ABSTRACT

2D QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP MODELING OF METHYLPHENIDATE ANALOGUES USING GENETIC ALGORITHM AND PARTIAL LEAST SQUARE REGRESSION

by
Noureen Wadhwaniya

Quantitative Structure-Activity Relationship (QSAR) analysis attempts to develop a predictive model of biological activity based on molecular descriptors. 2D QSAR uses descriptors, such as topological indices, that are independent of molecular conformation. A genetic algorithm - partial least squares (GA-PLS) approach was used to identify the molecular descriptors that correlate to the biological activity (binding affinity) of a set of 80 methylphenidate analogues and to construct a predictive model. The GA code was implemented using the fitness function $(1 - (n - 1)(1- q^2)/ (n - c))$, where n is the number of compounds, c is the optimal number of components, and q^2 is the cross-validated regression coefficient. Partial Least Squares Regression was then applied to the selected descriptors to create a predictive model of biological activity ($q^2 = 0.78$, fitness = 0.77). This model can be used to assist in the design of improved methylphenidate analogues for the treatment of cocaine abuse. The GA-PLS program was tested on the benchmark Selwood dataset of antifilarial antimycin analogues and identified several molecular descriptors in common with other 2D QSAR models.

2D QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP MODELING OF METHYLPHENIDATE ANALOGUES USING GENETIC ALGORITHM AND PARTIAL LEAST SQUARE REGRESSION

by
Noureen Wadhwaniya

A Master's Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computational Biology

Department of Computer Science

January 2005

APPROVAL PAGE

2D QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP MODELING OF METHYLPHENIDATE ANALOGUES USING GENETIC ALGORITHM AND PARTIAL LEAST SQUARE REGRESSION

Noureen Wadhwaniya

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

Dr. Michael L. Recce, Committee Member — Date
Associate Professor of Computer and Information Science, NJIT

Dr. Qun Ma, Committee Member — Date
Assistant Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Noureen Wadhwaniya

Degree: Master of Science

Date: January 2005

Undergraduate and Graduate Education:

- Master of Science in Computational Biology
  New Jersey Institute of Technology, Newark, NJ, 2005

- Bachelor of Engineering in Computer Engineering
  University of Mumbai (Bombay), India, 2003

Major: Computational Biology

To my father, for making me strive for excellence.
To my mother, for her belief in me.
To my sisters, for their love and support.
To my brother, for taking pride in me.

ACKNOWLEDGMENT

I would like to express my deep gratitude to Dr. Carol A. Venanzi, for serving as my thesis advisor. Her support and guidance made this thesis possible. Special thanks are due to Dr. Michael Recce and Dr. Qun Ma for actively participating in my committee and assisting me in completing my thesis.

I would like to thank everyone in Dr. Venanzi's research lab in the Department of Chemistry and Environmental Science for their help and insight into this project. I wish to especially thank Milind Misra for all of the data and help he provided me.

Additional thanks are due to Dr. Howard Deutsch (Georgia Institute of Technology) and Dr. Margaret Schweri (Mercer University School of Medicine) for providing data for this thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

CHAPTER 1

INTRODUCTION

Cocaine abuse and addiction continues to be a problem that plagues our society. In 2002, an estimated 1.5 million Americans could be classified as dependent on or abusing cocaine, according to the National Survey on Drug Use and Health.[1] The high social and economic costs associated with the treatment of cocaine abuse are a motivation for the development of an effective therapeutic drug for the treatment of cocaine dependence.

Cocaine in the brain binds to the dopamine transporter blocking the reuptake of dopamine. The accumulation of dopamine is believed to produce a feeling of elation. A therapeutic agent for the treatment of cocaine dependence would be a selective dopamine reuptake inhibitor in that it would exhibit a high binding affinity for the dopamine transporter and simultaneously permit some degree of dopamine reuptake without the addictive side effects. Methylphenidate (MP) analogues exhibit these characteristics and are believed to be potential therapeutic drugs for the treatment of cocaine abuse.

The design of such therapeutic MP analogues with desired properties and biological activity is a challenging task. The traditional approach requires a trial and error procedure involving synthesis and testing of a large number of potential candidate molecules. This is a laborious, time consuming and expensive process. Therefore there is an incentive to develop computer-aided molecular design methods that can be useful in the design of molecules with improved bioactivity.

1.1 Objective

The broad objective of this thesis is to aid in developing a model that predicts the biological activity of a MP analogue given its structure; that is, to develop a Quantitative Structure-Activity Relationship (QSAR) model of MP analogues. This is important because MP analogues have a similar mechanism of action to that of cocaine [2-5] and is believed to be of therapeutic use for cocaine abuse. The purpose of this study is to identify the two-dimensional topological and electrotopological state (E-state) descriptors that significantly correlate changes in the molecular structure to the changes in the biological activity (binding affinity) of the MP analogues. This is done by using genetic algorithm (GA) for selection of the molecular-level variables that describe the structure. A program implementing GA to develop a predictive model based on the most significant descriptors was developed. The partial least square regression (PLSR) method was applied to the selected descriptors in order to identify a predictive model for biological activity. The model was tested on the benchmark Selwood antifilarial antimycin analogues dataset.

This 2D QSAR approach is in contrast to 3D QSAR studies, where 3D steric and electrostatic descriptors that depend on the molecular conformation are utilized. A large number of conformers are generated and a choice of a representative conformer is made before the analysis. In contrast, 2D QSAR methods like GA-PLS rely only upon 2D topological descriptors of chemical structures and are computationally less taxing than 3D QSAR methods. The 2D QSAR model of MP analogues developed here may help in designing a more effective drug to treat cocaine dependence.

1.2 Background Information

1.2.1 Cocaine

Cocaine, $C_{17}H_{21}NO_4$ (Figure 1.1) is among one of the most heavily abused stimulant drugs producing euphoria, alertness, excitement and rapid flow of thoughts. Its devastating effects include severe psychological disturbances, paranoia, auditory hallucinations and cardiac arrhythmias. The duration of cocaine's immediate euphoric effects depends on the dosage and the route of administration.

The drug induces a sense of exhilaration in the user primarily by blocking the reuptake of the neurotransmitter dopamine (DA) in the brain.[6]

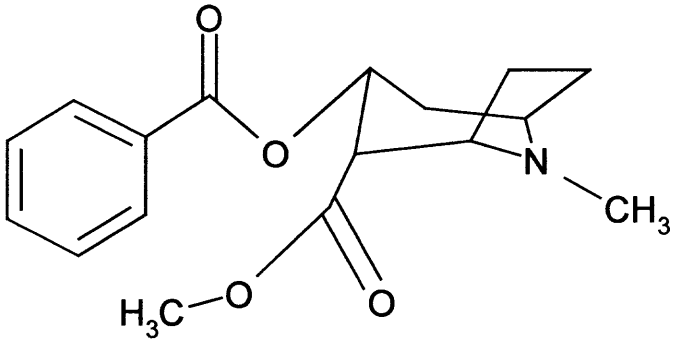


Figure 1.1 Cocaine.

The mesolimbic pathway (reward pathway) of the brain is considered to be the 'pleasure center' of the brain and dopamine a 'pleasure neurotransmitter'.[7] The mesolimbic dopamine neurons are naturally triggered during accomplishments or victories to release dopamine producing a feeling of 'natural high'. In the normal communication process, dopamine is released by a neuron into the synapse (the small gap between two neurons), where it binds to specialized proteins (called dopamine receptors) on the neighboring neuron, thereby sending a signal to that neuron.

Drugs of abuse like cocaine are able to interfere with this normal communication process. According to the "Dopamine Hypothesis",[8] cocaine binds to the dopamine transporter (DAT) blocking the reuptake of dopamine from the synapse, resulting in an accumulation of dopamine. This buildup of dopamine causes continuous stimulation of receiving neurons, which is associated with the euphoria commonly reported by cocaine abusers.[9]

One approach to find an effective treatment for cocaine abuse is to develop an antagonist of cocaine action, which does not block the reuptake of dopamine. But the drawback of this approach is that it does not reduce the craving for cocaine and the patient could annul the effect by simply administering more cocaine. Another approach is to develop a non-competitive inhibitor of cocaine that selectively and strongly binds to but dissociates slowly from the DAT. By partially inhibiting DA reuptake the ideal agent would provide sufficient DA to minimize cocaine craving, yet insufficient to produce euphoria.

1.2.2 The Dopamine Transporter

The dopamine transporter is a 12 membrane-spanning protein located on the plasma membrane of nerve terminals. The DAT is believed to contain a specific binding site for cocaine.[10]

Dopamine plays an important role in the control of movement, cognitive functions, and neuroendocrine systems. The dopamine transporter is a membrane-bound protein that functions to release dopamine into presynaptic terminals. The dopamine transporter is dependent on the presence of Na^+ and Cl^- in the extracellular fluid. Though substances such as cocaine can inhibit dopamine, norepinephrine and serotonin reuptake,

it is thought that the dopamine reuptake inhibition is responsible for the euphoric effect of cocaine.

1.2.3 Methylphenidate Analogues

Methylphenidate (MP, Figure 1.2) analogues are believed to be potential cocaine abuse therapeutic drugs. MP (Ritalin®) has a similar mechanism of action to that of cocaine.[2-5] It is already prescribed to children with Attention Deficit Hyperactivity Disorder (ADHD) and has limited abuse potential.

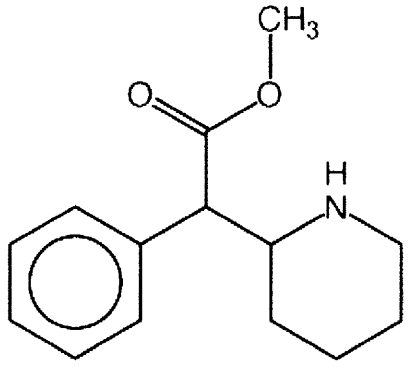


Figure 1.2 Methylphenidate.

1.2.4 QSAR Studies

QSAR modeling is a branch of Chemistry that attempts to use statistical modeling principles to estimate biological activity of molecules. The QSAR work of Hansch [11] initiated the use of computer based technology in the discovery, design and development of pharmaceutical agents. The QSAR methodology assumes that the change in the biological activity that is observed within a series of similar compounds is a function of the changes in chemical structure within the series.

An example of a QSAR data set correlation analysis is given below. The X-data represents the dependent descriptor matrix and Y-data the independent biological activity

matrix. In the example below there are L compounds, M different biological activity types and N descriptors.

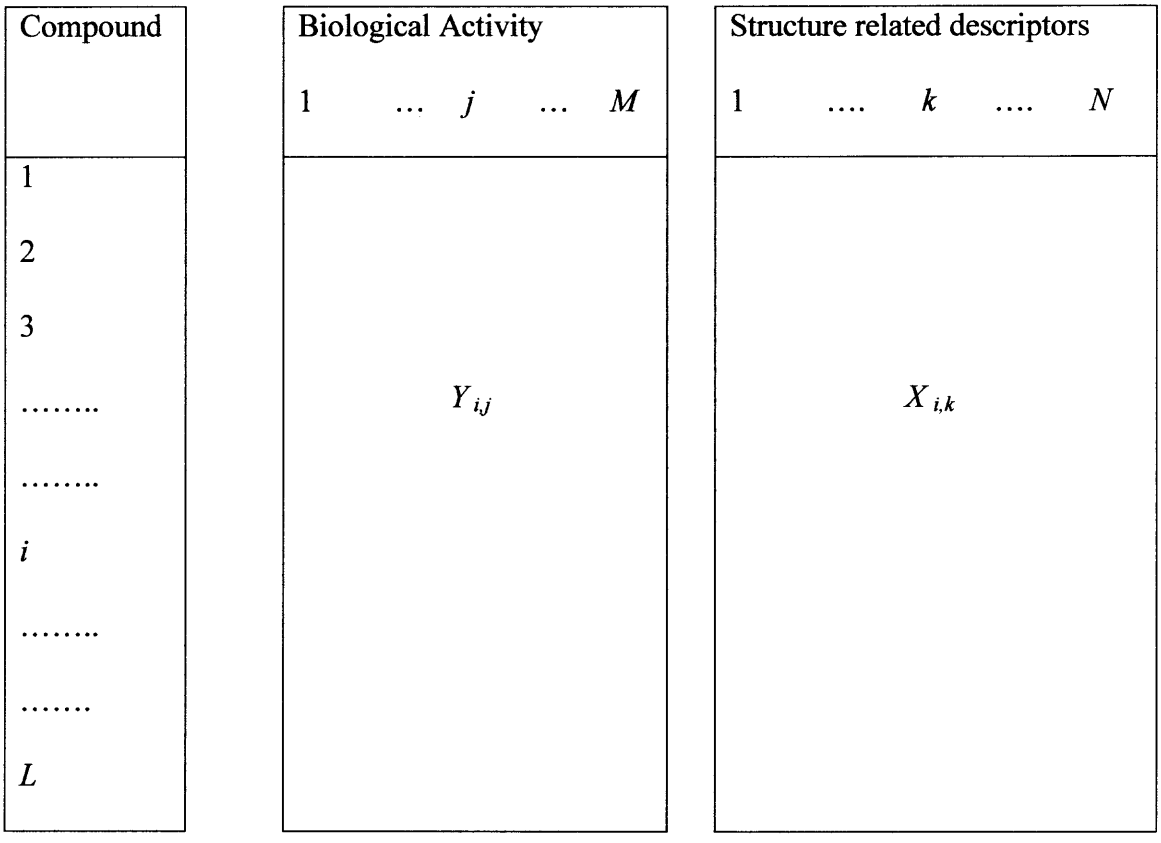


Figure 1.3 Standard QSAR data matrix.

The aim of a QSAR study is to find a predictive relationship between the independent X-data and dependent Y-data for the data set, or in the context of present study, to develop a model that relates the biological activity to the structural descriptors.

CHAPTER 2

THEORY

2.1 Genetic Algorithm

Search algorithms look to optimize a function in the search space by selecting sample points. There are basically three types of search methods viz., calculus-based, enumerative schemes and random searches. Calculus-based methods employ direct or indirect techniques to look for a point in the search space where the function is maximized. However this method assumes that the slopes are well defined and lacks robustness. [12, 13]

Enumerative schemes perform an exhaustive search by looking at each point in the specified space. These schemes look at all the points in the space and are thus inefficient by increasing the computational overhead.[12] Guided random searches like genetic algorithm (GA) use random choice to initiate the search by choosing a random set of points in the search space and evaluating the optimization function.

Genetic algorithms [14] are search algorithms based on the Darwinian principles of natural selection and natural genetics.[12] They make use of natural processes like selection, crossover and mutation of a population of strings. The strings are evaluated based on a fitness function that plays the role of the environmental pressure of Darwinian evolution. The parent strings are selected in such a way that the fittest individuals get more reproductive chances. In every new generation, a new set of individuals (strings) is created using parts of the fittest of the old generation and performing random mutation with a low probability. The less fit parents are replaced by fitter offspring. By repeating

the process over a number of generations it is expected that the fitness of the population as a whole keeps improving.

The GA approach is computationally simple yet powerful in the search for improvement because it does not restrict the search space. It codes the parameter set, searches from a population of points, uses an objective function to evaluate the fitness of points, and uses probabilistic transition rules to obtain a new generation of individuals. It does not use a point-to-point method of transition because a false optimum of the fitness function may be located but instead it uses a rich database of points.

The GA tries to include the adaptive processes of natural systems into artificial systems. In the present study the molecular structure is described by a number of variables (i.e. descriptors) obtained from a software program. In the present application GA is to select those variables that are the most influential in affecting the molecule's biological activity. In the computer program implementation of GA a binary digit represents the presence or absence of a particular descriptor. The program works by looking for binary strings with high fitness value and tries to find a set of descriptors that are more significant in predicting the molecule's activity by evolving a population of strings. The fitness of a particular string is evaluated by the Partial Least Squares regression method.

The GA is used for variable selection of descriptors as it picks descriptors randomly at first but later fine-tunes its selection by choosing descriptors that yield a higher fitness value. In its search for a better optimization of the fitness function, GA first chooses random points in space but once it finds a relatively good measure, it narrows the search area.

A simple genetic algorithm consists of three main procedures:

1. Reproduction
2. Crossover
3. Mutation

2.1.1 Reproduction

In the process of reproduction strings are copied based on their fitness values. Strings (individuals) for mating may be selected randomly or by using a technique that is biased towards selecting fitter individuals. Strings with a higher fitness value have a higher probability of contributing to the next generation. The selection process for mating may be implemented using a roulette wheel that is biased towards fitter individuals. In the roulette wheel selection process, each individual's chance of being selected for mating is directly proportional to its fitness.

Consider an example of four strings in an initial population where the fitness is defined by the square of the string's decimal equivalent.

Table 2.1 Fitness Table [12]

No.	String	x	Fitness (x^2)	% of total fitness
1	01101	13	169	14.4
2	11000	24	576	49.2
3	01000	8	64	5.5
4	10011	19	361	30.9
		Total =	1170	100.0

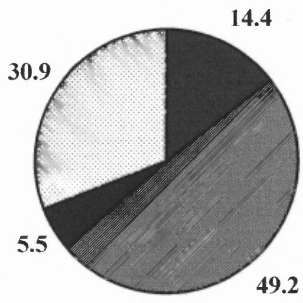


Figure 2.1 Roulette wheel.

As seen from the roulette wheel above, fitter individuals have more area on the wheel and hence, a higher chance of being selected. A simple spin of the weighted roulette wheel yields a reproduction candidate. Therefore highly fit strings have a higher number of offspring in succeeding generations.

Individuals for mating may also be selected by a scheme called 'Tournament selection'. In this scheme, m individuals from a total population size of n ($m<n$) are selected. The fitness of all individuals is compared and the fittest one wins the tournament. The fittest of the m individuals is chosen as a mate.

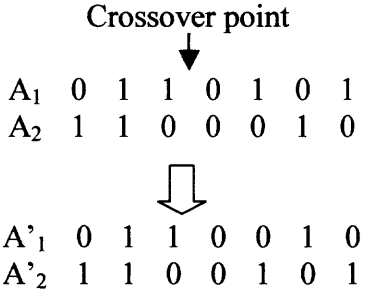
For the current application, individuals for mating are chosen randomly. Roulette wheel selection was not implemented as the fitness of the strings as obtained from Partial Least Squares Regression maybe negative. A population where many individuals' fitness is negative yields a negative value for the sum of all fitness. In such a case implementing roulette wheel becomes difficult as the fitness values vary over a wide range. Also, Tournament Selection method picks the winner of a tournament based on the fitness

values. Therefore, it never picks the least fit individual for mating and the individual is never replaced in the population.

2.1.2 Crossover

Crossover is the next step in the algorithm where strings are crossed at a random position to create two new strings (offspring). This follows the natural process where each parent's chromosomes contribute to create new individual, which has genetic material from both the parents.

Consider an example of a one-point crossover where two strings undergo crossover at a single position to create two new strings.



2.1.3 Mutation

Mutations induce sporadic and random alterations in the genetic material. Mutation plays a secondary role in the process of evolution. A binary string mutation just flips one or more of its bits to its complement. Even though reproduction and crossover search and combine extant notions, they may lose potentially useful genetic information. Mutation insures against such premature loss of important notions.

Both crossover and reproduction are fairly simple and computationally trivial operations but their combined emphasis makes GA a powerful technique. The natural

selection procedure is not purely based on chance but is guided by directed serendipity. It builds new solutions from the best partial solutions of previous trials.

Consider strings coded so that each is a complete idea or guide to performing a particular task and substrings contain notions of what is important or relevant to the task. Therefore a population of *n* strings contains not just a sample of *n* ideas but also a multitude of notions and rankings of notions for task performance.

GA makes maximum use of information by reproducing high quality notions and crossing these notions with other high performance notions from other strings. Thus the exchanging of notions to form new ideas leads to innovative ideas. Innovation is a juxtaposition of things that have worked well in the past. Reproduction and crossover are analogous to exchange of ideas and notions to come up with more innovative ideas that yield better performance.

2.1.4 Schema Theory

The Schema theory and building block hypothesis describe the basic mechanism of GA[12,14]. The aim behind using GA in a simulated environment is to look for a particular point(s) in space where the function is at its optimum value. In order to fine-tune the search, the space around the point where the best result has been achieved so far is scanned. In terms of strings, relationships are sought between similarities among strings with high fitness in the population. A schema or similarity template describing a subset of strings with similarities at certain string positions is devised.

A schema may be represented as a pattern-matching device. For instance a schema consisting of binary digits, 0 and 1 and a wild card symbol (*) can be given as,

{1*0*} = {1000, 1100, 1001, 1101}

For a string of length 4 at each position there are three choices (0,1, *). Hence, there are a total of $3^4 = 81$ combinations. A schema in strings with high fitness values are searched for. That is, a high fitness value is associated with a certain pattern in the string. When such useful schemas are passed from one generation to another via reproduction, the new population consists of fitter individuals.

Schemata may be destroyed by the process of crossover where parents' strings are crossed to create new individuals consisting of genetic material of both the parents. Certain schemata are more likely to propagate than others depending upon their lengths and the number of fixed positions in the schemata. For instance a schema like **11* is less likely to be disrupted by crossover then a schema like 0***1.

The GA promotes highly fit individuals with short length schemata to proliferate. Those schemata that encourage fitter individuals form the building blocks to build a new generation.

For a string of length l at each position there are three choices (0,1, *). Hence, there is a possibility of total 3^l schemata in all. For a representation of cardinality of k, there is a total of $(k+1)^l$ schemata.

Consider two schemata H_1 and H_2 given by,

H_1 = *1****0

H_2 = ***10**

The order of schema H, O(H) is defined as the number of fixed positions in the template.

$O(H_1) = 2$

$O(H_2) = 2$

The defining length of schema H, $\delta(H)$ is the distance between the first and the last position specified.

$\delta(H_1) = 7 - 2 = 5$

$\delta(H_2) = 5 - 4 = 1$

According to the Schema Theorem, the fundamental theorem of GA, reproduction allocates an exponentially increasing number of trials to the above average schemata and decreasing trials to the below average schemata. Simply stated, crossover creates new individuals (strings) with minimum disruption to favorable schemata in a population.

Consider the above example of two schemas H_1 and H_2. If A is a chosen mate given as,

$A = 0\ 1\ 1\ 1\ 0\ 0\ 0$
$\quad\ 1\ 2\ 3\ 4\ 5\ 6\ 7$

Crossover point

String A represents both the schemas H_1 and H_2. The crossover point destroys the schema H_1 but the schema H_2 is untouched. Schema H_1 is less likely to survive crossover and propagate to the next generation than schema H_2 because on average crossover point is more likely to separate two distant fixed positions.

In the above string A of length, $l = 7$ there are l-1 (i.e. 7-1 =6) possible crossover sites. The probability of a schema being destroyed by crossover and the probability of a schema surviving is given as follows.

For schema H_1,

Probability of destruction = $P_d(H_1)$
$= \delta(H_1) / (l\text{-}1)$
$= 5/6$

Survival probability = $1 - P_s(H_1)$
$= 1 - 5/6$
$= 1/6$

For Schema H_2,

Probability of destruction = $P_d(H_2)$
$= \delta(H_2) / (l-1)$
$= 1/6$

Survival probability = $1 - P_s(H_1)$
$= 1 - 1/6$
$= 5/6$

Therefore, schemas with short defining length (δ) are more likely to propagate.

2.2 Partial Least Squares Regression

The aim of a regression problem is to model one or more dependent variables (i.e. responses, Y) by means of a set of predictor variables (X). In this study, the biological activity, IC_{50}, is the inverse of the binding affinity of the methylphenidate (MP) analogues to the cocaine binding site on the DAT. A low value of IC_{50} corresponds to high binding affinity. The IC_{50} values of all MP analogues represent the Y variable. The IC_{50} values are related to the chemical structure of MP analogues. The structures of MP analogues are coded by the descriptor values derived from the Molconn-Z module in SYBYL®[15] software and stored as X-variables. These are called QSAR (Quantitative Structure Activity Relationship) models as an attempt is made to relate X, the quantitative description of variation in the structure of the investigated molecules to Y, their biological activity.

Partial Least Squares (PLS) is a regression technique developed by H. Wold[16] and is used to for modeling linear relationships between multivariate structure-activity measurements.[17] PLSR uses a two-block predictive PLS model to model the relationship between two matrices, X and Y. In its simplest form, it specifies the (linear) relationship

between a dependent (response) variable Y, and a set of predictor variables, the X's, so that

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_pX_p \quad (1)$$

where: b_0 = the regression coefficient for the intercept

b_i = the regression coefficients computed from the data.

The reason MLR (Multiple Linear Regression) cannot be used for QSAR applications is because it cannot analyze strongly collinear (correlated) data or noisy data and cannot handle numerous X-variables and model several Y responses simultaneously. The QSAR table usually contains hundreds of descriptors (X-variables), which are highly collinear. PLS method works well in cases where the data set is large, collinear and even when some of the data is missing.[18]

The descriptors of the MP analogues' structure are placed in a QSAR table where they are denoted by the X-variable matrix. PLS assumes that these descriptors represent most of the variation in the chemical structure. It also assumes that there are a small number of "intrinsic" variables, called latent variables (LVs). The descriptors can be represented as the combination of these LVs plus some noise. The result of a PLS analysis is an equation that describes or predicts the differences of the dependent Y-variable (biological activity in this case) from the differences in the values of the descriptor X-variables.

Before the PLS analysis, X and Y are transformed to make their distributions fairly symmetrical. The results of PLSR depend upon how the data are scaled. If some X-variables are known to be more important in predicting Y, then they should be weighted more heavily. If there is no prior knowledge about the relative importance of

the variables then each variable is usually scaled to unit variance by dividing them by their standard deviations or auto scaling them by giving each variable the same weight.[19]

Consider a training set of *N* observations with *K* X-variables and *M* Y-variables is formulated. The matrices X of dimensions (N×K) and Y of dimensions (N×M) are thus calculated. The linear PLSR model finds few "new" variables, which are estimates of LV's. These new variables are called X-scores denoted by t_a (a =1,2...A). These X-scores, which are predictors of Y and also model the X data, are collected in the (N×A) matrix. The X-scores are orthogonal and are linear combinations of the original scores x_k with weights w^*_{ka} as coefficients. A single latent variable, t is expressed as:

$$t = Xw \tag{2}$$

PLS simultaneously summarizes the Y data variation by a number of scores u_a. The bilinear model equations are given as

$$X = \sum_{i=1}^{A} t_i p_i + E \tag{3}$$

$$Y = \sum_{i=1}^{A} u_i q_i + F \tag{4}$$

where: u = latent variable for response (Y) variable

p = loadings corresponding to t

q = loadings corresponding to u

E = model residuals for X

F = model residuals for Y

A = number of components in the PLS model equation.

2.3 Cross-validation

The number of significant components *A*, is estimated by the cross-validation (CV) technique. It is essential to determine the correct complexity (*A*) of the model to avoid the problem of over-fitting the model where the data fits the model perfectly but the model cannot predict any new data.

Cross validation is performed by dividing the data into a number of groups and then developing a number of parallel models from the reduced data with one of the groups deleted. After developing a model, the differences between actual and predicted Y-values are calculated for the deleted data. The sum of squares of these differences is computed and collected to form the predicted residual sum of squares (PRESS), which estimates the predictive ability of the model. PRESS, the cross-validated correlation coefficient (q^2), and the cross validated standard error of estimate (s_{cv}) are computed as shown below:

$$PRESS = \sum_{Y} (Y_{pred} - Y_{actual})^2 \quad (5)$$

$$SS = \sum_{Y} (Y_{actual} - Y_{mean})^2 \quad (6)$$

$$q^2 = 1 - [PRESS / SS] \quad (7)$$

$$s_{cv} = \sqrt{[PRESS / (n - A - 1)]} \quad (8)$$

where: Y = set of all samples

Y_{pred} = a predicted value

Y_{actual} = an actual or experimental value

Y_{mean} = the mean of all values in the training set

n = the number of analogues in the set (i.e. number of rows)

A = the number of components

SS = Sum of square of actual residuals

s_{cv} = Cross-validated standard error of estimation

The optimal number of components, *A* is chosen to be the one that yields the minimum standard error of estimation (s_{cv}) or maximum q^2. The process of cross-validation is repeated for each of the division groups of the data and the optimal number of components for the entire model is estimated.

2.4 Genetic Algorithm – Partial Least Squares Regression

In the present study GA takes care of variable selection of independent X-variables for PLS analysis. In order to understand the role of variable selection in PLS, it is necessary to consider the idea of 'noise'. PLS components are extracted from the X- and Y-data which are along the axes of greatest variation and are optimally correlated. This means that systematic variation in the X-data that is not correlated with the variation in the Y-data is 'noise'. GA, by using subsets of X-data with cross-validated r^2 (q^2), considers models with different variable combinations. This sub-selection process filters out noise and improves the PLS model.

2.5 Topological Indices

The topological indices encode structural information about size, shape or branching of the molecules. The topological indices in this study are the molecular descriptors that form the independent X-variables. The Molconn-Z module in SYBYL® [15] software is designed to carry out the computation of a wide range of topological indices of molecular

structure. A molecular descriptor is a quantity that describes a molecule in terms of its physicochemical properties.

2.5.1 Molconn-Z

Molconn-Z is a computer program that calculates topological indices using a 2D sketcher to draw molecular structure. Molconn-Z parameters are used in a large variety of unique studies including environmental property modeling, molecular classification analysis and biological QSAR studies. The topological indices obtained from the program represent important elements of the molecular structure information, which are useful in relating structure to properties. The standard Molconn-Z program typically computes the following parameters:

- Molecular Connectivity Chi Indices: $^{m}\chi t$ and $^{m}\chi t^{v}$
- Kappa Shape Indices: $^{m}\kappa$ and $^{m}\kappa x$
- Electrotopological State (E-State) Indices: Si
- Molecular Connectivity Difference Chi Indices: d $^{m}\chi t$ and d $^{m}\chi t^{v}$
- Atom-type E-State Indices
- Group-type E-State Indices
- Topological Equivalence Classification of Atoms
- Other Topological Indices:
    - Shannon Index
    - Information Indices
    - Wiener Number
    - Platt Number
    - Bonchev-Trinajstiç
    - Total Topological Index
- Counts of Subgraphs: paths, rings, clusters, etc.

Some of the important parameters computed by the Molconn-Z program are explained below.

2.5.2 Electrotopological States

Electrotopological States (E-states) characterize an atom's properties due to its local and global molecular environment.[20] A molecule is a complex system formed by union of different atoms who give up their own identity and functions in a process called dissolvence.[21] The resulting molecule has properties that are different from constituent atoms and a linear combination of atoms' properties will not predict the molecule's functions. Hence a system of encoding this complexity of a molecule where some fragments are more influential than others must be developed. E-state indices help in quantifying structure information about fragments within the context of the entire molecule.

An atom can be described in terms of its electronic structure and distribution of valence electrons among various orbitals in hybrid states. An atom in a molecule is acted upon by its internal field as well as the fields of other bonded and non-bonded atoms. The intrinsic properties of an atom or a functional group (like a methyl group) are basic attributes that do not change significantly when that atom or functional group occurs in different molecules. These common attributes influence chemical, physical and biological properties. Attributes like elemental content, electronic organization and local topological state of an atom or group dictate its intrinsic state.

The local atomic properties are based on the atom's hybridization & electronic configuration. The local properties are defined by what Kier and Hall call an I-state,

which is related to a valence state. The global environment is dictated by the influence of all the atoms and sub-groups in the molecule.

The I-state is given by:

$$I = \frac{(2/N)^2 \delta_v + 1}{\delta}$$

where :

δ : Count of adjacent atoms except hydrogen atom

δ_v: Difference of valence electrons and hydrogen binding electrons

N: Principal quantum number

The δ value, which gives the count of adjacent atoms other than hydrogen, in essence describes the sigma bond skeleton. The δ_v value is obtained by subtracting hydrogen-binding electrons from the total number of valence electrons. The values of δ and δ_v are rich sources of information about atoms in structure description. They can thus be encoded as follows:

$$\delta = \sigma - h$$

where

σ : count of electrons in sigma orbital.

h : count of bonded hydrogen atoms.

$$\delta_v = Z_v - h$$

where

Z_v : Total number of valence electrons

Hence,

$$\delta_v - \delta = \pi + n$$

where

π: Count of π electrons.

n: count of lone-pair electrons.

The value of $\delta_v - \delta$ has been demonstrated to be highly correlated with Keir-Hall electronegativity and also with Mulliken-Jaffe electronegativity.[20]

Consider a molecule like ethyl acetate. The structure can also be represented in terms of a hydrogen-suppressed graph by not displaying the hydrogen atoms associated with carbon atoms to form a graph. The diagram below shows the molecule along with the delta values.

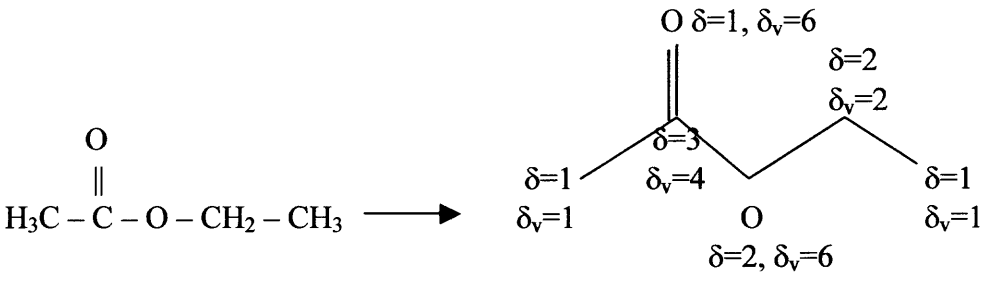


Figure 2.2 Hydrogen-suppressed graph of ethyl acetate.

The Intrinsic State equation encodes for the availability of atom(s) for molecular interactions and the influence of atom on bonds. However, the intrinsic state equation does not show the atom's or groups' position or influence within the field of other atoms in a molecule. The effect of the field and the surrounding molecular topology must be included in the atom description. To include such supplemental information we may consider the electronegativities of other atoms and the separation of two atoms in a molecule.

The separation between two atoms may be represented as

$$r_{ij} = d_{ij} + 1$$

where

r_{ij}: Count of atoms in minimum path length

d_{ij}: Usual graph distance

The difference between intrinsic states, the perturbation is given as:

$$\Delta I_{ij} = (I_i - I_j) / r_{ij}^{m}$$

where m is a constant usually taken as 2.

The total perturbation (S_i) of atom i is due to the influence of all atoms in the molecule and is the sum of its own internal state I_i along with the sum of all perturbations.

$$S_i = I_i + \Sigma_j \Delta I_{ij}$$

S_i, the total perturbation, is called the E-state for atom i.[20]

It can be easily demonstrated that the sum of all E-state values in the molecule, ΣS_i, is equal to the sum of all intrinsic states, ΣI_i, that is $\Sigma S_i = \Sigma I_i$. This shows that all valence electrons remain in the molecule despite the perturbation imposed by internal reorganization.

The local and global properties together shape the molecule's structure. The electrons in various orbitals, influencing the atom's internal structure and position, typify the local properties of an atom. On the other hand the global properties are influenced by the interactions of various atomic charges, bonds and the shape of the molecule. The structure describes the molecule in quantitative terms. Since the E-State is based on the total effect of steric bulk and electronegativities of atoms it is a good representative of molecular structure. It indicates the relationship between activity and structural attributes.

E-state values depict a trend in the chemical graph showing the structure of a molecule. When a part or functional group of the molecule is altered, a corresponding change in E-values occurs. Hence by changing a group in the molecule, the structure of the resulting molecule including the unaltered part changes. This change can affect the binding affinity of the molecule.

E-State encodes electronic and topological information of the atoms within the molecule. The intrinsic state (I state) of an atom is based on the available free valence electrons of the atom. Since the E-state is based on the I-state, the free valence electrons thus correlate with the E-state. There is a strong parallel between the structure and E-state.[20]

The intrinsic states of atoms in alkanes are affected by branching, leading to a significant change of E-state values of atoms at branch points. Introduction of double or triple bonds in organic molecule also alters the E-state value and ultimately influences the molecular structure. The presence of heteroatoms in alkanes produces an effect on adjacent atoms since a higher I_i value of the heteroatom implies greater Si and lower S_j value of the surrounding atoms.

E-state values tend to reflect common chemical intuition regarding the structure of organic molecules. E-state values for large molecules can be calculated using a computer program like Molconn-Z that finds the intrinsic state and computes it rapidly.

2.5.3 Molecular Connectivity

Molecular connectivity[22] is a method of molecular structure quantitation in which weighted counts of substructure fragments are incorporated into numerical indices. Structural features such as size, branching, unsaturation, heteroatom content and cyclicity are encoded.

The calculation of the indices begins with the reduction of the molecule to the hydrogen-suppressed skeleton or graph. Each atom other than those bonded to hydrogen atoms, is assigned two atom descriptors (δ and δ_v) based upon the count of sigma

electrons or valence electrons present. The valence delta values are used in calculating the valence molecular connectivity indices.

$$\delta = \sigma - h$$

$$\delta_v = Z_v - h$$

The molecular connectivity indices or chi indices are symbolized ${}^m\chi t$. Substructures for a molecular skeleton are defined by the decomposition of the skeleton into fragments of:

a) atoms (zero order, m = 0)

b) one bond paths (first order, m = l)

c) two bond fragments (second order, m = 2)

d) three contiguous bond fragments (third order Path, m = 3, t = P) and so forth. Other fragments include the cluster (three atoms attached to a central atom, m = 3, t = C); the path/cluster (equivalent to the isopentane skeleton, m = 4, t = PC); the chain fragment (cycles of 3, 4, 5 . . . atoms, m = 3, 4, 5 . . ., t = CH).

For each order and fragment type, a connectivity index may be calculated. This calculation is made by multiplying the δ (or δ_v) values for each atom in a fragment within a molecule. This product is then converted to the reciprocal square root and called the connectivity subgraph term c_i. These terms are then summed over all the subgraphs (of order m and type t) in the entire molecule, Ns, to calculate the molecular connectivity index ${}^m\chi t$ of order m and type t.

$$ {}^mc_i = \prod_{k=1}^{m+1} (\delta_k)^{-0.5} \quad \text{and} \quad {}^m\chi t = \sum_{i=1}^{Ns} {}^mc_i $$

The valence molecular connectivity indices ${}^m\chi t^v$ are calculated in the same way using δ_v values throughout. The calculations are made from input information, which

includes the connection matrix, designation of the atom type, and the count of hydrogen bonded atoms to each atom.

Consider an example of acetylsalicylic acid shown below along with its H-suppressed graph.

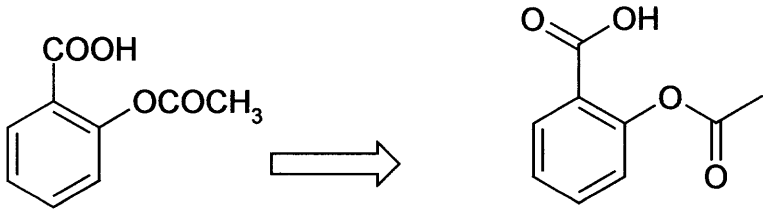


Figure 2.3 Acetylsalicylic acid.

To calculate $^{1}\chi^{v}$, dissect the molecule into all its one-bond fragments, shown below; each bond fragment is labeled with the δ^{v} values.

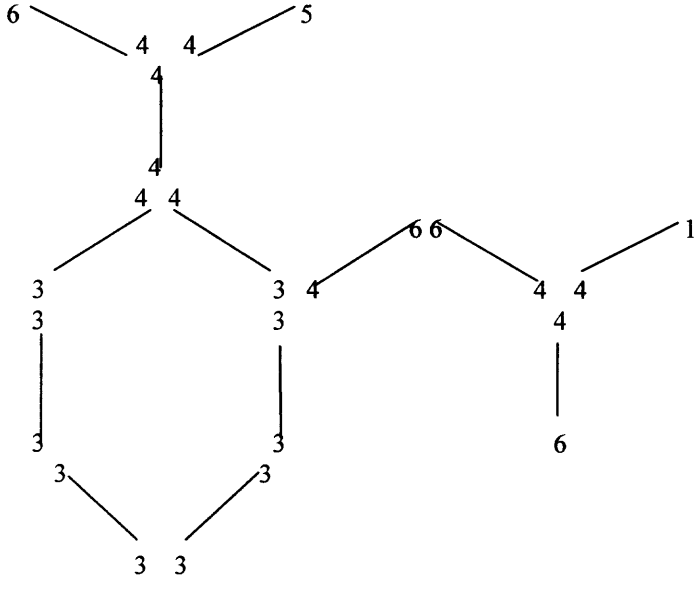


Figure 2.4 Delta values of Acetylsalicylic acid.

For each fragment compute the subgraph contribution, $(\delta^{v}_{i} * \delta^{v}_{j})^{-0.5}$, and sum over all the bond fragments:

$$^{1}\chi^{v} = 3.6175.$$

In an analogous manner the other connectivity indices can be calculated for path, cluster, path/cluster and ring type subgraphs. The finding of all the subgraphs of a given type becomes a difficult computational problem by hand. That is, of course, the reason for the development of Molconn-Z.

2.5.4 Molecular Shape Indices

The Molecular shape indices[22] are the basis of a method of molecular structure quantitation in which attributes of molecular shape are encoded into three indices (Kappa values). These Kappa values are derived from counts of one-bond, two-bond and three-bond fragments, each count being made relative to fragment counts in reference structures which possess a maximum and minimum value for that number of atoms.

The calculation of the indices begins with the reduction of the molecule to the hydrogen-suppressed skeleton. The count of one-, two-, and three-bond fragments, ^{1}P, ^{2}P and ^{3}P, respectively, is made. These values are used in calculating the Kappa indices $^{1}\kappa$, $^{2}\kappa$, and $^{3}\kappa$. The calculation of each index is made using the value of $^{m}P_{i}$ and the $^{m}P_{min}$ and $^{m}P_{max}$ counts for graphs with the same number of atoms, A. These latter counts are the minimum and maximum numbers of paths of that order (m) that can be found in a real or hypothetical structure. Specifically, the $^{m}P_{min}$ for all orders of Kappa is the count of paths of length m in the linear skeleton. The $^{m}P_{max}$ structure is, for the m = 1, a complete graph (a); for m = 2, a star graph (b); and for m = 3, a twin star graph (c).

Below are the graphs for the structures corresponding to the maximum count of paths for orders one, two, and three in the six-atom molecules.

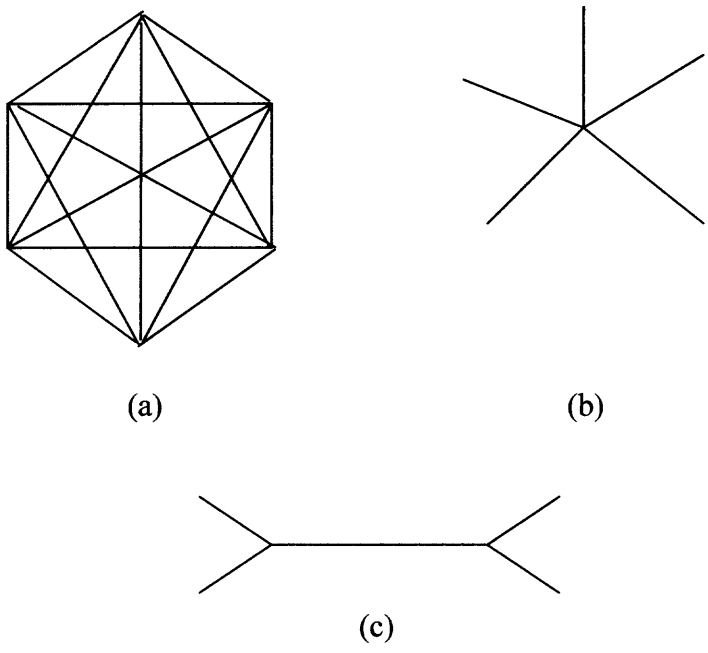


Figure 2.5 Graph of orders one, two and three for six-atom molecules.

The values of the Kappa indices can be calculated directly from the path count ${}^{m}P_i$ and the number of atoms using the equations:

$${}^{1}\kappa = A(A\text{-}1)^2/({}^{1}P_i)^2$$

$${}^{2}\kappa = (A\text{-}1)(A\text{-}2)^2/({}^{2}P_i)^2$$

$${}^{3}\kappa = (A\text{-}3)(A\text{-}2)^2/({}^{3}P_i)^2 \text{ when A is even}$$

$${}^{3}\kappa = (A\text{-}1)(A\text{-}3)^2/({}^{3}P_i)^2 \text{ when A is odd}$$

The presence of atoms other than $C(sp^3)$ is taken into consideration in each order of Kappa index by modifying each A and ${}^{m}P_i$ in the equations above with a value:

$\alpha = r(x)/r[C(sp^3)] - 1$

where r(x) is the covalent radius of atom x and $r[C(sp^3)]$ is the covalent radius of carbon in the sp^3 hybrid state. In the relation for the Kappa indices, A is replaced by $A + \alpha$ and ${}^{m}P_i$ is replaced by ${}^{m}P_i + \alpha$. The resulting Kappa values are designated as ${}^{m}\kappa^{\alpha}$.

2.5.5 Topological State Indices

The topological state indices[23] are numerical values associated with each atom in a molecule, which encode information about the topological environment of that atom due to all other atoms in the molecule. The topological relationship to each other atom is based on the encoding of atom information in all the paths emanating from that atom. Topologically-equivalent atoms have identical values of the topological state index and nonequivalent atoms have different values.

Each atom in the skeleton structure of a molecule is identified by the valence delta values δ and δ_v. Beginning with any atom i, all contiguous paths of atoms, emanating from that atom to each other atom j, are identified. The lowest order path is the atom itself. This process is followed by finding all first order paths (bonds containing atom i) and so on ultimately including the longest paths(s) terminating on atom i. A numerical value is calculated for each of these paths and is an entry in the Topological State Matrix T. Each entry is calculated according to the formula

$$t_{ij} = (GM_{ij})^a(d_{ij})^b$$

GM_{ij} is the geometric mean of the delta values of the atoms in the path of atoms of length d_{ij} atoms between atoms i and j. These path fragment values are then summed to give a topological state value T_i for atom i. Subsequent calculations produce T_i values for every atom. Several values of a and b have been investigated but the Molconn-Z program uses a = +1 and b = -2.

Consider an example of 2-Propanol,

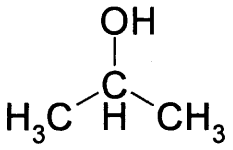


Figure 2.6 2-Propanol.

Table 2.2 Topological State Index

Paths Between Atoms					Topological State Matrix, T				Topological State Index, T_i
	1	2	3	4					
1	CH_3-	CH_3CH-	CH_3CHCH_3	CH_3CHOH	1.000	1.154	2.080	1.216	5.451
2		-CH<	>$CHCH_3$	>CHOH	1.154	0.333	1.154	0.516	3.159
3			CH_3	CH_3CHOH	2.080	1.154	1.000	1.216	5.451
4				-OH	1.216	0.516	1.216	0.200	3.149

In the above example topological equivalence is indicated by the T_i values. In 2-propanol the topological equivalence of the two methyl groups is shown by the fact that $T_1 = T_3$; no other values are equal in accordance with the fact that no other atoms are topologically equivalent. In this sense, the topological state index values represent the topological equivalence (topological symmetry) of the molecule. The pattern of T_i values for a portion of a molecule appears to be characteristic of that fragment and may be used as a basis for quantitative measures of fragment similarity.

CHAPTER 3

METHODS

3.1 Biological Activity Data

The methylphenidate (MP) data was provided to Professor Venanzi's group by Dr. Howard Deutsch of the Georgia Institute of Technology, who synthesized the analogues, and Dr. Margaret Schweri of the Mercer University School of Medicine, who measured the DAT binding affinity (IC_{50}) of the analogues. A total of 80 MP analogues included in this study with their IC_{50} are shown in Table A.1 in Appendix A.

3.2 Descriptor Data

The molecular descriptors of all 80 MP analogues were calculated using the Molconn-Z module in SYBYL®. There were 104 descriptors including topological indices and molecular descriptors that were calculated by Milind Misra of the Venanzi group. The categories of the 104 descriptors are described below in Table 3.1. The objective of running the GA part of the program is to reduce the number of descriptors and use only significant descriptors in building the final predictive model.

Table 3.1 MP Descriptor Types

Descriptor category	Number of descriptors
Molecular flexibility index	1
Shannon information index	1
Kappa indices	7
Topological state indices	10
Hydrogen bond related counts and E-state indices	14
Chi indices	32
Atom Type Electrotopological State Indices	39

3.3 Scaling

The descriptor and the activity data must be scaled before it is utilized to form a predictive model. The results of projection methods like PLS regression heavily depend upon scaling.[19] Scaling can be used to focus on more important Y-variables and increase the weights of more informative X-variables. In the absence about the relative importance of the variables, the standard approach is to auto-scale the data. This is done by dividing the variables by their standard deviations and centering them by subtracting their averages.

In the present study the descriptors were scaled according to the following formula [24]:

$$X_{ij}\ (\text{scaled}) = (X_{ij} - X_{j,min}) / (X_{j,max} - X_{j,min})$$

where

X_{ij} : The value of the *j*th descriptor for compound *i*

$X_{j,min}$: The minimum value of the *j*th descriptor

$X_{j,max}$: The maximum value of the *j*th descriptor

Thus, for all descriptors the minimum scaled X_{ij} is 0 and the maximum scaled X_{ij} is 1.

3.4 GA-PLS Algorithm

The GA part of the program that performs variable selection of descriptors was implemented in Matlab® [25]. The multivariate regression was implemented by using the PLS Toolbox® [26].

The GA-PLS algorithm is as follows:

1. The Molconn-Z module in SYBYL® is applied to each of the MP analogues to obtain 104 descriptors.

2. A population of 300 random combinations of the descriptors is generated. To apply the genetic algorithm, these combinations are considered as a binary string where a 'one' denotes the inclusion while a 'zero' shows the exclusion of that descriptor in that particular combination. Each of these combinations is considered to be a parent. The number of crossovers is initialized to zero.

3. Using the QSAR table containing the descriptors and the biological activity, the q^2 value of each parent is determined using PLSR. The optimal number of components of the model is obtained from the PLS routine.

4. The fitness of each individual is determined using the equation [27],
   fitness = $1-(n-1)(1-q^2)/(n-c)$
   where: q^2 = correlation coefficient
   n = number of analogues
   c = optimal number of components

5. Two parents are selected randomly for mating.

6. The two selected parents undergo crossover at a random point to reproduce two offspring. The number of crossovers is incremented by 1.

7. Each offspring is subjected to random point mutation depending upon the probability specified. Random point mutation is applied by flipping the bit in the binary string from 0 to 1 or vice versa.

8. The fitness of each offspring is evaluated using the PLS routine as described in steps 3 and 4.

9. If the fitness of the offspring is of a higher value than that of its parent, then the parent is replaced by the fitter offspring in the mating pool.

10. Steps 5 to 9 are repeated until a maximum number of crossovers are reached.

11. A predictive model is developed using the descriptors selected by the fittest individual when the maximum number of crossovers is reached.

12. The predictive model is validated using the leave-one-out cross-validation technique.

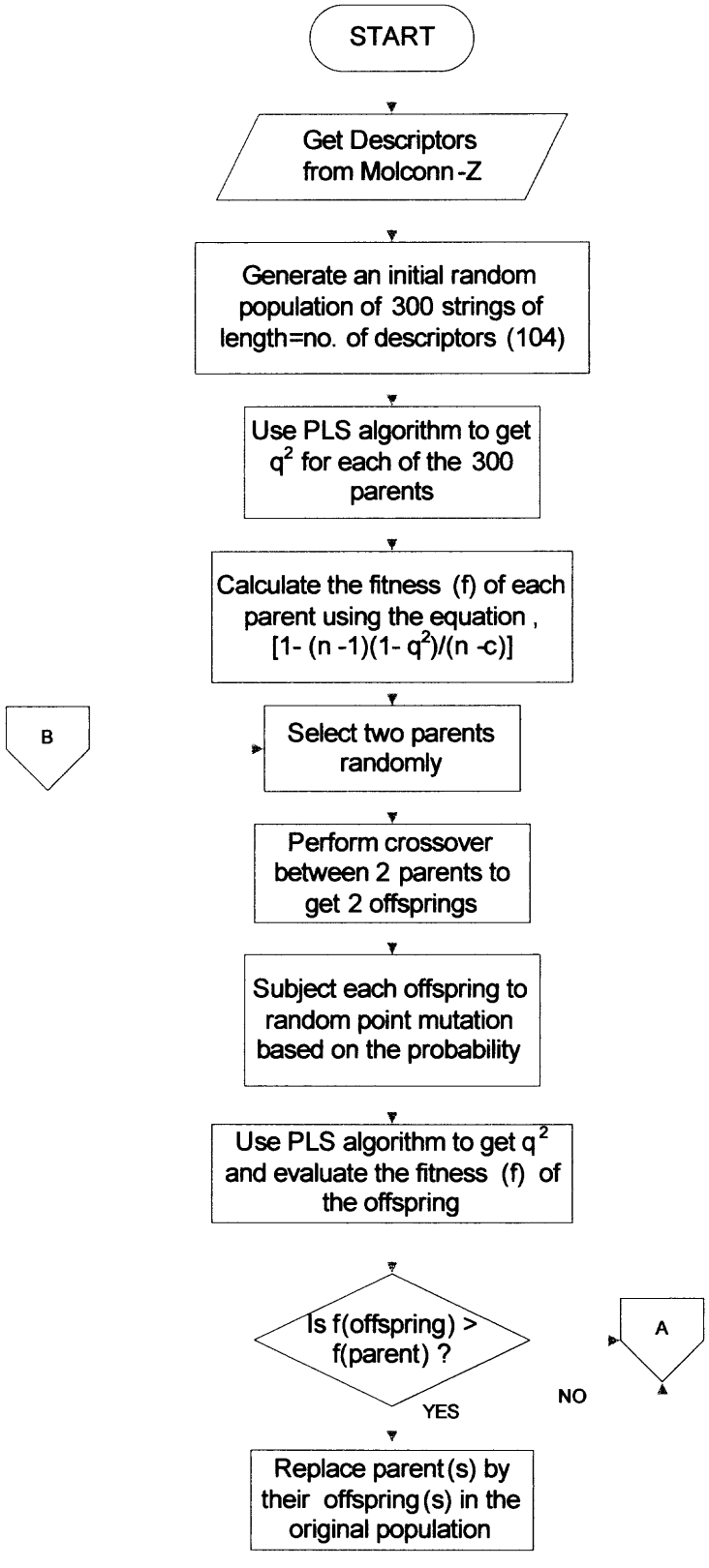


Figure 3.1 GA-PLS Flowchart.

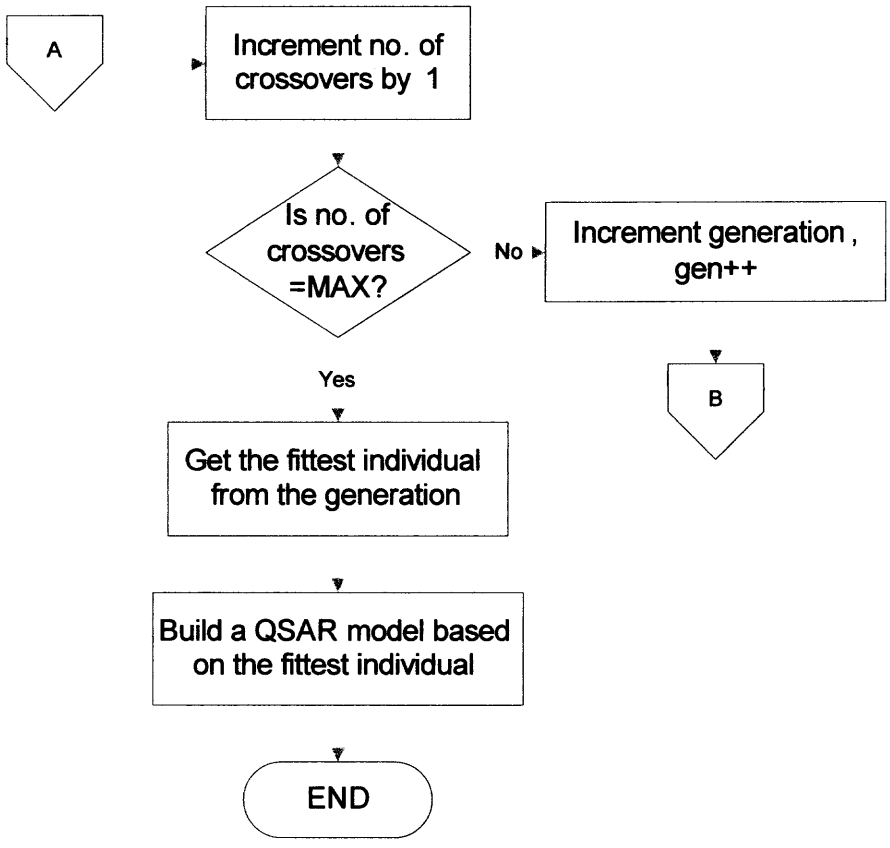


Figure 3.1 GA-PLS Flowchart (continued).

CHAPTER 4

RESULTS

4.1 Testing

The GA-PLS program was tested using the Selwood dataset [28] that contains a set of 31 antifilarial antimycin analogues described by 53 X-variables. The analogues are of the general form as shown in Figure 4.1. The Selwood dataset is considered as a benchmark for these types of studies because it illustrates a number of issues that arise when analyzing real-world data. This dataset is of particular interest because it has a very large ratio of number of descriptors to number of compounds. Also this dataset has been well studied and it consists of clear outliers, i.e. for a particular descriptor an analogue has a value that is far from the range of other analogues.

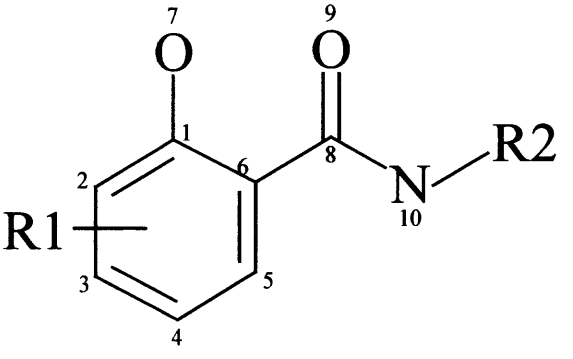


Figure 4.1 Structure of antimycin analogues.

Table 4.1 Selwood Descriptors

ATCH1-ATCH10	Partial atomic charges for atoms 1-10
DIPMOM	Dipole moment
DIPV_X, DIPV_Y, DIPV_Z	Dipole vector components in X.Y and Z
ESDL1- ESDL10	Electrophilic superdelocalizability for atoms 1-10
LOGP	Partition coefficient
M_PNT	Melting point
MOLWT	Molecular weight
MOFI_X, MOFI_Y, MOFI_Z	Principal moments of inertia in X, Y and Z
NSDL1- NSDL10	Nucleophilic superdelocalizability for atoms 1-10
PEAX_X, PEAX_Y, PEAX_Z	Principal ellipsoid axes in X, Y, and Z
S8_1DX, S8_1DY, S8_1DZ	Substituent on atom 8 dimensions in X, Y and Z
S8_1CX, S8_1CY, S8_1CZ	Substituent on atom 8 center in X, Y and Z
SURF_A	Surface area
SUM_F and SUM_R	Sums of F and R substituent constants
VDWVOL	Van der Waals volume

The descriptors consist of electronic properties like atomic charges, dipole moments, electronic and nucleophilic delocalizabilities for the atoms in the analogues. Bulk properties like Van der Waals volume, moments of inertia, molecular weight and ellipsoid axes are also included as descriptors. The descriptors SUM_F and SUM_R denote the electronic effects of change in substituents. The electronic effects can be represented as a combination of a field (inductive) effect, F, and a resonance effect, R.

The method of GA-PLS was applied on the Selwood dataset with input parameters as follows:

Number of individuals in a population: 300

Total number of crossovers: 10000

Probability of mutation of an individual: 0.1

The final population of GA-PLS consists of the same individual (string) with only nine of the descriptors selected. PLS regression was then applied to the fittest individual

of the final population to yield a model with predictive cross-validated coefficient, q^2 as 0.78 and a fitness of 0.75. The QSAR equation given by the GA-PLS technique is:

$$-\log(IC_{50}) = 0.4205\ ATCH4 - 0.3822\ DIPV\_X - 0.1777\ DIPV\_Z - 0.0891\ DIPMOM + 0.0650\ MOFI\_X - 0.8194\ MOFI\_Y - 0.1784\ S8\_1CY + 0.8828\ LOGP + 0.8694\ SUM\_F \quad (1)$$

Selwood[28] used multivariate regression to develop a QSAR model of the dataset. Selwood [28] proposed a QSAR model of three features: LOGP, M_PNT and ESDL10 with a regression only q^2 value of 0.44 [29]. The Selwood model is given by the equation:

$$-\log(IC_{50}) = -3.93 + 0.44 * LOGP + 0.008 * M\_PNT - 0.30 * ESDL1O \quad (2)$$

Wikel and Dow[30] proposed a QSAR model using neural network to select appropriate number of variables. The model has a regression only q^2 value of 0.46 [29] and the equation for their model is given as:

$$-\log(IC_{50}) = -1.63 + 0.231 * LOGP + 4.415 * ATCH4 + 0.01 * MOFI\_X \quad (3)$$

Rogers and Hopfinger[31] used the technique of GFA to generate 300 predictive models. Their top four models are given as:

$$-\log(IC_{50}) = -2.501 + 0584\ LOGP + 1.513\ SUM\_F - 0.000075\ MOFI\_Y \quad (4)$$

$$-\log(IC_{50}) = 2.871 + 0.568\ LOGP - 0.013\ SURF\_A + 0.810\ ESDL3 \quad (5)$$

$$-\log(IC_{50}) = -0.805 + 0589\ LOGP + 0.736\ ESDL3 - 0.000077\ MOFI\_Y \quad (6)$$

$$-\log(IC_{50}) = 1.791 + 0.5000\ LOGP + 0.842\ ESDL3 - 0.2\ PEAX\_X + 0.2807\ ATCH4 \quad (7)$$

The q^2 of the top-rated GFA model is 0.65 The cross-validation was conducted only after the features were selected using the full-data set.[29]

The different methods mentioned above give a list of the most useful features for building activity molecules. The significant descriptors used by these methods along with those used in the GA-PLS method are summarized in Table 4.2. The different

techniques select different variables to build a QSAR model with a one common feature, LOGP. This is likely caused by different selection pressures under each technique.[31] The common descriptor selected by all the techniques, LOGP is the ratio of concentration of compound in aqueous phase to the concentration in an immiscible solvent, as the neutral molecule.

Table 4.2 Significant Descriptors for Selwood Dataset

	Selwood	Wikel	GFA	GA-PLS
ATCH1			●	
ATCH2	●	●		
ATCH4		●	●	●
ATCH5			●	
ATCH6			●	
DIPMOM				●
DIPV_X		●		●
DIPV_Y	●			
DIPV_Z	●			●
ESDL3			●	
ESDL5	●			
ESDL10	●			
LOGP	●	●	●	●
M_PNT	●	●		
MOFI_X		●		●
MOFI_Y		●	●	●
NSDL2	●			
PEAX_X			●	
PEAX_Y		●		
S8_1CY				●
S8_1CZ	●			
SUM_F			●	●
SUM_R	●			
SURF_A			●	
VDWVOL		●		

The GA-PLS technique picks out nine significant descriptors from a total of 53 descriptors. Seven out of the nine descriptors are also chosen by the other techniques. The GA-PLS technique selects two new descriptors (DIPMOM and S8_1CY) not frequently picked by the other techniques. The table below summarizes the number of common descriptors selected by each of the techniques mentioned above.

Table 4.3 Common Descriptors for Selwood Dataset

	Selwood	Wikel	GFA	GA-PLS
Selwood	10	3	1	2
Wikel		9	3	5
GFA			10	4
GA-PLS				9

The above table shows that the GA-PLS technique selects about half of the same descriptors as selected by GFA and Wikel method. Thus, GA-PLS discovers a set of significant descriptors and builds a predictive model that depends on a few significant descriptors.

4.2 Application to Methylphenidate

The GA-PLS method was applied to the MP dataset. The dataset consists of 80 MP analogues described by 104 topological descriptors. The input parameters for the GA-PLS program were:

Number of individuals in a population: 300

Total number of crossovers: 10000

Probability of mutation of an individual: 0.1

The QSAR equation for the MP dataset is given as:

$$-\log(IC_{50}) = -0.0218\ Xvp6 - 0.0523\ SHsOH - 0.0910\ SHaaCH - 0.1768\ SHother + 0.8612\ SdssC - 0.1633\ SHBa + 0.1573\ SHBint2 - 0.3192\ SHBint5 + 0.7079\ Hmin \quad (8)$$

The model has a cross-validated correlation coefficient, q^2 of 0.78 and a fitness of 0.77. The model picked up only nine of the 104-descriptor variables and developed the model with a high predictive ability. A table explaining the descriptors picked by the GA-PLS routine is given below.

Table 4.4 Significant Methylphenidate Descriptors

Variable	Descriptor
Xvp6	Valence Chi path index -6 contiguous bond fragments.
SHsOH	Atom type electrotopological state index values for atom types. *s* stands for the bond in the group and *OH* stands for the Hydroxyl group.
SHaaCH	Atom type electrotopological state index value for the *:CH:* group.
SHother	General non-polar *H* descriptor (sum of *H* E-State values for all non-polar *C-H* bonds).
SdssC	Atom type electrotopological state index value for the *=C<* group.
SHBa	Descriptor for weak hydrogen bond acceptor.
SHBint2	E-State descriptors of potential internal H bond strength
SHBint5	E-State descriptors of potential internal H bond strength
Hmin	Smallest hydrogen E-State value

The above table shows that the GA-PLS method picks significant descriptors that depend upon the electrotoplogical state and Chi-indices. The QSAR model built from the above nine descriptors has high prediction accuracy due to the above average value of the

fitness. Therefore, the approach of using GA-PLS to build predictive QSAR models is reasonable.

CHAPTER 5

CONCLUSION

The method of GA-PLS was used to develop a predictive model that predicts the biological activity of MP analogues based on a few significant structural descriptors. The method uses GA to select the variables and estimate how good these variables are in predicting the activity. The prediction ability of a set of variables is determined by performing PLS regression. An individual in a population consisting of a set of descriptors represents a possible model.

The GA-PLS method was tested on Selwood dataset and the results were compared with previous studies. The model developed by GA-PLS includes about half the common descriptors as selected by the other studies. The method of GA-PLS thus builds reasonably predictive models.

The QSAR equation of the MP dataset using GA-PLS was determined. The cross-validated correlation coefficient, q^2 of the model was 0.78 and it had a fitness of 0.77. The above than average value of q^2 indicate that the GA-PLS technique develops reasonable good predictive models. The final population at the end of the GA-PLS routine consists of a single individual that is the fittest one.

The method of GA-PLS heavily depends upon the initial conditions and scaling of the data. A scaling technique that scales the data between 0 and 1 was employed.

The method of GA-PLS may further be improved upon by changing the mutational probability to increase the chances of a fitter individual and at the same time limiting the number of variables used.

APPENDIX A

IC_{50} VALUES

The Table A.1 contains the experimental IC_{50} values of 80 Methylphenidate analogues. The scaled log(IC_{50}) value is used as the Y variable in PLS regression analysis.

Table A.1 IC_{50} Values of Methylphenidate Analogues*

Number	GIT Number	Compound Name	IC_{50} (nM)	Y = -logIC_{50}(M)
1	AL34.1	threo-N-Benzyl-methylphenidate hydrochloride	52.9	7.27654
2	AN-1-68.2	threo-N-(4-isothiocyanatobenzyl)-methylphenidate	422.3	6.37438
3	BO-1-119.1	threo-N-(3-phenylpropyl)ritalinol	193.5	6.71332
4	BO-1-12.1	threo-N-propargyl-methylphenidate hydrochloride	820.5	6.08592
5	BO-1-120.1	threo-N-(4-phenylbutyl)ritalinol	622.5	6.20586
6	BO-1-122.1	threo-N-(2-phenylethyl)ritalinol	1431	5.84436
7	BO-1-128.1	threo-N-(3-phenylpropyl)methylphenidate HCl	267	6.57349
8	BO-1-13.1	threo-N-(3-chlorobenzyl)-methylphenidate hydrochloride	105.9	6.97510
9	BO-1-131.1	threo-N-(2-phenylethyl)methylphenidate HCl	677.5	6.16909
10	BO-1-144.1	threo-N-(4-phenylbutyl)methylphenidate HCl	205.3	6.68761
11	BO-1-145.1	threo-N-(5-phenylpentyl)methylphenidate HCl	1572.5	5.80341
12	BO-1-146.1	threo-N-(6-phenylhexyl)methylphenidate HCl	656	6.18310
13	BO-1-15.1	threo-N-(2-chlorobenzyl)-methylphenidate hydrochloride	242.6	6.61511
14	BO-1-17.1	threo-N-allyl-methylphenidate hydrochloride	597.2	6.22388
15	BO-1-19.1	threo-N-(4-chlorobenzyl)-methylphenidate hydrochloride	31.2	7.50585

* All the IC50 values were measured by Dr. Margaret Schweri of Mercer University School of Medicine and all the compounds were synthesized by Dr. Howard Deutsch of Georgia Institute of Technology unless marked †.

† Compounds synthesized in the laboratory of Dr. John Gatley.

Table A.1 IC_{50} Values of Methylphenidate Analogues (continued)

Number	GIT Number	Compound Name	IC_{50} (nmol)	Y = -logIC_{50}(mol)
16	BO-1-21.1	threo-N-(4-nitrobenzyl)-methylphenidate hydrochloride	112.9	6.94731
17	BO-1-23.1	threo-N-(4-methoxybenzyl)-methylphenidate hydrochloride	79.1	7.10182
18	BO-1-30.1	threo-N-methyl-[2-(5-chlorothiophene)]-methylphenidate hydrochloride	391.5	6.40727
19	BO-1-37.1	threo-N-(2-methylpyridyl)-methylphenidate,dihydrochloride	368.5	6.43356
20	BO-1-43.1	threo-N-(3-methylpyridyl)-methylphenidate,dihydrochloride	173.2	6.76145
21	BO-1-44.1	threo-N-(4-methylpyridyl)-methylphenidate,dihydrochloride	127.9	6.89313
22	BO-1-45.1	threo-N-(methyl-2-furan)-methylphenidate hydrochloride	535.7	6.27108
23	BO-1-46.1	threo-N-(methyl-3-thiophene)-methylphenidate hydrochloride	142.8	6.84527
24	BO-1-47.1	threo-N-(methyl-2-thiophene)-methylphenidate hydrochloride	223.7	6.65033
25	BO-1-48.1	threo-N-(methyl-3-furan)-methylphenidate hydrochloride	459.3	6.33790
26	BO-1-96	N-ethyl-threo-ritalinol	4602	5.33705
27	BO-2-28.1	threo-3,5-dimethylmethylphenidate hydrochloride	4685	5.32929
28	BO-2-40.1	threo-3,5-dichloromethylphenidate hydrochloride	65.6	7.18310
29	BO-2-57.1	threo-N-(3-chlorobenzyl)-ritalinol hydrochloride	25.8	7.58838
30	CE101.1	threo-ritalinol	447.5	6.34921
31	EGK-266/1	threo 4-hydroxymethylphenidate hydrochloride	98	7.00877
32	EGK-276-A	threo-N-methyl-3-hydroxymethyl-4-(O, hydroxymethyl)methylphenidate hydrochloride	620	6.20761
33	EGK-276-B	threo N-methy-4-hydroxymethylphenidate	1215	5.91542
34	LL81.2	4-nitromethylphenidate hydrochloride	493.8	6.30645
35	QS-1-114.1	3-aminomethylphenidate dihydrochloride	265	6.57675
36	QS-1-128.1	threo-4-aminomethylphenidate dihydrochloride	34.5	7.46218
37	QS-1-138.1	threo-4-methoxymethylphenidate	83	7.08092

Table A.1 IC_{50} Values of Methylphenidate Analogues (continued)

Number	GIT Number	Compound Name	IC_{50} (nmol)	Y = -logIC_{50}(mol)
38	QS-1-142.1	threo-4-chloromethylphenidate hydrochloride	20.6	7.68613
39	QS-1-89.4	threo-methylphenidate hydrochloride	84.3	7.07417
40	QS-2-116.3	threo-4-t-butylmethylphenidate hydrochloride	13450	4.87128
41	QS-2-124.2	threo-N-methyl-3-chloromethylphenidate	160.5	6.79452
42	QS-2-125.1	threo-4-iodomethylphenidate hydrochloride †	14	7.85387
43	QS-2-125.2	threo-3-bromomethylphenidate hydrochloride †	4.2	8.37675
44	QS-2-125.3	threo-N-methyl-methylphenidate	499	6.30190
45	QS-2-133.1	threo-N-methyl-4-methylmethylphenidate	139.8	6.85449
46	QS-2-147.2	threo-2-bromomethylphenidate hydrochloride †	1865	5.72932
47	QS-2-15.1	threo-2-methoxymethylphenidate hydrochloride	100666.7	3.99711
48	QS-2-29.4	threo-3-methoxymethylphenidate hydrochloride	287.5	6.54136
49	QS-2-40.1	threo/erythro-2-hydroxymethylphenidate hydrochloride	23050	4.63733
50	QS-2-41.2	threo-3-hydroxymethylphenidate hydrochloride	321	6.49349
51	QS-2-61.4	threo-3-chloromethylphenidate hydrochloride	5.1	8.29243
52	QS-2-71.3	threo-4-fluoromethylphenidate hydrochloride	35	7.45593
53	QS-2-81.4	threo-3,4-dichloromethylphenidate hydrochloride	5.3	8.27572
54	QS-2-84.4	threo-3,4-dimethoxymethylphenidate hydrochloride	810	6.09151
55	QS-2-88.1	threo-4-bromoromethylphenidate hydrochloride †	6.9	8.16115
56	QS-2-99.3	threo-2-chloromethylphenidate hydrochloride	1946.7	5.71070
57	WB47.4	threo-2-fluoromethylphenidate hydrochloride	1415	5.84924
58	WB48.4	threo-3-fluoromethylphenidate hydrochloride	40.5	7.39254
59	WB61.4	threo-4-methylmethylphenidate hydrochloride	33	7.48149

Table A.1 IC_{50} Values of Methylphenidate Analogues (continued)

Number	GIT Number	Compound Name	IC_{50} (nmol)	Y = -logIC_{50}(mol)
60	WB71.5	threo-3-methylmethylphenidate hydrochloride	21.4	7.66959
61	WB77.2	threo-3-fluororitalinol	281	6.55129
62	XY-1-102.3	threo-3,4-dichlororitalinol	4.2	8.37675
63	XY-1-127.5	threo-3,4-dichlororitalinol methyl ether, hydrochloride	1.7	8.76955
64	XY-1-129.2	threo-N-benzyl-3-chloromethylphenidate	41.2	7.38510
65	XY-1-144.4	threo-N-benzylmethylphenidate dimethylamide hydrochloride	1732.5	5.76133
66	XY-1-147.4	threo-N-benzylmethylphenidate amide hydrochloride	384	6.41567
67	XY-1-30.3	threo-N-benzylritalinol hydrochloride	23.7	7.62525
68	XY-1-44.5	threo-N-benzylritalinol methyl ether, hydrochloride	17.8	7.74958
69	XY-1-47.1	threo-ritalinol methyl ether hydrochloride	97.1	7.01278
70	XY-1-85.7	threo-N-benzyl-3,4-dichloromethylphenidate hydrochloride	76.3	7.11748
71	XY-1-86.2	threo-N-benzyl-3,4-dichlororitalinol hydrochloride	2.7	8.56864
72	XY-1-89.5	threo-N-benzyl-3,4-dichlororitalinol methyl ether hydrochloride	4.2	8.37675
73	XY-2-74.3	threo-3,4-dichloromethylphenidate amide hydrochloride	16.4	7.78516
74	ZL102.3	threo-benzylritalinate hydrochloride	1024.3	5.98957
75	ZL105.1	threo-N-methyl-3-methylmethylphenidate	107.7	6.96778
76	ZL21.1	threo-ritalinylacetate	690	6.16115
77	ZL26.1	threo-methylphenidate, amide	1728	5.76246
78	ZL38.1	threo-4-ethylmethylphenidate hydrochloride	736.7	6.13271
79	ZL68.3	threo-(2-naphthyl)methylphenidate hydrochloride	11	7.95861
80	ZL77.2	threo-4-(trifluoromethyl)methylphenidate, hydrochloride	615	6.21112

APPENDIX B

METHYLPHENIDATE DESCRIPTORS

The values of all 104 descriptors for each of the 80 Methylphenidate (MP) analogues are given in the following Tables.

Table B.1 Simple Chi Path Indices for MP Analogues

Analogue	X0	X1	X2	Xp3	Xp4	Xp5	Xp6	Xp7	Xp8	Xp9	Xp10
AL34.1	16.778	11.792	9.809	8.318	6.914	5.811	3.421	2.509	1.778	1.082	0.64
AN-1-68.2	19.062	13.224	10.98	9.257	7.655	6.419	3.95	2.939	2.059	1.279	0.839
BO-1-119.1	16.615	11.881	9.614	8.157	6.851	5.45	3.069	2.191	1.5	0.988	0.679
BO-1-12.1	14.372	9.774	7.858	6.617	5.5	4.524	2.662	1.811	1.089	0.536	0.279
BO-1-120.1	17.322	12.381	9.967	8.407	7.022	5.605	3.139	2.218	1.509	0.972	0.675
BO-1-122.1	15.908	11.381	9.26	7.915	6.632	5.351	3.031	2.179	1.523	0.995	0.644
BO-1-128.1	18.192	12.792	10.504	8.869	7.273	5.962	3.486	2.489	1.717	1.108	0.728
BO-1-13.1	17.648	12.186	10.443	8.639	7.265	5.95	3.7	2.592	1.854	1.172	0.696
BO-1-131.1	17.485	12.292	10.151	8.628	7.054	5.862	3.45	2.468	1.746	1.131	0.693
BO-1-144.1	18.899	13.292	10.858	9.119	7.444	6.116	3.556	2.514	1.731	1.088	0.712
BO-1-145.1	19.606	13.792	11.211	9.369	7.621	6.237	3.666	2.564	1.75	1.098	0.697
BO-1-146.1	20.313	14.292	11.565	9.619	7.798	6.362	3.751	2.641	1.785	1.111	0.704
BO-1-15.1	17.648	12.203	10.327	8.852	7.195	6.036	3.641	2.605	1.865	1.143	0.665
BO-1-17.1	14.372	9.774	7.858	6.617	5.5	4.524	2.662	1.811	1.089	0.536	0.279
BO-1-19.1	17.648	12.186	10.431	8.729	7.056	6.102	3.636	2.603	1.869	1.148	0.726
BO-1-21.1	19.225	13.097	11.33	9.428	7.582	6.287	4.052	2.879	2	1.284	0.827
BO-1-23.1	18.355	12.724	10.6	9.137	7.367	6.199	3.881	2.766	1.943	1.227	0.783
BO-1-30.1	16.941	11.686	10.089	8.397	7.03	5.552	3.38	2.545	1.773	1.082	0.639
BO-1-37.1	16.778	11.792	9.809	8.318	6.914	5.811	3.421	2.509	1.778	1.082	0.64

Table B.1 Simple Chi Path Indices for MP Analogues (continued)

Analogue	X0	X1	X2	Xp3	Xp4	Xp5	Xp6	Xp7	Xp8	Xp9	Xp10
BO-1-43.1	16.778	11.792	9.809	8.318	6.914	5.811	3.421	2.509	1.778	1.082	0.64
BO-1-44.1	16.778	11.792	9.809	8.318	6.914	5.811	3.421	2.509	1.778	1.082	0.64
BO-1-45.1	16.071	11.292	9.456	8.068	6.737	5.198	3.276	2.426	1.671	0.978	0.555
BO-1-46.1	16.071	11.292	9.456	8.068	6.737	5.198	3.276	2.426	1.671	0.978	0.555
BO-1-47.1	16.071	11.292	9.456	8.068	6.737	5.198	3.276	2.426	1.671	0.978	0.555
BO-1-48.1	16.071	11.292	9.456	8.068	6.737	5.198	3.276	2.426	1.671	0.978	0.555
BO-1-96	12.088	8.364	6.587	5.769	4.855	3.782	2.054	1.28	0.752	0.404	0.226
BO-2-28.1	13.828	9.113	8.039	6.254	5.429	3.968	2.663	1.521	1	0.47	0.204
BO-2-40.1	13.828	9.113	8.039	6.254	5.429	3.968	2.663	1.521	1	0.47	0.204
BO-2-57.1	16.071	11.275	9.552	7.926	6.843	5.436	3.293	2.294	1.614	1.03	0.643
CE101.1	10.51	7.415	5.868	4.982	4.176	3.179	1.575	1.083	0.648	0.272	0.144
EGK-266/1	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
EGK-276-A	16.82	11.117	9.179	7.882	6.389	5.183	3.184	2.215	1.516	0.897	0.437
EGK-276-B	13.828	9.13	7.908	6.578	5.096	4.283	2.422	1.565	1.032	0.476	0.201
LL81.2	14.535	9.63	8.28	6.799	5.302	4.128	2.519	1.864	1.273	0.586	0.241
QS-1-114.1	12.958	8.72	7.393	6.016	4.936	3.859	2.289	1.436	0.941	0.428	0.178
QS-1-128.1	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
QS-1-138.1	13.665	9.258	7.55	6.508	5.087	4.034	2.378	1.735	1.142	0.512	0.213
QS-1-142.1	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
QS-1-89.4	12.088	8.326	6.759	5.689	4.624	3.689	1.935	1.337	0.85	0.368	0.144
QS-2-116.3	15.458	9.931	9.339	7.036	5.475	4.21	2.637	1.972	1.378	0.646	0.263
QS-2-124.2	13.828	9.13	7.92	6.495	5.256	4.219	2.502	1.5	0.956	0.527	0.2
QS-2-125.1	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
QS-2-125.2	12.958	8.72	7.393	6.016	4.936	3.859	2.289	1.436	0.941	0.428	0.178
QS-2-125.3	12.958	8.736	7.287	6.167	4.944	4.047	2.165	1.365	0.88	0.454	0.174

Table B.1 Simple Chi Path Indices for MP Analogues (continued)

Analogue	X0	X1	X2	Xp3	Xp4	Xp5	Xp6	Xp7	Xp8	Xp9	Xp10
QS-2-133.1	13.828	9.13	7.908	6.578	5.096	4.283	2.422	1.565	1.032	0.476	0.201
QS-2-147.2	12.958	8.736	7.287	6.167	4.944	4.047	2.165	1.365	0.88	0.454	0.174
QS-2-15.1	13.665	9.274	7.478	6.482	5.277	4.288	2.491	1.58	0.937	0.496	0.249
QS-2-29.4	13.665	9.258	7.562	6.44	5.158	4.102	2.426	1.713	1.039	0.523	0.237
QS-2-40.1	12.958	8.736	7.287	6.167	4.944	4.047	2.165	1.365	0.88	0.454	0.174
QS-2-41.2	12.958	8.72	7.393	6.016	4.936	3.859	2.289	1.436	0.941	0.428	0.178
QS-2-61.4	12.958	8.72	7.393	6.016	4.936	3.859	2.289	1.436	0.941	0.428	0.178
QS-2-71.3	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
QS-2-81.4	13.828	9.13	7.889	6.678	5.044	4.064	2.466	1.701	1.018	0.449	0.201
QS-2-84.4	15.242	10.206	8.271	7.269	5.855	4.554	2.777	2.064	1.353	0.666	0.288
QS-2-88.1	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
QS-2-99.3	12.958	8.736	7.287	6.167	4.944	4.047	2.165	1.365	0.88	0.454	0.174
WB47.4	12.958	8.736	7.287	6.167	4.944	4.047	2.165	1.365	0.88	0.454	0.174
WB48.4	12.958	8.72	7.393	6.016	4.936	3.859	2.289	1.436	0.941	0.428	0.178
WB61.4	12.958	8.72	7.381	6.1	4.776	3.924	2.189	1.562	0.952	0.401	0.173
WB71.5	12.958	8.72	7.393	6.016	4.936	3.859	2.289	1.436	0.941	0.428	0.178
WB77.2	11.38	7.809	6.502	5.309	4.484	3.372	1.905	1.149	0.739	0.313	0.178
XY-1-102.3	12.251	8.22	6.998	5.971	4.592	3.573	2.113	1.384	0.763	0.346	0.201
XY-1-127.5	12.958	8.72	7.378	6.122	4.73	3.912	2.318	1.564	0.937	0.472	0.201
XY-1-129.2	17.648	12.186	10.443	8.645	7.226	5.983	3.76	2.628	1.867	1.163	0.68
XY-1-144.4	17.648	12.165	10.527	8.676	7.019	5.985	3.581	2.613	1.867	1.151	0.659
XY-1-147.4	16.071	11.254	9.668	7.787	6.794	5.574	3.195	2.366	1.652	0.977	0.614
XY-1-30.3	15.2	10.881	8.918	7.605	6.492	5.297	3.014	2.212	1.533	0.945	0.595

Table B.1 Simple Chi Path Indices for MP Analogues (continued)

Analogue	X0	X1	X2	Xp3	Xp4	Xp5	Xp6	Xp7	Xp8	Xp9	Xp10
XY-1-44.5	15.908	11.381	9.299	7.756	6.639	5.587	3.291	2.387	1.687	1.073	0.627
XY-1-47.1	11.217	7.915	6.249	5.133	4.314	3.513	1.81	1.249	0.766	0.39	0.144
XY-1-85.7	18.518	12.597	10.939	9.307	7.334	6.188	3.941	2.874	1.97	1.186	0.721
XY-1-86.2	16.941	11.686	10.048	8.594	6.907	5.693	3.541	2.515	1.672	1.043	0.675
XY-1-89.5	17.648	12.186	10.428	8.745	7.054	5.989	3.787	2.703	1.882	1.179	0.708
XY-2-74.3	13.121	8.592	7.747	6.147	4.931	3.786	2.299	1.554	0.876	0.346	0.201
ZL102.3	15.908	11.343	9.432	7.8	6.519	5.246	2.839	2.097	1.471	0.878	0.537
ZL105.1	13.828	9.13	7.92	6.495	5.256	4.219	2.502	1.5	0.956	0.527	0.2
ZL21.1	12.795	8.771	7.431	5.588	4.507	3.698	2.214	1.535	0.967	0.533	0.28
ZL26.1	11.38	7.788	6.618	5.158	4.512	3.417	1.742	1.201	0.753	0.272	0.144
ZL38.1	13.665	9.258	7.55	6.508	5.087	4.034	2.378	1.735	1.142	0.512	0.213
ZL68.3	14.656	10.292	8.739	7.62	6.365	5.283	3.259	2.483	1.812	1.126	0.585
ZL77.2	15.458	9.931	9.339	7.036	5.475	4.21	2.637	1.972	1.378	0.646	0.263

Table B.2 Valence Chi Path Indices for MP Analogues

Analogues	Xv0	Xv1	Xv2	Xvp3	Xvp4	Xvp5	Xvp6	Xvp7	Xvp8	Xvp9	Xvp10
AL34.1	14.227	8.701	6.431	4.891	3.711	2.809	1.551	1.003	0.625	0.348	0.186
AN-1-68.2	16.158	9.589	6.961	5.233	3.941	2.977	1.665	1.09	0.678	0.387	0.221
BO-1-119.1	14.479	9.32	6.888	5.307	4.03	2.933	1.627	1.059	0.655	0.4	0.248
BO-1-12.1	11.918	7.079	5.149	3.957	3.01	2.229	1.172	0.688	0.371	0.178	0.083
BO-1-120.1	15.187	9.82	7.242	5.557	4.188	3.084	1.717	1.119	0.698	0.431	0.27
BO-1-122.1	13.772	8.82	6.534	5.084	3.817	2.805	1.541	0.998	0.611	0.369	0.212
BO-1-128.1	15.642	9.701	7.1	5.416	4.105	3.053	1.726	1.131	0.707	0.426	0.258
BO-1-13.1	15.284	9.179	7.011	5.173	3.976	2.921	1.694	1.064	0.679	0.395	0.217
BO-1-131.1	14.935	9.201	6.746	5.193	3.892	2.926	1.643	1.065	0.664	0.399	0.222
BO-1-144.1	16.349	10.201	7.453	5.666	4.263	3.204	1.816	1.189	0.754	0.457	0.278
BO-1-145.1	17.056	10.701	7.807	5.916	4.44	3.316	1.923	1.253	0.795	0.49	0.299
BO-1-146.1	17.763	11.201	8.16	6.166	4.617	3.441	2.002	1.328	0.84	0.519	0.323
BO-1-15.1	15.284	9.185	6.944	5.363	3.934	2.983	1.691	1.086	0.684	0.388	0.202
BO-1-17.1	12.125	7.253	5.258	4.026	3.089	2.299	1.228	0.729	0.392	0.186	0.09
BO-1-19.1	15.284	9.179	7.008	5.214	3.844	3.008	1.652	1.061	0.682	0.392	0.222
BO-1-21.1	15.414	9.2	6.869	5.196	3.851	2.921	1.656	1.054	0.662	0.38	0.213
BO-1-23.1	15.558	9.224	6.793	5.208	3.859	2.92	1.661	1.056	0.662	0.381	0.213
BO-1-30.1	15.19	9.329	7.664	5.663	4.452	3.174	1.809	1.273	0.828	0.483	0.27
BO-1-37.1	14.097	8.561	6.266	4.754	3.593	2.709	1.493	0.957	0.589	0.324	0.171
BO-1-43.1	14.097	8.551	6.306	4.771	3.619	2.726	1.509	0.968	0.593	0.323	0.17
BO-1-44.1	14.097	8.551	6.301	4.798	3.616	2.734	1.515	0.971	0.596	0.324	0.169
BO-1-45.1	13.481	8.185	6.024	4.602	3.496	2.532	1.438	0.923	0.553	0.294	0.148
BO-1-46.1	14.133	8.926	6.7	5.214	4.038	2.79	1.655	1.127	0.72	0.412	0.219
BO-1-47.1	14.133	8.888	6.849	5.29	4.079	2.86	1.697	1.155	0.731	0.403	0.216
BO-1-48.1	13.481	8.172	6.077	4.629	3.51	2.556	1.453	0.933	0.557	0.291	0.147
BO-1-96	10.679	6.763	4.891	3.959	3.015	2.155	1.094	0.612	0.306	0.164	0.081
BO-2-28.1	12.032	7.01	5.478	3.809	3.012	2.034	1.154	0.641	0.361	0.175	0.076
BO-2-40.1	12.299	7.144	5.633	3.886	3.121	2.08	1.21	0.663	0.376	0.185	0.082
BO-2-57.1	14.122	8.798	6.8	5.063	3.901	2.795	1.599	0.997	0.625	0.365	0.208
CE101.1	9.024	5.808	4.256	3.259	2.387	1.667	0.762	0.462	0.23	0.096	0.045
EGK-266/1	10.556	6.323	4.649	3.458	2.505	1.832	0.886	0.542	0.296	0.125	0.049
EGK-276-A	13.696	7.874	5.843	4.501	3.291	2.373	1.271	0.746	0.445	0.238	0.107
EGK-276-B	11.503	6.702	5.136	3.919	2.846	2.099	1.061	0.58	0.342	0.162	0.066
LL81.2	11.373	6.688	4.906	3.669	2.626	1.893	0.956	0.6	0.335	0.15	0.062
QS-1-114.1	10.686	6.388	4.727	3.476	2.575	1.847	0.915	0.546	0.3	0.133	0.052
QS-1-128.1	10.686	6.388	4.724	3.502	2.526	1.857	0.906	0.556	0.305	0.13	0.052
QS-1-138.1	11.517	6.711	4.83	3.681	2.634	1.892	0.958	0.603	0.337	0.151	0.062
QS-1-142.1	11.243	6.666	5.045	3.687	2.619	1.964	0.995	0.615	0.342	0.148	0.066
QS-1-89.4	10.186	6.188	4.468	3.365	2.483	1.796	0.845	0.52	0.283	0.12	0.045

Table B.2 Valence Chi Path Indices for MP Analogues (continued)

Analogues	Xv0	Xv1	Xv2	Xvp3	Xvp4	Xvp5	Xvp6	Xvp7	Xvp8	Xvp9	Xvp10
QS-2-116.3	13.609	7.849	6.929	4.342	3.013	2.092	1.183	0.787	0.459	0.228	0.101
QS-2-124.2	12.19	7.045	5.536	4.11	3.063	2.218	1.182	0.658	0.367	0.2	0.079
QS-2-125.1	12.644	7.367	5.854	4.154	2.853	2.234	1.218	0.764	0.436	0.195	0.099
QS-2-125.2	12.073	7.081	5.528	3.907	2.94	2.105	1.158	0.668	0.383	0.189	0.082
QS-2-125.3	11.134	6.568	4.955	3.825	2.825	2.064	1.02	0.562	0.321	0.159	0.062
QS-2-133.1	12.056	6.978	5.455	4.103	2.938	2.206	1.149	0.637	0.392	0.18	0.079
QS-2-147.2	12.073	7.087	5.433	4.168	3.057	2.245	1.117	0.625	0.369	0.187	0.063
QS-2-15.1	11.517	6.717	4.799	3.68	2.722	1.972	0.982	0.585	0.314	0.149	0.061
QS-2-29.4	11.517	6.711	4.834	3.659	2.658	1.923	0.961	0.603	0.326	0.151	0.065
QS-2-40.1	10.556	6.329	4.616	3.478	2.549	1.846	0.88	0.524	0.289	0.129	0.046
QS-2-41.2	10.556	6.323	4.652	3.435	2.541	1.823	0.892	0.535	0.292	0.128	0.05
QS-2-61.4	11.243	6.666	5.049	3.649	2.722	1.95	1.013	0.595	0.333	0.155	0.064
QS-2-71.3	10.487	6.288	4.609	3.435	2.493	1.818	0.875	0.535	0.292	0.123	0.048
QS-2-81.4	12.299	7.15	5.545	4.248	2.835	2.102	1.136	0.712	0.385	0.179	0.083
QS-2-84.4	12.848	7.24	5.172	3.973	2.869	2.048	1.059	0.678	0.382	0.181	0.08
QS-2-88.1	12.073	7.081	5.524	3.964	2.757	2.124	1.127	0.704	0.398	0.176	0.085
QS-2-99.3	11.243	6.672	4.986	3.79	2.779	2.027	0.987	0.57	0.325	0.155	0.054
WB47.4	10.487	6.294	4.579	3.446	2.526	1.828	0.869	0.52	0.285	0.126	0.045
WB48.4	10.487	6.288	4.612	3.414	2.523	1.81	0.88	0.529	0.288	0.125	0.048
WB61.4	11.109	6.599	4.968	3.642	2.597	1.938	0.974	0.601	0.333	0.144	0.062
WB71.5	11.109	6.599	4.971	3.607	2.686	1.925	0.989	0.583	0.325	0.15	0.062
WB77.2	9.325	5.907	4.4	3.308	2.424	1.686	0.795	0.468	0.236	0.1	0.048
XY-1-102.3	11.137	6.769	5.333	4.142	2.737	1.989	1.067	0.635	0.313	0.151	0.083
XY-1-127.5	12.098	7.15	5.606	4.291	2.896	2.188	1.19	0.738	0.393	0.191	0.083
XY-1-129.2	15.284	9.179	7.011	5.176	3.95	2.964	1.713	1.092	0.679	0.394	0.209
XY-1-144.4	15.266	9.207	7.14	5.143	3.876	2.995	1.696	1.095	0.691	0.384	0.197
XY-1-147.4	13.397	8.377	6.31	4.743	3.645	2.711	1.464	0.953	0.584	0.322	0.178
XY-1-30.3	13.065	8.32	6.219	4.782	3.636	2.684	1.456	0.937	0.568	0.319	0.178
XY-1-44.5	14.026	8.701	6.492	4.931	3.796	2.876	1.61	1.032	0.639	0.36	0.191
XY-1-47.1	9.985	6.188	4.529	3.408	2.547	1.866	0.891	0.547	0.286	0.131	0.045
XY-1-85.7	16.341	9.662	7.508	5.774	4.063	3.116	1.838	1.203	0.745	0.422	0.235
XY-1-86.2	15.178	9.282	7.296	5.665	3.985	3.007	1.756	1.118	0.668	0.389	0.228
XY-1-89.5	16.139	9.662	7.569	5.814	4.145	3.199	1.905	1.232	0.764	0.436	0.241
XY-2-74.3	11.469	6.826	5.424	4.099	2.771	1.991	1.074	0.658	0.337	0.151	0.083
ZL102.3	13.28	8.333	6.038	4.387	3.175	2.254	1.074	0.689	0.411	0.217	0.116
ZL105.1	12.056	6.978	5.459	4.068	3.028	2.193	1.159	0.644	0.359	0.194	0.076
ZL21.1	10.893	6.688	4.876	3.528	2.574	1.886	0.971	0.598	0.32	0.155	0.069
ZL26.1	9.355	5.865	4.347	3.216	2.419	1.686	0.774	0.474	0.251	0.096	0.045
ZL38.1	11.816	7.16	5.152	3.953	2.784	2	1.063	0.682	0.389	0.181	0.079
ZL68.3	12.341	7.593	5.664	4.339	3.224	2.383	1.261	0.83	0.513	0.281	0.139
ZL77.2	11.743	6.916	5.177	3.803	2.702	1.936	1.003	0.638	0.36	0.165	0.07

Table B.3 Simple Cluster and Path Cluster Chi Path Indices for MP Analogues

Analogue	Xc3	Xc4	Xpc4	Xch5	Xch6	Xvc3	Xvc4	Xvpc4	Xvch5	Xvch6
AL34.1	1.02	0	2.706	0	0.287	0.58	0	1.415	0	0.129
AN-1-68.2	1.224	0	3.139	0	0.269	0.655	0	1.538	0	0.124
BO-1-119.1	0.81	0	2.299	0	0.287	0.566	0	1.478	0	0.129
BO-1-12.1	0.816	0	2.321	0	0.185	0.462	0	1.226	0	0.097
BO-1-120.1	0.81	0	2.299	0	0.287	0.566	0	1.478	0	0.129
BO-1-122.1	0.81	0	2.299	0	0.287	0.566	0	1.478	0	0.129
BO-1-128.1	1.02	0	2.754	0	0.287	0.58	0	1.472	0	0.129
BO-1-13.1	1.309	0	3.05	0	0.269	0.769	0	1.609	0	0.124
BO-1-131.1	1.02	0	2.754	0	0.287	0.58	0	1.472	0	0.129
BO-1-144.1	1.02	0	2.754	0	0.287	0.58	0	1.472	0	0.129
BO-1-145.1	1.02	0	2.754	0	0.287	0.58	0	1.472	0	0.129
BO-1-146.1	1.02	0	2.754	0	0.287	0.58	0	1.472	0	0.129
BO-1-15.1	1.219	0	3.298	0	0.269	0.728	0	1.81	0	0.124
BO-1-17.1	0.816	0	2.321	0	0.185	0.462	0	1.236	0	0.097
BO-1-19.1	1.309	0	3.114	0	0.269	0.769	0	1.633	0	0.124
BO-1-21.1	1.52	0	3.746	0	0.269	0.692	0	1.605	0	0.124
BO-1-23.1	1.224	0	3.199	0	0.269	0.648	0	1.561	0	0.124
BO-1-30.1	1.309	0	3.05	0.118	0.185	1.026	0	1.93	0.088	0.097
BO-1-37.1	1.02	0	2.706	0	0.287	0.553	0	1.372	0	0.121
BO-1-43.1	1.02	0	2.706	0	0.287	0.58	0	1.399	0	0.121
BO-1-44.1	1.02	0	2.706	0	0.287	0.58	0	1.415	0	0.121
BO-1-45.1	1.02	0	2.706	0.144	0.185	0.546	0	1.359	0.039	0.097
BO-1-46.1	1.02	0	2.706	0.144	0.185	0.58	0	1.472	0.102	0.097
BO-1-47.1	1.02	0	2.706	0.144	0.185	0.679	0	1.573	0.102	0.097

Table B.3 Simple Cluster and Path Cluster Chi Path Indices for MP Analogues (continued)

Analogue	Xc3	Xc4	Xpc4	Xch5	Xch6	Xvc3	Xvc4	Xvpc4	Xvch5	Xvch6
BO-1-48.1	1.02	0	2.706	0.144	0.185	0.58	0	1.395	0.039	0.097
BO-1-96	0.605	0	1.915	0	0.185	0.449	0	1.297	0	0.097
BO-2-28.1	1.257	0	2.492	0	0.17	0.679	0	1.183	0	0.096
BO-2-40.1	1.257	0	2.492	0	0.17	0.723	0	1.227	0	0.096
BO-2-57.1	1.098	0	2.596	0	0.269	0.755	0	1.615	0	0.124
CE101.1	0.469	0	1.421	0	0.204	0.332	0	0.875	0	0.104
EGK-266/1	0.969	0	2.276	0	0.185	0.42	0	0.95	0	0.1
EGK-276-A	1.219	0	3.329	0	0.151	0.677	0	1.64	0	0.089
EGK-276-B	1.174	0	2.789	0	0.167	0.59	0	1.385	0	0.092
LL81.2	1.18	0	2.909	0	0.185	0.457	0	1.054	0	0.1
QS-1-114.1	0.969	0	2.217	0	0.185	0.442	0	0.96	0	0.1
QS-1-128.1	0.969	0	2.276	0	0.185	0.442	0	0.975	0	0.1
QS-1-138.1	0.884	0	2.361	0	0.185	0.414	0	1.011	0	0.1
QS-1-142.1	0.969	0	2.276	0	0.185	0.535	0	1.082	0	0.1
QS-1-89.4	0.68	0	1.868	0	0.204	0.346	0	0.864	0	0.104
QS-2-116.3	2.191	0.289	3.73	0	0.185	1.679	0.25	2.076	0	0.1
QS-2-124.2	1.174	0	2.73	0	0.167	0.705	0	1.495	0	0.092
QS-2-125.1	0.969	0	2.276	0	0.185	0.768	0	1.352	0	0.1
QS-2-125.2	0.969	0	2.217	0	0.185	0.673	0	1.209	0	0.1
QS-2-125.3	0.885	0	2.381	0	0.185	0.516	0	1.299	0	0.097
QS-2-133.1	1.174	0	2.789	0	0.167	0.682	0	1.491	0	0.092
QS-2-147.2	0.885	0	2.381	0	0.185	0.616	0	1.484	0	0.1
QS-2-15.1	0.816	0	2.37	0	0.185	0.392	0	1.024	0	0.1
QS-2-29.4	0.884	0	2.313	0	0.185	0.414	0	0.998	0	0.1

Table B.3 Simple Cluster and Path Cluster Chi Path Indices for MP Analogues (continued)

Analogue	Xc3	Xc4	Xpc4	Xch5	Xch6	Xvc3	Xvc4	Xvpc4	Xvch5	Xvch6
QS-2-40.1	0.885	0	2.381	0	0.185	0.397	0	0.978	0	0.1
QS-2-41.2	0.969	0	2.217	0	0.185	0.42	0	0.937	0	0.1
QS-2-61.4	0.969	0	2.217	0	0.185	0.535	0	1.06	0	0.1
QS-2-71.3	0.969	0	2.276	0	0.185	0.409	0	0.937	0	0.1
QS-2-81.4	1.152	0	2.954	0	0.17	0.673	0	1.593	0	0.096
QS-2-84.4	1.013	0	2.865	0	0.17	0.463	0	1.154	0	0.096
QS-2-88.1	0.969	0	2.276	0	0.185	0.673	0	1.242	0	0.1
QS-2-99.3	0.885	0	2.381	0	0.185	0.496	0	1.207	0	0.1
WB47.4	0.885	0	2.381	0	0.185	0.387	0	0.955	0	0.1
WB48.4	0.969	0	2.217	0	0.185	0.409	0	0.925	0	0.1
WB61.4	0.969	0	2.276	0	0.185	0.512	0	1.057	0	0.1
WB71.5	0.969	0	2.217	0	0.185	0.512	0	1.036	0	0.1
WB77.2	0.758	0	1.77	0	0.185	0.395	0	0.935	0	0.1
XY-1-102.3	0.941	0	2.506	0	0.17	0.659	0	1.604	0	0.096
XY-1-127.5	0.941	0	2.467	0	0.17	0.659	0	1.599	0	0.096
XY-1-129.2	1.309	0	3.055	0	0.269	0.769	0	1.611	0	0.124
XY-1-144.4	1.31	0	3.317	0	0.287	0.808	0	1.701	0	0.129
XY-1-147.4	1.118	0	2.616	0	0.287	0.6	0	1.402	0	0.129
XY-1-30.3	0.81	0	2.251	0	0.287	0.566	0	1.421	0	0.129
XY-1-44.5	0.81	0	2.211	0	0.287	0.566	0	1.416	0	0.129
XY-1-47.1	0.469	0	1.381	0	0.204	0.332	0	0.87	0	0.104
XY-1-85.7	1.492	0	3.792	0	0.253	0.907	0	2.144	0	0.121
XY-1-86.2	1.281	0	3.337	0	0.253	0.894	0	2.15	0	0.121
XY-1-89.5	1.281	0	3.297	0	0.253	0.894	0	2.145	0	0.121

Table B.3 Simple Cluster and Path Cluster Chi Path Indices for MP Analogues (continued)

Analogue	Xc3	Xc4	Xpc4	Xch5	Xch6	Xvc3	Xvc4	Xvpc4	Xvch5	Xvch6
XY-2-74.3	1.249	0	2.864	0	0.17	0.693	0	1.581	0	0.096
ZL102.3	0.884	0	2.232	0	0.306	0.463	0	1.034	0	0.136
ZL105.1	1.174	0	2.73	0	0.167	0.682	0	1.471	0	0.092
ZL21.1	0.878	0	1.669	0	0.204	0.415	0	0.929	0	0.104
ZL26.1	0.778	0	1.778	0	0.204	0.365	0	0.852	0	0.104
ZL38.1	0.884	0	2.361	0	0.185	0.463	0	1.118	0	0.1
ZL68.3	1.013	0	2.768	0	0.253	0.512	0	1.235	0	0.124
ZL77.2	2.191	0.289	3.73	0	0.185	0.563	0.013	1.179	0	0.1

Table B.4 Kappa and Phi Indices for MP Analogues

Analogues	k0	k1	k2	k3	ka1	ka2	ka3	phia
AL34.1	30.717	18.781	9.63	5.019	16.873	8.224	4.123	5.782
AN-1-68.2	36.239	21.703	11.253	5.99	19.588	9.675	4.964	7.019
BO-1-119.1	30.717	18.781	10.222	5.497	17.192	9.003	4.691	6.449
BO-1-12.1	24.816	16.372	8.444	4.25	14.772	7.232	3.487	5.342
BO-1-120.1	32.54	19.753	10.983	6	18.16	9.742	5.166	7.077
BO-1-122.1	28.911	17.811	9.475	4.989	16.226	8.279	4.213	5.841
BO-1-128.1	34.381	20.727	11.111	5.997	18.81	9.652	5.033	6.983
BO-1-13.1	33.744	19.753	9.796	5.258	18.121	8.613	4.487	6.243
BO-1-131.1	32.54	19.753	10.364	5.49	17.84	8.931	4.559	6.373
BO-1-144.1	36.239	21.703	11.87	6.5	19.782	10.386	5.504	7.61
BO-1-145.1	38.112	22.68	12.64	7.039	20.756	11.134	6.012	8.253
BO-1-146.1	40.001	23.659	13.42	7.571	21.732	11.893	6.516	8.912
BO-1-15.1	33.744	19.753	9.796	5.042	18.121	8.613	4.295	6.243
BO-1-17.1	24.816	16.372	8.444	4.25	14.949	7.364	3.568	5.504
BO-1-19.1	32.54	19.753	9.796	5.258	18.121	8.613	4.487	6.243
BO-1-21.1	35.637	21.703	10.684	5.758	19.354	8.992	4.657	6.445
BO-1-23.1	34.381	20.727	10.519	5.518	18.771	9.08	4.597	6.556
BO-1-30.1	31.921	18.781	9.087	4.803	17.618	8.256	4.268	6.06
BO-1-37.1	31.921	18.781	9.63	5.019	16.805	8.175	4.093	5.724
BO-1-43.1	31.921	18.781	9.63	5.019	16.805	8.175	4.093	5.724
BO-1-44.1	30.717	18.781	9.63	5.019	16.805	8.175	4.093	5.724
BO-1-45.1	30.116	17.811	8.909	4.545	16.12	7.684	3.778	5.385
BO-1-46.1	30.116	17.811	8.909	4.545	16.371	7.863	3.889	5.597
BO-1-47.1	30.116	17.811	8.909	4.545	16.371	7.863	3.889	5.597
BO-1-48.1	30.116	17.811	8.909	4.545	16.12	7.684	3.778	5.385
BO-1-96	19.714	13.432	6.805	3.263	12.593	6.175	2.878	4.575
BO-2-28.1	22.49	15.39	7.136	3.986	14.224	6.307	3.425	4.722
BO-2-40.1	22.49	15.39	7.136	3.986	14.792	6.708	3.694	5.222
BO-2-57.1	30.116	17.811	8.909	4.759	16.506	7.96	4.14	5.712
CE101.1	16.437	11.484	5.915	2.982	10.651	5.286	2.58	3.754
EGK-266/1	21.391	14.41	6.963	3.75	13.208	6.089	3.173	4.468
EGK-276-A	31.32	19.326	9.475	4.759	18.035	8.516	4.158	6.678
EGK-276-B	23.092	15.39	7.136	3.762	14.185	6.28	3.208	4.689
LL81.2	24.214	16.372	7.852	4.25	14.772	6.694	3.487	4.944
QS-1-114.1	22.595	14.41	6.963	3.75	13.208	6.089	3.173	4.468
QS-1-128.1	21.391	14.41	6.963	3.75	13.208	6.089	3.173	4.468
QS-1-138.1	23.092	15.39	7.695	3.986	14.185	6.797	3.407	5.075
QS-1-142.1	21.391	14.41	6.963	3.75	13.53	6.32	3.324	4.751
QS-1-89.4	19.714	13.432	6.805	3.484	12.272	5.938	2.928	4.286
QS-2-116.3	25.131	17.355	7.513	4.488	16.185	6.714	3.923	5.175
QS-2-124.2	24.296	15.39	7.136	3.762	14.508	6.507	3.353	4.968

Table B.4 Kappa and Phi Indices for MP Analogues (continued)

Analogues	k0	k1	k2	k3	ka1	ka2	ka3	phia
QS-2-125.1	21.391	14.41	6.963	3.75	13.96	6.633	3.53	5.144
QS-2-125.2	22.595	14.41	6.963	3.75	13.715	6.455	3.412	4.918
QS-2-125.3	21.391	14.41	6.963	3.526	13.247	6.117	2.993	4.501
QS-2-133.1	23.092	15.39	7.136	3.762	14.224	6.307	3.225	4.722
QS-2-147.2	22.595	14.41	6.963	3.526	13.715	6.455	3.204	4.918
QS-2-15.1	24.296	15.39	7.695	3.762	14.185	6.797	3.208	5.075
QS-2-29.4	24.296	15.39	7.695	3.986	14.185	6.797	3.407	5.075
QS-2-40.1	22.595	14.41	6.963	3.526	13.208	6.089	2.976	4.468
QS-2-41.2	22.595	14.41	6.963	3.75	13.208	6.089	3.173	4.468
QS-2-61.4	22.595	14.41	6.963	3.75	13.53	6.32	3.324	4.751
QS-2-71.3	21.391	14.41	6.963	3.75	13.178	6.068	3.159	4.442
QS-2-81.4	24.296	15.39	7.136	3.762	14.792	6.708	3.483	5.222
QS-2-84.4	27.767	17.355	8.585	4.26	16.107	7.657	3.683	5.873
QS-2-88.1	21.391	14.41	6.963	3.75	13.715	6.455	3.412	4.918
QS-2-99.3	22.595	14.41	6.963	3.526	13.53	6.32	3.12	4.751
WB47.4	22.595	14.41	6.963	3.526	13.178	6.068	2.963	4.442
WB48.4	22.595	14.41	6.963	3.75	13.178	6.068	3.159	4.442
WB61.4	21.391	14.41	6.963	3.75	13.247	6.117	3.191	4.501
WB71.5	22.595	14.41	6.963	3.75	13.247	6.117	3.191	4.501
WB77.2	19.266	12.457	6.074	3.25	11.552	5.414	2.814	3.909
XY-1-102.3	20.918	13.432	6.25	3.263	13.159	6.054	3.136	4.686
XY-1-127.5	22.595	14.41	6.963	3.75	14.136	6.762	3.615	5.31
XY-1-129.2	33.744	19.753	9.796	5.258	18.121	8.613	4.487	6.243
XY-1-144.4	31.938	19.753	9.796	5.258	17.84	8.413	4.359	6.004
XY-1-147.4	28.911	17.811	8.909	4.759	15.908	7.533	3.868	5.211
XY-1-30.3	27.125	16.844	8.741	4.521	15.263	7.571	3.777	5.252
XY-1-44.5	28.911	17.811	9.475	4.989	16.226	8.279	4.213	5.841
XY-1-47.1	18.062	12.457	6.667	3.495	11.62	6.019	3.068	4.371
XY-1-85.7	35.585	20.727	9.972	5.299	19.374	9.005	4.675	6.71
XY-1-86.2	31.921	18.781	9.087	4.803	17.753	8.352	4.329	6.178
XY-1-89.5	33.744	19.753	9.796	5.258	18.723	9.045	4.766	6.774
XY-2-74.3	22.595	14.41	6.438	3.526	13.813	6.022	3.249	4.621
ZL102.3	28.911	17.811	9.475	5.235	15.908	8.044	4.274	5.563
ZL105.1	24.296	15.39	7.136	3.762	14.224	6.307	3.225	4.722
ZL21.1	21.391	14.41	7.556	4.566	13.247	6.665	3.917	4.905
ZL26.1	18.062	12.457	6.074	3.25	11.3	5.233	2.696	3.696
ZL38.1	23.092	15.39	7.695	3.986	14.224	6.826	3.425	5.11
ZL68.3	27.767	15.879	7.513	3.673	14.237	6.373	2.99	4.321
ZL77.2	25.131	17.355	7.513	4.488	15.979	6.576	3.826	5.004

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 1-8)

Analogues	SHsOH	SHsNH2	SHssNH	SHtCH	SHdCH2	SHdsCH	SHaaCH	SHCsats
AL34.1	0	0	0	0	0	0	12.359	4.609
AN-1-68.2	0	0	0	0	0	0	11.852	4.756
BO-1-119.1	2.472	0	0	0	0	0	11.743	5.624
BO-1-12.1	0	0	0	1.57	0	0	6.26	4.486
BO-1-120.1	2.47	0	0	0	0	0	11.694	6.166
BO-1-122.1	2.475	0	0	0	0	0	11.814	5.03
BO-1-128.1	0	0	0	0	0	0	12.133	6.015
BO-1-13.1	0	0	0	0	0	0	11.486	4.689
BO-1-131.1	0	0	0	0	0	0	12.223	5.38
BO-1-144.1	0	0	0	0	0	0	12.07	6.589
BO-1-145.1	0	0	0	0	0	0	12.025	7.128
BO-1-146.1	0	0	0	0	0	0	11.99	7.646
BO-1-15.1	0	0	0	0	0	0	11.397	4.715
BO-1-17.1	0	0	0	0	1.007	1.115	6.222	4.382
BO-1-19.1	0	0	0	0	0	0	11.532	4.672
BO-1-21.1	0	0	0	0	0	0	12.152	4.831
BO-1-23.1	0	0	0	0	0	0	11.707	5.432
BO-1-30.1	0	0	0	0	0	0	8.896	4.662
BO-1-37.1	0	0	0	0	0	0	11.412	4.693
BO-1-43.1	0	0	0	0	0	0	11.53	4.67
BO-1-44.1	0	0	0	0	0	0	11.583	4.655
BO-1-45.1	0	0	0	0	0	0	10.252	4.694
BO-1-46.1	0	0	0	0	0	0	9.959	4.578
BO-1-47.1	0	0	0	0	0	0	9.893	4.581
BO-1-48.1	0	0	0	0	0	0	10.401	4.659
BO-1-96	2.444	0	0	0	0	0	5.865	5.023
BO-2-28.1	0	0	1.625	0	0	0	3.826	4.156
BO-2-40.1	0	0	1.653	0	0	0	4.175	4.299
BO-2-57.1	2.487	0	0	0	0	0	11.076	4.402
CE101.1	2.434	0	1.538	0	0	0	5.825	3.798
EGK-266/1	2.57	0	1.636	0	0	0	5.385	4.213
EGK-276-A	5.135	0	0	0	0	0	4.404	6.231
EGK-276-B	2.573	0	0	0	0	0	5.405	4.938
LL81.2	0	0	1.664	0	0	0	5.806	4.363
QS-1-114.1	0	1.597	1.631	0	0	0	5.154	4.188
QS-1-128.1	0	1.57	1.628	0	0	0	5.212	4.171
QS-1-138.1	0	0	1.638	0	0	0	5.426	4.946
QS-1-142.1	0	0	1.631	0	0	0	5.27	4.185
QS-1-89.4	0	0	1.617	0	0	0	6.132	4.113
QS-2-116.3	0	0	1.628	0	0	0	5.161	5.714

Table B.5 5 Atom Type Electrotopological State Indices for MP Analogues (Columns 1-8. Continued)

Analogue	SHsOH	SHsNH2	SHssNH	SHtCH	SHdCH2	SHdsCH	SHaaCH	SHCsats
QS-2-124.2	0	0	0	0	0	0	5.226	4.93
QS-2-125.1	0	0	1.624	0	0	0	5.121	4.149
QS-2-125.2	0	0	1.629	0	0	0	5.113	4.175
QS-2-125.3	0	0	0	0	0	0	6.155	4.823
QS-2-133.1	0	0	0	0	0	0	5.057	4.843
QS-2-147.2	0	0	1.633	0	0	0	5.012	4.196
QS-2-15.1	0	0	1.655	0	0	0	5.219	5.073
QS-2-29.4	0	0	1.645	0	0	0	5.355	4.997
QS-2-40.1	2.639	0	1.65	0	0	0	5.185	4.286
QS-2-41.2	2.597	0	1.642	0	0	0	5.316	4.242
QS-2-61.4	0	0	1.635	0	0	0	5.208	4.206
QS-2-71.3	0	0	1.644	0	0	0	5.559	4.254
QS-2-81.4	0	0	1.648	0	0	0	4.153	4.278
QS-2-84.4	0	0	1.666	0	0	0	4.357	5.937
QS-2-88.1	0	0	1.626	0	0	0	5.168	4.161
QS-2-99.3	0	0	1.641	0	0	0	5.093	4.238
WB47.4	0	0	1.664	0	0	0	5.323	4.359
WB48.4	0	0	1.652	0	0	0	5.479	4.295
WB61.4	0	0	1.62	0	0	0	5.038	4.129
WB71.5	0	0	1.621	0	0	0	4.991	4.134
WB77.2	2.469	0	1.573	0	0	0	5.226	4.001
XY-1-102.3	2.465	0	1.569	0	0	0	3.94	3.98
XY-1-127.5	0	0	1.575	0	0	0	3.955	4.699
XY-1-129.2	0	0	0	0	0	0	11.438	4.702
XY-1-144.4	0	0	0	0	0	0	12.281	3.744
XY-1-147.4	0	1.702	0	0	0	0	12.215	3.694
XY-1-30.3	2.479	0	0	0	0	0	11.925	4.318
XY-1-44.5	0	0	0	0	0	0	11.958	5.047
XY-1-47.1	0	0	1.544	0	0	0	5.848	4.492
XY-1-85.7	0	0	0	0	0	0	10.386	4.774
XY-1-86.2	2.51	0	0	0	0	0	10.046	4.5
XY-1-89.5	0	0	0	0	0	0	10.072	5.254
XY-2-74.3	0	1.689	1.623	0	0	0	4.084	3.374
ZL102.3	0	0	1.656	0	0	0	12.445	3.523
ZL105.1	0	0	0	0	0	0	5.01	4.848
ZL21.1	0	0	1.584	0	0	0	6.015	4.158
ZL26.1	0	1.658	1.592	0	0	0	6.032	3.233
ZL38.1	0	0	1.623	0	0	0	5.079	4.613
ZL68.3	0	0	1.646	0	0	0	9.046	4.266
ZL77.2	0	0	1.683	0	0	0	6.084	4.467

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 9-16. Continued)

Analogue	SHCsatu	SHother	SsCH3	SdCH2	SssCH2	StCH	SdsCH	SaaCH
AL34.1	2.086	12.359	1.489	0	5.287	0	0	20.549
AN-1-68.2	2.163	11.852	1.473	0	5.037	0	0	17.992
BO-1-119.1	1.619	11.743	0	0	8.631	0	0	21.303
BO-1-12.1	2.031	6.26	1.451	0	4.795	5.461	0	9.853
BO-1-120.1	1.589	11.694	0	0	10.007	0	0	21.355
BO-1-122.1	1.674	11.814	0	0	7.294	0	0	21.239
BO-1-128.1	1.885	12.133	1.501	0	7.742	0	0	20.757
BO-1-13.1	2.134	11.486	1.475	0	5.065	0	0	17.941
BO-1-131.1	1.949	12.223	1.495	0	6.469	0	0	20.667
BO-1-144.1	1.849	12.07	1.505	0	9.068	0	0	20.828
BO-1-145.1	1.825	12.025	1.509	0	10.43	0	0	20.886
BO-1-146.1	1.809	11.99	1.512	0	11.818	0	0	20.935
BO-1-15.1	2.158	11.397	1.472	0	4.98	0	0	17.9
BO-1-17.1	1.924	8.344	1.471	3.828	5.232	0	1.914	9.961
BO-1-19.1	2.121	11.532	1.477	0	5.118	0	0	17.954
BO-1-21.1	2.202	12.152	1.432	0	4.612	0	0	16.406
BO-1-23.1	2.14	11.707	3.158	0	5.135	0	0	18.189
BO-1-30.1	2.121	8.896	1.479	0	5.205	0	0	14.074
BO-1-37.1	2.155	11.412	1.476	0	5.06	0	0	17.794
BO-1-43.1	2.127	11.53	1.479	0	5.141	0	0	17.749
BO-1-44.1	2.114	11.583	1.48	0	5.183	0	0	17.743
BO-1-45.1	2.165	10.252	1.47	0	4.999	0	0	15.547
BO-1-46.1	2.068	9.959	1.493	0	5.396	0	0	16.694
BO-1-47.1	2.072	9.893	1.494	0	5.428	0	0	16.564
BO-1-48.1	2.123	10.401	1.474	0	5.119	0	0	15.44
BO-1-96	0.874	5.865	2.22	0	6.347	0	0	10.456
BO-2-28.1	2.315	3.826	5.61	0	4.392	0	0	6.335
BO-2-40.1	1.159	4.175	1.407	0	4.122	0	0	5.228
BO-2-57.1	1.844	11.076	0	0	5.778	0	0	18.515
CE101.1	0.853	5.825	0	0	5.056	0	0	10.322
EGK-266/1	1.123	5.385	1.415	0	4.195	0	0	6.78
EGK-276-A	2.216	4.404	3.421	0	3.416	0	0	5.235
EGK-276-B	1.135	5.405	3.485	0	4.324	0	0	6.848
LL81.2	1.168	5.806	1.365	0	3.949	0	0	6.119
QS-1-114.1	1.117	5.154	1.435	0	4.268	0	0	7.502
QS-1-128.1	1.109	5.212	1.44	0	4.295	0	0	7.475
QS-1-138.1	1.127	5.426	3.081	0	4.3	0	0	7.648
QS-1-142.1	1.113	5.27	1.439	0	4.29	0	0	7.436
QS-1-89.4	1.089	6.132	1.459	0	4.4	0	0	9.893
QS-2-116.3	1.107	5.161	8.065	0	4.373	0	0	8.43

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 9-16. Continued)

Analogue	SHCsatu	SHother	SsCH3	SdCH2	SssCH2	StCH	SdsCH	SaaCH
QS-2-124.2	1.136	5.226	3.521	0	4.39	0	0	7.53
QS-2-125.1	1.101	5.121	1.463	0	4.39	0	0	8.128
QS-2-125.2	1.112	5.113	1.455	0	4.348	0	0	7.908
QS-2-125.3	1.102	6.155	3.573	0	4.529	0	0	9.971
QS-2-133.1	1.68	5.057	5.643	0	4.525	0	0	8.237
QS-2-147.2	1.125	5.012	1.453	0	4.333	0	0	7.866
QS-2-15.1	1.172	5.219	3.065	0	4.238	0	0	7.659
QS-2-29.4	1.143	5.355	3.073	0	4.275	0	0	7.664
QS-2-40.1	1.164	5.185	1.387	0	4.059	0	0	6.969
QS-2-41.2	1.137	5.316	1.404	0	4.14	0	0	6.852
QS-2-61.4	1.124	5.208	1.433	0	4.261	0	0	7.466
QS-2-71.3	1.136	5.559	1.391	0	4.094	0	0	6.086
QS-2-81.4	1.148	4.153	1.412	0	4.151	0	0	5.311
QS-2-84.4	1.18	4.357	4.616	0	4.175	0	0	5.586
QS-2-88.1	1.105	5.168	1.455	0	4.358	0	0	7.909
QS-2-99.3	1.143	5.093	1.426	0	4.218	0	0	7.49
WB47.4	1.195	5.323	1.346	0	3.891	0	0	6.417
WB48.4	1.157	5.479	1.372	0	4.012	0	0	6.203
WB61.4	1.665	5.038	3.516	0	4.396	0	0	8.169
WB71.5	1.694	4.991	3.514	0	4.396	0	0	8.151
WB77.2	0.921	5.226	0	0	4.501	0	0	6.555
XY-1-102.3	0.912	3.94	0	0	4.703	0	0	5.608
XY-1-127.5	0.924	3.955	1.743	0	5.521	0	0	5.878
XY-1-129.2	2.134	11.438	1.463	0	5.114	0	0	18.018
XY-1-144.4	3.429	12.281	3.727	0	5.451	0	0	20.882
XY-1-147.4	2	12.215	0	0	5.268	0	0	20.438
XY-1-30.3	1.795	11.925	0	0	6.028	0	0	21.156
XY-1-44.5	1.812	11.958	1.82	0	6.904	0	0	21.711
XY-1-47.1	0.866	5.848	1.79	0	5.873	0	0	10.712
XY-1-85.7	2.169	10.386	1.442	0	4.977	0	0	15.779
XY-1-86.2	1.878	10.046	0	0	5.615	0	0	16.255
XY-1-89.5	1.895	10.072	1.773	0	6.491	0	0	16.69
XY-2-74.3	1.08	4.084	0	0	4.135	0	0	5.257
ZL102.3	2.21	12.445	0	0	4.653	0	0	19.809
ZL105.1	1.71	5.01	5.645	0	4.525	0	0	8.215
ZL21.1	1.658	6.015	1.469	0	5.179	0	0	10.327
ZL26.1	1.021	6.032	0	0	4.384	0	0	9.815
ZL38.1	1.719	5.079	3.605	0	5.419	0	0	8.343
ZL68.3	1.141	9.046	1.471	0	4.346	0	0	14.49
ZL77.2	1.199	6.084	1.294	0	3.642	0	0	4.74

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 17-24. Continued)

Analogue	SsssCH	SddC	StsC	SdssC	SaasC	SaaaC	SssssC	SsNH2
AL34.1	-0.033	0	0	-0.133	2.343	0	0	0
AN-1-68.2	-0.137	2.39	0	-0.163	3.023	0	0	0
BO-1-119.1	0.713	0	0	0	2.702	0	0	0
BO-1-12.1	-0.121	0	2.704	-0.176	1.008	0	0	0
BO-1-120.1	0.73	0	0	0	2.724	0	0	0
BO-1-122.1	0.687	0	0	0	2.67	0	0	0
BO-1-128.1	0.035	0	0	-0.115	2.452	0	0	0
BO-1-13.1	-0.117	0	0	-0.158	2.951	0	0	0
BO-1-131.1	0.009	0	0	-0.122	2.409	0	0	0
BO-1-144.1	0.052	0	0	-0.109	2.481	0	0	0
BO-1-145.1	0.063	0	0	-0.104	2.503	0	0	0
BO-1-146.1	0.072	0	0	-0.101	2.52	0	0	0
BO-1-15.1	-0.145	0	0	-0.165	2.902	0	0	0
BO-1-17.1	-0.002	0	0	-0.142	1.042	0	0	0
BO-1-19.1	-0.099	0	0	-0.154	2.992	0	0	0
BO-1-21.1	-0.293	0	0	-0.221	2.055	0	0	0
BO-1-23.1	-0.097	0	0	-0.15	3.126	0	0	0
BO-1-30.1	-0.067	0	0	-0.146	3.199	0	0	0
BO-1-37.1	-0.101	0	0	-0.154	2.073	0	0	0
BO-1-43.1	-0.082	0	0	-0.149	2.216	0	0	0
BO-1-44.1	-0.069	0	0	-0.146	2.268	0	0	0
BO-1-45.1	-0.111	0	0	-0.16	1.962	0	0	0
BO-1-46.1	0.009	0	0	-0.123	2.417	0	0	0
BO-1-47.1	0.017	0	0	-0.121	2.473	0	0	0
BO-1-48.1	-0.082	0	0	-0.153	2.178	0	0	0
BO-1-96	0.789	0	0	0	1.278	0	0	0
BO-2-28.1	0.006	0	0	-0.135	3.461	0	0	0
BO-2-40.1	-0.28	0	0	-0.252	1.878	0	0	0
BO-2-57.1	0.561	0	0	0	3.268	0	0	0
CE101.1	0.706	0	0	0	1.249	0	0	0
EGK-266/1	-0.174	0	0	-0.222	1.095	0	0	0
EGK-276-A	-0.328	0	0	-0.275	1.758	0	0	0
EGK-276-B	-0.11	0	0	-0.204	1.118	0	0	0
LL81.2	-0.39	0	0	-0.307	0.774	0	0	0
QS-1-114.1	-0.119	0	0	-0.194	1.61	0	0	5.795
QS-1-128.1	-0.078	0	0	-0.181	1.675	0	0	5.682
QS-1-138.1	-0.078	0	0	-0.176	1.774	0	0	0
QS-1-142.1	-0.083	0	0	-0.184	1.643	0	0	0
QS-1-89.4	0.028	0	0	-0.143	1.041	0	0	0
QS-2-116.3	-0.018	0	0	-0.141	2.334	0	0.127	0
QS-2-124.2	-0.061	0	0	-0.18	1.6	0	0	0

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 17-24. Continued)

Analogue	SsssCH	SddC	StsC	SdssC	SaasC	SaaaC	SssssC	SsNH2
QS-2-125.1	0.013	0	0	-0.143	2.221	0	0	0
QS-2-125.2	-0.034	0	0	-0.16	2	0	0	0
QS-2-125.3	0.093	0	0	-0.125	1.058	0	0	0
QS-2-133.1	0.083	0	0	-0.123	2.278	0	0	0
QS-2-147.2	-0.059	0	0	-0.165	1.969	0	0	0
QS-2-15.1	-0.182	0	0	-0.205	1.64	0	0	0
QS-2-29.4	-0.117	0	0	-0.187	1.716	0	0	0
QS-2-40.1	-0.393	0	0	-0.295	0.795	0	0	0
QS-2-41.2	-0.254	0	0	-0.25	0.985	0	0	0
QS-2-61.4	-0.126	0	0	-0.198	1.575	0	0	0
QS-2-71.3	-0.271	0	0	-0.263	0.515	0	0	0
QS-2-81.4	-0.237	0	0	-0.238	1.801	0	0	0
QS-2-84.4	-0.223	0	0	-0.221	2.167	0	0	0
QS-2-88.1	-0.018	0	0	-0.156	2.038	0	0	0
QS-2-99.3	-0.199	0	0	-0.22	1.477	0	0	0
WB47.4	-0.598	0	0	-0.375	0.073	0	0	0
WB48.4	-0.39	0	0	-0.306	0.36	0	0	0
WB61.4	0.018	0	0	-0.141	2.255	0	0	0
WB71.5	0.017	0	0	-0.139	2.235	0	0	0
WB77.2	0.288	0	0	0	0.661	0	0	0
XY-1-102.3	0.441	0	0	0	2.17	0	0	0
XY-1-127.5	0.806	0	0	0	2.421	0	0	0
XY-1-129.2	-0.187	0	0	-0.188	2.852	0	0	0
XY-1-144.4	0.159	0	0	0.205	2.449	0	0	0
XY-1-147.4	-0.052	0	0	-0.221	2.32	0	0	5.79
XY-1-30.3	0.645	0	0	0	2.62	0	0	0
XY-1-44.5	1.011	0	0	0	2.808	0	0	0
XY-1-47.1	1.072	0	0	0	1.392	0	0	0
XY-1-85.7	-0.299	0	0	-0.228	3.056	0	0	0
XY-1-86.2	0.38	0	0	0	3.493	0	0	0
XY-1-89.5	0.745	0	0	0	3.79	0	0	0
XY-2-74.3	-0.256	0	0	-0.326	1.771	0	0	5.533
ZL102.3	-0.074	0	0	-0.139	2.05	0	0	0
ZL105.1	0.082	0	0	-0.121	2.26	0	0	0
ZL21.1	0.677	0	0	-0.202	1.248	0	0	0
ZL26.1	0.009	0	0	-0.231	1.023	0	0	5.521
ZL38.1	0.018	0	0	-0.136	2.354	0	0	0
ZL68.3	-0.049	0	0	-0.148	1.042	2.362	0	0
ZL77.2	-0.666	0	0	-0.422	-0.166	0	-4.373	0

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 25-32. Continued)

Analogues	SssNH	SdsN	SaaN	SsssN	SddsN	SsOH	SdO	SssO
AL34.1	0	0	0	2.452	0	0	12.561	5.153
AN-1-68.2	0	4.002	0	2.412	0	0	12.648	5.166
BO-1-119.1	0	0	0	2.625	0	10.026	0	0
BO-1-12.1	0	0	0	2.234	0	0	12.261	5.029
BO-1-120.1	0	0	0	2.641	0	10.043	0	0
BO-1-122.1	0	0	0	2.602	0	10.008	0	0
BO-1-128.1	0	0	0	2.515	0	0	12.604	5.177
BO-1-13.1	0	0	0	2.406	0	0	12.594	5.151
BO-1-131.1	0	0	0	2.491	0	0	12.583	5.166
BO-1-144.1	0	0	0	2.531	0	0	12.623	5.186
BO-1-145.1	0	0	0	2.542	0	0	12.642	5.195
BO-1-146.1	0	0	0	2.551	0	0	12.658	5.202
BO-1-15.1	0	0	0	2.388	0	0	12.602	5.151
BO-1-17.1	0	0	0	2.352	0	0	12.287	5.055
BO-1-19.1	0	0	0	2.418	0	0	12.588	5.151
BO-1-21.1	0	0	0	2.305	-0.393	0	34.31	5.12
BO-1-23.1	0	0	0	2.43	0	0	12.628	10.414
BO-1-30.1	0	0	0	2.431	0	0	12.55	5.142
BO-1-37.1	0	0	4.447	2.389	0	0	12.545	5.137
BO-1-43.1	0	0	4.211	2.412	0	0	12.548	5.141
BO-1-44.1	0	0	4.089	2.424	0	0	12.551	5.143
BO-1-45.1	0	0	0	2.358	0	0	12.487	5.115
BO-1-46.1	0	0	0	2.467	0	0	12.516	5.143
BO-1-47.1	0	0	0	2.476	0	0	12.517	5.144
BO-1-48.1	0	0	0	2.392	0	0	12.492	5.119
BO-1-96	0	0	0	2.525	0	9.718	0	0
BO-2-28.1	3.471	0	0	0	0	0	12.172	5.021
BO-2-40.1	3.385	0	0	0	0	0	12.086	4.935
BO-2-57.1	0	0	0	2.517	0	10.012	0	0
CE101.1	3.509	0	0	0	0	9.492	0	0
EGK-266/1	3.384	0	0	0	0	9.313	11.972	4.909
EGK-276-A	0	0	0	2.196	0	18.42	12.385	10.105
EGK-276-B	0	0	0	2.234	0	9.37	12.131	4.97
LL81.2	3.336	0	0	0	-0.449	0	33.385	4.885
QS-1-114.1	3.411	0	0	0	0	0	12.02	4.94
QS-1-128.1	3.415	0	0	0	0	0	12.003	4.94
QS-1-138.1	3.429	0	0	0	0	0	12.073	10.116
QS-1-142.1	3.414	0	0	0	0	0	12.001	4.938
QS-1-89.4	3.432	0	0	0	0	0	11.949	4.941
QS-2-116.3	3.477	0	0	0	0	0	12.231	5.039
QS-2-124.2	0	0	0	2.256	0	0	12.177	5

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 25-32. Continued)

Analogues	SssNH	SdsN	SaaN	SsssN	SddsN	SsOH	SdO	SssO
QS-2-125.1	3.445	0	0	0	0	0	12.033	4.969
QS-2-125.2	3.436	0	0	0	0	0	12.045	4.966
QS-2-125.3	0	0	0	2.29	0	0	12.108	5.003
QS-2-133.1	0	0	0	2.297	0	0	12.193	5.033
QS-2-147.2	3.438	0	0	0	0	0	12.081	4.975
QS-2-15.1	3.428	0	0	0	0	0	12.162	10.361
QS-2-29.4	3.429	0	0	0	0	0	12.108	10.207
QS-2-40.1	3.348	0	0	0	0	9.92	11.991	4.885
QS-2-41.2	3.37	0	0	0	0	9.542	11.979	4.899
QS-2-61.4	3.408	0	0	0	0	0	12.018	4.938
QS-2-71.3	3.353	0	0	0	0	0	11.941	4.878
QS-2-81.4	3.39	0	0	0	0	0	12.07	4.935
QS-2-84.4	3.426	0	0	0	0	0	12.232	15.576
QS-2-88.1	3.435	0	0	0	0	0	12.023	4.96
QS-2-99.3	3.401	0	0	0	0	0	12.043	4.937
WB47.4	3.293	0	0	0	0	0	11.935	4.829
WB48.4	3.329	0	0	0	0	0	11.938	4.858
WB61.4	3.447	0	0	0	0	0	12.034	4.971
WB71.5	3.452	0	0	0	0	0	12.061	4.981
WB77.2	3.406	0	0	0	0	9.46	0	0
XY-1-102.3	3.467	0	0	0	0	9.576	0	0
XY-1-127.5	3.58	0	0	0	0	0	0	5.366
XY-1-129.2	0	0	0	2.418	0	0	12.629	5.15
XY-1-144.4	0	0	0	4.255	0	0	13.04	0
XY-1-147.4	0	0	0	2.432	0	0	12.193	0
XY-1-30.3	0	0	0	2.562	0	9.989	0	0
XY-1-44.5	0	0	0	2.662	0	0	0	5.583
XY-1-47.1	3.622	0	0	0	0	0	0	5.372
XY-1-85.7	0	0	0	2.392	0	0	12.681	5.147
XY-1-86.2	0	0	0	2.502	0	10.073	0	0
XY-1-89.5	0	0	0	2.602	0	0	0	5.578
XY-2-74.3	3.356	0	0	0	0	0	11.702	0
ZL102.3	3.496	0	0	0	0	0	12.754	5.619
ZL105.1	0	0	0	2.299	0	0	12.22	5.043
ZL21.1	3.55	0	0	0	0	0	11.02	5.233
ZL26.1	3.398	0	0	0	0	0	11.581	0
ZL38.1	3.461	0	0	0	0	0	12.104	4.998
ZL68.3	3.48	0	0	0	0	0	12.286	5.053
ZL77.2	3.255	0	0	0	0	0	12.009	4.816

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Columns 33-39. Continued)

Analogues	SaaO	SsF	SdS	SaaS	SsCl	SsBr	SsI
AL34.1	0	0	0	0	0	0	0
AN-1-68.2	0	0	4.655	0	0	0	0
BO-1-119.1	0	0	0	0	0	0	0
BO-1-12.1	0	0	0	0	0	0	0
BO-1-120.1	0	0	0	0	0	0	0
BO-1-122.1	0	0	0	0	0	0	0
BO-1-128.1	0	0	0	0	0	0	0
BO-1-13.1	0	0	0	0	6.138	0	0
BO-1-131.1	0	0	0	0	0	0	0
BO-1-144.1	0	0	0	0	0	0	0
BO-1-145.1	0	0	0	0	0	0	0
BO-1-146.1	0	0	0	0	0	0	0
BO-1-15.1	0	0	0	0	6.359	0	0
BO-1-17.1	0	0	0	0	0	0	0
BO-1-19.1	0	0	0	0	5.999	0	0
BO-1-21.1	0	0	0	0	0	0	0
BO-1-23.1	0	0	0	0	0	0	0
BO-1-30.1	0	0	0	1.092	6.096	0	0
BO-1-37.1	0	0	0	0	0	0	0
BO-1-43.1	0	0	0	0	0	0	0
BO-1-44.1	0	0	0	0	0	0	0
BO-1-45.1	5.499	0	0	0	0	0	0
BO-1-46.1	0	0	0	1.265	0	0	0
BO-1-47.1	0	0	0	1.287	0	0	0
BO-1-48.1	5.187	0	0	0	0	0	0
BO-1-96	0	0	0	0	0	0	0
BO-2-28.1	0	0	0	0	0	0	0
BO-2-40.1	0	0	0	0	12.048	0	0
BO-2-57.1	0	0	0	0	6.126	0	0
CE101.1	0	0	0	0	0	0	0
EGK-266/1	0	0	0	0	0	0	0
EGK-276-A	0	0	0	0	0	0	0
EGK-276-B	0	0	0	0	0	0	0
LL81.2	0	0	0	0	0	0	0
QS-1-114.1	0	0	0	0	0	0	0
QS-1-128.1	0	0	0	0	0	0	0
QS-1-138.1	0	0	0	0	0	0	0
QS-1-142.1	0	0	0	0	5.884	0	0
QS-1-89.4	0	0	0	0	0	0	0
QS-2-116.3	0	0	0	0	0	0	0
QS-2-124.2	0	0	0	0	6.046	0	0

Table B.5 Atom Type Electrotopological State Indices for MP Analogues (Continued. Columns 33-39)

Analogues	SaaO	SsF	SdS	SaaS	SsCl	SsBr	SsI
QS-2-125.1	0	0	0	0	0	0	2.269
QS-2-125.2	0	0	0	0	0	3.453	0
QS-2-125.3	0	0	0	0	0	0	0
QS-2-133.1	0	0	0	0	0	0	0
QS-2-147.2	0	0	0	0	0	3.525	0
QS-2-15.1	0	0	0	0	0	0	0
QS-2-29.4	0	0	0	0	0	0	0
QS-2-40.1	0	0	0	0	0	0	0
QS-2-41.2	0	0	0	0	0	0	0
QS-2-61.4	0	0	0	0	6.003	0	0
QS-2-71.3	0	12.943	0	0	0	0	0
QS-2-81.4	0	0	0	0	11.961	0	0
QS-2-84.4	0	0	0	0	0	0	0
QS-2-88.1	0	0	0	0	0	3.413	0
QS-2-99.3	0	0	0	0	6.204	0	0
WB47.4	0	13.856	0	0	0	0	0
WB48.4	0	13.29	0	0	0	0	0
WB61.4	0	0	0	0	0	0	0
WB71.5	0	0	0	0	0	0	0
WB77.2	0	13.129	0	0	0	0	0
XY-1-102.3	0	0	0	0	11.924	0	0
XY-1-127.5	0	0	0	0	12.073	0	0
XY-1-129.2	0	0	0	0	6.176	0	0
XY-1-144.4	0	0	0	0	0	0	0
XY-1-147.4	0	0	0	0	0	0	0
XY-1-30.3	0	0	0	0	0	0	0
XY-1-44.5	0	0	0	0	0	0	0
XY-1-47.1	0	0	0	0	0	0	0
XY-1-85.7	0	0	0	0	12.275	0	0
XY-1-86.2	0	0	0	0	12.238	0	0
XY-1-89.5	0	0	0	0	12.387	0	0
XY-2-74.3	0	0	0	0	11.884	0	0
ZL102.3	0	0	0	0	0	0	0
ZL105.1	0	0	0	0	0	0	0
ZL21.1	0	0	0	0	0	0	0
ZL26.1	0	0	0	0	0	0	0
ZL38.1	0	0	0	0	0	0	0
ZL68.3	0	0	0	0	0	0	0
ZL77.2	0	37.788	0	0	0	0	0

Table B.6 Topological State Indices for MP Analogues

Analogues	sumdelI	sumI	tets1	tets2	tets3	htets1	htets2	htets3	Qv	totop
AL34.1	11.061	49.667	135.71	28.427	8.568	119.906	23.91	6.817	1.226	149.146
AN-1-68.2	13.821	58.5	169.54	33.699	9.952	148.012	27.694	7.669	1.163	173.688
BO-1-119.1	7.45	46	148.832	30.401	9.124	112.75	22.129	6.367	1.365	134.329
BO-1-12.1	11.748	44.5	75.347	18.594	6.113	74.302	17.007	5.165	1.177	117.11
BO-1-120.1	7.516	47.5	154.753	31.379	9.437	116.067	22.521	6.474	1.392	137.278
BO-1-122.1	7.37	44.5	142.881	29.424	8.81	109.379	21.726	6.255	1.337	131.57
BO-1-128.1	11.306	52.667	146.257	30.083	9.114	126.997	24.74	7.058	1.279	154.173
BO-1-13.1	13.3	53.444	140.546	28.902	8.613	129.76	25.279	7.113	1.199	155.124
BO-1-131.1	11.199	51.167	140.886	29.228	8.835	123.456	24.325	6.942	1.253	151.463
BO-1-144.1	11.393	54.167	151.754	30.969	9.399	130.484	25.143	7.169	1.304	157.092
BO-1-145.1	11.467	55.667	157.34	31.872	9.688	133.904	25.535	7.275	1.328	160.133
BO-1-146.1	11.531	57.167	162.994	32.788	9.979	137.255	25.916	7.379	1.351	163.25
BO-1-15.1	13.421	53.444	138.723	28.647	8.567	130.252	25.461	7.177	1.199	155.422
BO-1-17.1	10.788	43	76.622	18.722	6.097	74.169	16.956	5.143	1.261	114.381
BO-1-19.1	13.195	53.444	141.612	29.071	8.652	129.623	25.239	7.104	1.199	155.041
BO-1-21.1	19.178	65.333	131.574	26.795	8.043	149.045	27.998	7.763	0.969	184.082
BO-1-23.1	13.207	54.833	157.397	31.741	9.393	139.084	26.527	7.399	1.23	166.887
BO-1-30.1	12.956	51.056	128.406	27.034	8.134	120.48	24.006	6.808	1.213	143.533
BO-1-37.1	12.291	50.667	133.991	28.263	8.564	119.881	23.912	6.821	1.178	152.564
BO-1-43.1	12.15	50.667	135.536	28.539	8.644	119.605	23.806	6.777	1.178	152.37

Table B.6 Topological State Indices for MP Analogues (continued)

Analogues	sumdelI	sumI	tets1	tets2	tets3	htets1	htets2	htets3	Qv	totop
BO-1-44.1	12.07	50.667	136.146	28.633	8.664	119.544	23.792	6.776	1.178	152.323
BO-1-45.1	12.642	49.167	119.782	26.112	8.036	110.602	22.642	6.525	1.149	147.44
BO-1-46.1	11.063	47.278	121.666	26.205	7.985	110.681	22.669	6.535	1.243	138.276
BO-1-47.1	10.946	47.278	121.949	26.256	8.002	110.622	22.643	6.523	1.243	137.965
BO-1-48.1	12.431	49.167	122.09	26.552	8.165	110.341	22.532	6.477	1.149	147.204
BO-1-96	6.183	33.333	71.987	17.65	5.688	57.46	13.487	4.149	1.453	88.179
BO-2-28.1	10.437	40.333	70.376	17.365	5.61	67.478	15.627	4.752	1.369	108.035
BO-2-40.1	13.79	44.556	62.168	15.458	5.019	67.792	15.731	4.791	1.122	107.213
BO-2-57.1	9.462	46.778	140.014	28.588	8.444	115.338	22.675	6.439	1.269	134.991
CE101.1	6.026	30.333	56.445	14.436	4.777	44.461	10.982	3.516	1.284	77.847
EGK-266/1	13.046	42.667	55.288	13.948	4.583	61.138	14.526	4.515	1.049	107.544
EGK-276-A	18.73	56.333	85.276	19.742	6.201	96.243	20.945	6.181	1.044	146.011
EGK-276-B	13.007	44.167	63.038	15.642	5.09	68.646	15.942	4.861	1.142	113.422
LL81.2	17.323	52.667	52.858	13.61	4.608	73.66	16.876	5.151	0.883	129.246
QS-1-114.1	11.574	40.667	61.392	15.471	5.081	61.146	14.518	4.504	1.154	105.066
QS-1-128.1	11.478	40.667	61.534	15.5	5.09	61.005	14.475	4.493	1.154	104.953
QS-1-138.1	11.643	42.167	71.29	17.638	5.756	67.058	15.581	4.787	1.189	112.756
QS-1-142.1	11.565	40.778	61.314	15.445	5.072	61.049	14.492	4.501	1.148	101.584
QS-1-89.4	9.51	37	56.814	14.672	4.912	55.007	13.389	4.224	1.179	96.353
QS-2-116.3	11.371	43.917	77.661	18.261	5.737	79.632	17.921	5.399	1.479	117.899

Table B.6 Topological State Indices for MP Analogues (continued)

Analogues	sumdelI	sumI	tets1	tets2	tets3	htets1	htets2	htets3	Qv	totop
QS-2-124.2	11.615	42.278	68.951	17.098	5.563	68.72	15.957	4.858	1.246	107.533
QS-2-125.1	10.059	38.787	63.663	16.037	5.27	60.935	14.448	4.482	1.269	100.101
QS-2-125.2	10.563	39.417	63.227	15.929	5.232	61.105	14.503	4.497	1.229	100.506
QS-2-125.3	9.408	38.5	63.874	16.243	5.389	62.27	14.806	4.579	1.288	102.101
QS-2-133.1	9.868	40.167	71.132	17.653	5.753	68.48	15.867	4.826	1.381	107.843
QS-2-147.2	10.59	39.417	63.104	15.924	5.238	61.578	14.688	4.564	1.229	100.767
QS-2-15.1	11.894	42.167	70.765	17.59	5.757	67.984	15.916	4.897	1.189	113.561
QS-2-29.4	11.756	42.167	71.133	17.613	5.749	67.278	15.649	4.804	1.189	112.93
QS-2-40.1	13.512	42.667	51.625	13.249	4.424	61.877	14.796	4.608	1.049	108.135
QS-2-41.2	13.239	42.667	54.272	13.736	4.525	61.311	14.579	4.528	1.049	107.669
QS-2-61.4	11.666	40.778	61.152	15.411	5.061	61.201	14.539	4.512	1.148	101.671
QS-2-71.3	14.645	44.667	54.719	13.879	4.58	61.271	14.578	4.537	0.957	109.996
QS-2-81.4	13.629	44.556	61.849	15.362	4.991	67.901	15.802	4.831	1.122	107.218
QS-2-84.4	13.958	47.333	82.847	19.981	6.422	80.825	18.12	5.424	1.197	130.523
QS-2-88.1	10.521	39.417	63.307	15.948	5.239	60.971	14.462	4.488	1.229	100.435
QS-2-99.3	11.79	40.778	60.74	15.36	5.06	61.717	14.738	4.585	1.148	101.99
WB47.4	15.392	44.667	53.385	13.756	4.602	62.115	14.882	4.643	0.957	110.631
WB48.4	14.946	44.667	53.001	13.597	4.528	61.475	14.639	4.553	0.957	110.13
WB61.4	9.972	38.667	63.676	16.04	5.271	60.871	14.424	4.471	1.277	101.983
WB71.5	9.99	38.667	63.489	15.997	5.257	60.982	14.458	4.479	1.277	102.074

Table B.6 Topological State Indices for MP Analogues (continued)

Analogues	sumdelI	sumI	tets1	tets2	tets3	htets1	htets2	htets3	Qv	totop
WB77.2	11.286	38	38.288	10.174	3.478	50.835	12.293	3.881	0.994	91.341
XY-1-102.3	10.008	37.889	55	13.726	4.468	56.803	13.41	4.158	1.194	88.278
XY-1-127.5	8.607	37.389	74.18	18.168	5.848	61.144	14.201	4.351	1.366	92.725
XY-1-129.2	13.396	53.444	140.68	29.024	8.656	129.411	25.272	7.107	1.199	155.275
XY-1-144.4	9.973	50.167	140.5	29.517	8.924	126.462	25.063	7.106	1.36	151.941
XY-1-147.4	10.598	48.167	130.146	27.273	8.19	113.315	22.722	6.502	1.197	141.136
XY-1-30.3	7.262	43	136.917	28.465	8.496	105.995	21.319	6.134	1.307	129.173
XY-1-44.5	5.551	42.5	162.992	33.536	9.938	111.852	22.229	6.335	1.466	134.001
XY-1-47.1	4.599	29.833	72.346	18.283	6.01	48.48	11.758	3.715	1.51	82.146
XY-1-85.7	15.505	57.222	140.488	28.716	8.518	139.479	26.768	7.437	1.175	161.631
XY-1-86.2	11.569	50.556	131.778	26.988	8.011	124.939	24.178	6.787	1.237	141.223
XY-1-89.5	9.884	50.056	165.363	33.12	9.667	131.153	25.109	6.983	1.367	146.199
XY-2-74.3	13.021	43.056	57.034	14.181	4.59	62.821	14.733	4.527	1.088	99.642
ZL102.3	11.87	48.167	102.619	23.486	7.416	96.635	20.9	6.276	1.141	139.354
ZL105.1	9.885	40.167	70.942	17.61	5.74	68.588	15.902	4.833	1.381	107.939
ZL21.1	9.014	38.5	69.038	17.123	5.57	57.393	13.703	4.307	1.217	98.89
ZL26.1	8.991	35.5	53.16	13.712	4.564	50.315	12.349	3.917	1.139	88.951
ZL38.1	10.382	40.167	70.714	17.343	5.613	66.562	15.363	4.686	1.311	106.858
ZL68.3	10.807	44.333	127.044	26.058	7.56	108.854	21.63	6.051	1.178	138.069
ZL77.2	24.05	61.917	100.779	26.828	9.245	80.498	18.246	5.535	0.744	142.56

Table B.7 Shannon Information Index for MP Analogues

Analogue	Si
AL34.1	1.28
AN-1-68.2	1.342
BO-1-119.1	1.28
BO-1-12.1	1.241
BO-1-120.1	1.302
BO-1-122.1	1.257
BO-1-128.1	1.322
BO-1-13.1	1.35
BO-1-131.1	1.302
BO-1-144.1	1.342
BO-1-145.1	1.361
BO-1-146.1	1.379
BO-1-15.1	1.35
BO-1-17.1	1.241
BO-1-19.1	1.302
BO-1-21.1	1.32
BO-1-23.1	1.322
BO-1-30.1	1.33
BO-1-37.1	1.33
BO-1-43.1	1.33
BO-1-44.1	1.28
BO-1-45.1	1.309
BO-1-46.1	1.309
BO-1-47.1	1.309
BO-1-48.1	1.309
BO-1-96	1.16
BO-2-28.1	1.184

Analogue	Si
BO-2-40.1	1.184
BO-2-57.1	1.309
CE101.1	1.096
EGK-266/1	1.188
EGK-276-A	1.362
EGK-276-B	1.215
LL81.2	1.211
QS-1-114.1	1.255
QS-1-128.1	1.188
QS-1-138.1	1.215
QS-1-142.1	1.188
QS-1-89.4	1.16
QS-2-116.3	1.197
QS-2-124.2	1.279
QS-2-125.1	1.188
QS-2-125.2	1.255
QS-2-125.3	1.188
QS-2-133.1	1.215
QS-2-147.2	1.255
QS-2-15.1	1.279
QS-2-29.4	1.279
QS-2-40.1	1.255
QS-2-41.2	1.255
QS-2-61.4	1.255
QS-2-71.3	1.188
QS-2-81.4	1.279
QS-2-84.4	1.322

Analogue	Si
QS-2-88.1	1.184
QS-2-99.3	1.309
WB47.4	1.096
WB48.4	1.188
WB61.4	1.362
WB71.5	1.215
WB77.2	1.211
XY-1-102.3	1.255
XY-1-127.5	1.188
XY-1-129.2	1.215
XY-1-144.4	1.188
XY-1-147.4	1.16
XY-1-30.3	1.197
XY-1-44.5	1.279
XY-1-47.1	1.188
XY-1-85.7	1.255
XY-1-86.2	1.188
XY-1-89.5	1.215
XY-2-74.3	1.255
ZL102.3	1.279
ZL105.1	1.279
ZL21.1	1.255
ZL26.1	1.255
ZL38.1	1.255
ZL68.3	1.188
ZL77.2	1.279

Table B.8 Hydrogen Bond-Related Counts and Estate Indices (Columns 1-7)

Analogue	SHBd	SHBa	SwHBa	SHBint2	SHBint4	SHBint5	SHBint6
AL34.1	0	20.166	22.758	0	0	0	0
AN-1-68.2	0	28.884	23.242	0	0	0	0
BO-1-119.1	0	12.652	24.005	0	6.49	0	0
BO-1-12.1	0	19.524	18.85	0	0	0	0
BO-1-120.1	0	12.684	24.079	0	6.525	0	0
BO-1-122.1	0	12.61	23.909	0	6.438	0	0
BO-1-128.1	0	20.295	23.094	0	0	0	0
BO-1-13.1	0	26.289	20.734	0	0	0	0
BO-1-131.1	0	20.239	22.954	0	0	0	0
BO-1-144.1	0	20.34	23.201	0	0	0	0
BO-1-145.1	0	20.379	23.285	0	0	0	0
BO-1-146.1	0	20.411	23.354	0	0	0	0
BO-1-15.1	0	26.5	20.638	0	0	0	0
BO-1-17.1	0	19.694	16.604	0	0	0	0
BO-1-19.1	0	26.156	20.792	0	0	0	0
BO-1-21.1	0	41.735	18.24	0	0	0	0
BO-1-23.1	0	25.472	21.164	0	0	0	0
BO-1-30.1	0	27.312	17.127	0	0	0	0
BO-1-37.1	0	24.518	19.713	0	0	0	0
BO-1-43.1	0	24.312	19.817	0	0	0	0
BO-1-44.1	0	24.207	19.866	0	0	0	0
BO-1-45.1	0	25.459	17.35	0	0	0	0
BO-1-46.1	0	21.391	18.988	0	0	0	0
BO-1-47.1	0	21.424	18.916	0	0	0	0
BO-1-48.1	0	25.19	17.465	0	0	0	0
BO-1-96	0	12.244	11.734	0	6.171	0	0
BO-2-28.1	0	20.664	9.662	0	19.784	0	0
BO-2-40.1	0	32.453	6.854	0	19.972	0	0
BO-2-57.1	0	18.655	21.784	0	6.26	0	0
CE101.1	0	13	11.571	0	14.598	0	0
EGK-266/1	0	29.578	7.654	0	19.585	0	0
EGK-276-A	0	43.106	6.718	13.05	13.063	0	24.435
EGK-276-B	0	28.705	7.762	0	0	0	0
LL81.2	0	41.606	6.586	0	20.024	0	0
QS-1-114.1	0	26.166	8.917	0	19.61	0	19.197
QS-1-128.1	0	26.041	8.969	0	19.542	0	0
QS-1-138.1	0	25.618	9.246	0	19.78	0	0
QS-1-142.1	0	26.237	8.896	0	19.571	0	0
QS-1-89.4	0	20.323	10.79	0	19.324	0	0
QS-2-116.3	0	20.747	10.623	0	19.909	0	0

Table B.8 Hydrogen Bond-Related Counts and Estate Indices (Columns 1-7. Continued)

Analogue	SHBd	SHBa	SwHBa	SHBint2	SHBint4	SHBint5	SHBint6
QS-2-124.2	0	25.478	8.95	0	0	0	0
QS-2-125.1	0	20.447	10.205	0	19.541	0	0
QS-2-125.2	0	20.447	9.748	0	19.621	0	0
QS-2-125.3	0	19.401	10.904	0	0	0	0
QS-2-133.1	0	19.523	10.393	0	0	0	0
QS-2-147.2	0	20.495	9.67	0	19.73	0	0
QS-2-15.1	0	25.951	9.095	0	20.123	0	0
QS-2-29.4	0	25.744	9.192	0	19.915	0	0
QS-2-40.1	0	30.144	7.469	0	19.79	31.646	0
QS-2-41.2	0	29.791	7.587	0	19.666	0	31.111
QS-2-61.4	0	26.367	8.843	0	19.647	0	0
QS-2-71.3	0	33.115	6.338	0	19.627	0	0
QS-2-81.4	0	32.356	6.874	0	19.895	0	0
QS-2-84.4	0	31.233	7.532	0	20.378	0	0
QS-2-88.1	0	20.417	9.791	0	19.551	0	0
QS-2-99.3	0	26.585	8.748	0	19.765	0	0
WB47.4	0	33.913	6.114	0	19.864	0	0
WB48.4	0	33.415	6.257	0	19.72	0	0
WB61.4	0	20.452	10.284	0	19.499	0	0
WB71.5	0	20.493	10.247	0	19.553	0	0
WB77.2	0	25.995	7.216	0	14.878	0	32.414
XY-1-102.3	0	24.966	7.778	0	15.026	0	14.842
XY-1-127.5	0	21.019	8.299	0	8.45	0	0
XY-1-129.2	0	26.373	20.682	0	0	0	0
XY-1-144.4	0	17.295	23.535	0	0	0	0
XY-1-147.4	0	20.414	22.537	20.755	4.14	0	0
XY-1-30.3	0	12.551	23.776	0	6.351	0	0
XY-1-44.5	0	8.246	24.519	0	0	0	0
XY-1-47.1	0	8.994	12.104	0	8.291	0	0
XY-1-85.7	0	32.495	18.607	0	0	0	0
XY-1-86.2	0	24.813	19.748	0	6.28	0	15.543
XY-1-89.5	0	20.567	20.479	0	0	0	0
XY-2-74.3	0	32.474	6.702	19.765	18.99	0	10.129
ZL102.3	0	21.869	21.72	0	21.121	0	0
ZL105.1	0	19.561	10.354	0	0	0	0
ZL21.1	0	19.802	11.373	0	8.289	0	17.457
ZL26.1	0	20.5	10.607	19.2	18.433	0	0
ZL38.1	0	20.562	10.561	0	19.642	0	0
ZL68.3	0	20.819	17.746	0	20.221	0	0
ZL77.2	0	57.868	4.152	0	20.214	0	0

Table B.8 Hydrogen Bond-Related Counts and Estate Indices (Columns 8-14.Continued)

Analogue	SHBint7	SHBint9	Hmax	Gmax	Hmin	Gmin	Hmaxpos
AL34.1	0	0	1.323	12.561	0.622	-0.222	1.323
AN-1-68.2	0	0	1.372	12.648	0.644	-0.268	1.372
BO-1-119.1	0	0	2.472	10.026	0.559	0.239	2.472
BO-1-12.1	0	0	1.57	12.261	0.601	-0.255	1.57
BO-1-120.1	0	0	2.47	10.043	0.556	0.245	2.47
BO-1-122.1	0	0	2.475	10.008	0.563	0.229	2.475
BO-1-128.1	0	0	1.317	12.604	0.606	-0.196	1.317
BO-1-13.1	0	0	1.344	12.594	0.634	-0.259	1.344
BO-1-131.1	0	0	1.319	12.583	0.612	-0.206	1.319
BO-1-144.1	0	0	1.315	12.623	0.602	-0.189	1.315
BO-1-145.1	0	0	1.314	12.642	0.599	-0.184	1.314
BO-1-146.1	0	0	1.313	12.658	0.582	-0.181	1.313
BO-1-15.1	0	0	1.334	12.602	0.637	-0.269	1.334
BO-1-17.1	0	0	1.3	12.287	0.584	-0.207	1.3
BO-1-19.1	0	0	1.33	12.588	0.632	-0.251	1.33
BO-1-21.1	0	0	1.452	12.605	0.654	-0.393	1.452
BO-1-23.1	0	0	1.339	12.628	0.637	-0.25	1.339
BO-1-30.1	0	0	1.483	12.55	0.63	-0.237	1.483
BO-1-37.1	0	0	1.331	12.545	0.633	-0.25	1.331
BO-1-43.1	0	0	1.373	12.548	0.631	-0.242	1.373
BO-1-44.1	0	0	1.328	12.551	0.629	-0.238	1.328
BO-1-45.1	0	0	1.352	12.487	0.633	-0.253	1.352
BO-1-46.1	0	0	1.331	12.516	0.616	-0.204	1.331
BO-1-47.1	0	0	1.386	12.517	0.617	-0.201	1.386
BO-1-48.1	0	0	1.407	12.492	0.629	-0.242	1.407
BO-1-96	0	0	2.444	9.718	0.463	0.255	2.444
BO-2-28.1	0	0	1.625	12.172	0.546	-0.191	1.625
BO-2-40.1	0	0	1.653	12.086	0.562	-0.36	1.653
BO-2-57.1	0	0	2.487	10.012	0.58	0.177	2.487
CE101.1	0	0	2.434	9.492	0.501	0.238	2.434
EGK-266/1	0	0	2.57	11.972	0.553	-0.295	2.57
EGK-276-A	31.816	31.784	2.569	12.385	0.59	-0.46	2.569
EGK-276-B	0	0	2.573	12.131	0.565	-0.282	2.573
LL81.2	0	0	1.664	12.034	0.572	-0.449	1.664
QS-1-114.1	0	0	1.631	12.02	0.549	-0.265	1.631
QS-1-128.1	0	0	1.628	12.003	0.548	-0.24	1.628
QS-1-138.1	0	0	1.638	12.073	0.554	-0.24	1.638
QS-1-142.1	0	0	1.631	12.001	0.549	-0.243	1.631
QS-1-89.4	0	0	1.617	11.949	0.541	-0.179	1.617
QS-2-116.3	0	0	1.628	12.231	0.515	-0.204	1.628

Table B.8 Hydrogen Bond-Related Counts and Estate Indices (Columns 8-14.Continued)

Analogue	SHBint7	SHBint9	Hmax	Gmax	Hmin	Gmin	Hmaxpos
QS-2-124.2	0	0	1.381	12.177	0.564	-0.256	1.381
QS-2-125.1	0	0	1.624	12.033	0.545	-0.188	1.624
QS-2-125.2	0	0	1.629	12.045	0.548	-0.215	1.629
QS-2-125.3	0	0	1.284	12.108	0.553	-0.165	1.284
QS-2-133.1	0	0	1.297	12.193	0.555	-0.17	1.297
QS-2-147.2	0	0	1.633	12.081	0.55	-0.231	1.633
QS-2-15.1	0	0	1.655	12.162	0.562	-0.304	1.655
QS-2-29.4	0	0	1.645	12.108	0.558	-0.264	1.645
QS-2-40.1	0	0	2.639	11.991	0.559	-0.434	2.639
QS-2-41.2	0	0	2.597	11.979	0.556	-0.345	2.597
QS-2-61.4	0	0	1.635	12.018	0.551	-0.269	1.635
QS-2-71.3	0	0	1.644	12.943	0.558	-0.351	1.644
QS-2-81.4	0	0	1.648	12.07	0.56	-0.333	1.648
QS-2-84.4	0	0	1.666	12.232	0.571	-0.324	1.666
QS-2-88.1	0	0	1.626	12.023	0.546	-0.205	1.626
QS-2-99.3	0	0	1.641	12.043	0.554	-0.316	1.641
WB47.4	0	0	1.664	13.856	0.567	-0.559	1.664
WB48.4	0	0	1.652	13.29	0.562	-0.425	1.652
WB61.4	0	0	1.62	12.034	0.543	-0.184	1.62
WB71.5	0	0	1.621	12.061	0.543	-0.185	1.621
WB77.2	0	0	2.469	13.129	0.522	-0.23	2.469
XY-1-102.3	0	0	2.465	9.576	0.52	0.102	2.465
XY-1-127.5	0	0	1.575	6.104	0.524	0.338	1.575
XY-1-129.2	0	0	1.419	12.629	0.632	-0.312	1.419
XY-1-144.4	0	0	1.309	13.04	0.613	-0.097	1.309
XY-1-147.4	0	0	1.702	12.193	0.605	-0.236	1.702
XY-1-30.3	0	0	2.479	9.989	0.569	0.213	2.479
XY-1-44.5	0	0	1.25	5.583	0.573	0.449	1.25
XY-1-47.1	0	0	1.544	5.372	0.504	0.492	1.544
XY-1-85.7	0	0	1.474	12.681	0.641	-0.376	1.474
XY-1-86.2	0	0	2.51	10.073	0.593	0.059	2.51
XY-1-89.5	0	0	1.4	6.277	0.597	0.295	1.4
XY-2-74.3	0	0	1.689	11.702	0.547	-0.347	1.689
ZL102.3	0	0	1.656	12.754	0.566	-0.236	1.656
ZL105.1	0	0	1.306	12.22	0.556	-0.171	1.306
ZL21.1	0	0	1.584	11.02	0.528	-0.202	1.584
ZL26.1	0	0	1.658	11.581	0.527	-0.231	1.658
ZL38.1	0	0	1.623	12.104	0.47	-0.184	1.623
ZL68.3	0	0	1.646	12.286	0.559	-0.222	1.646
ZL77.2	0	0	1.683	12.596	0.585	-4.373	1.683

APPENDIX C

MATLAB SOURCE CODE

The main GA-PLS script is written in Matlab and stored as GA_PLS.m. The functional routine are stored in files evaluate.m, statistics.m and count_ones.m.

GA_PLS.m

```
%  This is a Matlab implementation of Genetic Algorithm that
%  selects descriptors from a group of descriptors. It uses
%  Partial Least Squares Regression to evaluate the fitness of a
%  particular combination of descriptors.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   Genetic Algorithm - Partial Least Squares (GA-PLS)
%   Written by: Noureen Wadhwaniya
%   Date: 11/01/2004
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%  This program generates a population of randomly generated
%  strings of 0 and 1.
%  The strings are treated as parents who undergo crossover to
%  yield offspring. The offspring are subjected to point
%  mutations based on probability.
%  The fitness of the resulting offspring is evaluated using PLS
%  and compared to that of its parents. If the fitness of the
%  offspring is better than its parent it replaces it(Survival of
%  the fittest).
%  The process of crossover, mutation and survival of fittest is
%  repeated till a specified number of crossoves are reached

clear           % Initialize memory

% Load the data
% The data is scaled in such a way that its value falls between 0
% and 1
load mp_x_scaled;
load mp_y_scaled;

% Assign the x & y blocks
x_block=mp_x_scaled;
y_block=mp_y_scaled;

% Define constants
MAX_CROSSOVERS = 10000;         % Maximum number of crossovers
POP_SIZE =300;                  % Individuals in a population
CHROM_LENGTH= size(x_block,2); % No. of descriptors
COMPOUNDS= size(x_block,1);     % No. of compounds
```

```
PROBABILITY=5/CHROM_LENGTH;     % Probability of selecting a
descriptor
MUTATE_PROB=0.1;                % Prob of mutation

% Here Crossover probability is kept as 1 since a different
% generation of new individuals is not obtained. Rather as soon
% as new offspring are reproduced they are added to the mating
% pool. This ensures that fit offspring are given a high chance
% of reproducing. Single point mutation occurs based on
% probability in offspring. The location of mutation is
% random.

% Offspring will be stored at the two positions after the
POP_SIZE
% The offspring will be accessed by index m & m+1
m = POP_SIZE+1;

% Initialize the generation counter
gen=1;

% Initialize number of crossovers
ncross=0;

% Initialize an array of all zeroes of CHROM_LENGTH
for lchrom = 1:CHROM_LENGTH
    zero_array(lchrom)=0;
end;

% We want to make the initial population of size = POP_SIZE
% called oldpop.
% Each column represents a chromosome and each element in that
% column representing a gene.  There will be POP_SIZE chromosomes
% with each of them having CHROM_LENGTH genes.
disp('Running ..........');

% Initialize a population
for i = 1:POP_SIZE
    for lchrom = 1:CHROM_LENGTH
           % the probability of including or not including a
           % descriptor is not equally likely but is
           % 5/CHROM_LENGTH.
      oldpop(gen).individual(i).chrom(lchrom)=
      (rand(1,1)<=PROBABILITY);
           % This starts the population off with random bits
    end

    % Finding fitness, q^2 and no. of components of the old
    % population
        [oldpop(gen).individual(i).fitness,
        oldpop(gen).individual(i).q2,
```

```
        oldpop(gen).individual(i).ncomp]=
        evaluate(oldpop(gen).individual(i),CHROM_LENGTH,
        x_block, COMPOUNDS);

    % Since the initial population has no parents, assigning zero
    oldpop(gen).individual(i).parent1=0;
    oldpop(gen).individual(i).parent2=0;

    % Since there is no crossover, assigning zero as the site of
    % crossover
    oldpop(gen).individual(i).xsite=0;

    % Counting no. of descriptors present, i.e. No. of 1s in
    % chromosome
    oldpop(gen).individual(i).count=
    count_ones(oldpop(gen).individual(i),CHROM_LENGTH);

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Display initial population
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
fid = fopen('output_mp2.txt','w');
fprintf(fid,'Initial Population \n');
for i=1:POP_SIZE
    fprintf(fid,' P%d:  V=%d  F=%f
    \n',i,oldpop(gen).individual(i).count,
    oldpop(gen).individual(i).fitness);
end;

fprintf(fid,'Begin Evolution \n');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% In this loop two individuals are selected for mating.
% The individuals undergo crossover to yield two offspring.
% The offspring undergo mutation after which the fitness is
% calculated.
% The process of selection, crossover & mutation is repeated till
% the maximum no. of crossovers are reached.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
while (ncross< MAX_CROSSOVERS)
    % Initialize site of crossover
    jcross=0;

    % Initialize statistical parameters
    average(gen)=0;
    maximum(gen)=oldpop(gen).individual(1).fitness;
    minimum(gen)=oldpop(gen).individual(1).fitness;
    sumfitness(gen)=oldpop(gen).individual(1).fitness;

    % Get the statistics of each generation
```

```
  [maximum(gen),minimum(gen),average(gen),sumfitness(gen)]=
  statistics(POP_SIZE,maximum(gen),minimum(gen),
  average(gen),sumfitness(gen),oldpop(gen).individual);

% Find the fittest individual among the old generation
% Initializing the first individual as the fittest one
   fittest(gen).person=oldpop(gen).individual(1);
   index=1;

   for i=2:POP_SIZE
    if(oldpop(gen).individual(i).fitness>
    fittest(gen).person.fitness)
         fittest(gen).person=oldpop(gen).individual(i);
         index=i;
     else
         if((oldpop(gen).individual(i).fitness==
         fittest(gen).person.fitness)&&...
         oldpop(gen).individual(i).count<
         fittest(gen).person.count))
             fittest(gen).person=oldpop(gen).individual(i);
             index=i;
         end;
     end
 end
% Display fittest individual
fprintf(fid,'Fittest Individual of generation %d is the %d
  individual\n',gen,index);
for i=1:CHROM_LENGTH
    fprintf(fid,'%d',fittest(gen).person.chrom(i));
    if (mod(i,5)==0)
        fprintf(fid,' ');
    end;
end;
fprintf(fid,'\n');
fprintf(fid,'ncomp: %d \n',fittest(gen).person.ncomp);
fprintf(fid,'q2: %f \n',fittest(gen).person.q2);
fprintf(fid,'fitness: %f \n',fittest(gen).person.fitness);
fprintf(fid,'No. of variables: %d \n\n',
        fittest(gen).person.count);

% Selection of mate for crossover randomly
mate1=round(rand*(POP_SIZE - 1) + 1);
mate2=round(rand*(POP_SIZE - 1) + 1);

% Perform crossover
% Pick random integer position between specified limits
jcross=round(rand*(CHROM_LENGTH - 1) + 1);
ncross = ncross+1;
```

```
 % Get first part of the new string (individual's chromosome)
 % from parent.
  for j=1:jcross
      oldpop(gen).individual(m).chrom(j)=
      oldpop(gen).individual(mate1).chrom(j);
      oldpop(gen).individual(m+1).chrom(j)=
         oldpop(gen).individual(mate2).chrom(j);
  end

% Get the other part of the string (individual's chromosome)
% from another parent
  if (jcross~=CHROM_LENGTH)
      for j=jcross+1:CHROM_LENGTH
         oldpop(gen).individual(m).chrom(j)=
         oldpop(gen).individual(mate2).chrom(j);
         oldpop(gen).individual(m+1).chrom(j)=
         oldpop(gen).individual(mate1).chrom(j);
      end
  end
% If the probability is less then or equal to the mutational
% probability, then perform mutation
  if (rand<=MUTATE_PROB)

      % This mutates a string by flipping one of the bits. For
      % offspring1 Pick a random position in the string to
      % mutate
      bit=round(rand*(CHROM_LENGTH - 1) + 1);

      % Flip the bit at the random position
      if (oldpop(gen).individual(m).chrom(bit)==0)
          oldpop(gen).individual(m).chrom(bit)=1;
      else
          oldpop(gen).individual(m).chrom(bit)=0;
      end
  end;

  if (rand<=MUTATE_PROB)

      % For offspring2 pick a random position to mutate
      bit=round(rand*(CHROM_LENGTH - 1) + 1);

      % Flip the bit at the random position
      if (oldpop(gen).individual(m+1).chrom(bit)==0)
          oldpop(gen).individual(m+1).chrom(bit)=1;
      else
          oldpop(gen).individual(m+1).chrom(bit)=0;
      end
  end

    % Evaluate fitness of the new offspring
    [oldpop(gen).individual(m).fitness,
```

```
  oldpop(gen).individual(m).q2,
  oldpop(gen).individual(m).ncomp] =
  evaluate(oldpop(gen).individual(m),CHROM_LENGTH,x_block,
  COMPOUNDS);
  oldpop(gen).individual(m+1).fitness,
  oldpop(gen).individual(m+1).q2,
  oldpop(gen).individual(m+1).ncomp]=
  evaluate(oldpop(gen).individual(m+1),CHROM_LENGTH,
  x_block,COMPOUNDS);

% Record parentage of new offspring
oldpop(gen).individual(m).parent1=mate1;
oldpop(gen).individual(m).parent2=mate2;

oldpop(gen).individual(m+1).parent1=mate1;
oldpop(gen).individual(m+1).parent2=mate2;

% Record the site of crossover
oldpop(gen).individual(m).xsite=jcross;
oldpop(gen).individual(m+1).xsite=jcross;

% Count the number of 1s in the chromosome
  oldpop(gen).individual(m).count=count_ones
  (oldpop(gen).individual(m),CHROM_LENGTH);

  oldpop(gen).individual(m+1).count=count_ones
  (oldpop(gen).individual(m+1),CHROM_LENGTH);

  % Survival of the fittest. Set up an array of the 2
  % selected parents, their 2 offspring along with their
  % position
array=[mate1,oldpop(gen).individual(mate1).fitness;
    mate2,oldpop(gen).individual(mate2).fitness;
    m , oldpop(gen).individual(m).fitness;
    m+1,oldpop(gen).individual(m+1).fitness];

% Sort the 4 individuals as per their fitness. The parents
% & 2 offspring are sorted to find the fittest two
array2=sortrows(array,2);

% Get the position of the fittest two individuals in the old
% population
newindex1 = array2(3,1);
newindex2 = array2(4,1);

% Setup temporary individuals that keep track of the two
% fittest ones
temp1=oldpop(gen).individual(newindex1);
temp2=oldpop(gen).individual(newindex2);

% Replace the parents by the fittest two of the four
```

```
    oldpop(gen).individual(mate1)=temp1;
    oldpop(gen).individual(mate2)=temp2;

    % Pass the individuals from current to next generation
    for i=1:POP_SIZE
        oldpop(gen+1).individual(i)=oldpop(gen).individual(i);
    end;

    % Increment the generation counter
    gen=gen+1;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Display the evolving population
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:gen-1
    fprintf(fid,' #%1.1d:  L=%f  H=%f  A=%f \n',
    i,minimum(i),maximum(i),average(i));
end;

fprintf(fid,'Final Population \n');
fprintf(fid,'Individual Fitness \t No of ones \n');
% Display results of the final population
for i=1:POP_SIZE
    fprintf(fid,'\t %1.1d \t \t %f \t\t %d \n',
    i,oldpop(gen).individual(i).fitness,
    oldpop(gen).individual(i).count);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Creating a QSAR model.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Select the columns corresponding to a particular combination of
% descriptors
% Start off with the first column of x-block
k=1;
% Repeat the process from the first to last row of x-block
for i=1:CHROM_LENGTH
    if (fittest(gen-1).person.chrom(i)==1)
        % Select column from the x-block
        % Insert the selected column into another array
        for j=1:COMPOUNDS
            new_array(j,k)=x_block(j,i);
        end;
        k=k+1;
    end;
end;

myoptions=pls('options');
myoptions.display='off';
myoptions.plots='none';
% Calibration. Create a QSAR model.
```

```
model = pls(new_array,y_block,
        fittest(gen-1).person.ncomp,myoptions);
disp('Done!');
% Calculate the coefficients
[b,ssq,p,q,w,t,u,bin] = pls(new_array,y_block,
                       fittest(gen-1).person.ncomp);

fclose(fid);
```

evaluate.m

```
function [fitness,q2,ncomp]=
evaluate(individual,CHROM_LENGTH,x_block,COMPOUNDS)
% This function evaluates the fitness of an individual

% Sum of squares for scaled y-data for MP-dataset
ss=2.928518907;

% The max factor for choosing an additional component during PLS
MAX_FACTOR=2.00;

% Load y-block
load selwood_y_scaled;
y_block=selwood_y_scaled;

%Initialize the array to Null
new_array=[];

% Start off with the first column of x-block
k=1;
% Repeat the process from the first to last row of x-block
for i=1:CHROM_LENGTH
    if (individual.chrom(i)==1)
        % Select column from the x-block
        % Insert the selected column into another array
        for j=1:COMPOUNDS
            new_array(j,k)=x_block(j,i);
        end;
        k=k+1;
    end;
end;

% Restrict the initial number of components to 5 or less
if (size(new_array,2)>5)
   n=5;
else
    n=size(new_array,2);
end;

if (isempty(new_array)==0)
    % Perform PLS & Cross-validation
    [press,cumpress,rmsecv,rmsec,cvpred,misclassed]=
    crossval(new_array,y_block,'sim',{'loo'},n,0,0);
    % Initialize Optimal no. of components to 1
    ncomp=1;
    for i=1:n-1
        factor=((rmsecv(i)-rmsecv(i+1))/rmsecv(i))*100;
        % Choose to increase no. of components only if the error
        % reduces and the reduction in error is atleast
        % MAX_FACTOR
        if ((rmsecv(i+1)<rmsecv(i))&& factor>MAX_FACTOR)
```

```
            ncomp=i+1;
        end;
    end;

    % find q2
    q2 = (1 - (cumpress(ncomp)/ss));

    % Evaluate fitness
    fitness=( 1 - ((COMPOUNDS-1)*(1-q2)/(COMPOUNDS-ncomp)));
else
    q2=0;
    fitness=0;
    ncomp=0;
end;

%Clear out the arrays
clear new_array;
clear k;
clear cumpress;
```

count_ones.m

```
function [count]=count_ones(person,CHROM_LENGTH)
 % This function counts the number of 1s in the chromosome.
 % i.e. it counts the number of descriptors included.
 count=0;
 for i=1:CHROM_LENGTH
     if (person.chrom(i)==1)
         count=count+1;
     end;
 end;
```

statistics.m

```
function [maximum,minimum,average,sumfitness]=
statistics(POP_SIZE,maximum,minimum,average,sumfitness,abc)
% This function finds the minimum, maximum, average & sum of
% fitnesses per generation

% The first individual's fitness is assigned as initial max, min
% & sumfitness.
% For the entire population, find max, min, & sum
for j=2:POP_SIZE
    sumfitness=sumfitness+abc(j).fitness;
    if (abc(j).fitness>maximum)
        maximum=abc(j).fitness;
    end
    if (abc(j).fitness<minimum)
        minimum=abc(j).fitness;
    end
end

% Find average
average=sumfitness/POP_SIZE;
```

REFERENCES

1. National Survey On Drug Use and Health: National Findings. 2002. United States Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

2. Volkow N.D., Ding Y. S., Fowler J.S., Wang G. J., Logan J., Gatley S. J., Dewey S., Ashby C., Libermann J., Hitzemann R., and Wolf A. P. 1995. Is Methylphenidate Like Cocaine? Studies on Their Pharmacokinetics and Distribution in the Human Brain. *Arch. Gen. Psych.* **52** 456-463.

3. Volkow N.D., Wang G. J., Gatley S. J., Fowler J.S., Ding Y. S., Logan J., Hitzemann R., Angrist B. and Libermann J. 1996. Temporal Relationships Between the Pharmacokinetics of Methylphenidate in the Human Brain and its Behavioral and Cardiovascular Effects. *Psychopharm.* **123** 26-33.

4. Volkow N.D., Wang G. J., Fowler J.S., Gatley S. J., Ding Y. S., Logan J., Hitzemann R., Angrist B. and Liebermann J. 1996. *Relationship Between Psychostimulant-Induced "High" and Dopamine Transporter Occupancy.* in *Proc. Natl. Acad. Sci. USA.*

5. Volkow N.D., Wang G. J., Fowler J.S., Logan J., Angrist B., Hitzemann R., Liebermann J. and Pappas N. 1997. Effects of Methylphenidate on Regional Barin Glucose Metabolism in Humans: Relationship to Dopamine D2 Receptors. *Am. J. Psychiatry* **154** 50-55.

6. Lowinson, J, Ruiz, P., Millman, R., and Langrod, J. 1997. *Substance Abuse: A Comprehensive Textbook.* 3rd ed: Williams L. and Wilkens.

7. Stahl, Stephen M. 2000. *Essential Psychopharmacology.* 2 ed: Cambridge University Press.

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

9. Ritz M. C., Lamb R. J., Goldberg S. R. and Kuhar M. J. 1987. Cocaine Receptors on Dopamine Transporters Are Related to Self-Administration of Cocaine, in *Science.* 1219-1223.

10. Madras K., Fahey M. A., Bergman J., Canfield D. R. and Spealman R. D. 1989. Effects of Cocaine and Related Drugs in Nonhuman Primates. I. [^{3}H] Cocaine Binding Sites in Caudate Putamen. *J. Pharm. Expt. Ther.* **251** 131-141.

11. Hansch C. 1968. A Quantitative Approach to Biochemical Structure-Activity Relationships. *Accts. Chem. Res.* **2** 232-239.

12. Goldberg D. E. 1989. *Genetic Algorithm in Search, Optimization, and Machine Learning*. Reading, MA.: Addison-Wesley.

13. Riberio Filho J.L., Treleaven P.C. and Alippi C. 1994. Genetic-algorithm Programming Eenvironments. *IEEE Computer* **27** 28-43.

14. Holland J. H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press.

15. SYBYL ® 6.9 Tripos Inc. 1699 South Hanley Rd., St. Loius, Missouri, 63144, USA.

16. Wold H. 1982. Soft Modeling, The Basic Design and Some Extensions, in *Systems Under Indirect Observation*: Amsterdam.

17. Miyashita Y., Li Z. and Sasaki S. 1993. Chemical Pattern Recognition and Multivariate Analysis for QSAR Studies. *Trends Anal. Chem.* **12** 50-60.

18. Wold S., Johannson E. and Cocchi M. 1993. PLS-Partial Least Squares Projection to Latent Structures, in *3D QSAR in Drug Design, Theory, Methods, and Applications*ESCOM Science Publishers: Leiden. 523-550.

19. Wold, S., Sjostrom, M., and Eriksson, L. 2001. PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58** 109-130.

20. Kier L. B. and Hall L. H. 1999. *Molecular Structure Description, The Electrotopological State*: Academic Press.

21. Testa B., Kier L B. and Carrupt P.A. 1997. A Systems Approach to Molecular Structure, Intermolecular Recognition, and Emergence-Dissolvence in Medicinal Research. *Med. Res. Rev.* **17** 303-326.

22. Kier L. B. and Hall L. H. 1991. The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling, in *Reviews of Computational Chemistry*, Boyd, D. B. and Lipkowitz, K.

23. Hall L. H. 1989. Computational Aspects of Molecular Connectivity and its Role in Structure-Activity Modeling, in *Computational Chemical Graph Theory*, Rouvray D. H., Nova Press.

24. Golbraikh A. and Tropsha A. 2002. Beware of q^2! *Journal of Molecular Graphics and Modelling* **20** 269-276.

25. MatLab Copyright by The Mathworks, Inc. 1994-2004. 3 Apple Hill Drive, Natick, MA 01760-2098 USA.

26. PLS Toolbox Copyright by Eigenvector Research Inc. 2004. 830 Wapato Lake Road, Manson, WA 98831.

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

28. Selwood D.L., Livingstone D.J., Comley J.C.W., O'Dowd A.B., Hudson A.T., Jackson P., Jandu K.S., Rose V.S. and Stables J.N. 1990. Structure-Activity Relationships of Antifilarial Antimycin Analogues: A Multivariate Pattern Recognition Study. *J. Med. Chem.* **33** 136-142.

29. Dunn W.J. and Rogers D. 1996. Genetic Partial Least Squares in QSAR, in *Genetic Algorithms in Molecular Modeling*, Devillers, Academic Press. 109-130.

30. Wikel J. and Dow E. 1993. The Use of Neural-Networks for Variable Selection in QSAR. *Bioorg. Med. Chem. Lett.* **3** 645-651.

31. Rogers D. and Hopfinger A. J. 1994. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **34** 854-866.