Spring 2013

QSAR modeling of chemical penetration enhancers using novel replacement algorithms

Hui Qiu
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

QSAR MODELING OF CHEMICAL PENETRATION ENHANCERS USING NOVEL REPLACEMENT ALGORITHMS

by
Hui Qiu

The applications of transdermal delivery are limited because of the resistance of the skin to drug diffusion. Only potent drugs, with molecular weight less than 500 Da, are suitable to cross the skin barrier. Chemical Penetration Enhancers (CPEs) are used to promote the absorption of solutes across the dermal layers. In this investigation, a Quantitative Structure-Activity Relationship (QSAR) model is applied to relate chemical penetration enhancer structures with the flux enhancement ratio through a statistical approach.

A database, consisting of 61 non-polar CPEs, is selected for the study. Each compound is represented by 777 QSAR descriptors, which encode the physical characteristics of the CPE and its structure. Selection replacement techniques are used to choose the eight most important descriptors. The enhancement ratio, an evaluation of the effect of the CPE, correlates well with this subset of features. The QSAR model can be adopted to predict factors that need to be adjusted to improve permeation of the drug through the skin.

Three QSAR models are developed using different algorithms: forward stepwise regression (FSR), replacement (RM) and enhanced replacement (ERM) techniques. The first two methods yield equations with poor predictive power. The enhanced replacement method gives the best results, which meet cross-validation criteria: q^2 = 0.79, 0.63 and 0.76 for the training set, test set and combined data, respectively. These results meet the predetermined criteria.

QSAR MODELING OF CHEMICAL PENETRATION ENHANCERS USING NOVEL REPLACEMENT ALGORITHMS

by
Hui Qiu

A 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 Pharmaceutical Bioprocessing

Department of Chemical, Biological & Pharmaceutical Engineering

May 2013

APPROVAL PAGE

QSAR MODELING OF CHEMICAL PENETRATION ENHANCERS USING NOVEL REPLACEMENT ALGORITHMS

Hui Qiu

Dr. Laurent Simon, Thesis Advisor Date
Associate Professor of Chemical, Biological & Pharmaceutical Engineering, NJIT

Dr. Piero M. Armenante, Committee Member Date
Distinguished Professor of Chemical, Biological & Pharmaceutical Engineering, NJIT

Dr. Carol A. Venanzi, Committee Member Date
Distinguished Professor of Chemistry and Environmental Science, NJIT

Dr. Norman Loney, Committee Member Date
Professor and Chair of Chemical, Biological & Pharmaceutical Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Hui Qiu

Degree: Master of Science

Date: May 2013

Undergraduate and Graduate Education:

- Master of Science in Pharmaceutical Bioprocessing,
  New Jersey Institute of Technology, Newark, NJ, 2013

- Bachelor of Science in Pharmaceutics,
  Jiangsu University, Zhenjiang, P. R. China, 2007

Major: Pharmaceutical Bioprocessing

ACKNOWLEDGMENT

I wish to express my deepest gratitude to Professor Laurent Simon, my thesis advisor, for his valuable guidance, which makes this thesis possible. His help opened my eyes to specialty areas that I have never explored before. My special thanks to committee members, Dr. Piero M. Armenante, Dr. Carol A. Venanzi and Dr. Norman Loney for taking time out of their busy schedules to participate in my defense.

I would also like to thank to Professor Laurent Simon for assisting me with the collection of data which facilitates the completion of the thesis in two semesters.

Last, I would like to thank my parents who always encourage me to tackle difficult problems when I am in doubt and want to avoid challenges. Without them, I would not have the opportunity to develop QSAR models and finish this work.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

0D	Zero-Dimensional
1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
AD	Applicability Domain
ANN	Artificial Neural Network
CPEs	Chemical Penetration Enhancers
DMSO	Dimethyl Sulfoxide
ER	Enhancement Ratio
ERM	Enhanced Replacement Method
FDA	Food and Drug Administration
FSR	Forward Stepwise Regression
GA	Genetics Algorithm
IC_{50}	50% Inhibitory Concentration
MMP+	Molecular Modeling Pro™ Plus
$Mold^2$	Molecular 2D Descriptors
MRM	Modified Replacement Method
PG	Propylene Glycol
RG	Radius of Gyration
RM	Replacement Method
SD	Standard Deviation

SMILES	Simplified Molecular Input Line Entry System
Tg	Glass Transition Temperature
QSAR	Quantitative Structure-Activity Relationship
TDDS	Transdermal Drug Delivery Systems

LIST OF DEFINITIONS

Applicability Domain	Region of application of a QSAR model.
Nonpolar	Property of a compound that does not have a dipole.
Penetration Enhancer	Pharmaceutically inert chemicals which can partition into and interact with the stratum corneum when used in a transdermal formulation. These compounds can lessen the skin resistance to drug transport.
QSAR	Quantitative structure–activity relationships connecting a compound's chemical structure to its pharmacological activity.
Surfactant	A surface-active agent.

CHAPTER 1

INTRODUCTION

1.1 Objective

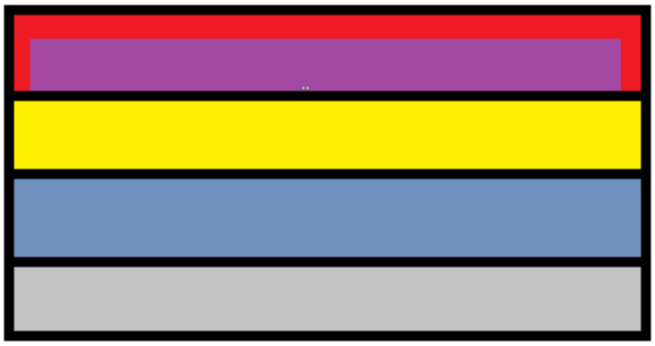
Chemical penetration enhancers (CPEs) are useful because of their ability to overcome skin barrier resistance and to stimulate drug transport across the skin. However, the task of choosing CPEs, for a particular drug, can be challenging as it may involve expensive and time-consuming trial-and-error experiments. The objective of this thesis is to build a quantitative structure-activity relationship (QSAR) model from a set of skin penetration enhancers. Mathematical expressions are derived that relate the Enhancement Ratio (ER), a measure of the effectiveness of the CPEs, to the enhancers' molecular descriptors. Replacement methods are applied to select a subset of structural features. One of the main advantages of the study is the creation of a platform to help connect the drug release rate to a few QSAR molecular descriptors. This procedure can be conducted to screen the best CPEs from a large database.

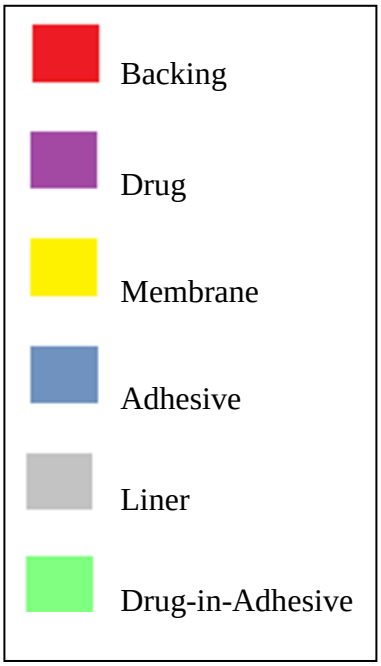
1.2 Background Information

Transdermal drug-delivery systems (TDDS) are developed to transport drugs through the skin to the systemic circulation. The most common preparation is a medicated adhesive patch that is placed on the skin to administer a specific dose of medication. Traditional TDDS have a porous membrane covering a reservoir of medication embedded in an adhesive. Figure 1.1 presents three different types of TDDS drug release mechanisms.

After traversing the dermal layers, the molecules enter the bloodstream through the capillary walls and reach an effective concentration at the target site.

a. Reservoir system.





b. Matrix system without a rate-controlling membrane.

c. Matrix system with a rate-controlling membrane.

Figure 1.1 TDDS different drug release mechanism.

TDDS can be used as local therapies or to produce systemic effects. It is especially useful for chronic disease because the method is noninvasive, self-administered and provides the patient with a controlled release of the medication. Because the skin prevents the entry of foreign substances and guards against nutrient and water loss, TDDS are also applied for topical skin disease.

Apart from these benefits, TDDS have unique advantages when compared with other dosage forms. They can prevent metabolism of the active pharmaceutical ingredient (API) and avoid first-pass effect of the liver when the medication is taken orally. Risks, such as cross-infections through reuse of needles, posed by hypodermic injections are averted. In addition, the patches are not painful and especially safe for low-income and developing countries [1,2].

The first transdermal patch approved by the FDA as a prescription drug can be traced back to December 1979. Scopolamine was the API used to treat motion sickness [3-5]. Because of the slow delivery rate and the fact that only a small amount of the agent can be delivered through the skin, side effects, such as dry mouth, dizziness and hallucination, associated with potent scopolamine were decreased [6,7]. During the mid-1980’s, a nicotine patch was studied and a patent issued. This new dosage form attracted the public attention because of its efficacy in reducing craving for cigarettes [8-10]. Today, there are more than twenty different types of transdermal delivery systems, such as nitroglycerin, estradiol and lidocaine, which are approved by the FDA [11]. The research environment spurs the emergence of combination patches, containing multiple drugs, iontophoretic and ultrasonic delivery systems [1].

The transdermal route is only appropriate for a few drugs because of the presence of the stratum corneum, mainly composed of dead cells. A rational selection is based on key factors: physico-chemical properties of the drug (such as solubility, crystallinity, molecular weight, polarity, melting point), pharmacokinetic parameters (such as half-life, volume of distribution, total body clearance, therapeutic plasma concentration, bioavailable factor) and biological factors (skin toxicity, site of application, allergic reactions, skin metabolism, skin permeability) [12]. All available drugs formulated in a patch have three fundamental characteristics, which allow them to cross the skin effectively: low molecular mass (<500 Da), high lipophilicity (oil soluble) and small required dose (up to milligrams) [13]. These properties are found in first-generation delivery systems. The emergence of second-generation (e.g., devices with CPEs and iontophoresis) and third-generation delivery systems (e.g., microneedles) make a significant contribution to medicine [1]. Physical or chemical enhancement techniques are very effective delivery systems that are used to overcome the resistance of the stratum corneum. One of the popular methods is the use of CPEs. CPEs are chemical compounds added to the formulation in TDDS to increase skin permeability [14].

In the past, less attention was paid to hydrophilic solutes in academic settings because of the lipid bilayer structure of skin. The situation has changed with the introduction of CPEs. More investigations on mathematical modeling of transdermal transport of hydrophilic drugs are conducted [15,16]. Passive transport these solutes also allow diffusion through the epidermis [13].

Penetration enhancers should be nontoxic, nonirritating, and deprived of any pharmacological activity. In addition, compatibility among the CPEs, the drug and other

materials present in the system is desirable. The skin can recover its barrier function quickly after removal of the CPE. Although it is difficult to find a CPE with all the desired properties, many compounds show most of these features.

CPEs are classified as terpenes, terpenoids and essential oils, pyrrolidones and their derivatives (such as N-methyl-2-pyrrolidone), fatty acids and esters, sulfoxide and similar compounds (e.g., dimethyl sulfoxide (DMSO), alcohols, glycols and glycerides, (e.g., ethanol, propylene glycol); azones (e.g., laurocapram) and other miscellaneous enhancers (such as phospholipids, lipid synthesis inhibitors, cyclodextrin complexes, amino acid derivatives, clofibric acid, dodecyl-N,N-dimethylamino acetate and enzymes) [17]. Given a particular active pharmaceutical ingredient, it would be a massive project to conduct experiments using compounds in each group to select the most effective CPE. Therefore, modern computational techniques are applied to assist in the process and suggest CPEs that can be tested in the laboratory.

Quantitative structure–activity relationships (QSAR) are empirical linear models that have been widely used in the pharmaceutical, chemical sciences, environment protection policy and health research. These models map a set of physico-chemical properties or theoretical molecular descriptors of compounds (X) to a response variable (Y) where Y represents a biological activity. Examples of response variables are the molar concentration of a compound that inhibits 50% growth of bacteria (IC_{50}) [18], the octanol-water partition coefficient [19] and the cytotoxic activity [20]. Therefore, QSAR models are only one kind of regression models. In addition, such an approach can also be used for classification purposes where the predictor variable is assigned a categorical value. Regardless of the specific application of the QSAR model, the number of

descriptors can be very large (i.e., 1000). As a result, it is important to choose particular descriptors that best explain the response variable.

The goal of this project is to derive a QSAR model to relate structures of a set of nonpolar CPEs to the flux enhancement ratio using hydrocortisone as a control.

Hydrocortisone, also known as Cortisol, is a steroid hormone. Its structure is shown in Figure 1.2. The main glucocorticoid is secreted by the zona fasciculata of the adrenal cortex [21]. It has multiple effects and is used to treat inflammation, allergy, collagen diseases, asthma, adrenocortical deficiency, shock and some neoplastic conditions [22]. The main function is to increase blood sugar, stimulate gluconeogenesis, suppress the immune system and aid in fat, protein and carbohydrate metabolism [23].

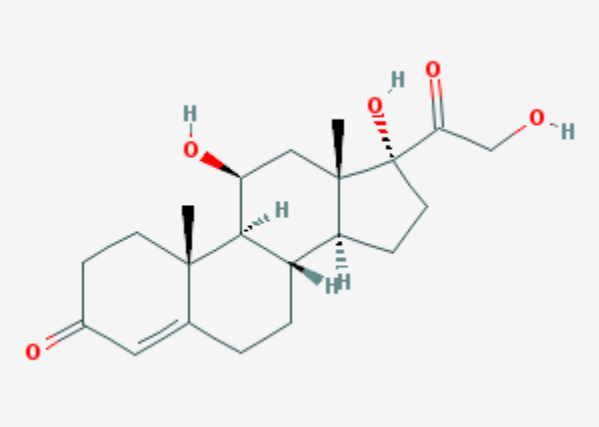


Figure 1.2 2D structure of hydrocortisone.

Source: http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=5754 Retrieved on March 30, 2013.

Several dosage forms of hydrocortisone are ready for use: lotion, cream, gel, ointment and enema. It is available both as an over-the-counter (low-strength, 0.5% or 1%) or prescription drug (high-strength). Ghafourian et al. investigated the enhancing activities of terpenes towards hydrocortisone [24,25]. Simon et al. studied the influence of 61 non-polar CPEs on the transdermal delivery rate of hydrocortisone [14].

The descriptors are usually produced by software packages such as Molecular Modeling Pro™ Plus (MMP+). Three different methods for deciding an optimal subset of descriptors were studied by Simon and Abdelmalek [14]: the replacement method (RM), the enhanced replacement method (ERM) and the traditional forward stepwise regression (FSR). The hydrogen bond acceptor, polar surface area, moment of inertia, glass transition temperature (T_g), molar volume, radius of gyration (RG), dipole moment and polarity correlates well with the ER [14]. However, these eight properties were selected from thirty-one descriptors previously used in CPE studies. The present study will use a larger database to build the QSAR model. The applicability domain, which was not determined in the previous study, will be defined.

CHAPTER 2

CHEMICAL PENETRATION ENHANCERS AND MOLECULAR DESCRIPTORS

2.1 Overview

Chemical penetration enhancers are widely used in dermal related products, such as transdermal drug-delivery systems and cosmetics. Common CPEs are classified as terpenes, terpenoids and essential oils; pyrrolidones and their derivatives; fatty acids and esters; sulfoxide and similar compounds; alcohols, glycols and glycerides; azones and other miscellaneous enhancers [17]. Each group has specific advantages and disadvantages. For example, DMSO, a common chemical penetration enhancer, is a sulfoxide. It increases penetration because of its interaction with the stratum corneum and solubilization with the drug itself. However, DMSO might cause skin irritation, or may be fatal to a human when exposed to high concentrations [26].

Azone, also named laurocapram, is a clear and colorless liquid that dissolves poorly in water but mixes well with most organics. It affects the absorption of hydrophilic drugs more significantly than lipophilic agents. Laurocapram has an effective concentration in the range of 1-6%. Although the CPE's action on the skin is slow, about two to ten hours, its impact could last several days because azone accumulates gradually in the stratum corneum [27].

Except for the amino-acid derivatives, most CPEs would provoke skin irritation after traversing the epidemic cells. These derivatives are absorbed into the stratum corneum lipid barrier, increase the mobility of the barrier lipids and improve drug

permeation, as a result. Janůšová et al. studied the proline derivative L-Pro2 and found that it could increase ER by up to 40% and yielded better results than azone. Contrary to single-chain lipid-like substances, double-chain lipid-like compounds did not increase the flux, as noted in [28].

A higher permeation can be achieved when different CPEs are combined [5]. Propylene glycol (PG) is a good example. When used alone, PG mildly improves the transport of estradiol and 5-fluorouracil [29]. It exhibits a synergistic action when combined with other CPEs, such as ethanol and oleic acid. Janůšová et. al found that a mixture of proline derivative L-Pro2 and PG increased the theophylline flux 40 times higher than that of PG acting alone and 2.5 times greater than that of L-Pro2 in water [28].

2.2 Chemical Penetration Enhancers in This Study

Codes representing the molecular structures of the sixty-one CPEs used in the study are listed in Appendix A.

2.3 Molecular Descriptors

Computational applications in chemoinformatics and toxicoinformatics are becoming popular in drug research and discovery phases. Many predictive models are based on the compound structure or activity. Since the structure alone cannot be represented by the model, it is necessary to extract structural information and convert it into numerical or digital representations. “The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some

standardized experiment" [30]. Therefore, molecular descriptors can be experimental or theoretical. Examples of experimental measurements are log P, dipole moment, any physico-chemical properties or biological activities. Theoretical descriptors are symbolic representations of the molecules and can be further classified as:

- Zero-dimensional (0D) descriptors, obtained from the chemical formula and independent of the molecular structure (i.e. atom type counts, the molecular mass, atomic charge)
- 1D-descriptors, which represent a list of structural fragments (i.e. list of structural fragments, functional group count)
- 2D-descriptors, which are topological indices that are derived by converting molecular structures into graphs (i.e. graph invariants)
- 3D-descriptors, calculated from a geometrical or 3D representation of a molecule (i.e. 3D-MoRSE descriptors, WHIM descriptors, GETAWAY descriptors, quantum-chemical descriptors, size, steric, surface and volume descriptors)
- 4D-descriptors, which come from a stereo-electronic or lattice representation (i.e. those stemmed from GRID or CoMFA methods, Volsurf) [31,32].

Molecular descriptors help interpret properties of existing CPEs as well as guide the design of new molecules. With the emergence of these new metrics, a descriptor may fail to describe important characteristics of a compound while providing a wealth of information on the structure of another molecule. Descriptors are instrumental in assessing the importance of theoretically established models [30]. When they are the result of an optimization strategy, these numbers are expected to correlate well with at least one property and be sensitive to gradual change in molecular structures [33].

In this study, SMILES, a linear notation system is used. Such notations are two-dimensional representation alternatives to the molecular graph [30].

2.4 Screening of Molecular Descriptors

The software Molecular 2D Descriptors (Mold^2) Generator Software, version 2 (center for bioinformatics, NCTR, FDA, USA) is used to transform a CPE's structure into 777 descriptors. Part of them listed in Appendix B. MATLAB® (The Mathworks Inc., MA, USA) is later used. Only the eight most relevant descriptors, a number much less than the supplied data, are selected to represent the CPEs. The output of the QSAR model is the ER.

CHAPTER 3

ENHANCEMENT RATIO

3.1 Enhancement Ratio

Skin penetration enhancement techniques were developed to improve bioavailability and increase the range of drugs that can be administered topically and transdermally [31]. An enhancement ratio (ER) is usually calculated to give an accurate measure of the effectiveness of the accelerant. This number is defined as the ratio of the flux in the presence of a fixed concentration of the CPE to the delivery rate when the CPE is not added to the formulation [14, 34]:

$$\textit{Enhancement Ratio (ER)} = \frac{Delivery\ rate\ after\ enhancer\ treatment}{Delivery\ rate\ before\ enhancer\ treatment} \tag{3.1}$$

3.2 Enhancement Ratio Test Method

Two methods are available for the evaluation of skin permeation *in vitro*: the static and flow-through cell experiments [35-37]. The static Franz diffusion cell test is widely used for calculating ER values. This apparatus contains two compartments: a receptor cell (a static receptor solution reservoir with a side-arm sampling port) and a donor cell [38] (Figure 3.2). In flow-type Franz cells, the receptor medium is continuously circulated through the receptor compartment.

In the original work, data were obtained using static Franz diffusion cells [38]. These systems are also recommended by the FDA [39]. CPEs were first applied to the membranes for 24 hours and placed between the two compartments. Samples were removed from the receptor cell and replaced with the same volume of fresh solution at

pre-set time intervals. Drug concentration was determined by spectrophotometry and the diffusion coefficient was estimated after applying to Higuchi equation [40,41]. The early-time release data (less than 60%) can be modeled by the following equation [42]:

$$\frac{Q}{A} = 2C_0 \cdot \left(\frac{Dt}{\pi}\right)^{1/2} \tag{3.2}$$

where Q is the cumulative amount of drug collected in the receiver chamber (mg), A is the area available for diffusion (cm^2), D is the diffusion coefficient (cm^2/h), C_0 is the initial concentration (mg/cm^3) and t is the time.

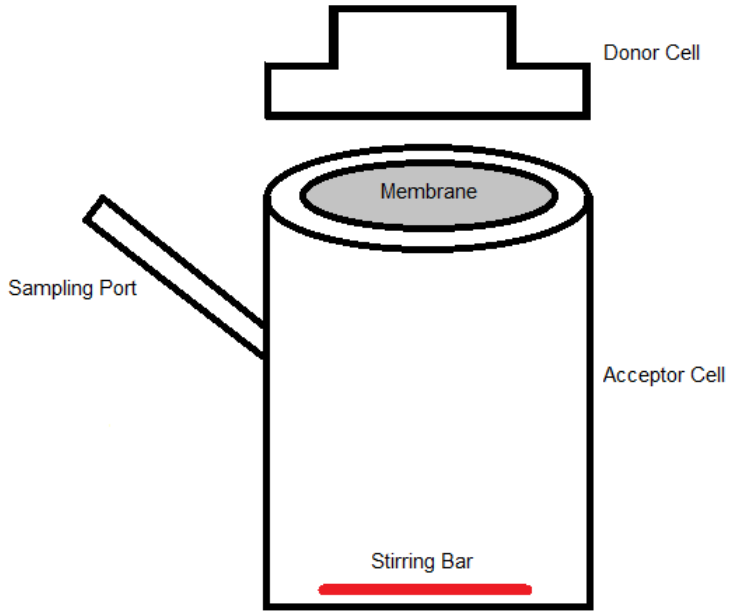


Figure 3.2 Scheme of the static Franz diffusion cell.

CHAPTER 4

DESIGN OF SKIN PENETRATION ENHANCERS USING REPLACEMENT METHODS FOR THE SELECTION OF MOLECULAR DESCRIPTORS

In Section 1.2, three feature-selection methods in QSAR studies were briefly discussed: RM, ERM and FSR. Another option is to implement these algorithms first to extract the best descriptors. The second phase would consist of building nonlinear input-output mappings, such as an artificial neural network (ANN), to represent the data [14]. Although this approach produced superior results, when compared to a traditional QSAR, this study is mainly focused on the use of Mold^2 with advanced replacement techniques to produce linear relationships between descriptors and the ER.

4.1 Multiple Linear Regression

Multiple linear regressions are commonly used in QSAR studies to develop the relationship between descriptors (independent variable) and the dependent variable (biological activity). Patel et al. (2002) derived a correlation between skin permeability and the compound's hydrophobicity, molecular size and hydrogen bonding ability [43]. Mercader et al. (2010) credited changes in 50% inhibitory concentration (IC_{50}) to the mean topological charge index of order 1, the average molecular weight, 3D – MoRSE – signal 30 weighted by atomic masses and the first component symmetry directional WHIM index weighted by atomic polarizabilities [44].

The generic form of a multiple linear regression model, given p observations, is:

$$y = \alpha + \beta_1 X_i + \beta_2 X_2 + \cdots + \beta_p X_p \tag{4.1}$$

where α is the intercept, Xi is a feature, also called a descriptor (e.g., hydrophobicity, molecular weight) and β is the partial regression coefficient.

4.2 Replacement Method

Techniques, such as FSR and Genetic Algorithm (GA), can be used to conduct the multi-parameter, non-linear inverse analysis [45]. The latter strategy is inspired by the process of Darwinian evolution because the mechanism selects the fittest individuals over several generations. Mercader et al. proposed the replacement method (RM) which yields systems with better statistical parameters than the GA and the FSR. The GA replaces the descriptors randomly and does not calculate the error in the regression coefficient as in the case of RM. In addition, tuning of the mutation probability, crossover probability and generation gap complicates the process [46].

The response (y) may be related to a host of molecular descriptors (X). The challenge is to decide on a subset of relevant input variables $d=\{X_1, X_2,\ldots, X_p\}$ from a large pool of dimension N. An effective algorithm should be applicable to problems with the size of d in the range of 0 to 10 and $N > 1000$. The RM procedure, based on the minimum standard deviation (SD), can select a subset of the population. First, descriptors are chosen randomly and replaced one at a time by the remaining elements. The set that produces the smallest SD is kept. Second, the input variable with the coefficient showing the largest relative deviation is replaced. The procedure continues until there is no need to make additional substitutions. In the end, the best variables of the first path are obtained. The process is repeated for all possible paths and the predictor variables with the smallest SD are kept [46, 47].

4.3 Enhanced Replacement Method

The enhanced replacement method (ERM), an improved version of RM, was developed later by Mercader et al. It yields even better statistical parameters than the RM. The technique is similar to that of the RM because it uses the minimum standard deviation for identifying the descriptors [48]. The main difference is that the ERM combines two algorithms in the sequence: RM and MRM (modified replacement method). The latter technique is similar to the RM except that, in this case, the descriptor with the largest error is replaced in each step regardless of the standard deviation value [46].

4.4 Stepwise Regression Method

Stepwise regression creates a mathematical representation by adding or removing variables continuously until a subset based on the F-ratio statistic is selected. There are three main approaches: forward selection, backward elimination and bidirectional elimination. In the forward selection, variables are added to an “empty” model. The regression terms, that are not used, are examined one by one to see whether they could improve the model. The procedure continues until further improvement is no longer viable. Backward elimination starts from all candidate predictors and removes the input variable with the smallest F-ratio [49,50]. Bidirectional elimination is a combination of the above two schemes.

4.5 Classification of Training and Testing Data

QSARs make it possible to interpret molecular properties and/or to predict other unknown compounds. Ideally, the tool should be able to predict chemical properties or

biological activities of those compounds that are not included in the model development [51]. The most common method of achieving this goal is to divide compounds into training and test samples. While the learning phase relies on the training data, validation uses the test set. Selection of data belonging to these two categories is not arbitrary. The training data should cover the whole descriptor space and the test set could be within the range of (or close to) points in the training set. Therefore, the training examples should be diverse and representative of the compounds studied.

4.6 Applicability Domain

QSAR modeling assumes that all compounds share similar properties and influence the dependent variable using analogous mechanisms. The model is expected to perform well on samples that are within a domain fixed by the training data. This "applicability domain" (AD) is an important concept in building a QSAR and is based on the compound's characteristics or biological activities (e.g., physical, chemical and biological features). The QSAR prediction cannot be considered reliable if the data falls outside of this domain. Eriksson et al. introduced several techniques of defining AD to assess the predictive power of the model [52]. In this study, the determination of leverage values (the William's plot) is adopted. Leverage h is a scale of the influence of each compound on the formula. The leverage of a compound in the original variable space is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i \qquad (i=1,2,...n) \tag{4.2}$$

where x_i is the feature vector of the considered compound and X is the descriptor matrix constructed from the training set [53]. The warning leverage (h^*) is defined as:

$$h^* = \frac{3P}{n} \tag{4.3}$$

where *n* is the number of training compounds, *P* is the number of explanatory variables plus one [54].

William's plot is the graph of standardized residual versus leverage values. Other than the influence of a point, the plot can also show the Euclidean distances of all compounds to the model through standardized residuals for both training and test data.

The plot can be analyzed as follows: If h is greater than h*, the compound has a large influence on the model. If the cross-validated standardized residual is larger than three standard deviations, the compound is outside of the range and is considered a response outlier [54]. Therefore, h is less than h* for a point in the domain and the standardized residual is within three standard deviations.

4.7 Evaluation of QSAR Results

It is important to have a method to validate QSARs. The leave-one-out (LOO) or leave-some-out (LSO) cross-validation procedures are the common assessment tools. In this study, LSO is chosen. The coefficient of determination q^2 is used to evaluate the accuracy of the model, which is calculated according to the formula:

$$q^2 = 1 - \frac{SS_{err}}{SS_{tot}} \tag{4.4}$$

where SS_{err} is the sum of squares of residuals and SS_{tot} is the total sum of squares, which is proportional to the sample variance:

$$SS_{err} = \sum (y_i - \hat{y}_i)^2 \tag{4.5}$$

$$SS_{tot} = \sum (y_i - \bar{y})^2 \tag{4.6}$$

with

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{4.7}$$

where y_i, $\hat{y}_i$ and $\bar{y}$ are the experimental, predicted and averaged activities, respectively.

According to Golbraihk et al., a QSAR is predictive if the following conditions are satisfied for the test set [51]:

$$R^2 > 0.6 \tag{4.8}$$

$$q^2 > 0.5 \tag{4.9}$$

$$\frac{R^2 - {R_0}^2}{R^2} < 0.1 \; or \; \frac{R^2 - {R_0'}^2}{R^2} < 0.1 \tag{4.10}$$

$$0.85 \ll k \ll 1.15 \; or \; 0.85 \ll k' \ll 1.15 \tag{4.11}$$

where R is the correlation coefficient between $\hat{y}_i$ and y_i; R_0^2 is the coefficient of determination between $\hat{y}_i$ and y_i.and $R_0'^2$ is the coefficient of determination of y_i versus $\hat{y}_i$; k and k' are the slopes of the regression lines (i.e., between $\hat{y}_i$ and y_i and y_i versus $\hat{y}_i$, respectively) through the origin.

CHAPTER 5

IMPLEMENTATION

In previous contributions, QSAR models were developed to predict the effects of sixty-one (61) nonpolar enhancers on the transdermal delivery of hydrocortisone [14, 55]. In [14, 34], the software MOLECULAR MODELING PRO™ Plus (MMP+) was used to calculate the CPEs. In [14], it estimated nearly 114 descriptors of the molecular structures. However, this number was later reduced to 31 features commonly used in studies of skin penetration enhancers. A linear model based on eight input variables could predict the ER because of the pretreatment with the accelerants. R^2 values of 0.683, 0.683 and 0.671, corresponding to ERM, RM and FSR algorithms, respectively, were computed. In this study, the QSAR was based on 777 variables calculated using the Mold^2 software. This large and diverse set of descriptors encodes 1D and 2D chemical structure information [56].

5.1 Software and the Flow Diagram

The following tools are applied in this study:

- Online SMILES Code Translator
- PUBCHEM
- Molecular 2D Descriptors (Mold^2) Generator Software, version 2 (center for bioinformatics, NCTR, FDA, USA)
- CHEMBENCH
- MATLAB® (The Mathworks Inc., MA, USA)

Some of the websites relevant to this work are provided in [57–59]. A flow diagram, describing how these tools are implemented in sequence, is given in Figure 5.1.

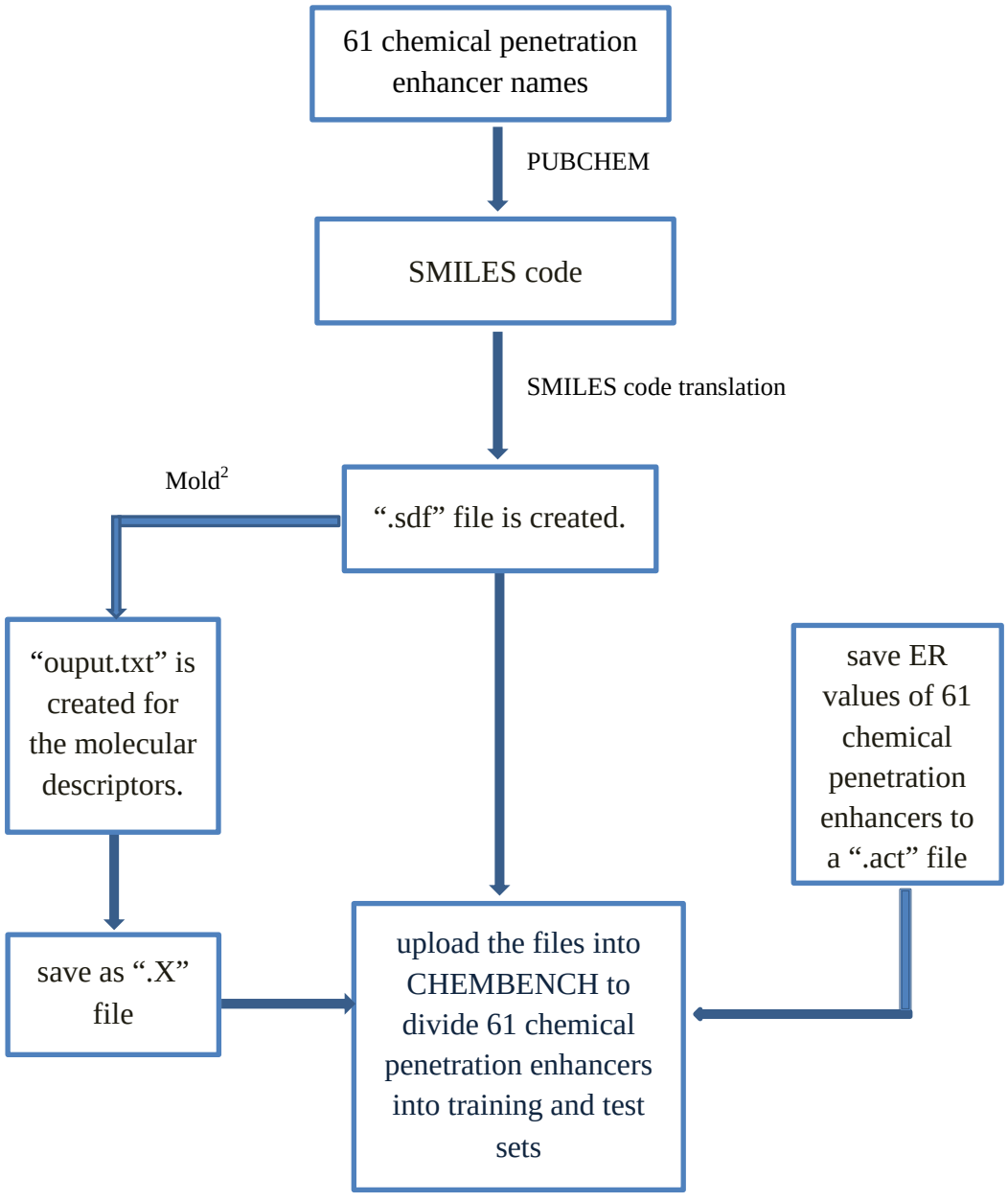


Figure 5.1 Flow diagram of software.

5.2 SMILES Code

Simplified Molecular Input Line Entry System (SMILES) (http://www.daylight.com, 2006; http://www.epa.gov, 2006) is a chemical language that helps explain the structure of a compound. The system encodes information in a way that the molecular data can be integrated for QSAR applications [60]. A SMILES string, though produced by a canonicalization algorithm, is unique for each structure.

There are five basic syntax rules for SMILES:

1 atoms and bonds

2 simple chains

3 branches

4 rings

5 charged atoms

The U.S. Environmental Protection Agency provides more details on these five rules on its website [61].

5.2.1 SMILES Code Acquisition

The PUBCHEM website [57] provides the SMILES code. Figure 5.2 is the screenshot of a “PUBCHEM structure search." Once a compound name is entered, a search is launched and a code is created.

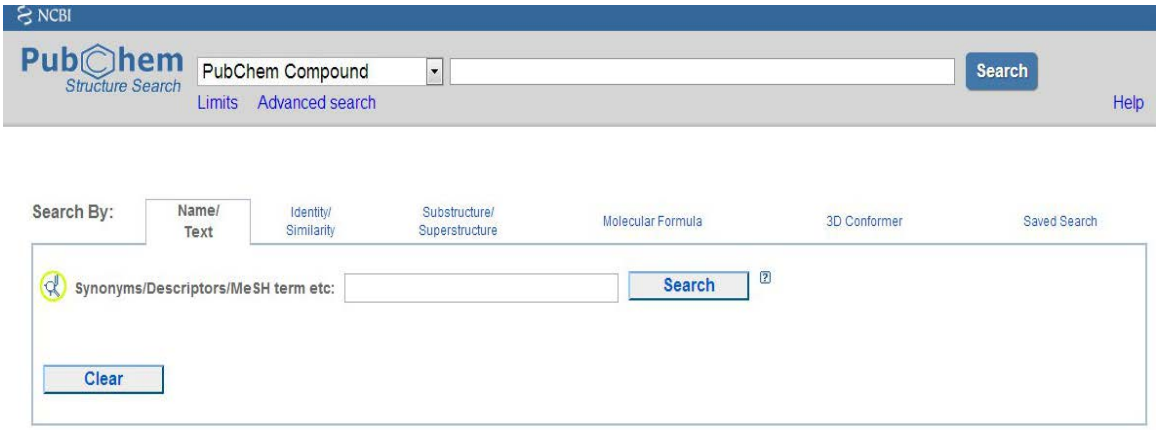


Figure 5.2 Screenshot of "PUBCHEM structure search".

5.2.2 SMILES Code Translation

Given a specific compound, the following steps are carried out to translate SMILES strings to an SDF file:

1 From the PUBCHEM website, enter the specific compound name to generate the SMILES code.

2 From the online code translator [58], enter the string in "Input Format".

3 Choose SDF, kekule and 2D, and click "Translate".

4 The SDF file is shown after selecting the "Result" option.

For a batch of compounds, a TXT file containing the SMILES strings should be created.

Figure 5.3 is the screenshot of "Online SMILES Translator", where "A" is the field to copy the string for the compound and "B" is the location for a batch of compounds.

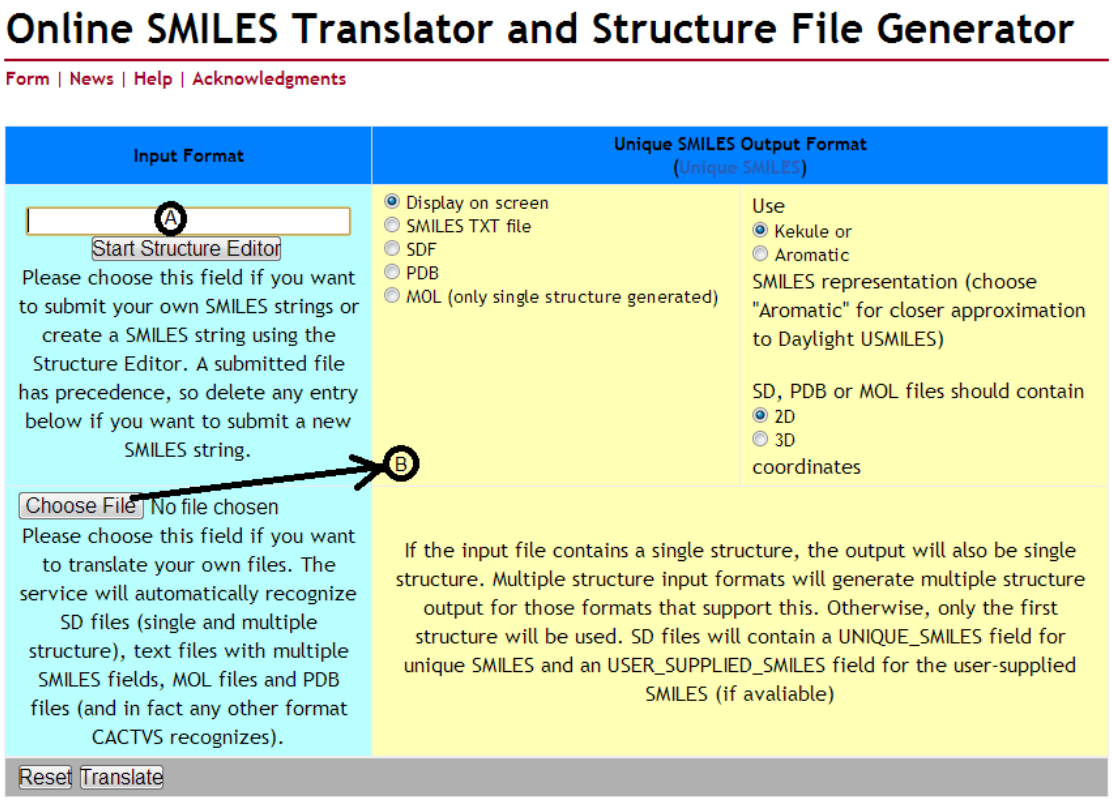


Figure 5.3 Screenshot of "online SMILES translator".

5.3 Molecular Descriptor

As discussed in Chapter 2, compound structures are difficult to interpret. For the QSAR analysis, it is necessary to identify a series of predictors that describe the molecule in an accurate and comprehensive way. 2D chemical structures for molecular descriptors are presented in this study. Such representations are simpler and, sometimes lead to better predictions than 3D designs [62-64]. Ceriums, Dragon, and Molconn-Z are common options for forming 2D-based descriptors. However, Mold^2 was selected in this work because of the low computing cost. In addition, compared to the other tools, Mold^2 is applicable to smaller datasets and provides similar information [65]. The software is also freely available to the public and can generate 777 descriptors.

The procedure followed to produce the predictor variables consists of these steps:

1 Double-click “Mold2.bat”, then enter SDF file in “Input SDF file” to obtain an output file and a report file.

2 The “descriptors.txt” is created and lists the descriptors in the output file, while “report.txt” contains an error report.

5.4 Random Selection of Training and Testing Sets

The sixty-one data are randomly divided into training and testing set in order to validate the QSAR created. The testing set consists of 20% of the data, i.e., 12 CPEs; the remaining patterns were selected for training. A reliable QSAR model depends on the quality of the data points in the training phase. It may be necessary to randomize the data several times to make sure that the examples provided are representative of the total population. A number of alternative techniques, such as the sphere exclusion method, are implemented in the literature [51, 66]. In this work, the randomization step was repeated several times.

This procedure was pursued in CHEMBECH:

1 The ACT, SDF and X files were uploaded at the “Modeling with Descriptor” under the “Dataset” Tab.

2 The “continuous” option and the “standardized structures” boxes were selected. At the “New Type” option, Mold2 was entered.

3 In “Random split”, “Use activity binning” was chosen and twenty percent was entered for the set size.

4 The name of the dataset was added.

5 The result was shown in “My bench” Tab. A zip file, which contained the results, was available in the Datasets section.

5.5 MATLAB®

MATLAB®, developed by MATHWORKS, is widely accepted in industry and academia. More than a million engineers and scientists use it for numerical computation, visualization and programming. It allows users to manipulate a matrix, analyze and plot functions and data, develop algorithms and create models. It can be further extended after interfacing with programs, including C, C++, Java and Fortran [67].

5.5.1 Removal of Redundant "0" and Test Set from Data

A program was written in Matlab to remove unnecessary variables. Columns of zeroes were deleted from the data.

5.5.2 Application of Replacement Method, Enhanced Replacement Method and Forward Stepwise Regression

Three Matlab functions, "rmt.m", "erm.m" and "stepwise.m", written by Mercader et al. [48], were applied to perform the RM, the ERM and the FSR, respectively. Eight features were extracted using these techniques.

5.5.3 Calculation of the Regression Coefficients

A Matlab function, "ls.m" written by Mercader et al. (2007) [48], estimated the statistical parameters of a model, including the regression coefficients for the eight descriptors. This function yields the correlation coefficient (R), standard deviation (S), F-statistic, average of the squared residuals, Akaike Information Criterion (AIC), a fit index (FIT) for the model and the errors in the coefficients [48].

CHAPTER 6

RESULTS AND VALIDATION OF QSAR MODELS

SMILES codes for the 61 CPEs are listed in Appendix A. Part of the descriptors are presented in Appendix B. The full database is provided in [65].

6.1 Selection Result of Training and Test Sets

The test set, randomly selected in CHEMBENCH, was determined by the lowest MAE (mean absolute error) value (9.625) and was composed of the following 12 compounds: enhancers 8, 11, 12, 17, 29, 30, 35, 37, 43, 45, 46 and 50. The remaining data formed the training set (see Appendix A).

6.2 Applicability Domain Plot

Figure 6.1 shows the applicability domain (AD) of the model. The dots and triangles represent the training and test sets, respectively. All points fall within three standard deviation units while two dots are outside of the AD (i.e., greater than the warning leverage h^*). The CPEs are within the response range; there is no response outlier. Two training samples have great influence on the model: compound 58 (leverage: 0.83069) and compound 60 (leverage: 0.556). The test set is within the AD. Therefore, its prediction is reliable.

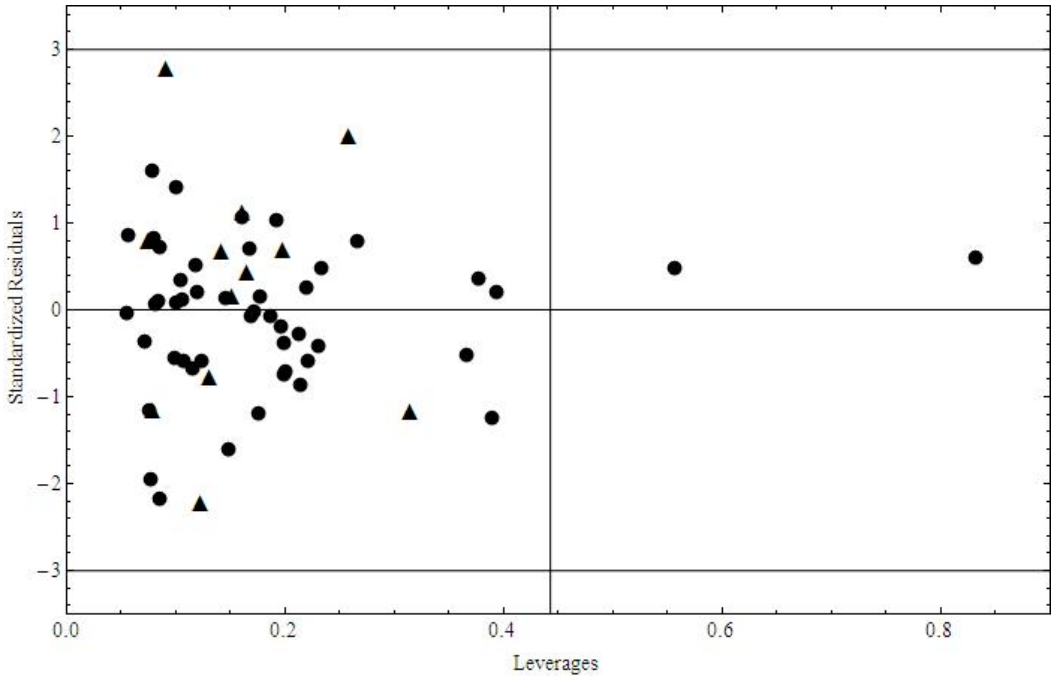


Figure 6.1 Applicability domain plot.

6.3 RM Result

Eight relevant descriptors are identified using the RM algorithm:

D712: number of group donor atoms for H-bonds (with N and O)

D223: average valence vertex connectivity order-5 Index

D498: Moran topological structure autocorrelation length-4 weighted by atomic Sanderson electronegativities

D139: topological distance count order-3

D361: ratio of convention bonds with total path counts

D563: lowest eigenvalue from Burden matrix weighted by polarizabilities order-8

D468: Geary topological structure autocorrelation length-6 weighted by atomic Sanderson electronegativities

D144: mean atomic van der Waals Carbon-scale

The standard deviation of the resulting model is 8.9870. The enhancement ratio (ER) equation can be written as:

$$\begin{aligned} \textit{Enhancement Ratio (ER)} = {} & \textit{-492.3785709 - 8.334892401} \times \textit{(number of group donor atoms for H-bonds (with N and O))} - \\ & \textit{971.1163824} \times \textit{(average valence vertex connectivity order-5 Index)} - \textit{178.7675311} \times \textit{(Moran topological structure autocorrelation length-4 weighted by atomic Sanderson electronegativities)} - \\ & \textit{0.529379308} \times \textit{(Topological distance count order-3)} - \textit{39.18517223} \times \textit{(ratio of convention bonds with total path counts)} - \\ & \textit{78.08649465} \times \textit{(Lowest eigenvalue from Burden matrix weighted by polarizabilities order-8)} + \textit{83.89527801} \times \\ & \textit{(Geary topological structure autocorrelation length-6 weighted by atomic Sanderson electronegativities)} + \textit{1110.128698} \times \textit{(mean atomic van der Waals Carbon-scale)} \end{aligned} \quad (6.1)$$

The q^2 value is 0.76.

6.4 ERM Result

Application of the ERM methodology yields the following descriptors:

D595: highest eigenvalue from Burden matrix weighted by polarizabilities order-8

D583: highest eigenvalue from Burden matrix weighted by electronegativities Sanderson-Scale order-4

D148: mean atomic electronegativity Sanderson-scaled on Carbon

D719: number of group CH2RX”

D487: Moran topological structure autocorrelation length-1 weighted by atomic van der Waals volumes

D252: structure centric index

D202: vertex connectivity order-2 index

D146: mean atomic electronegativities Pauling-scaled on Carbon

The standard deviation is 8.2596. As a result, the ER equation is given by:

$$\begin{aligned} &\textit{Enhancement Ratio (ER)} = 3475.644081 - 32.3125901 \times (\textit{Highest eigenvalue from Burden matrix weighted by polarizability order-8}) + 135.7830714 \times (\textit{Highest eigenvalue from Burden matrix weighted by electronegativities Sanderson-Scale order-4}) \\ &+ 10260.89116 \times (\textit{mean atomic electronegativity Sanderson-scaled on Carbon}) + 10.88119291 \times (\textit{number of group CH2RX}) + \\ &80.43358564 \times (\textit{Moran topological structure autocorrelation length-1 weighted by atomic van der Waals volumes}) + \\ &1.3463313 \times (\textit{structure centric index}) - 8.755968019 \times (\textit{vertex connectivity order-2 index}) - \\ &14528.47706 \times (\textit{mean atomic electronegativities Pauling-scaled on Carbon}) \end{aligned} \quad (6.2)$$

The q^2 value is 0.79.

6.5 FSR Result

The descriptors obtained after using the traditional FSR method are:

D252: structure centric index

D026: number of Oxygen

D563: lowest eigenvalue from Burden matrix weighted by polarizabilities order-8

D253: structure lopping centric group index

D492: Moran topological structure autocorrelation length-6 weighted by atomic van der Waals volumes

D621: number of group esters (aliphatic)

D494: Moran topological structure autocorrelation length-8 weighted by atomic van der Waals volumes

D503: Moran topological structure autocorrelation length-1 weighted by atomic polarizabilities

The standard deviation for the model is 9.4732. The ER equation becomes:

$$\begin{aligned} \textit{Enhancement Ratio (ER)} = {} & \textit{18.57872893-3.67957} \times \textit{(structure centric index)} + \textit{13.39025} \times \textit{(number of Oxygen)} - \textit{108.095} \times \\ & \textit{(Lowest eigenvalue from Burden matrix weighted by polarizabilities order-8)} + \textit{29.18622} \times \textit{(structure lopping centric group index)} - \textit{94.0105} \times \textit{(Moran topological structure autocorrelation length-6 weighted by atomic van der Waals volumes)} - \textit{24.3658} \times \textit{(number of group esters (aliphatic))} + \\ & \textit{25.98054} \times \textit{(Moran topological structure autocorrelation length-8 weighted by atomic van der Waals volumes)} - \textit{11.544} \times \textit{(Moran topological structure autocorrelation length-1 weighted by atomic polarizabilities)} \end{aligned} \quad (6.3)$$

The q^2 value is 0.73.

6.6 Validation of Test Set Result

6.6.1 Validation of the RM Result

Using the results from the RM algorithm, the following statistics are obtained for the testing set: $\bar{y}$ = 14.4267, SS_{tot} = 3563.0367, SS_{err} = 4288.7209, q^2 = -0.2037. Table 6.1 lists the test set and the corresponding descriptors.

6.6.2 Validation of the ERM Result

The following statistics are achieved for the test set using the ERM results: $\bar{y}$ = 14.4267, SS_{tot} = 3563.0367, SS_{err} = 1331.6290, q^2 = 0.6263. Table 6.2 lists the test set and the corresponding descriptors.

6.6.3 Validation of the FSR Result

Application of the Forward step wise method gives: $\bar{y}$ = 14.4267, SS_{tot} = 3563.0367, SS_{err} = 3796.6429, q^2 = -0.0656. Table 6.3 lists the test set data and the corresponding descriptors.

Table 6.1 Predicted ER Values for the Test Set. The RM Algorithm was Applied

	D712	D223	D498	D139	D361	D563	D468	D144	Experimental	Predicted
compound8	0	0.067	0.461	24	1.186	-0.025	0.947	0.589	25.78	36.23288
compound11	0	0.058	0.559	23	1.2	-0.025	0.797	0.592	12.67	18.18059
compound12	0	0.061	0.531	24	1.193	-0.025	0.816	0.591	36.44	20.50153
compound17	0	0.068	0.459	31	1.165	-0.025	0.786	0.587	1.96	17.00913
compound29	0	0.065	0.565	22	1.216	-0.023	0.915	0.591	8.6	18.84592
compound30	0	0.066	0.49	19	1.233	-0.023	0.899	0.593	12.8	33.0823
compound35	1	0.06	0.359	34	1.056	-0.016	1	0.589	5	56.47404
compound37	0	0.087	0.255	18	1.13	-0.025	0.948	0.578	60.1	46.87978
compound43	0	0.073	0.282	13	1	0.601	0.9	0.575	2.17	7.150154
compound45	0	0.069	0.403	12	1	0.599	0.839	0.576	1.03	-13.9182
compound46	0	0.076	0.334	14	1	0.599	0.887	0.575	2.4	-6.52295
compound50	0	0.073	0.285	18	1	0.599	0.915	0.574	4.17	4.271429

D stands for descriptor; the number represents the number of the descriptor. For example, D712 is the descriptor 712, “number of group donor atoms for H-bonds (with N and O)”.

Table 6.2 Predicted ER Values for the Test Set. The ERM Algorithm was Applied

	D595	D583	D148	D719	D487	D252	D202	D146	Experimental	Predicted
compound8	2.126	4.322	0.962	3	0.221	20	9.04	0.953	25.78	17.33268
compound11	1.922	4.232	0.968	3	0.256	16	7.98	0.957	12.67	21.86659
compound12	1.96	4.26	0.966	3	0.243	17	8.333	0.955	36.44	30.18565
compound17	2.229	4.323	0.959	3	0.202	20	9.748	0.95	1.96	19.21556
compound29	1.74	4.138	0.96	1	0.152	14	7.166	0.951	8.6	-5.62518
compound30	1.617	4.135	0.962	1	0.161	14	6.813	0.952	12.8	7.749986
compound35	2.28	4.268	0.946	1	0.149	16	10.505	0.938	5	13.01073
compound37	1.642	4.138	0.947	2	0.16	15	6.992	0.94	60.1	38.35765
compound43	1.077	3.954	0.955	1	0.239	11	5.694	0.948	2.17	-1.05753
compound45	1.064	3.876	0.959	2	0.341	9	5.182	0.951	1.03	7.105402
compound46	1.224	3.876	0.955	2	0.296	11	5.889	0.948	2.4	-2.64006
compound50	1.517	4.028	0.951	1	0.247	14	7.237	0.944	4.17	3.015228

Table 6.3 Predicted ER Values for the Test Set. The FSR Algorithm was Applied

	D252	D026	D563	D253	D492	D621	D494	D503	Experimental	Predicted
compound8	20	3	-0.025	3.821	0.348	1	0.391	1.636	25.78	33.57198
compound11	16	3	-0.025	3.347	0.554	1	0.489	1.803	12.67	15.70806
compound12	17	3	-0.025	3.446	0.519	1	0.451	1.744	36.44	17.90213
compound17	20	3	-0.025	3.616	0.505	1	0.345	1.54	1.96	12.74228
compound29	14	2	-0.023	3.13	0.385	0	0.455	1.948	8.6	40.82374
compound30	14	2	-0.023	3.243	0.419	0	0.503	2.019	12.8	41.35286
compound35	16	1	-0.016	2.972	0.188	0	0.21	0.91	5	38.84379
compound37	15	1	-0.025	3.354	0.228	0	0.27	1.592	60.1	44.57074
compound43	11	2	0.601	2.865	0.319	0	0.41	0.324	2.17	0.459715
compound45	9	2	0.599	2.562	0.444	0	0.828	0.377	1.03	-2.31167
compound46	11	2	0.599	2.865	0.346	0	0.445	0.324	2.4	-0.95306
compound50	14	2	0.599	3.243	0.28	0	0.336	0.269	4.17	3.04837

CHAPTER 7

DISCUSSION

7.1 Model Analysis of the Replacement Methods

The RM, ERM and FSR methods extracted different descriptors from the database and resulted in distinct QSAR models.

7.1.1 The Replacement Method Algorithm

The q^2 values for the RM-generated model are: 0.76 (for training data), -0.20 (for test vector) and 0.55 (for all 61 CPEs). Based on Eqs 4.8 – 4.11, the following results are calculated for the test data (Figure 7.1): $R^2 = 0.28$, $R_0{}^2 = 0.24$, $\frac{R^2 - R_0{}^2}{R^2} = 0.17$ and $k = 0.59$. The prediction fails to meet the requirements specified in Section 4.7. The results for the 61 CPEs are shown in Figure 7.2.

7.1.2 The Enhanced Replacement Method Algorithm

The q^2 values for the ERM-based model are: 0.79 (for training data), 0.63 (for test set) and 0.76 (for the full dataset). Additional statistics are (Figure 7.3): $R^2 = 0.64$, $R_0{}^2 = 0.64$, $\frac{R^2 - R_0{}^2}{R^2} = 4.36 \times 10^{-3}$ and $k = 1.11$. The prediction meets the conditions outlined in Section 4.7. Figure 7.4 shows a correlation plot with the 61 CPEs.

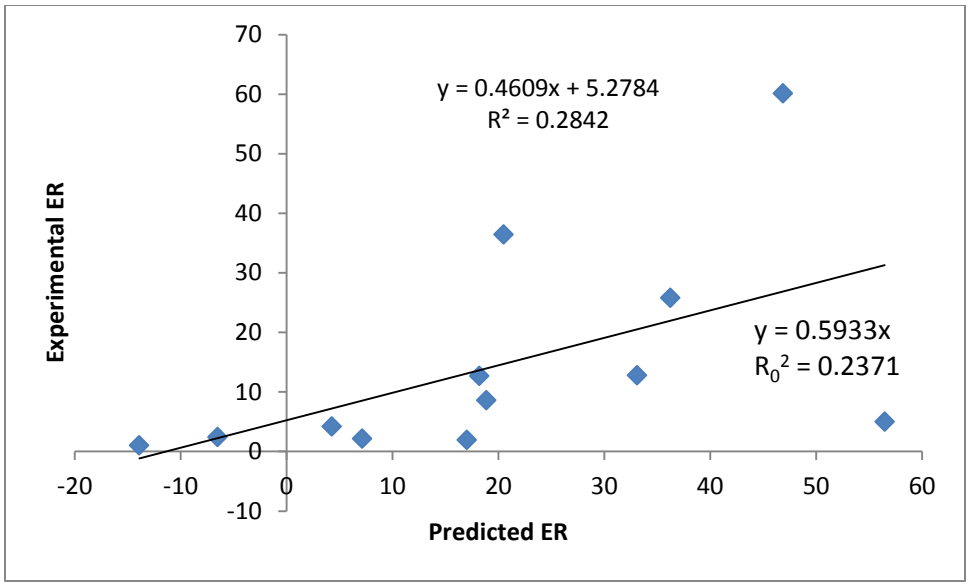


Figure 7.1 Correlation between the experimental and predicted flux ratios for the test set. The RM algorithm was applied.

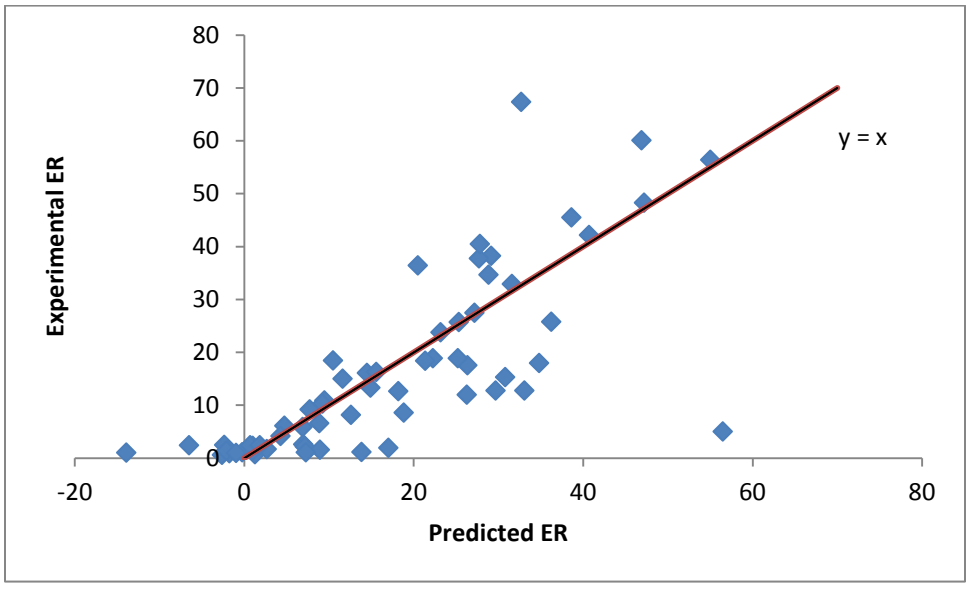


Figure 7.2 Correlation between the experimental and predicted flux ratios for the full dataset. The RM algorithm was applied.

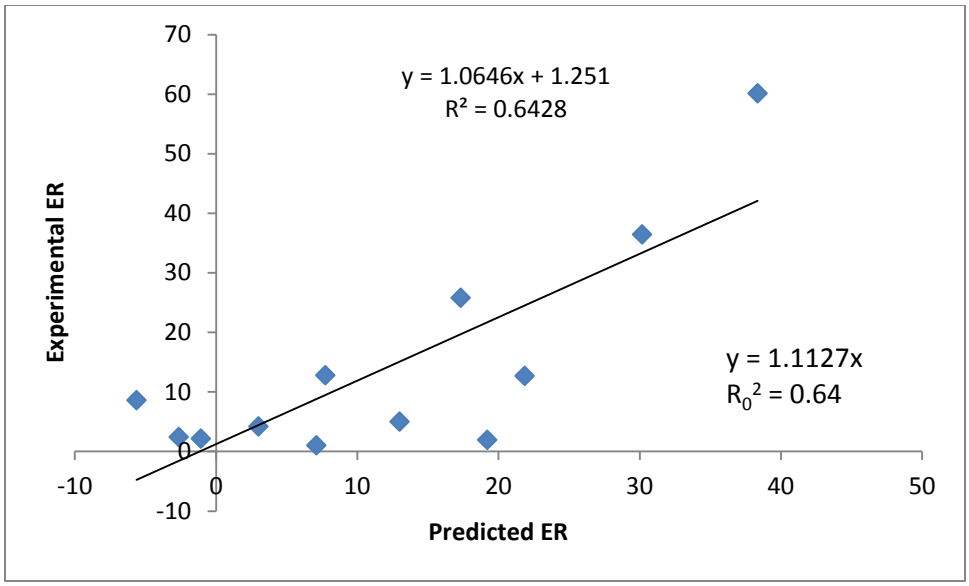


Figure 7.3 Correlation between the experimental and predicted flux ratios for the test set. The ERM algorithm was applied.

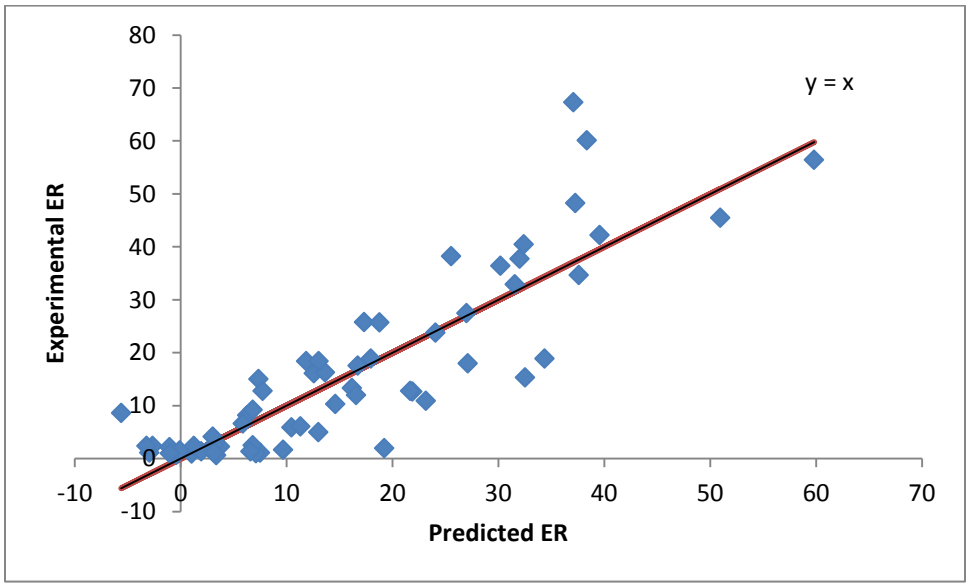


Figure 7.4 Correlation between the experimental and predicted flux ratios for the full dataset. The ERM algorithm was applied.

7.1.3 The Forward Stepwise Regression Algorithm

The q^2 values calculated when the FSR algorithm is applied are: 0.73 (for training set), -0.066 (for test set) and 0.56 (for the 61 CPEs). Other test statistics are (Figure 7.5): R^2 = 0.29, $R_0{}^2$ = 0.27, $\frac{R^2-R_0{}^2}{R^2}$ = 0.064 and k = 0.63. The prediction is poor. The correlation plot is shown for the 61 CPEs (Figure 7.6).

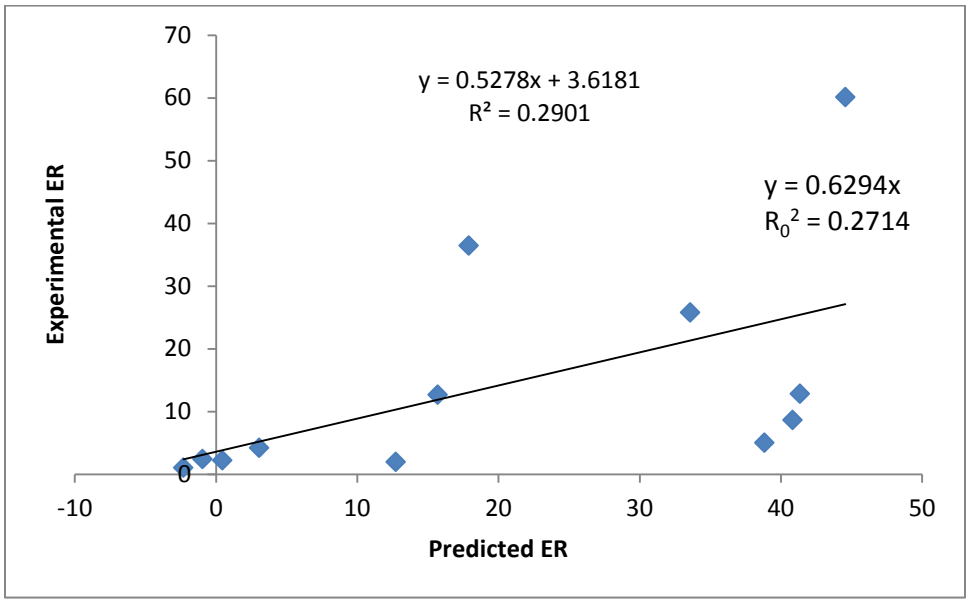


Figure 7.5 Correlation between the experimental and predicted flux ratios for the test set. The FSR algorithm was applied.

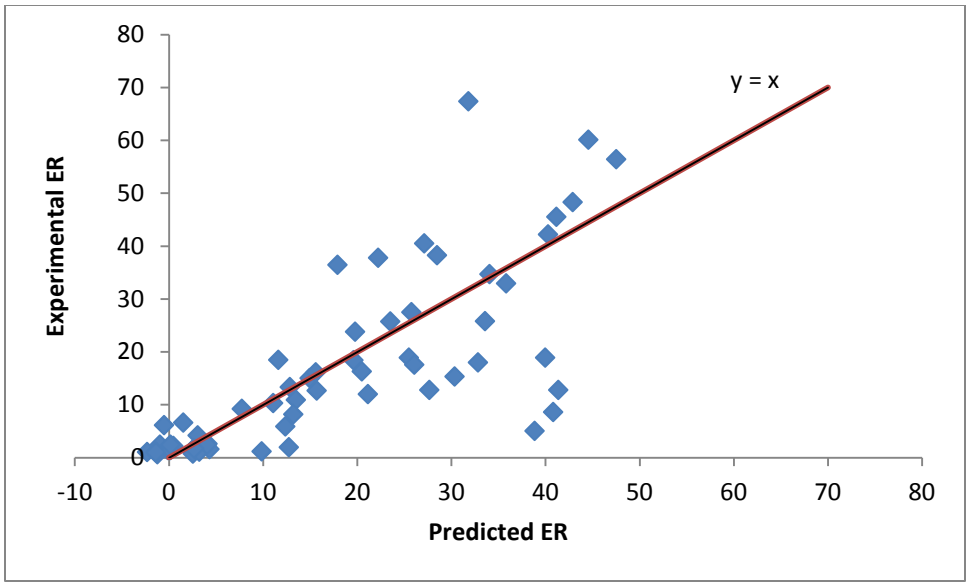


Figure 7.6 Correlation between the experimental and predicted flux ratios for the full dataset. The FSR algorithm was applied.

7.2 Evaluation of Models

The RM and FSR approaches yield negative q^2 values. Based on Equation 4.4, the mean of the data outperforms the QSAR models. Therefore, the predictor variables selected by both methods cannot predict ER when a linear trend is assumed. The ERM algorithm produces the best result, in accordance with previous observations. Simon and Abdelmalek (2012) analyzed the same 61 non-polar CPEs by extracting eight different descriptors using MMP+. Multiple linear regression models were also derived. After applying the ERM technique, the q^2 values were 0.683 and 0.74 for the full dataset and training data, respectively. The q^2 values for the 61 CPEs were 0.683 and 0.671 with the RM and FSR algorithms, respectively. Those findings show that the ERM performed more effectively when the test results were considered and could serve as a viable

modeling tool for the selection for molecular structures [14]. This observation is consistent with the three cases studied in this work. In addition, the ERM results in Section 7.1.2 outperformed those recorded in [14]. The main reason is the application of the public domain software Mold^2 which can create more descriptors at a low computational cost while conveying sufficient structural information. In summary, the combination of ERM and Mold^2 predictors is a promising approach for designing CPEs.

7.3 Interpretation of Eight Enhanced Replacement Method Descriptors

D595 and D583 belong to the burden eigenvalues, a subclass of eigenvalues-based descriptors [30]. Polarizability, applied in D595, is a fundamental property and is defined as the ability of particles to be polarized. It controls the dynamical response of a bound system to an external field and can shed light on a molecule's internal structure [69]. As a chemical property, electronegativity (in D583) is related to the tendency of an atom, or a functional group, to pull electrons towards itself. Sanderson electronegativity's model can compute parameters, such as bond energies, molecular geometry and NMR spin-spin constants [70].

D148 and D146 belong to electronegativity scales, a subclass of quantum-chemical descriptors. The Sanderson scales χ_s are based on covalent radii. Sanderson group electronegativity (ESG) is calculated by the equation [30]:

$$ESG_i = (\chi_{S,1}, \chi_{S,2}, \dots, \chi_{S,m})^{1/m} \quad (7.1)$$

where χ_s is the Sanderson electronegativity and m the number of atoms of the ith molecular group.

Pauling defined electronegativity as “the power of an atom in a molecule to attract electrons to itself”. Pauling-scale χ_{PA} is estimated using the ionization potential (IP) and the electronic affinity of valence state (EA) [30]:

$$\chi_{PA} = 0.303 \times \frac{IP + EA}{2} \tag{7.2}$$

D719 belongs to the subclass of structural features while D487 is a Moran topological structure autocorrelation descriptor, which is expressed as $I_d\,(w)$ with d=1,2,..,10). The function $I_d\,(w)$ is defined as [30, 71]:

$$I_d(w) = \frac{\frac{1}{\Delta}\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_{ij}(w_i - \overline{w})(w_j - \overline{w})}{\frac{1}{N}\sum_{i=1}^{N}(w_i - \overline{w})^2} \tag{7.3}$$

where the weight *w* is any atomic property and $\overline{w}$ is its average value; N is the number of atoms; *d* is the considered topological distance and δ_{ij} is the Kronecker delta. The sum of the Kronecker deltas is given by Δ.

D252 is a centric index and is calculated using a partition of graph vertices based on its position relative to the center. Details can be found in [30]. D202 is a connectivity index that can be calculated from the vertex degree δ of the atoms in a molecular graph where all the hydrogen atoms are excluded. The vertex degree represents the count of its σ electrons in the plot and can be estimated by the following equation:

$$\delta_i = \sum_{j=1}^{A} a_{ij} = {}^1f_i = [\mathbf{A}^2]_{ii} \tag{7.5}$$

where δ_i stands for the vertex degree of the *i*th atom, a_{ij} marks the *i*th row of the adjacency matrix **A** and ${}^1f_i$ is the *i* th row of the distance matrix **D** [30].

7.4 Correlation of the Enhanced Replacement Method-based Descriptors

A correlation study was conducted to detect possible relationships among ERM-based properties (Table 7.1). If the correlation coefficient is greater than 0.8 (or less than -0.8), the two variables are strongly correlated and the label “s” is used. Values between 0.3 and 0.8 (or -0.8 and -0.3) describe weak correlation and the symbol “w” is assigned. If the value is smaller than 0.3 (or greater than -0.3), there is no correlation, a case that is described by “n." Such analyses help to identify which variables can be omitted in future modeling efforts.

Table 7.1 Correlation between the Eight Descriptors

	D595	**D583**	**D148**	**D719**	**D487**	**D252**	**D202**	**D146**
D595		0.9067	0.1961	0.5775	-0.2103	0.3042	0.9329	0.1465
D583	s		0.3578	0.6497	-0.2101	0.4372	0.8954	0.3183
D148	n	w		0.6827	0.6708	0.5722	0.3416	0.9968
D719	w	w	w		0.4301	0.6652	0.6560	0.6712
D487	n	n	w	w		0.4798	-0.0436	0.6920
D252	w	w	w	w	w		0.4602	0.5752
D202	s	s	w	w	n	w		0.2987
D146	n	w	s	w	w	w	n	

7.5 Application of the Enhanced Replacement Method Model

Although The ERM algorithm is effective at selecting the best CPEs, compounds to be tested should be within the AD of the model (Figure 6.1). The leverage is expected to be smaller than a warning value and the ER response not to exceed the three-standard deviation threshold. This is especially important for the design of new CPEs for a certain drug.

The sign of the regression coefficients of the QSAR is instructive for designing CPEs. To increase the transdermal flux, variables with positive coefficients should have

large values. An opposite effect is observed when the coefficient is negative. For the ERM-based selections, D595, D202 and D146 need to be as small as possible while the other five variables should take on large values. The magnitudes of coefficients can help evaluate the contribution of each feature: the larger the coefficient absolute value, the more dominant the variable. The influence level is expressed as:

$$D252 < D202 < D719 < D595 < D487 < D583 < D146 < D148 \quad (7.1)$$

Reversible decoding (or inverse QSAR) is a process for re-constructing the structure, or fragment, by using calculated molecular descriptors. Without performing experiments, the response of a new molecule can be predicted using Equation 6.2. The influence level is an important consideration in screening for the best CPEs. The difficult task in reversible decoding is the ability to transform these descriptors into structures, a step necessary for guiding the design or synthesis of new molecules. In fact, this could be a limiting factor in choosing software that produces structural descriptors. Reversible decoding is an important research area that deserves increased attention [72].

CHAPTER 8

CONCLUSIONS

In this study, three selection algorithms, RM, ERM and FSR, were applied to a database composed of 61 nonpolar CPEs. Eight features, which affected the enhancement ratio the most, were chosen from 777 molecular descriptors created by Mold[2]. QSAR models were developed using hydrocortisone as a control. Although the RM and FSR method yielded high q^2 values during training (0.73 and 0.76, respectively), the test set results were unsatisfactory.

The ERM produced the best QSAR model, with high predictive power. The q^2 values were: 0.79 (training), 0.63 (testing) and 0.76 (full database). These statistics suggest that the approach can be used to explain the relationship between ER and molecular descriptors. Another potential application is in the area of reverse decoding to promote the design of new CPEs. For the technique to be effective, the compound needs to belong to the model's AD.

This work focused on a method to fabricate CPEs that would improve the permeation of hydrocortisone through the skin. A strategy that is also applicable to other drugs. Research on CPEs to aid the transdermal transport of insulin is continuing in our laboratory. The response variable is the permeability instead of the ER [73].

APPENDIX A

SMILES CODES FOR THE SIXTY-ONE CPES

SMILES codes representing the molecular structures of the sixty-one CPEs are listed.

compound1	C1CCCC(=O)N(CCCCCCCCCCCC)C1
compound2	C1CCC(=O)N1CC(=O)OCCCCCCCC
compound3	C1CCC(=O)N1CC(=O)OCCCCCCCCC
compound4	C1CCC(=O)N1CC(=O)OCCCCCCCCCC
compound5	C1CCC(=O)N1CC(=O)OCCCCCCCCCCC
compound6	C1CCC(=O)N1CC(=O)OCCCCCCCCCCCC
compound7	C1CCC(=O)N1CC(=O)OCCCCCCCCCCCCC
compound8	C1CCC(=O)N1CC(=O)OCCCCCCCCCCCCCC
compound9	C1CCC(=O)N(CC(=O)OCCCCCCCC)C1
compound10	C1CCC(=O)N(CC(=O)OCCCCCCCCC)C1
compound11	C1CCC(=O)N(CC(=O)OCCCCCCCCCC)C1
compound12	C1CCC(=O)N(CC(=O)OCCCCCCCCCCC)C1
compound13	C1CCC(=O)N(CC(=O)OCCCCCCCCCCCC)C1
compound14	C1CCC(=O)N(CC(=O)OCCCCCCCCCCCCC)C1
compound15	C1CCC(=O)N(CC(=O)OCCCCCCCCCCCCCC)C1
compound16	C1CCCC(=O)N(CC(=O)OCCCCCCCCCCCC)C1
compound17	C1CCCC(=O)N(CC(=O)OCCCCCCCCCCCCCC)C1
compound18	C1CCC(N2CCCC2=O)C(=O)N(CCCCCCCCCC)C1
compound19	C1CCC(N2CCCC2=O)C(=O)N(CCCCCCCCCCC)C1
compound20	C1CCC(N2CCCC2=O)C(=O)N(CCCCCCCCCCCCC)C1
compound21	C1CCC(N2CCCC2=O)C(=O)N(CCCCCCCCCCCCCC)C1
compound22	C1CCC(N2CCCC2=O)C(=O)N(CC(=O)OCCCCCC)C1

compound23	C1CCC(N2CCCC2=O)C(=O)N(CC(=O)OCCCCCCCCC)C1
compound24	C1CCC(N2CCCC2=O)C(=O)N(CC(=O)OCCCCCCCCCCC)C1
compound25	C1CCC(N2CCCC2=O)C(=O)N(CC(=O)OCCCCCCCCCCCCC)C1
compound26	C1CCC(N2CCCC2=O)C(=O)N(CC(=O)OCCCCCCCCCCCCCCC)C1
compound27	C1CCN(C(CCCCCCCCCC)=O)CC1
compound28	C1CN(C(CCCCCCCCCC)=O)CC1
compound29	C1CC(=O)N(C(CCCCCCCCCC)=O)CC1
compound30	C1C(=O)N(C(CCCCCCCCCC)=O)CC1
compound31	C1CCC(=O)C(CCCCCCCCCC)C1
compound32	C1C(=O)C(CCCCCCCCCC)CC1
compound33	C1N(C(CCCCCCCCCC)=O)CCSC1
compound34	CC(N(CCCCCCCCCCCC)CCCCCCCCCCCC)=O
compound35	C1C(CC(C(=O)NCCCCCCCCCCCC)(C2)C3)CC2CC13
compound36	C1N(C(CCCCCCCCCC)=O)CCOC1
compound37	C1C(=O)N(CCCCCCCCCCCCC)CC1
compound38	C1CCC(=O)N(CCCCCCCCCCCCC)C1
compound39	O1C(CCCCCCC)OCC1
compound40	O1C(CCCCCCCCC)OCC1
compound41	O1C(CCCCCCCCCCC)OCC1
compound42	O1C(CCCCCCC)OC(C)C1
compound43	O1C(CCCCCCCCC)OC(C)C1
compound44	O1C(CCCCCCCCCCC)OC(C)C1
compound45	O1C(C)(CCCCCCC)OCC1
compound46	O1C(C)(CCCCCCCCC)OCC1
compound47	O1C(C)(CCCCCCCCCCC)OCC1
compound48	O1C(C)(CCCCCCC)OC(C)C1
compound49	O1C(C)(CCCCCCCCC)OC(C)C1
compound50	O1C(C)(CCCCCCCCCCC)OC(C)C1

compound51	O=C(NCCCCCCCCCCCC)N
compound52	O=C(NCCCCCCCCCCCC)NC
compound53	S=C(NCCCCCCCCCCCC)NC
compound54	O=C(NCCCCCCCCCCCC)NCCCCCCCCCCCC
compound55	S=C(NCCCCCCCCCCCC)NCCCCCCCCCCCC
compound56	O=C(NCCCCCCCCCCCC)Nc1=cc=cc=c1
compound57	S=C(NCCCCCCCCCCCC)Nc1=cc=cc=c1
compound58	C(C(=CCO)C)CC=C(C)C
compound59	N(CCCCCCCCCCCC)CCCCCCCCCCCC
compound60	NCCCCCCCCCCCC
compound61	NCCCCCCCCCCCCCCCCCC

APPENDIX B

A DATABASE OF 777 DESCRIPTORS

The molecular structures of the sixty-one CPEs are transformed into 777 molecular descriptors. Part of the molecular descriptor names with descriptor codes is listed.

For the full database, consult Mold2_SoftwareIntroduction_12012008.doc at http://www.fda.gov/downloads/ScienceResearch/BioinformaticsTools/Mold2/UCM161940.pdf [68].

Description Code	Descriptor Name
D001	Number of 6-membered aromatic rings only carbon atoms
D002	Number of 03-membered rings
D003	Number of 04-membered rings
D004	Number of 05-membered rings
D005	Number of 06-membered rings
D006	Number of 07-membered rings
D007	Number of 08-membered rings
D008	Number of 09-membered rings
D009	Number of 10-membered rings
D010	Number of 11-membered rings
D011	Number of 12-membered rings
D012	Number of multiple bonds
D013	Number of circuits structure
D014	Number of rotatable bonds
D015	Rotatable bond fraction
D016	Number of double bonds
D017	Number of aromatic bonds
D018	Sum of conventional bond orders (h-depleted)
D019	Number of hydrogen
D020	Number of helium

REFERENCE

[1] Prausnitz1,M.R.; Langer R. Transdermal drug delivery. *Nature Biotechnology.* 26: 1261-1268 (2008).

[2] Miller, M.A.; Pisani, E. The cost of unsafe injections. *Bull. World Helath Organ.* 77: 808-811 (1999).

[3] Scheindlin, S. Transdermal drug delivery: past, present, future. *Mol Interv.* 4(6):308-12 (December 2004).

[4] Tanner, T; Marks, R. Delivering drugs by the transdermal route: review and comment. *Skin Res Technol.* 14:249-60 (2008).

[5] Durand, C.; Alhammad, A.; Willett, K.C. Practical considerations for optimal transdermal drug delivery. *American Journal of Health-System Pharmacy.* 69(2): 116-124 (January 15, 2012).

[6] Segal, M. Patches, pumps and timed release: new ways to deliver drugs. *FDA Consumer*. 25(8) (October 1991).

[7] FDA approves scopolamine patch to prevent peri-operative nausea. *Food and Drug Administration.* November 10, 1997. Retrieved on February 22, 2013, from http://scienceblog.com/community/older/archives/M/1/fda0443.htm.

[8] Rose, J. E.; Jarvik, M. E.; Rose, K. D. Transdermal administration of nicotine. *Drug and alcohol dependence.* 13 (3): 209–213 (1984).

[9] Rose, J. E.; Herskovic, J. E.; Trilling, Y.; Jarvik, M. E. Transdermal nicotine reduces cigarette craving and nicotine preference. *Clinical pharmacology and therapeutics.* 38 (4): 450–456 (1985).

[10] Etscorn; F.T. Transcutaneous application of nicotine. US Patent 4597961.

[11] Sadrieh, N. Challenges in the development of transdermal drug delivery systems. Retrieved on March 22, 2013, from http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AdvisoryCommitteeforPharmaceuticalScienceandClinicalPharmacology/UCM178991.pdf.

[12] Chandrashekar, N.S.; Shobha Rani, R.H. Physicochemical and pharmacokinetic parameters in drug selection and loading for transdermal drug delivery. *Indian J Pharm Sci.* 70(1): 94–96 (January-February 2008).

[13] Prausnitz1, M.R.; Mitragotri, S.; Langer, R. Current status and future potential of transdermal drug delivery. *Nat. Rev. Drug Discov.* 3: 115-124 (2004).

[14] Simon, L; Abdelmalek B. Design of skin penetration enhancers using replacement methods for the selection of the molecular descriptors. *Pharmaceutics*. 4: 343-353 (2012).

[15] Peck, K.D.; Ghanem, A.H.; Higuchi, W. I. Hindered diffusion of polar molecules through and effective pore radii estimates of intact and ethanol treated human epidermal membrane. *Pharm. Res.* 11: 1306-1314 (1994).

[16] Mitragotri, S. Modeling skin permeability to hydrophilic and hydrophobic solutes based on four permeation pathways. *J. Control. Release.* 86: 69-92 (2003).

[17] Sinha, V. R.; Kaur, M.P. Permeation enhancers for transdermal drug delivery. *Drug Development and Industrial Pharmacy.* 26(11): 1131–1140 (2000).

[18] Kumar, B. Quantitative structure activity relationship (QSAR) modeling of 2-X-5,8-dimethoxy-1,4-naphthoquinones against L1210 cells. *International Journal of Pharmacy & Pharmaceutical Sciences.* 4: 445-448 (October 2012 Supplement).

[19] Hiren, P.; Berge, W.t.; Cronin, M.T.D. Quatntative structure-acitivity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere*. 48(6): 603 (August 2002).

[20] Kalani, K.; Yadav, D.; Khan, F.; Srivastava, S.; Suri. N. Pharmacophore, QSAR, and ADME based semisynthesis and in vitro evaluation of ursolic acid analogs for anticancer activity. *Journal of Molecular Modeling.* 18(7): 3389-3413 (July 2012).

[21] Scott, E. Cortisol and Stress: How to Stay Healthy. (Updated September 22, 2011) Retrieved on March 22, 2013, from http://stress.about.com/od/stresshealth/a/cortisol.htm.

[22] Hydrocortisone – compound summary. Retrieved on March 22, 2013, from http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=5754.

[23] Hoehn, K; Marieb, E.N. *Human Anatomy & Physiology.* Benjamin Cummings, San Francisco, California (2010).

[24] Ghafourian, T.; Zandastrar, P.; Hamishekar, H.; Nokhodchi, A. The effect of penetration enhancers on drug delivery through skin: a QSAR study. *Journal of controlled release.* 99: 113-125 (2004).

[25] El-Kattan, A.F.; Asbill, C.S.; Michniak, B.B. The effect of terpene enhancer lipophilicity on the percutaneous permeation of hydrocortisone formulated in HPMC gel systems. *International Journal of Pharmaceutics.* 198: 179–189 (2000).

[26] "DMSO" From American Cancer Society. Retrieved on March 22, 2013, from http://www.cancer.org/treatment/treatmentsandsideeffects/complementaryandalternativemedicine/pharmacologicalandbiologicaltreatment/dmso.

[27] Cui, F. (2003). *Pharmaceutics* (5^{th} edition.). People's Medical Publishing House, Beijing, China (2003).

[28] Janůšová, B.; Školová, B.; Tükörová, B.; Wojnarová, L.; Šimůnek, T.; Mladěnka, P.; Filipský, T.; Říha, M.; Roh, J.; Palát, K.; Hrabálek, A.; Vávrová, K. Amino acid derivatives as transdermal permeation enhancers. *Journal of Controlled Release*. 165: 91–100 (2013).

[29] Williams, A.C.; Barry, B.W. Penetration enhancers. *Advanced Drug Delivery Reviews.* 56: 603- 618 (2004).

[30] Todeschini, B.; Consonni, V. *Handbook of Molecular Descriptors.* Wiley-VCH, Weinheim, Germany (2000).

[31] Benson, H.A.E. Transdermal drug delivery: penetration enhancement techniques. *Current Drug Delivery*., 2(1): 23-33 (March 2005).

[32] Put, R.; Perrin, C.; Questier, F.; Coomans, D.; Massart, D.L.; Vander Heyden, Y. Classification and regression tree analysis for molecular descriptor selection and retention prediction in chromatographic quantitative structure–retention relationship studies. *Journal of Chromatography A*. 988: 261–276 (2003).

[33] Randic, M. Molecular bonding profiles. *Journal of Mathematical Chemistry*. 19: 375-392 (1996).

[34] Karande, P.; Jain, A.; Ergun, K.; Kispersky, V.; Mitragotri, S.; Israelachvili, J.N. Design principles of chemical penetration enhancers for transdermal drug delivery. *Proceedings of the National Academy of Sciences of the United States of America.* 102(13): 4688-4693 (March 29, 2005).

[35] Frantz, S.W. Instrumentation and methodology for in vitro skin diffusion cells in methodology for skin absorption. *Methods for Skin Absorption* (Kemppainen, B.W.; Reifenrath, W.G. Edition), CRC Press, Boca Raton, Florida (1990).

[36] Tojo, K. Design and calibration of in vitro permeation apparatus. *Transdermal Controlled Systemic Medications* (Chien, Y.W. Edition), Marcel Dekker, New York (1987).

[37] Barry, B.W. Methods for studying percutaneous absorption. *Dermatological Formulations: Percutaneous absorption*, Marcel Dekker, New York (1983).

[38] Patel, D.; Patel, N.; Parmar, M.; Kaur, N. Transdermal drug delivery system: review. *International Journal of biopharmaceutical and toxicological research*. 1(1) (May 2011).

[39] SUPAC-SS, FDA (1997). Guidance for industry. Nonsterile semisolid dosage forms. SUPAC-SS, CMC7.

[40] Higuchi, T. (1960). Physical chemical analysis of percutaneous absorption process from creams and ointments. *Journal of the Society of Cosmetic Chemists*. 11: 85–97.

[41] Higuchi, T. (1961). Rate of release of medicaments from ointment bases containing drugs in suspension. *Journal of Pharmaceutical Sciences.* 50: 874–875.

[42] Lucero, M.J.; Claro, C.; Casas, M.; Jiménez-Castellanos, M.R. Drug diffusion from disperse systems with a hydrophobically modified polysaccharide: Enhancer vs Franz cells. *Carbohydrate Polymers*. 92: 149-156 (2013).

[43] Patel, H.; Berge, W.T.; Cronin, M.T.D. Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere*. 48: 603–613 (2002).

[44] Mercader, A.G; Pomilio, A.B. (2010). QSAR study of flavonoids and bioflavonoid as influenza H1N1 virus neuraminidase inhibitors. *European J. Med. Chem.* 45: 1724-1730.

[45] Jang, C.; Youn, B.D.; Wang, P.F.; Han, B.; Ham, S.J. Forward-stepwise regression analysis for fine leak batch testing of wafer-level hermetic MEMS packages. *Microelectronics Reliability.* 50: 507–513 (2010).

[46] Mercader, A.G.; Duchowicz, P.R.; Fernández, F.M.; Castro, E.A. Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories. *Chemometrics and Intelligent Laboratory Systems.* 92: 138–144 (2008).

[47] Mercader, A.G.; Duchowicz, P.R.; Fernández, F.M.; Castro, E.A. Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. *J. Chem. Inf. Model.* 50: 1542-1548 (2010).

[48] Mercader, A.G. QSAR/QSPR Search Algorithms Toolbox. (April 14, 2008. Updated April 15, 2008) Retrieved on April 06, 2013, from http://www.mathworks.com/matlabcentral/fileexchange/19578-qsarqspr-search-algorithms-toolbox.

[49] About forward stepwise and forward selection. Retrieved on March 13, 2013, from http://www.biomedware.com/files/documentation/spacestat/Statistics/Regression/About_Foward_Stepwise_regression.htm.

[50] Hocking, R.R. The analysis and selection of variables in linear regression. *Biometrics.* 32: 1-49 (1976).

[51] Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Journal of Computer-Aided Molecular Design*. 16: 357 - 369 (2002).

[52] Eriksson, L.; Jaworska, J.; Worth, A.P.; Cronln, M.T.D; McDowell, R.M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspective.* 111(10): 1361-1375 (August 2003).

[53] Netzeva, T.I.; Worth, A.P.; Aldenberg, T.; Benigni, R.; Cronin, M.T.D.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G.Y.; Perkins, R.; Roberts, D.W.; Schultz, T.W.; Stanton, D.T.; vande Sandt, J.J.M.; Tong, W.; Veith, G.; Yang, C. Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern. Lab.* 33: 155–173 (2005).

[54] Khosrokhavar, R., Ghasemi, J.B.;Shiri, F. 2D quantitative structure-property relationship study of mycotoxins by multiple linear regression and support vector machine. *Int. J. Mol. Sci.* 11: 3052-3068 (2010).

[55] Iyer, M.; Zheng, T.; Hopfinger, A.J.; Tseng, Y.J. QSAR analyses of skin penetration enhancers. *J. Chem. Inf. Model.* 47 (3): 1130-1149 (2007).

[56] Dehmer, M.; Varmuza. K.; Bonchev, D. Statistical modeling of molecular descriptors in QSAR/QSPR. Wiley-Blackwell, Weinheim, Germany (2012).

[57] PUBCHEM website Access on April 13, 2013, from http://pubchem.ncbi.nlm.nih.gov/search.

[58] Online SMILES code translator website Access on April 13, 2013, from http://cactus.nci.nih.gov/translate.

[59] CHEMBENCH website Access on April 13, 2013, from http://chembench.mml.unc.edu.

[60] Toropov, A.A.; Benfenati, E. SMILES as an alternative to the graph in QSAR modelling of bee toxicity. *Computational Biology & Chemistry.* 31(1): 57-60. (February 2007).

[61] SMILES Tutorial: What is SMILES? Retrieved on March 24, 2013, from http://www.epa.gov/med/Prods_Pubs/smiles.htm.

[62] McGregor, M.J.; Pallai, P.V. Clustering large database of compounds: using the MDL “keys” as structural descriptors. *Journal of Chemical Information and Computer Sciences.* 37: 443-448 (1997).

[63] Brown, R.D.; Martin, Y.C. The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. *Journal of Chemical Information and Computer Sciences.* 37: 1-9 (1997).

[64] Matter, H.; Potter, T. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *Journal of Chemical Information and Computer Sciences.* 39: 1211-1225 (1999).

[65] Hong, H.; Xie, Q.; Ge, W.; Qian, F.; Fang, H.; Shi, L.; Su, Z.; Perkins, R.; Tong, W. Mold^2, Molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J. Chem. Inf. Model.* 48: 1337-1344 (2008).

[66] Jain, S.K.;Yadav, A.K.; Nayak, P. 2D QSAR analysis on oxadiazole derivatives as anticancer agents. *International Journal of Current Pharmaceutical Research.* 3(4) (2011).

[67] MATLAB Overview. Retrieved on March 24, 2013, from http://www.mathworks.com/products/matlab/index.html?sec=applications.

[68] Mold2_SoftwareIntroduction_12012008.doc Access on April 13, 2013, from http://www.fda.gov/downloads/ScienceResearch/BioinformaticsTools/Mold2/UCM161940.pdf.

[69] Zhou，L.；Lee，F.X.；Wilcox，W.；Christensen，J. Magnetic polarizability of hadrons from lattice QCD. *Nuclear Physics B (Proceeding. Supplements).* 119: 272-274 (2003).

[70] Zefirov, N. S.; Kirpichenok, M. A.； Izmailov, F. F.； Trofimov，M. I. (1987). *Doklady Akademii Nauk SSSR.* 296: 883.

[71] Section 2.4.37 in MoDeL Reference Manual. Retrieved on March 21, 2013, from http://jing.cz3.nus.edu.sg/model/findex.htm#h2.4.37.

[72] Quantitative Structure-Activity Relationships (QSAR). Retrieved on April 03, 2013, from http://www.issb.genopole.fr/~faulon/QSAR.php.

[73] Yerramsetty, K.M.; Neely, B.J.; Madihally, S.V.; Gasem, K.A.M. A skin permeability model of insulin in the presence of chemical penetration enhancer. *International Journal of Pharmaceutics.* 388: 13–23 (2010).