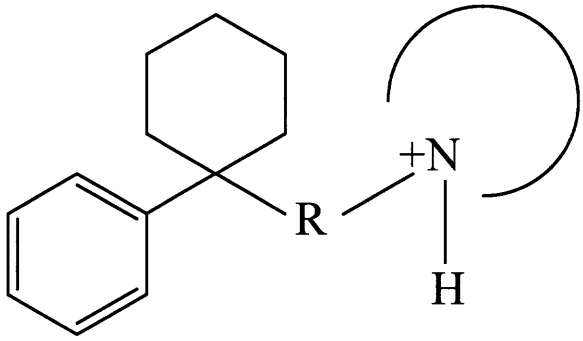
Figure 5.10 CoMFA electrostatic contour map for compound number 23 (PRE-084), derived by 27 PCP derivatives for sigma 1 receptor-ligands using various charge and HF/3-21G* geometry optimization methods.

[a-d] All geometries were optimized in HF/3-21G* and atomic charges are calculated in [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, [d] AM1 methods.

5.4 Conclusions in CoMFA Studies of PCP Derivatives for Sigma 1 Receptor-Ligands

This study derived a pharmacophore for pcp sigma 1 receptor-ligands using DISCOtech. Three points (nitrogen, the center of a phenyl ring, the center of an alkyl ring) was chosen for the pharmacophore points. This pharmacophore was successful in aligning AM1, or HF/3-21G* optimized geometries and was successful in explaining the experimental activity of various sigma1 ligands whose atomic chargeswere calculated by AM1, HF/3-21G*, B3LYP/3-21G*, MP2/3-21G* calculations on Mulliken populations. Furthermore the derived CoMFA model was successful in prediciting activities of new compounds. This study also proved that ComFA models from HF/3-21G* optimized geometries are more reliable in predicting activities of new compounds than CoMFA models obtained from AM1 optimized geometries after validation tests.

Table 5.6 Prediction of Bioactivity for New Ligands



Amine	R	HF//HF [a]	B3//HF [b]	MP2//HF [c]	AM1//HF [d]
			Predicted	pIC50	
Piperidine	C_2H_4	-1.83	-1.83	-1.86	-1.77
Piperidine	C_3H_6	-1.40	-1.40	-1.40	-1.41
Piperidine	C_4H_8	-2.54	-2.53	-2.54	-2.57
Pyrrolidine	C_3H_6	-1.54	-1.53	-1.54	-1.57
Pyrrolidine	C_4H_8	-2.35	-2.34	-2.35	-2.49
Morpholine	C_2H_4	-1.83	-1.88	-1.90	-1.79
Morpholine	C_3H_6	-1.61	-1.56	-1.60	-1.47
Morpholine	C_4H_8	-2.37	-2.57	-2.60	-2.6
$N(C_2H_5)_2$	C_2H_4	-1.95	-1.95	-1.95	-1.93
$N(C_2H_5)_2$	C_3H_6	-1.56	-1.54	-1.56	-1.57
$N(C_2H_5)_2$	C_4H_8	-2.34	-2.33	-2.34	-2.44
$N(CH_3)_2$	C_2H_4	-1.83	-1.84	-1.83	-1.81
$N(CH_3)_2$	C_3H_6	-1.52	-1.89	-1.52	-1.57
$N(CH_3)_2$	C_4H_8	-2.33	-2.32	-2.32	-2.46

[a-d] Geometries were optimized in HF/3-21G* and charges are calculated in [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, and [d] AM1 methods.

CHAPTER 6

COMFA STUDIES USING SEMI-EMPIRICAL, DENSITY FUNCTIONAL, AB INITIO METHODS AND PHARMACOPHORE DERIVATION USING DISCOTECH ON SIGMA 2 RECEPTOR-LIGANDS

6.1 Sigma 2 Receptor-Ligands

Sigma receptors have been classified as σ1, σ2, and σ3 receptor subtypes based upon the differential biochemical and pharmacological properties of structurally diverse ligands. However, at this time, the precise role of its function, and clinical relevance of three subtypes sigma 1, 2 and 3 receptors are not well known. The sigma receptor may represent a new approach for the development of therapeutic agents useful in treating various mental, motor, and other disorders [80-89]. The high density of sigma 1 and sigma 2 binding sites found at various cancer cells suggests important cellular functions of sigma receptors in cancer, as well as potential diagnostic utility for tumor imaging agents which target sigma sites [122,123]. Recently, it was reported that sigma 2 receptor agonists had ability to induce cancer cell death by a mechanism consistent with apoptosis. In breast tumor cell lines that are sensitive (MCF-7) and resistant (MCF-7/Adr-, T47D, and SKBr3) to antineoplastic agents, incubation with the sigma 2 subtype-selective agonists CB-64D and CB-184 produced dose-dependent cytotoxicity [124]. Although many high affinity σ1 ligands have been developed [80-89], very few selective σ2 ligands and their SARs have been reported. Examples include azaperol [125], related BMY-14802 (4-amino-1-arylbutanols) [125], vesamicol analogues [126], alkylamine derivatives [127], trishomocubane [128], N-alkylazacycloheptane derivatives [129], and 5-(3-hydroxyphenyl)-2-methylmorphan-7-one derivatives [130]. The DISCOtech [74-78]

was successful in finding initial conformers for Comparative Molecular Field Analysis (CoMFA) in previous chapters for the σ1 receptor-ligands. It may also enable us to find and design new selective and potent σ2 ligands. Usually semi-empirical AM1 calculation or molecular mechanics were used for CoMFA study, but in this study, ab initio HF calculation with 3-21G* basis set was used to optimize geometry and compared with AM1 method. Atomic charges were calculated AM1, HF/3-21G*, B3LYP/3-21G*, and MP2/3-21G* calculations.

6.2 Materials and Methods

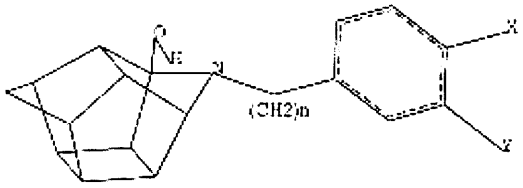
The initial conformer searching and pharmacophore study were analyzed using DISCOtech [74-78] on SYBYL6.9 [71] and the optimization of geometry and calculation of atomic charge were performed using Gaussian 98 [79]. All CoMFA models were derived using SYBYL6.9.

6.2.1 Selection of Ligands

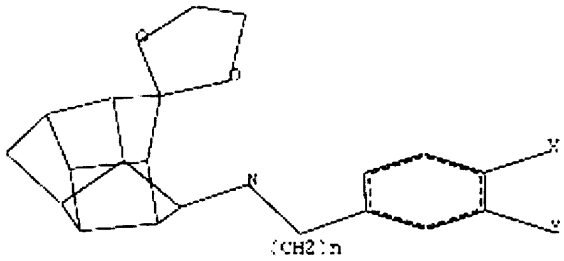
σ2 receptor affinity values for 24 compounds, found in the literature, were critically evaluated. All values had been obtained using [^{3}H]DTG in the presence of (+) pentazocine for the σ2 receptor, and [^{3}H](+)pentazocine used radiolabel for σ1 sites. The K_i values were converted to pK_i values ($pK_i = -\log K_i$). All the used compounds belonged to three structurally different families. These were the Trishomocubane derivatives [128] (shown in Table 6.1), Vesamicol analogues [126] (shown in Table 6.2) and 5-(3-hydroxyphenyl)-2-methylmorphan-7-one derivatives [130] (shown in Table 6.3). A

structural diversity and a homogeneous repatriation of the affinities are necessary to obtain meaningful results from a 3D-QSAR study using the CoMFA method [90-99].

Table 6.1 Binding and Functional Data of Trishomocubane Derivatives [128]



Compound 1-14

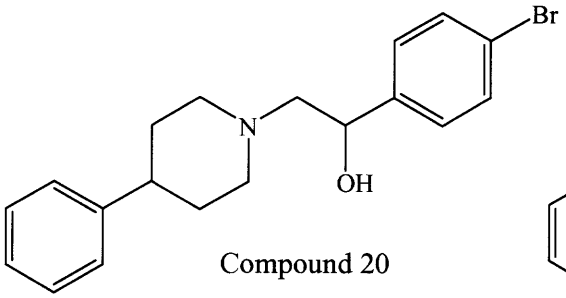


Compound 15-19

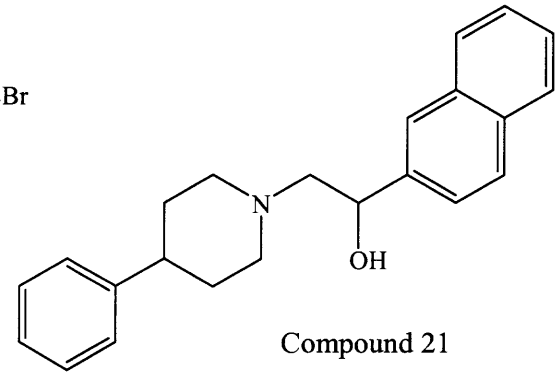
Compounds	n	X	Y	K_i (nM) σ_2	K_i (nM) σ_1	$\sigma1/\sigma2$
1	1	H	H	51 ± 8	103 ± 25	2.02
2	1	Br	H	176 ± 32	86 ± 3	0.49
3	1	I	H	246 ± 46	81 ± 4	0.33
4	1	OCH_3	H	136 ± 19	103 ± 1	0.76
5	1	H	F	20 ± 4	152 ± 1	7.60
6	1	H	Cl	30 ± 1	186 ± 8	0.62
7	1	H	Br	40 ± 22	208 ± 13	5.20
8 [a]	1	H	I	54 ± 18	169 ± 10	3.13
9	1	H	CH_3	108 ± 6	97 ± 6	0.90
10	2	H	H	307 ± 18	20 ± 4	0.07
11	2	Br	H	166 ± 32	10 ± 1	0.06
12	2	H	Cl	153 ± 35	21 ± 2	0.14
13	3	H	H	238 ± 7	21 ± 2	0.09
14	4	H	H	171 ±17	9 ± 3	0.05
15	1	H	H	864 ± 258	67 ± 11	0.08
16	1	H	Br	208 ± 36	17 ± 1	0.08
17 [a]	1	H	I	285 ± 46	124 ± 10	0.44
18	1	I	H	246 ± 9	72 ±10	0.29
19	2	H	H	608 ± 2	15 ± 2	0.02

[a] Compounds of test sets.

Table 6.2 Binding and functional data of vesamicol analogues [126]



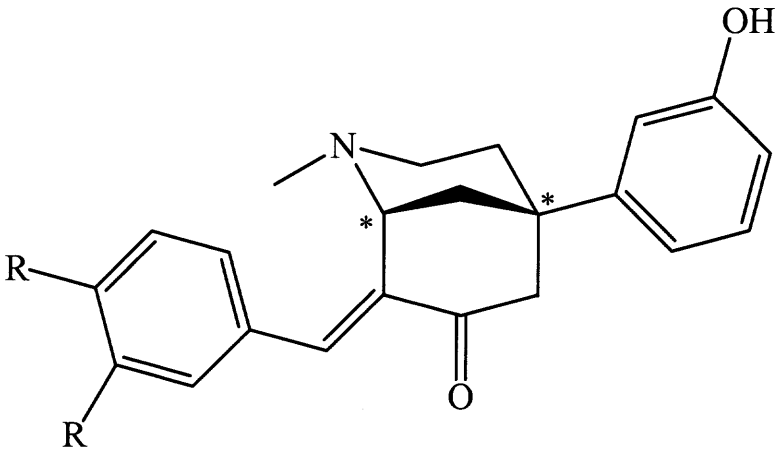
Compound 20



Compound 21

Compounds	K_i (nM) σ_2	K_i (nM) σ_1	σ_1/σ_2
20	7 ± 2	13 ± 0	1.86
21	43 ± 10	63 ± 20	1.47

Table 6.3 Binding and functional data of CB-64D, CB-182 and CB-184 [130]



Compounds	R	stereo *	Name	K_i (nM) σ_2	K_i (nM) σ_1	$\sigma1/\sigma2$
22	H	R	CB-64D [b]	16.5 ± 2.7	3063 ± 78	185.64
23 [a]	Cl	S	CB-182 [c]	35.5 ± 8.8	27.3 ± 2.8	0.77
24	Cl	R	CB-184 [d]	13.4 ± 2.0	7436 ± 308	554.93

[a] Compound of test sets. [b] (+)1R,5R–(E)-8-Benzylidene-5-(3hydroxyphenyl)-2-methylmorphan-7-one. [c] (–)1S,5S– and [d] (+)1R,5R–(E)-8-(3,4-dichlorobenzylidene)-5-(3hydroxyphenyl)-2-methylmorphan-7-one.

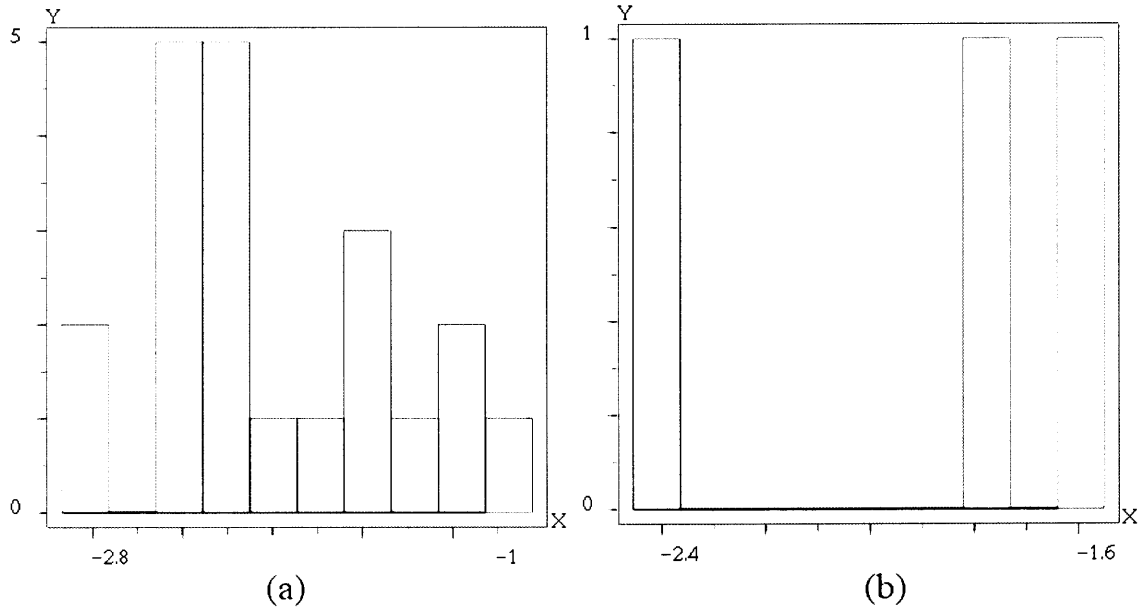


Figure 6.1. Histogram of pKi (abscissa) vs number of molecules (ordinate).

However, the current available selective σ2 ligands are limited. A training set containing 21 compounds and a test set of three compounds were used to assess the predictive power of the model. Histogram pictures of train and test sets are shown in Figure 6.1. The range of binding affinities for the training set was –2.94 to –0.85 log units, and –2.45 to –1.55 log units for the test set.

6.2.2 Choice of Initial Conformations

The CoMFA study began with the selection of the three-dimensional conformation for each compound. Initial structures were generated by building with SYBYL6.9 [71] default bond distances and angles, and minimized with the tools MAXIMIN2 in SYBYL6.9 in which the Tripos force field was applied with a distance-dependent dielectric function. Then, DISCOtech [74-78] in SYBYL6.9 was used to search possible conformations and proper pharmacophores using these initial molecules. The energy limit

was 35 kcal/mol. The reference structure in computing DISCOtech models was compound 5, because it is the most bioactive σ2 receptor ligand in main trishomocubane derivatives. The other important parameters are "Match_All" as structure requirements, Range of points consisting of a minimum of 3 to a maximum of 8 selected feature requirements.

6.2.3 Pharmacophore Information

DISCOtech [74-78] derived a pharmacophore model based on the Trishomocubane derivatives [128] Vesamicol analogues [126] and 5-(3-hydroxyphenyl)-2-methylmorphan-7-one derivatives [130] (shown in Figure 6.2). DISCOtech found possible conformations within reasonable energy boundaries (in this study, 35 kcal/mol) and suggested a proper pharmacophore model. The overall pharmacophore is a triangle that include a nitrogen and two centers of hydrophobic rings. This is a new trial to derive a pharmacophore for sigma 2 receptor-ligands.

6.2.4 Geometry Optimzation and Atomic Charges

The conformers, derived by DISCOtech [74-78] were optimized with ab initio HF/3-21G* [99,103] method or with semi-empirical AM1 [99-102] calculation. with Gaussian 98 [79]. Theses geometries were calculated for atomic charges in semi-empirical AM1, density functional B3LYP/3-21G*, ab initio HF/3-21G* and MP2/3-21G* levels according to Mulliken population using Gaussian 98.

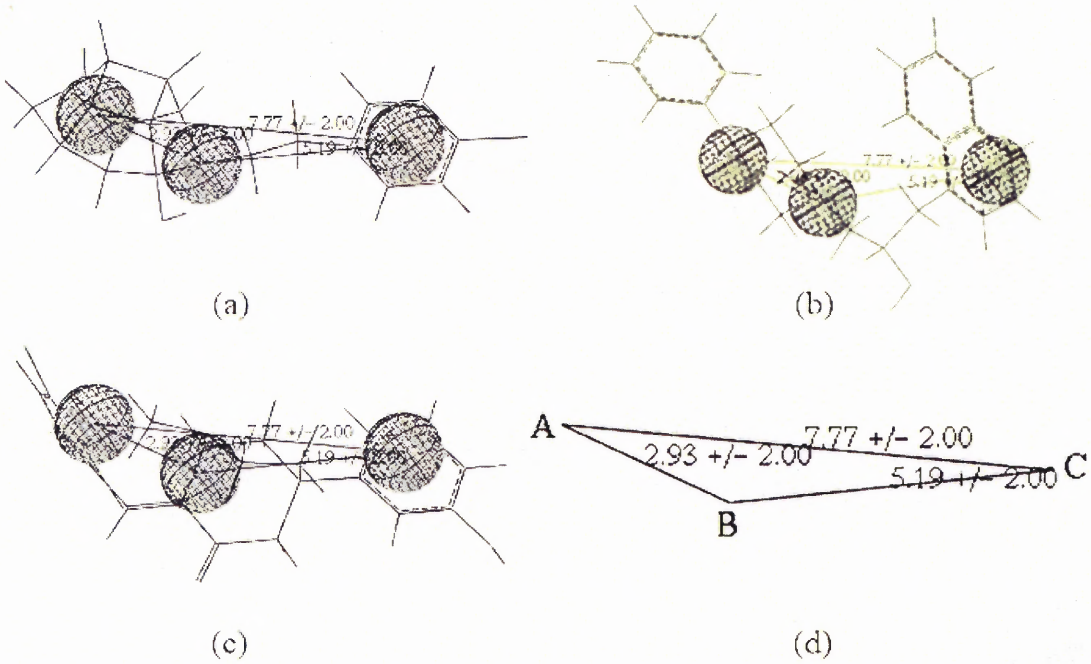


Figure 6.2 DISCOtech pharmacophore for sigma 2 receptor-ligands.

[a] Trishomocubane derivatives with compound number 5 (Table 6.1), [b] vesamicol analogues with compound number 21 (Table 6.2),. [c] CB-182, compound number 22 (Table 6.3), [d] DISCOtech model; A is the center of a hydrophobic ring. B is a nitrogen atom. C is the center of a phenyl ring.

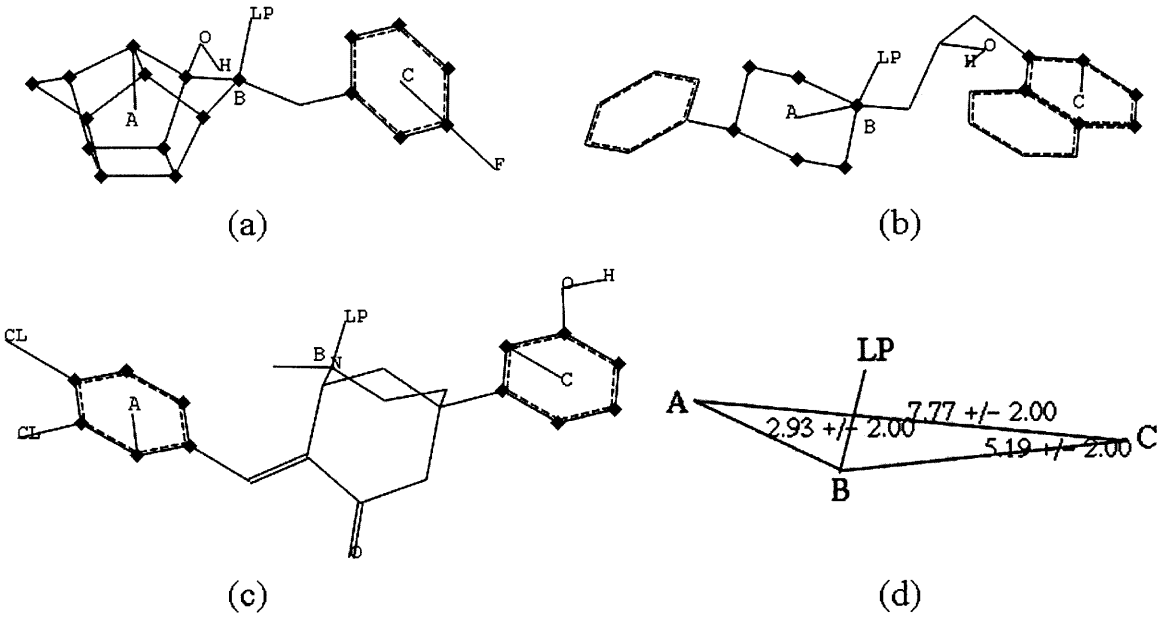


Figure 6.3 Alignments using modified DISCOtech pharmacophore.

[a] A template molecule of compound number 5 (Table 6.1); A is the center of bolded hydrophobic ring; B is a nitrogen atom and LP is the lone pair of electrons; the distance from N atom to a lone pair of electrons was scaled in 1.4Å because it was reported [110] that 1.4Å performed the best result on CoMFA study for the distance from N atom to a lone pair of electrons.; C is the center of bolded hydrophobic phenyl ring. [b] compound number 21 (Table 6.2), [c] compound number 24 (Table 6.3), [d] modified DISCOtech pharmacophore; three point triangle of (d) in Figure 6.2 and a lone pair of electron is added.

6.2.5 Alignment

Alignment of the presumed bound conformations of the training set compounds is also an essential prelude to the CoMFA study. The AM1 or HF/3-21G* optimized conformers were aligned by a fit function in SYBYL6.9 using a template compound (number 5 in Table 6.1) using a 4 point pharmacophore (two hydrophobic ring center A, B, nitrogen atom, and a lone pair of electrons in Figure 6.3). How these centers of hydrophobic ring were defined is explained in Figure 6.3, also. The aligned 24 molecules, used in training and test sets are shown by optimization methods in Figure 6.4 for (a) AM1 and (b) for HF/3-21G*.

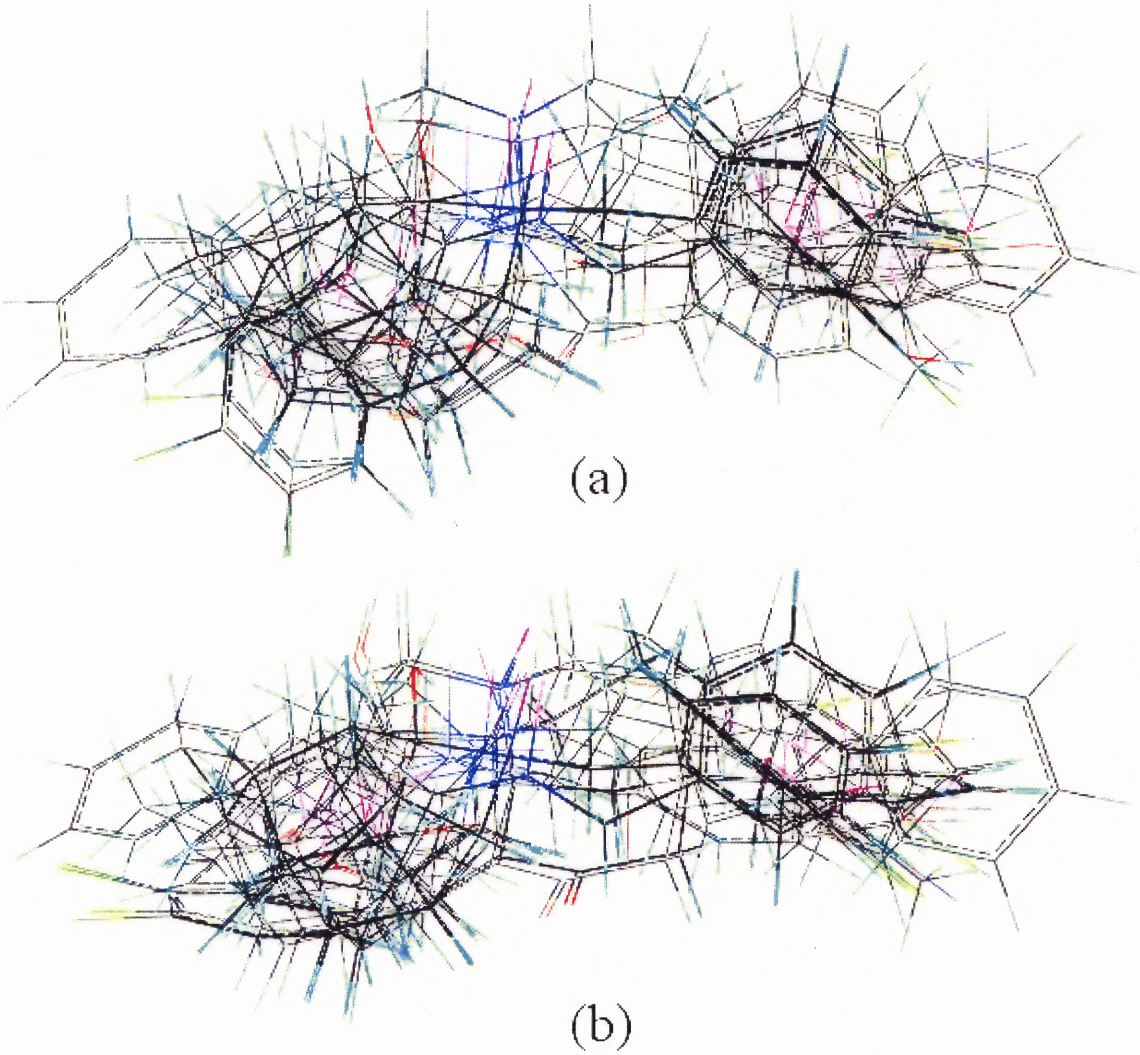


Figure 6.4 Alignments of all Molecules, optimized using (a) AM1 and (b) HF/3-21G* methods.

6.2.6 CoMFA Model

Auto CoMFA columns were calculated using the Tripos Standard CoMFA field class. It extended 4 Å beyond every molecule in all directions, and had a 2 Å spacing, and a probe atom of C.3 (sp^3 carbon) and a charge of +1 with a dielectric function of 1/r, a dielectric constant ε of 1 and the default of 30 kcal/mol energy cutoff for steric and electrostatic fields. Partial least squares analysis, regresses a target property against predictors calculated as steric and electrostatic components of the intermolecular interaction field. Scaling was used as the CoMFA standard. The SAMPLS (SAMple-distance PLS) algorithm developed by Bruce Bush [111] was used to determine "leave-one-out" cross-validation q^2. The method for cross-validation serves two purposes; (1) to find out whether the CoMFA model was productively useful, and (2) if useful, to decide how many components to use for the best model. The number of optical components was considered by the 5% rule; if the q^2 increases by at least 5% upon increasing the number of components by one, then it is justified to add an additional component. The Partial Least Squares (PLS) analysis was then repeated without cross-validation using the optimum number of components. This final analysis yielded a predictive model, and a CoMFA coefficient contour plot for the steric and electrostatic potentials contributions.

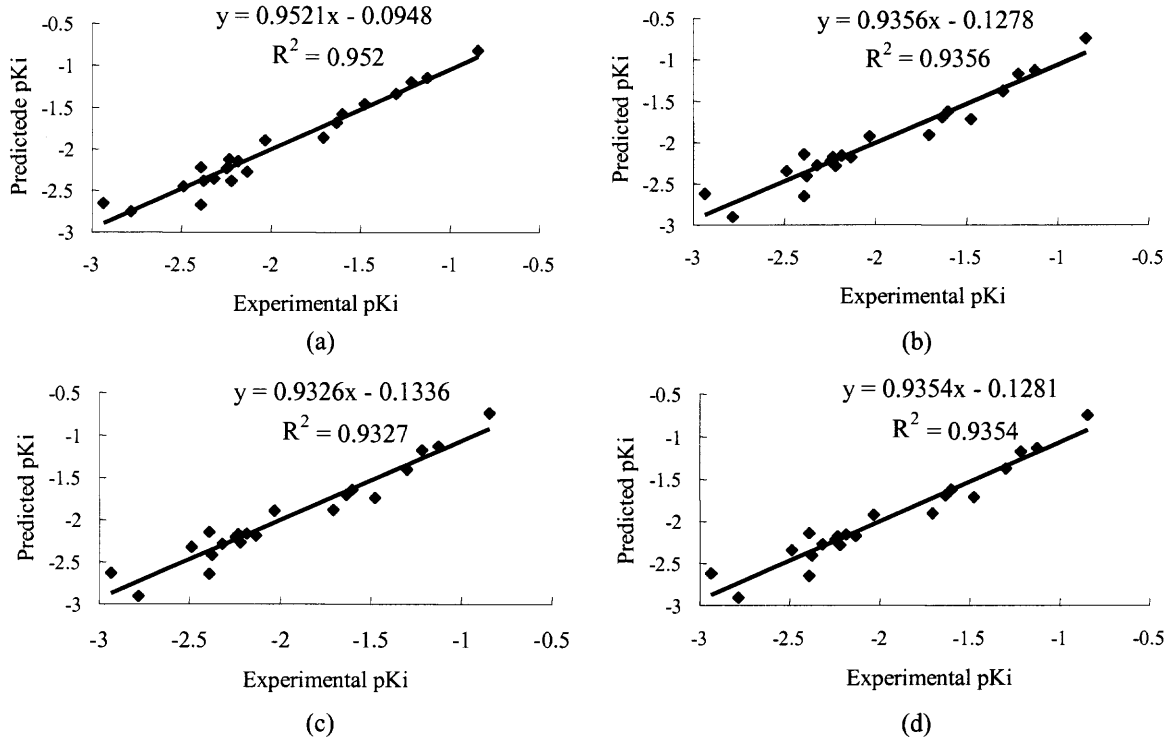


Figure 6.5 Graph of experimental pKi(-logKi) versus predicted bioactivity by the CoMFA model using different calculational methods.

[a-d] All geometries were optimized in AM1 and atomic charges were calculated in [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, [d] MP2/3-21G* methods.

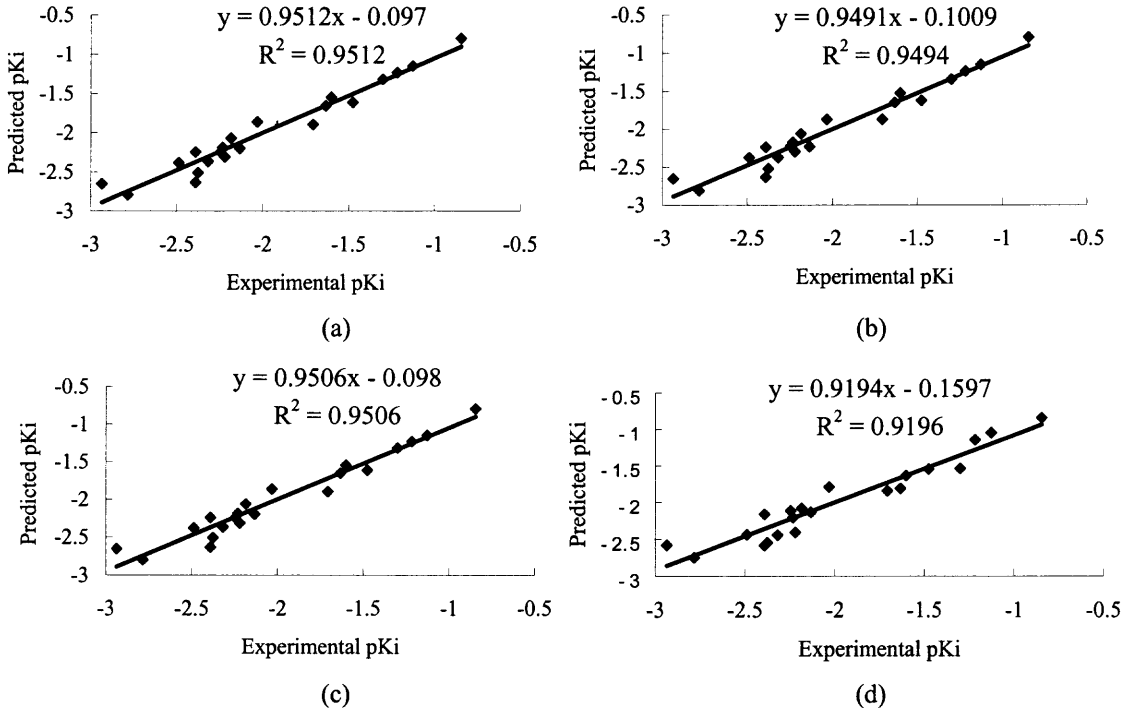


Figure 6.6 Graph of experimental pKi(-logKi) versus predicted bioactivity by the CoMFA Model using different calculational methods.

[a-d] All geometries were optimized in HF/3-21G* and atomic charges were calculated in [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, [d] AM1, methods.

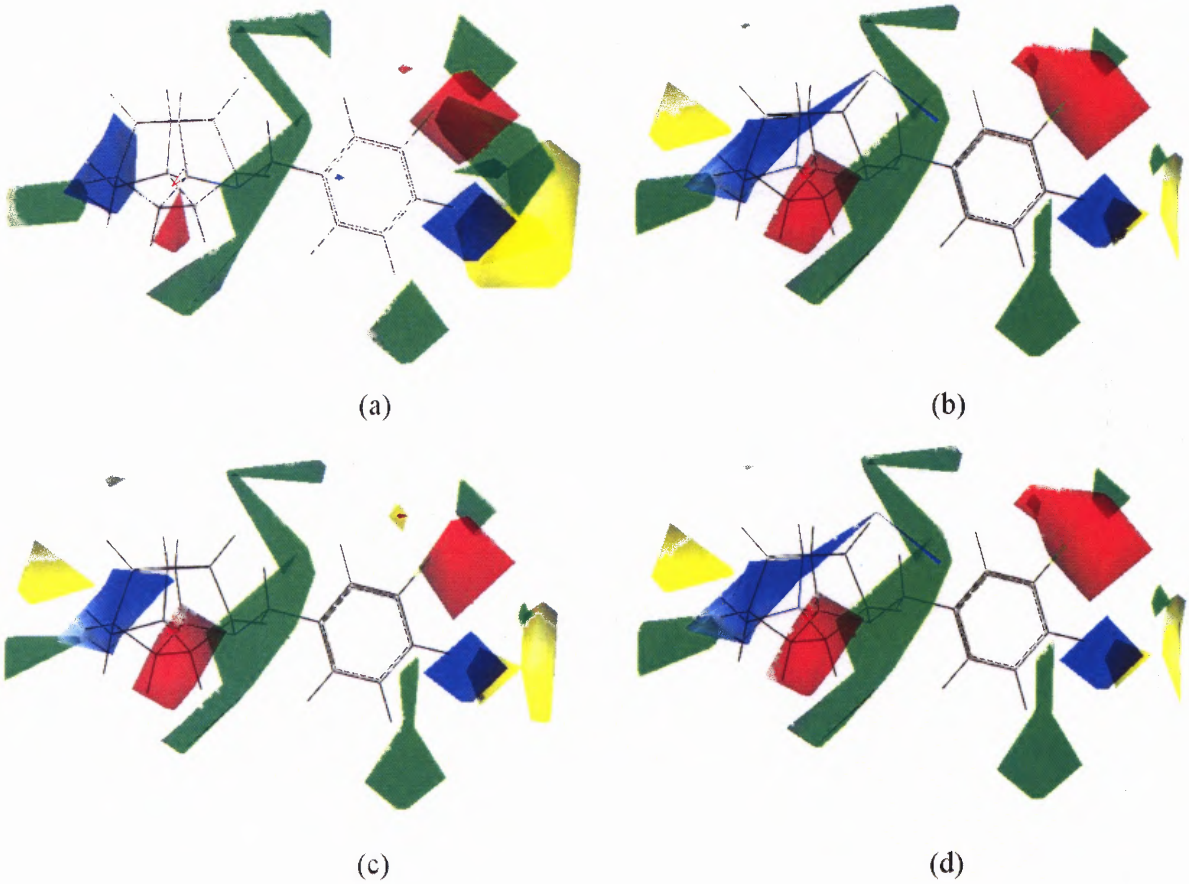


Figure 6.7 CoMFA contour map for compound number 5, derived by 21 Sigma 2 receptor-ligands using various charge and AM1 geometry optimization methods.

[a-d] All geometries were optimized in AM1 and atomic charges were calculated in a AM1, b HF/3-21G*, c B3LYP/3-21G*, d MP2/3-21G methods.

6.3 Results and Discussion

6.3.1 Comparative Molecular Field Analysis

The CoMFA model in this study, required four or five optimal components in different calculations to explain the variance in binding affinity to sigma 2 receptors in Table 6.4. All crossvalidated q^2 were more than 0.4 in Table 6.4, which surpassed the generally accepted criterion, 0.3 [112] for statistical validity. The highest q^2 (0.602) was for HF/3-21G* optimized geometries and MP2/3-21G* atomic charge calculations. The CoMFA models of AM1 optimized geometries produced lower q^2 (0.475-0.503) than those of HF/3-21G* optimized geometries (0.542-0.603). It suggests that CoMFA models of HF/3-21G* optimized geometries explain more correctly about the actual variance in activity among additional similar sigma 2 ligands than those of AM1 optimized geometries.

Having the crossvalidation to confirm the predictive ability, a PLS analysis was performed without any validation to derive the best predictive model for use in graphics and in numerical prediction. R^2 measures of fit were 0.920 to 0.952 and the standard errors of estimate were 0.142 to 0.179. The steric fields contributed 36.9 to 40.7% of the model's information, while the electrostatic fields represented the other 59.3 to 63.1% in Table 6.6. The relationship is shown between calculated and measured pK_i values (predicted) for the non-cross-validated analysis in Table 6.6 and Figure 6.5 and 6.6. CoMFA models obtained from AM1//AM1 calculations (AM1 for charge calculation and AM1 for geometry optimization) show the highest R^2 (0.952) values in Table 6.5, but AM1//HF calculations (AM1 for charge calculations and HF for geometry optimization) displays the lowest q^2 (0.920) value in Table 6.4. Otherwise, HF/3-21G* geometry

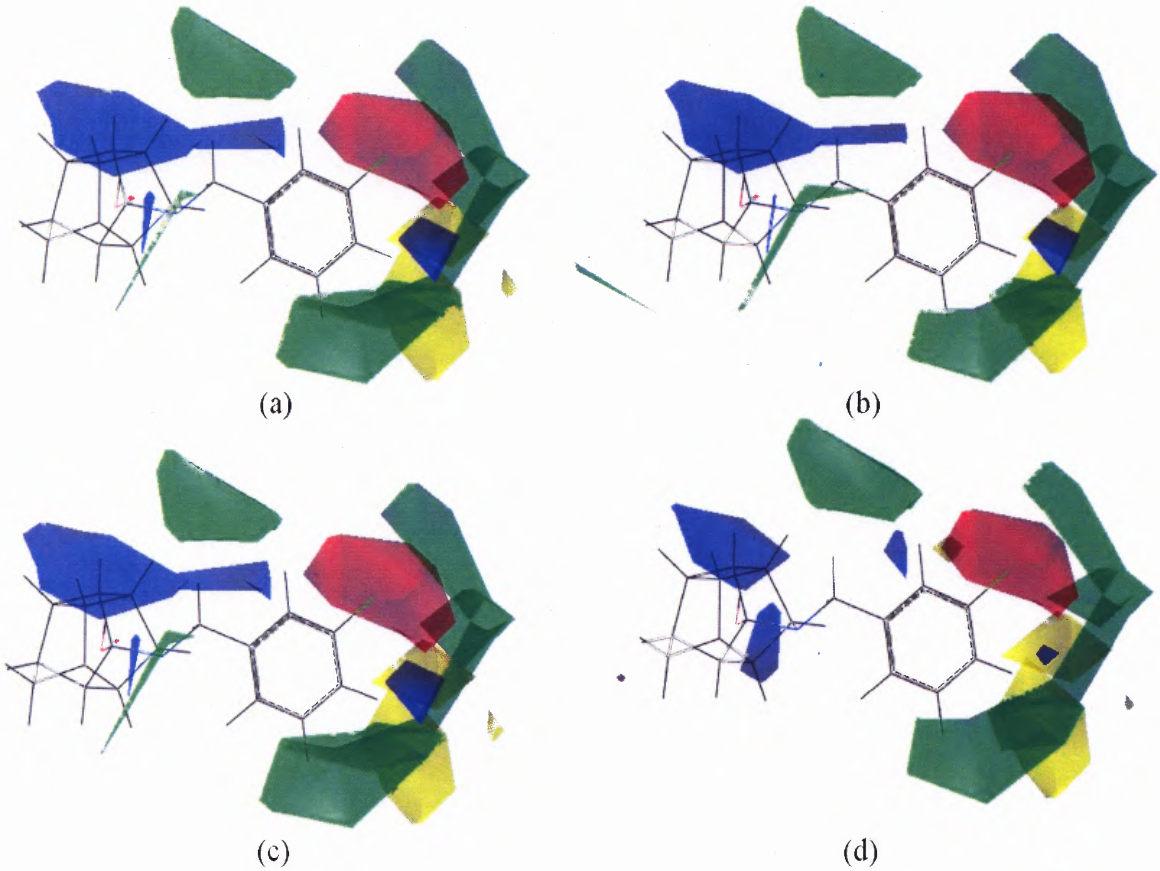


Figure 6.8 CoMFA Contour Map for Compound Number 5, Derived by 21 Sigma 2 receptor-ligands using Various Charge and HF/3-21G* Geometry Optimization Methods.

[a-d] All geometries were optimized in HF/3-21G* and atomic charges are calculated in [a] HF/3-21G*, [b] B3LYP/3-21G*, [c] MP2/3-21G*, [d] AM1 methods.

optimization has higher R^2 (0.951/HF/3-21G*, 0.949/B3LYP/3-21G*, 0.951/MP2/3-21G*) than that of AM1 geometry optimization using different various methods for atomic charge calculations (0.936/HF/3-21G*, 0.933/B3LYP/3-21G*, 0.935/MP2/3-21G*). From our results it is apparent that HF/3-21G* methods can be used successfully

to interpret QSAR (Quantum Structure-Activity Relationships) through PLS studies. The AM1 optimized geometries are interpreted well also using AM1 atomic charges

Table 6.4 The Number of Optimal Components [i] and q^2 by "Leave-One-Out" Using SAMPLS [112] by the Training Set of 21 Molecules

Charge// Geometry	Term	Comp. 1 j	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7
AM1//AM1 [a]	S.E.E. [k]	0.504	0.476	0.47	0.464	**0.464**	0.468	0.5
	q2 [l]	0.238	0.355	0.408	0.456	**0.491**	0.515	0.488
HF//AM1 [b]	S.E.E.	0.492	0.456	0.446	**0.444**	0.449	0.435	0.472
	q2	0.273	0.41	0.466	**0.503**	0.522	0.583	0.543
B3//AM1 [c]	S.E.E.	0.49	0.468	0.459	**0.456**	0.464	0.447	0.479
	q2	0.279	0.379	0.435	**0.475**	0.49	0.558	0.529
MP2//AM1 [d]	S.E.E.	0.493	0.456	0.447	**0.444**	0.449	0.435	0.473
	q2	0.273	0.409	0.465	**0.502**	0.522	0.582	0.542
HF//HF [e]	S.E.E.	0.511	0.424	0.422	0.418	**0.412**	0.429	0.46
	q2	0.218	0.489	0.521	0.559	**0.599**	0.594	0.565
B3//HF [f]	S.E.E.	0.503	0.423	0.425	0.419	**0.419**	0.436	0.464
	q2	0.24	0.491	0.515	0.557	**0.584**	0.581	0.558
MP2//HF [g]	S.E.E.	0.51	0.421	0.419	0.414	**0.41**	0.427	0.459
	q2	0.221	0.496	0.528	0.567	**0.602**	0.598	0.568
AM1//HF [h]	S.E.E.	0.518	0.426	0.43	**0.426**	0.444	0.468	0.495
	q2	0.195	0.485	0.503	**0.542**	0.532	0.517	0.498

[a-d] Charges were calculated using [a] AM1, [b] HF/3-21G*, [c] B3LYP/3-21G*, [d] MP2/3-21G* theory and geometries were optimized in AM1. [e-h] Geometry was optimized in HF/3-21G* and charges are calculated in [e] HF/3-21G*, [f] B3LYP/3-21G*, [g] MP2/3-21G*, and [h] AM1 methods.. [i] The number of optimal components is the bold one in Table in each level. [j] Component number, [k] standard error of estimate, and [l] leave-one-out cross-validated regression of co-efficient.

Table 6.5 QSAR Reports by Non-Crossvalidation Using SAMPLS [112] by the Training set of 21 molecules

Theory	S. E.[i]	R^2 [j]	F values	Steric. [k]	Electro. [l]
AM1//AM1 [a]	0.142	0.952	(n1=5,n2=15) 59.517	0.386	0.614
HF//AM1 [b]	0.160	0.936	(n1=4,n2=16) 58.157	0.369	0.631
B3//AM1 [c]	0.163	0.933	(n1=4,n2=16) 55.433	0.373	0.627
MP2//AM1 [d]	0.160	0.935	(n1=4,n2=16) 57.837	0.369	0.631
HF//HF [e]	0.144	0.951	(n1=5,n2=15) 58.418	0.388	0.612
B3//HF [f]	0.146	0.949	(n1=5,n2=15) 56.271	0.397	0.603
MP2//HF [g]	0.145	0.951	(n1=5,n2=15) 57.652	0.391	0.609
AM1//HF [h]	0.179	0.920	(n1=4,n2=16) 45.701	0.407	0.593

[a-h] See in Table 6.4. [i] Standard error of estimation. [j] R^2 of non-crossvalidation using training set of 21 molecules in Table 6.4.1-4.3. [k] Steric contribution to this CoMFA field. [l] Electrostatic contribution to this CoMFA field.

6.3.2 Validation of the CoMFA Model

The test compounds, selected were two Trishomocubane derivatives (shown in Table 6.1) and one 5-(3-hydroxyphenyl)-2-methylmorphan-7-one derivative (CB-182, shown in Table 6.3). The range of binding affinities for the test set was –2.45 to –1.55 log units and the predicted range of pKi for the test set was –2.90 to –1.71 log units. The predictive utility of the CoMFA model for the three ligands in the test set was considered satisfactory. Compounds 8 and 17 were predicted well by all calculational methods but compound 13 (CB-182) was not predicted well when geometries were optimized by the AM1 method. HF geometry optimizations and charges calculated by all methods produced good results for this compound however.

6.3.3 Design of New Ligands

CoMFA models (Figures 6.7 and 6.8) illustrate the spatial distributions of important steric and electrostatic properties affecting the activities of the 21 compounds used in the derivation of the models. The contour maps of the steric fields are shown in yellow and green. The green areas (80% contribution) are regions where more bulky substitutions are desirable, and yellow (20% contribution) areas are regions where less bulk is favorable for the higher σ2 activity. The CoMFA steric contour model derived by AM1 and HF/3-21G* optimized geometries, shows that the Y position of trishmocubane derivartives in Table 6.1 is greenish, where steric bulk is favored. The X position of substitutents falls in the yellow areas where less steric bulk is favored for the higher σ2 activity. The contour maps of the electrostatic fields are shown in red and blue. The red areas (80% contribution) are the regions where more negative charge is favorable and blue areas (20% contribution) are the regions where more negative charge is disfavorable for the higher σ2 activity. The CoMFA electrostatic contour model derived from AM1 and HF/3-21G* optimized geometries, shows that the Y position of trishomocubane derivartives in Table 6.1 falls in the red region where more negative charge is favorable for the higher σ2 activity. Through the investigation of steric and electrostatic contour maps the most desired substitutents are those that are sterically small and electron withdrawing. When NO_2 was substituted in the Y position σ2 activity increased (Table 6.7).

Table 6.6 Experimental and Predicted Bioactivities (pK_i) by the Training Set of 21 Molecules Using Various Calculation Methods

Compounds	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//[f]	MP2//[g]	AM1//[h]
	pIC50	AM1	AM1	AM1	Predicted pIC50		HF	HF	HF
1	-1.71	-1.86	-1.91	-1.88	-1.91	-1.9	-1.87	-1.9	-1.84
2	-2.25	-2.23	-2.22	-2.21	-2.22	-2.25	-2.22	-2.24	-2.12
3	-2.39	-2.22	-2.14	-2.14	-2.14	-2.24	-2.24	-2.24	-2.16
4	-2.13	-2.27	-2.18	-2.19	-2.17	-2.2	-2.23	-2.2	-2.14
5	-1.3	-1.34	-1.38	-1.41	-1.38	-1.32	-1.34	-1.32	-1.54
6	-1.48	-1.46	-1.72	-1.74	-1.72	-1.61	-1.63	-1.61	-1.55
7	-1.6	-1.58	-1.62	-1.64	-1.63	-1.54	-1.53	-1.55	-1.63
9	-2.03	-1.89	-1.92	-1.89	-1.92	-1.86	-1.87	-1.86	-1.79
10	-2.49	-2.45	-2.34	-2.32	-2.34	-2.38	-2.38	-2.38	-2.44
11	-2.22	-2.39	-2.28	-2.27	-2.28	-2.31	-2.3	-2.32	-2.41
12	-2.19	-2.14	-2.16	-2.16	-2.16	-2.07	-2.06	-2.06	-2.08
13	-2.38	-2.39	-2.41	-2.42	-2.41	-2.51	-2.52	-2.51	-2.55
14	-2.23	-2.12	-2.18	-2.17	-2.18	-2.19	-2.17	-2.19	-2.21
15	-2.94	-2.65	-2.62	-2.63	-2.62	-2.65	-2.65	-2.65	-2.58
16	-2.32	-2.36	-2.27	-2.28	-2.27	-2.37	-2.37	-2.37	-2.45
18	-2.39	-2.67	-2.65	-2.65	-2.65	-2.63	-2.63	-2.63	-2.58
19	-2.78	-2.75	-2.9	-2.91	-2.91	-2.8	-2.81	-2.8	-2.75
20	-0.85	-0.83	-0.75	-0.74	-0.75	-0.8	-0.79	-0.8	-0.85
21	-1.63	-1.68	-1.7	-1.71	-1.7	-1.66	-1.65	-1.65	-1.81
22	-1.22	-1.2	-1.17	-1.17	-1.17	-1.23	-1.24	-1.23	-1.14
24	-1.13	-1.15	-1.13	-1.13	-1.13	-1.15	-1.15	-1.15	-1.05

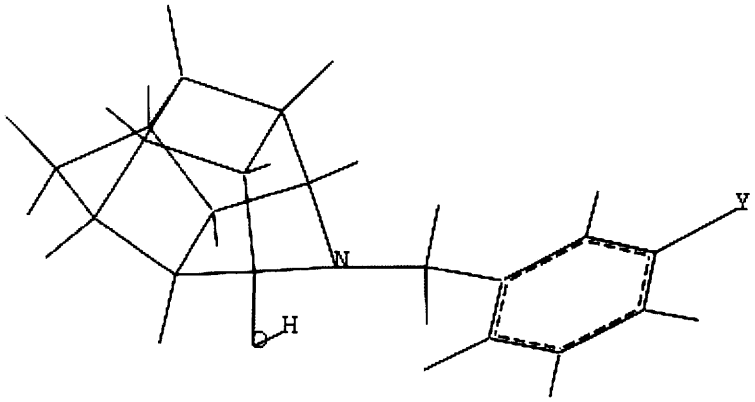
[a-h] See in Table 6.4.

Table 6.7 Experimental and Predicted Bioactivities (pK_i) by Test Set of three molecules using Various Calculation Methods

Compounds	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//[f]	MP2//[g]	AM1//[h]
	pKi	AM1	AM1	AM1	Predicted pIC50		HF	HF	HF
8	-1.73	-1.79	-1.84	-1.81	-1.85	-1.75	-1.74	-1.75	-1.73
17	-2.45	-2.58	-2.59	-2.57	-2.59	-2.6	-2.9	-2.6	-2.52
23	-1.55	-2.23	-2.26	-2.23	-2.26	-1.8	-1.83	-1.8	-1.71

[a-h] See in Table 6.4

Table 6.8 Prediction of Bioactivity for New Ligands



Y	Experiment	AM1//[a]	HF//[b]	B3//[c]	MP2//[d]	HF//[e]	B3//HF[f]	MP2//[g]	AM1//[h]
	pKi	AM1	AM1	AM1	Predicted pK_i		HF	HF	HF
F	-1.30	-1.34	-1.38	-1.41	-1.38	-1.32	-1.34	-1.32	-1.54
NO_2	Unknown	-1.04	-1.21	-1.00	-1.21	-1.08	-1.04	-1.08	-1.31

[a-h] See in Table 6.4

6.4 Conclusions of CoMFA Studies of Sigma 2 Receptor-Ligands

In this study a plausible pharmacophore for sigma 2 receptor-ligands was derived using DISCOtech from Sybyl. Four points (nitrogen, lone pair of electrons, and 2 centers of hydrophobic rings) were used to successfully align 24 compounds whose geometries were optimized using AM1 or HF/3-21G* calculations. A pharmacophore was derived which consists of four points. A triangle of two hydrophobic points (centroids of two phenyl rings A,C) and a nitrogen (B) and a lone pair of nitrogen (LP). The distances found are as follows: A-B (2.93Å) B-C (5.19Å) A-C (7.77Å) and B-LP (1.40Å). Using the derived pharmacophore ,CoMFA studies were performed. CoMFA models were derived by using several different levels of calculations for charge and geometry optimizations. Atomic charges were calculated using AM1, HF/3-21G*, B3LYP/3-21G*, MP2/3-21G* on Mulliken populations. Geometry optimizations were performed using AM1 or HF/3-21G*level calculations. The CoMFA models were successful in predicting activities of three new compounds. This study also shows that CoMFA models obtained with HF/3-21G* optimized geometries are more reliable in predicting activities of new compounds than CoMFA models that were derived from AM1 optimized geometries after the validation test.

CHAPTER 7

GENERAL CONCLUSIONS

A CoMFA study is composed of a training set and a test set. The training set is used of PLS study. The PLS study is composed of cross-validated analysis and noncross-validated analysis. The final result of PLS is noncross-validated analysis but the number of optimal components is needed for noncross-validated analysis. The number of components is decided by crossvalidated analysis. The final CoMFA model from these PLS steps should be validated using a test set. To compare calculational methods for CoMFA studies, Table 7.1 displayed the best and second result in cross-validated, noncross-validated PLS studies by training sets and validation steps by test sets.

Table 7.1 The Comparison of Calculational Methos by Chapters

	Cross-validated (q2,n) (Training Set)	Noncross-validated (R2, S) (Training Set)	Validation Test (Test Set)
Chapter 4	AM1//HF 0.569, 7	AM1//HF 0.989, 0.130	HF//HF
σ1	HF//HF 0.523, 6	HF//HF 0.977, 0.184	MP2//HF
Chapter 5	B3//HF 0.641, 3	B3//HF 0.906, 0.237	B3//HF
PCP σ1	MP2//HF 0.635, 3	MP2//HF 0.906, 0.238	MP2//HF
Chapter 6	MP2//HF 0.602, 5	AM1//AM1 0.952, 0.142	HF//HF, MP2//HF
σ2	HF//HF 0.599, 5	MP2//HF 0.951, 0.145	

Table 7.1 suggests that the best methods for obtaining most predictive ComFA maps are derived using HF/3-21G* optimized geometries and ab initio HF, MP2, or density functional B3LYP atomic charge calculations with a 3-21G* basi set. Sometimes AM1

optimized geometries or atomic charges work well but they are proved not as good methods by validation tests in Table 7.1.

The MP2 method was expensive but suggested almost the same result compared to HF calculations. The B3LYP calculation also took time but gave similar results compared to the HF method. Fom these studies it is suggest that the best calculation to use for CoMFA studies is HF//HF (using HF atomic charges and HF optimized geometries).

The automatic pharmacophore using DISCOtech showed good agreement with previously manually derived pharmacophores using similar types of ligands. These are shown in Figures 7.1 and 7.2. DISCOtech conformers using the Tripos field suggested good initial starting points for CoMFA studies. They were optimized using AM1 or HF.3-21G* methods. DISCOtech pharmacophores were used in aligning molecules. Aligned HF/3-21G* optimized geometries were successful to make proper CoMFA models to predict correct bioactivities for sigma receptor-ligands. These results are shown in Chapters 4 to 6.

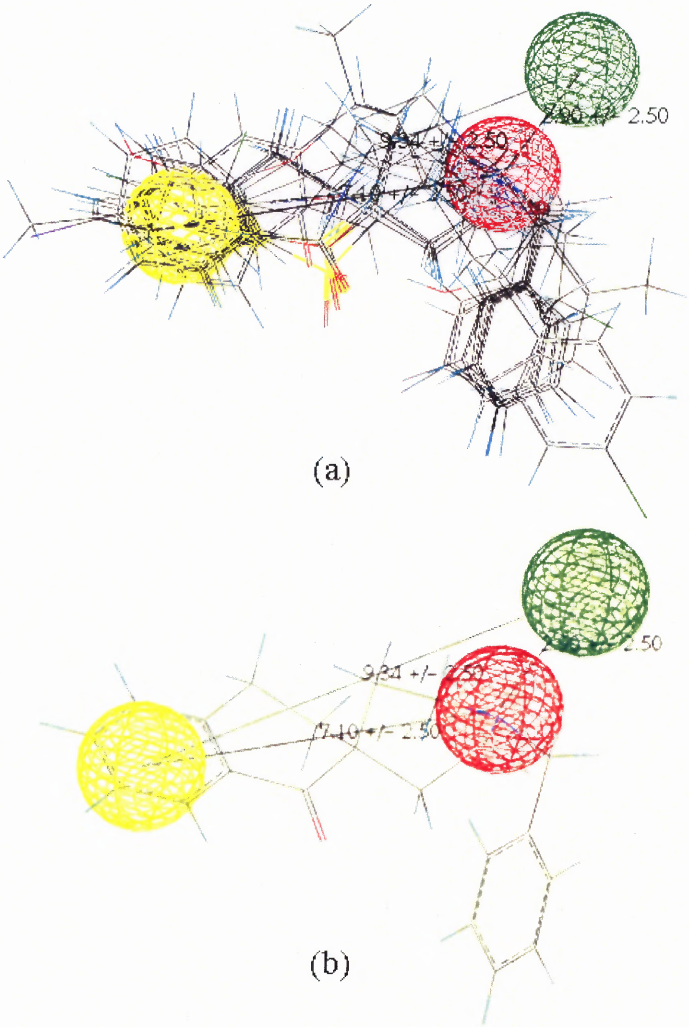
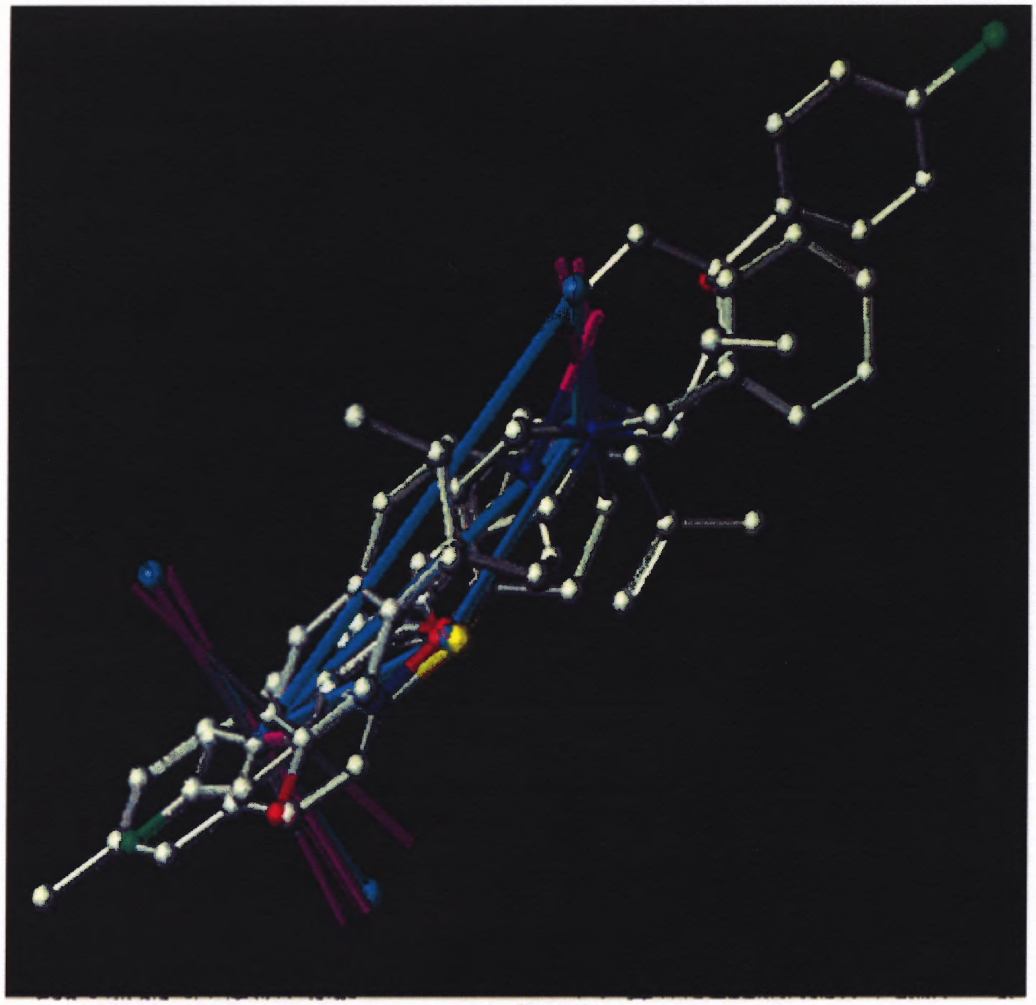
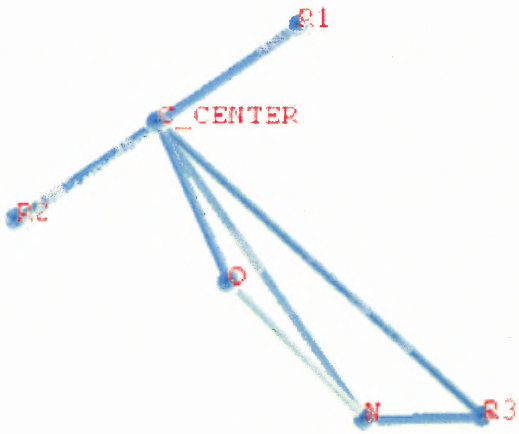


Figure 7.1 DISCOtech pharmacophore for Spipethiane and other ligands.

[a] all molecules in Table 4.1 and 4.4, [b] DISCOtech model with compound number 2 in Table 4.1.





Distances:
C_Center - N: 7.138 Å
C_Center - R3: 8.662 Å
O - N: 4.168 Å
Angles:
R1-C_Center-N: 90.21°
C-N-R3: 119.78°
O-C_Center: 3.682Å
C-O-N: 130.71°.
Torsion:
R1-C-N-R3: 12°

Figure 7.2 Manual Pharmacophore for sigma 1 receptor ligands for PD144418, Spipethiane, Haloperidol, and Pentazocine.

7.1 Suggestions for Further Work

The HF methods are expensive compared to semi-empirical calculations but are not so expensive compared to B3LYP calculations. BLYP methods with KMLYP is a proper calculational method to calculate bulky molecules. It can be applied in CoMFA theory with small basis set.

Unity is a database program in SYBYL6.9 [71]. It can be used to find new ligands which can fit our suggested pharmacophore by virtual screening. The database in Unity includes many commercially available ligands. These ligands could be tested for sigma activity and some could become new drugs.

DISCOtech pharmacophore was used as the initial point for derivation of CoMFA maps in this study. DISCOtech as a procedure for deriving pharmacophores is still in its infancy and pharmacophores derived by this method should be checked with manually derived pharmacophores. The conformations used by DISCOtech should also be checked with other calculations.

REFERENCES

1. Martin, W. R.; Eades, C. G.; Thompson, J. A.; Huppler R. E.; Gilbert, P. E. J Pharmacol Exp Ther 1976, 197, 517.

2. Martin, W. R.; Eades, C. G.; Thompson, J. A.; Huppler R. E.; Gilbert, P. E. J Pharmacol Exp Ther 1976, 197, 517.

3. Vaupel, D. B. Eur J Pharmacol 1983, 92, 269.

4. Quirion, R.; Chicheportiche, R.; Contreras, P. C.; Johnson, K. M.; Lodge, D.; Tam, S. W.; Woods, J. H.; Zukin, S. R.; Trends Neurosci 1987, 10, 444.

5. Maurice, T.; Lockhart, B. P. Prog Neuropsychopharmacol Biol Psychiatry 1997, 21, 69.

6. Maurice, T.; Phan, V. L.; Urani, A.; Kamei, H.; Noda Y.; Nabeshima, T.; Jpn J Pharmacol 1999, 81, 125.

7. Maurice, T.; Urani, A.; Phan V. L.; Romieu, P. Brain Res Rev 2001, 37, 116.

8. Walker, J. M.; Bowen, W. D.; Walker, F. O.; Matsumoto, R. R.; De Costa B.;and Rice, K. C. Pharmacol Rev 1990, 42 355.

9. Contreras, P. C.; Quirion R.; O'Donohue, T.L.; Neurosci Lett 1986, 67, 101.

10. Gundlach, A. L.; Largent B. L.; Snyder, S. H. J Neurosci 1986, 6, 1757.

11. Largent, B. L.; Gundlach, A. L.; Snyder, S. H. J Pharmacol Exp Ther 1986, 238, 739.

12. McLean, S.; Weber, E. Neuroscience 1988, 25, 259.

13. Tam, S. W.; Cook, L. Proc Natl Acad Sci USA 1984, 81, 5618.

14. Bowen, W. D. Pharm Acta Helv 2000, 74, 211.

15. Hellewell, S. B.; Bowen, W. B. Brain Res 1990, 527, 244.

16. Hellewell, S. B.; Bruce, A.; Feinstein, G.; Orringer, J.; Williams, W.; Bowen, W. B. Eur J Pharmacol 1994, 268, 9.

17. Quirion, R.; Bowen, W. D.; Itzhak, Y.; Junien, J. L.; Musacchio, J. M.; Rothman, R. B.; Su, T. P.; Tam, S. W.; Taylor, D. P. Trends Pharmacol Sci 1992, 13, 85.

18. Musacchio, J. M.; Klein, M.; Santiago, L. J. J Pharmacol Exp Ther 1988, 247, 424.

19. Itzhak, Y.; Stein, I. Life Sci 1990, 47, 1073.

20. Itzhak, Y. Mol Pharmacol 1989, 36, 512.

21. Monnet, F. P.; Blier, P.; Debonnel G.; de Montigny, C. Naunyn-Schmiedebergs Arch Pharmacol 1992, 346, 32.

22. Klein, M.; Cooper, T. B.; Musacchio, J. M. Eur J Pharmacol 1994, 254, 239.

23. Matsumoto, R. R.; Bowen, W. D.; Walker, J. M. Prog Clin Biol Res 1990, 328, 125.

24. Okuyama, S.; Imagawa, Y.; Sakagawa, T.; Nakazato, A.; Yamaguchi, K.; Katoh, M.; Yamada, S.; Araki H.; Otomo, S. Life Sci 1994, 55, PL133.

25. Taylor D. P.; Dekleva, J. Drug Dev Res 1987, 11, 65.

26. Taylor, D. P.; Eison, M. S;. Moon, S. L. Yocca, F. D. Adv Neuropsychiatr Psychopharmacol 1991, 1, 307.

27. Hanner, M.; Moebius, F. F.; Flandorfer, A.; Knaus, H. G.; Striessnig, J.; Kempner E.; Glossman, H. Proc Natl Acad Sci USA 1996, 93, 8072.

28. Jbilo, O.; Vidal, H.; Paul, R.; De Nys, N.; Bensaid, M.; Silve, S.; Carayon, P.; Davi, D.; Galiegue, S.; Bourrie, B.; Guillemot, J. C.; Ferrara, P.; Loison, G.; Maffrand, J. P.; Le Fur, G. Casellas, P. J Biol Chem 1997, 272, 27107.

29. Kekuda, R.; Prasad, P.D.; Fei, Y.J.; Leibach, F. H.; Ganapathy, V. Biochem Biophys Res Commun 1996, 229, 553.

30. Prasad, P. D.; Li, H. W.; Fei, Y. J.; Ganapathy, M. E.; Fujita, T.; Plumley, L. H.; Yang-Feng, T. L.; Leibach, F. H.; Ganapathy, V. J Neurochem 1998, 70, 443.

31. Pan, Y. X.; Mei, J.; Xu, J.; Wan, B. L.; Zuckerman, A.; Pasternak, G.W.; J Neurochem 1998, 70, 2279.

32. Seth, P.; Leibach, F. H.; Ganapathy, V. Biochem Biophys Res Commun 1997, 241, 535.

33. Mei, J.; Pasternak, G. W. Biochem Pharmacol 2001, 62, 349.

34. Fred Brouwer, Lecture note, Organic Chemistry, University of Amsterdam, (1995, Dec) http://www.caos.kun.nl/~borkent/compcourse/fred/ch5.html.

35. Levine, I. N. Quantum Chemistry, Prentice Hall, 4th ed., 1991.

36. Dykstra, C. E.; Augspurger, J. D.; Kirtman, B.; Malik, D. J., Reviews in Computational Chemistry, K.B. Lipkowitz and D.B. Boyd, eds., VCH, Vol. 1, 1990, 83.

37. Nàray-Szabò, G.; Surjàn, P. R.; Angyàn, J. G. Applied Quantum Chemistry, Reidel, 1987

38. Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. Ab initio molecular orbital theory, Wiley, 1986

39. Szabo and Ostlund, Modern Quantum Chemistry, McGraw-Hill, 1989.

40. Foresman, J. B.; Frisch, A. Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian, Gaussian Inc., 1993.

41. Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J., J Phys Chem, 1992, 96, 135-149.

42. Palmer, I. J.; Ragazos, I. N.; Bernardi, F.; Olivucci, M.; Robb, M. A., J Am Chem Soc 1993,115, 673-682.

43. Olivucci, M.; Ragazos, I. N.; Bernardi, F.; Robb, M. A., J Am Chem Soc 1993, 115, 3710-3721.

44. Roos, B. O.; Andersson, K.; Fülscher, M. P., Chem. Phys Lett, 1992, 192, 5-13.

45. Kohn, L. J. Self-Consistent Equations Including Exchange and Correlation Effects, Phys Rev, 140A, 1133 (1965).

46. Spartan User's Guide, version 3.0, Wavefunction, Inc., 1993.

47. Hehre, W. J.; Stewart, R. F.; Pople, J. A. J Chem Phys 1969, 51, 2657.

48. Binkley, J. S.; Pople, J. A.; Hehre, W. J. J Chem Soc 1980, 102, 939.

49. Hehre, W. J.; Ditchfield, R. Pople, J.A. J Chem Phys 1972, 56, 2257.

50. Krishnan, R.; Frisch, M. J.; Pople, J. A. J Chem Phys 1980, 71,4244.

51. Hariharan, P. C.; Pople, J. A. Chem Phys Lett 1972, 66, 217.

52. Richon; A. B. Yoing; S. S. Network Science Corporation, 412 Carolina Blvd, Isle of Palms, SC 29451 (1997, Oct) http://www.netsci.org/Science/Compchem/feature19.html.

53. Stewart, J. J. P., Reviews in Computational Chemistry, Lipkowitz, K. B.; Boyd, D. B. eds., VCH, Vol. 1, 1990, 45-82.

54. Walpole, C. S. J.; Wrigglerworth, R.; Bevan, S.; Campbell, E. A.; Dray, A.; James, I. F.; Masdin, K. J.; Perkins, M. N.; Winter, J. J Med Chem 1993, 36, 2381.

55. Rogers, D.; Hopfinger, A. J. J Chem Inf Comp Sci 1994, 34, 854-866.

56. Gould, R. F. (ed.), Biological Correlations -- The Hansch Approach, Advances in Chemistry Series, No. 114, American Chemical Society, Washington, D.C., 1972.

57. Martin, Y. C. Quantitative Drug Design, Marcel Dekker, New York, (1978).

58. Topliss, J. G. J Med Chem 1972, 15, 1006.

59. Wold, S. Pattern Recognition 1976, 8, 127.

60. Stuper, A. J.; Brugger, W. E.; Jurs, P. C. Chemometrics: Theory and Application, B.R. Kowalski (Ed.), American Chemical Society, Washington, D.C., 1977.

61. Klopman, G.; Dimayuga, M. I. J Comput-Aided Mol Design, 1990, 4, 117.

62. Ling, R. F (1973), "A Probability Theory of Cluster Analysis," Journal of the American Statistical Association, 68, 159-169.

63. Jurs, P. C. Chemometrics and Multivariate Analysis in Analytical Chemistry, in Reviews in Computational Chemistry, Volume 1, K.B. Lipkowitz and B.B. Boyd (Eds.), VCH Publishers, Inc., New York, 1990.

64. Jurs, P. C. J.; Chou, T.; Yuan, M. J Med Chem 1979, 22, 476.

65. Kier, L. B. Molecular Orbital Theory in Drug Research, Academic Press, New York, NY, 1971.

66. Crippen, G. M. J. Med. Chem., 1979, 22, 988.

67. Mabilia, M.; Pearlstein, R. A.; Hopfinger, A. J. Computer Graphics in Molecular Shape Analysis, in Molecular Graphics and Drug Design, Burgen, A. S. V.; Roberts G. C. K.; Tute, M. S. (Eds.), Elsevier Science Publishers, Amsterdam, 1986.

68. Weinstein, H. Chemical Applications of Molecular Electrostatic Potentials, Politzer, P; Truhlar, D. G. (Eds.), Plenum Press, New York, NY, 1981.

69. Potenzone, R.; Cavicchi, E.; Weintraub, H. J. R.; Hopfinger, A. J Comput Chem, 1997, 1, 187

70. Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug Design: The Active Analog Approach, in Computer Assisted Drug Design, ACS Symposia, 112, Olson, E. C.; Christofferson, R. E. (Eds.), American Chemical Society, Washington D.C., 1979.

71. SYBYL Version 6.9, Tripos Associates, Inc.: St. Louis, MO, 2002.

72. Hopfinger, A. J.; J Amer Chem Soc, 1980, 102, 7196.

73. Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. J Amer Chem Soc, 1988, 110, 5959..

74. Martin, Y.; Bures, M.; Dahaner, E.; DeLazzer, J.; Lico, I.; Pavlik, P. J Computer-Aided Mol Design 1993, 7, 83-102.

75. Marriott, D. P.; Dougall, I. G.; Meghani, P.; Liu, Y.; Flower, D. R. J Med Chem 1999, 42, 3210.

76. Bondavalli, F.; Botta, M.; Bruno, O.; Ciacci, A.; Corelli, F.; Fossa, P.; A. Lucacchini,; Manetti, F.; Martini, C.; Menozzi, G.; Mosti, L.; Ranise, A.; Schenone, S.; Tafi, A.; Trincavellic, M. L. J Med Chem 2002, 45, 4875.

77. Fossa, P.; Mosti, L.; Menozzi, G.; Manetti, F.; Dorigo, P.; Floreani, M.; Med Chem Res 2002, 11, 137.

78. Lamothe, M.; Pauwels, P. J.; Belliard, K.; Schambel, P.; Halazy, S.; J. Med. Chem. 1997, 40 3542.

79. Gaussian 98, Revision A.9, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, Jr. 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.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; 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.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.;

Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2001.

80. Abou-Gharbia, M.; Ablordeppey, S. Y.; Glennon, R. A. Ann Reports in Med Chem 1993, 28, 1.

81. Maeda, D. Y.; Williams, W.; Kim W. E.; Thatcher, L. N.; Bowen, W. D.; Coop, A. BIOORG AND MED CHEM LETT 2002, 12(3), 497.

82. Maier, C. A.; Wuensch, B. Novel J Med Chem 2002, 45(22), 4923.

83. Ronsisvalle, G.; Prezzavento, O.; Marrazzo, A.; Vittorio, F.; Massimino, M.; Murari, G.; Spampinato, S. J Med Chem 2002, 45(12), 2662.

84. Horan, B.; Gardner, E. L.; Dewey, S. L.; Brodie, J. D.; Ashby, C. R. Eur J Pharm 2001, 426(1,2), R1.

85. Maj, J.; Rogoz, Z.; Skuza, G.; Mazela, H. Eur J Pharm 1996, 315, 235.

86. Ogawa, S.; Okuyama, S.; Araki, H.; Otomo, S. Eur J Pharm 1994, 263, 9.

87. Okuyama, S.; Imagawa, Y.; Ogawa, S.; Araki, H.; Ajima, A.; Tanaka, M.; Muramatsu, M.; Nakazato, A.; Yamaguchi, K.; Yoshida, M.; Otomo, S. Life Sci 1993, 53(18), 285.

88. Okuyama, S.; Nakazato, A. CNS Drug reviews 1996, 2(2), 226.

89. Shepard, P. D.; Lehmann, H.; Connelly, S. T. Eur J Pharm 1994, 265, 141.

90. Brown, M. L.; Zha, C. C.; Van Dyke, C. C.; Brown, G. B.; and Brouillette, W. J. J Med Chem 1999, 42, 1537.

91. Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. J Am Chem Soc 1988, 110, 5959.

92. Debnath, A. K. J Med Chem 1999, 42, 249.

93. Gaillard, P.; Carrupt, P. A.; Testa, B.; Schambel, P. J Med Chem 1998, 39, 134.

94. Marot, C.; Chavatte, P.; Morin-Allory, L.; Viaud, M. C.; Guillaumet, G.; Renard, P.; Lesieur, D.; Michel, A. J Med Chem 1998, 41, 4453-4465.

95. Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Lonze, P.; Weithmann, K. U. J Med Chem 1999, 42, 1908.

96. Shim, J. Y.; Collantes, E. R.; Welsh, W. J. J Med Chem 1999, 41, 4521.

97. Wilcox, R. E; Tseng, T.; Brusniak, M.-Y. K.; Ginsburg, B.; Pearlman, R. S.; Teeter, M.; DuRand, C.; Starr, S.; Neve, K. A. J Med Chem 1998, 41, 4385.

98. Ablordeppey, S. Y.; El-Ashmaw, M.; Fischer, J. B.; Glennon, R. A. Eur. J Med Chem 1998, 33, 625.

99. Bursi, R.; Grootenhuis, P. D. J. J Computer-Aided Mol Design 1999, 13(3), 221.

100. Tonmunphean, S.; Kokpol, S.; Parasuk, V.; Wolschann, P.; Winger, R. H.; Liedl, K. R.; Rode, B. M. J Computer-Aided Mol Design 1998, 12(4), 397.

101. Grunewaled, G. L.; Caldwell, T. M.; Dahanukar, V. H.; Jalluri, R. K.; Criscione, K. R. Bioorg Med Chem Lett 1999, 9, 481.

102. Baurin, N.; Vangrevelinghe, E.; Morin- Allory, L.; Ménour, J.-Y.; Renard, P.; Payard, M.; Guilaumet, G.; Marot, C. J. Med. Chem. 2000, 43, 1109.

103. Hannongbua, S.; Nivesanond, K.; Lawtrakul, L.; Pungpo, P.; Wolschann, P. J Chem Inform and Comp Sci 2001, 41(3), 848.

104. Quaglia, W.; Giannella, M.; Piergentili, A.; Pigini, M.; Brasili, L.; Toro, R. D.; Hudkins, R. L.; Mailman, R. B.; DeHaven-Hudkins, D. L. J Med Chem 1994, 37, 1964.

105. Ucar, H.; Cacciaguerra, S.; Spampinato, S.; Van derpoorten, K.; Isa, M.; Kanyonyo, M.; Poupaert, J. H. Eur J Pharm 1997, 355, 267.

106. Akunne, H.C.; Whetzel, S.Z.; Wiley, J. N.; Corbin, A.E.; Ninteman, F.W.; Tecle, H.; Pei, Y.; Pugsley, T.A.; Heffner, T. G. Neuropharm 1997, 36(1), 51.

107. Shiue, C. Y.; Shiue, G. G.; Zhang, S. X.; Wilder, S.; Greenberg, J. H.; Benard, F.; Wortman, J. A.; Alavi, A. A. Nuc Med & Bio 1997, 24, 671.

108. Waterhouse, R. N.; Mardon, K.; Giles, K. M.; Collier, T. L.; O'Brien Joanne, C. J Med Chem 1997, 40, 1657.

109. Gund, T. M.; Shukla, K.; Su, T. P. J Math Chem 1991, 8(4), 309.

110. Kenny, P.W. J Chemical Soc Perkin Tran 2 1994, 2, 199.

111. Bush, B.L.; Nachbar, R.B. J Computer-Aided Mol Design 1993, 7, 587.

112. Clark, M.; Cramer III, R.D. Quant Struct-Act Relat 1993, 12, 137.

113. Su, T. P.; Wu, X. Z.; Cone, E. J.; Shukla, K.; Gund, T. M.; Dodge, A. L.; Parish, D. W.; J Pharmacol Exp Ther 1991, 259, 543.

114. Urani, A.; Romieu, P.; Roman, F. J.; Maurice, T. Behaviour Brain Res 2002, 134, 239.

115. Maurice, T.; Su, T. P.; Parish, D. W.; Nabeshima, T.; Privat, A. Pharmacol. Biochem. and Behav. 1994, 49, 859.

116. Maurice, T.; Roman, F. J.; Su, T. P.; Brain Res. 1996, 733, 219.

117. Phan, V.; Urani, A.; Sandillon, F.; Privat, A.; Maurice T. Neurobiology of Aging, Accepted for publication, Available online 21 January (2003)

118. Maurice, T. Roman, Su , T. P.; Privat, A. European Journal of Pharmacology 2001, 431, 223.

119. Hayashi, T.; Maurice, T.; Su , T. P. J. Pharmacol. Exp. Ther. 2000, 293. 788.

120. Maurice, T.; Martin-Fardon, R.; Romieu, P.; Matsumoto, R. R. Neurosci. & Biobehav. Review. 2002, 26, 499.

121. Romieu, P.; Phan, V.; Martin-Fardon, R.; Maurice, T. Neuropsychopharmacol 2002, 26, 444.

122. Vilner, B. J.; John, C. S.; Bowen ,W. D. Cancer Res 1995, 55, 408.

123. Brent, P. J.; Pang, G. T. Eur J Pharmacol 1995, 278, 151.

124. Crawford, K. W.; Bowen, W. D., Cancer Res. 2002, 62, 313.

125. Hudkins, R. L.; Mailman, R. B.; DeHaven-Hudkins, D. L. J Med Chem 1994, 37, 1964.

126. Efange, S. M. N.; Mach, R. H.; Smith, C. R.; Khare, A. B.; Foulon, C.; Biochem Pharmacol 1995, 49, 791.

127. Zhang, Y.; Williams, W.; Bowen, W. D.; Rice, K. C.; J Med Chem 1996, 39 3564.

128. Nguyen, V. H.; Kassiou, M.; Johnston, G. A. R.; Christie, M. J.; Eur J Pharmacol 1996, 311, 233.

129. Yamashita, A.; Takahashi, N.; Mochizuki, D.; Tsujita, R.; Yamada, S.; Kawakubo, H; Suzuki, Y; Watanabe, H. Bioorg & Med Chem Lett 1997, 7, 2303.

130. Wayne, D.; Bowen, C. M.; Bertha, B. J. V.; Kenner, C.R. Eur J Pharmacol 1995, 278, 257.