بسم الله الرحمن الرحيم

Razi University
Faculty of Science
Department of Chemistry

PhD Thesis

Quantitative Structure–Property Relationships (QSPRs) Studies of Aromatic Acids, Refrigerants, Cationic Surfactants, Diverse Drugs and Macromolecules Using Chemometrics Methods

Supervisor:

Prof. Jahanbakhsh Ghasemi

By:

Saadi Saaidpour

July 2008

دانشگاه رازی

دانشکده علوم

گروه شیمی

پایان نامه جهت اخذ درجه دکتری تخصصی

رشته ی شیمی گرایش تجزیه

مطالعات روابط کمی ساختار ملکولی ترکیبات آلی اسیدهای آروماتیک، خنک کننده ها ، دترجنت های کاتیونی ،ترکیبات شبه دارویی و دارویی، وماکروملکولها با خواص آنها بوسیله روشهای مختلف کمومتریکس

استاد راهنما:

پروفسور جهانبخش قاسمی

نگارش:

سعدی سعیدپور

تیر ماه ۱۳۸۷

Razi University

Faculty of Science
Department of Chemistry

PhD Thesis

Quantitative Structure–Property Relationships (QSPRs) Studies of Aromatic Acids, Refrigerants, Cationic Surfactants, Diverse Drugs and Macromolecules Using Chemometrics Methods

By:

Saadi Saaidpour

Evaluated and approved by thesis committee: as *EXCELLENT*

.......**Supervisor: Jahanbakhsh Ghasemi, Prof. of Analytical Chemistry (Chairman)**

.......**External Examiner: Abbas Afkhami, Prof. of Analytical Chemistry**

.......**External Examiner: Tayyebeh Madrakian, Assoc. Prof. of Analytical Chemistry**

.......**Internal Examiner: Mohammad Bagher Gholivand, Prof. of Analytical Chemistry**

.......**Internal Examiner: Mohsen irandoust, Assist. Prof. of Analytical Chemistry**

.......**Internal Examiner: Fahimeh Jalali, Assist. Prof. of Analytical Chemistry**

July 2008

DEDICATION

Dedicated to:

My Parents

My Brothers ***My Sister***

My Wife Mina

and

My Children

Soroush and Samanah

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my supervisor, Prof. Jahanbakhsh Ghasemi for his patience, guidance and encouragement all through the project. I admire him for his attitude to explore and learn new aspects of science and for his enthusiasm in teaching and mentoring. Prof. Ghasemi helped me identify my strengths and also guided me through difficult situations. This dissertation would not have been possible without his guidance and persistent help.

I would like to thanks the thesis committee, Prof. M.B. Gholivand, Dr. F. Jalali, Dr. M. irandoust, Prof. A. Afkhami, and Dr. T. Madrakian for their direction, assistance and guidance.

I want to give special thanks to the Dr. M. irandoust head of chemistry department for his contribution.

I would like to thank my parents and my family members who have been a constant source of support and encouragement. It is their confidence in me that has allowed me to achieve my goals.

Finally, I would like to express my gratitude to my wife, who as a friend helped me through my project with constant support, love and encouragement. Finally, I thank anyone teach me even one word.

ABSTRACT

A quantitative structure-activity/property relationship (QSAR/QSPR) has become an important branch of modern chemistry in past decades. A fundamental goal of QSAR/QSPR studies is to predict complex physical, chemical, biological, and technological properties of chemicals from simpler descriptors, preferably those calculated solely from molecular structure. This thesis focuses upon the methodology of QSPR, and the results of seven studies that implement that methodology. The goal of these works is to create predictive models that will link the molecular structures of sets of organic compounds to their physicochemical properties and/or biological activities.

The first study (chapter2) involves a very simple, strong, descriptive and interpretable model, based on the quantitative structure–property relationship (QSPR), is developed using multiple linear regression approach and quantum chemical descriptors derived from AM1-based calculations (MOPAC7.0) for determination of the acidity constants of some aromatic acids derivatives. A multiple linear regression (MLR) model with 74 molecules as training set has been developed for the prediction of the acidity constants of some aromatic acids using quantum chemical descriptors. The pK_a values of aromatic acids generally decreased with increasing positive partially charges of acidic hydrogen atom. This model was applied for the prediction of the pK_a of some aromatic acids (33 test acids), which were not used in the modeling procedure. The average relative error ($\overline{RE}\%$) of prediction is lower than 1% and square correlation coefficient (R^2) of 0.9882 is obtained.

In the second study (chapter3) of QSPR model for the estimation of boiling points of organic compounds containing halogens, oxygen, or sulfur without hydrogen bonding were established with the Molecular Modeling Pro Plus (MMPP) software. A QSPR study was performed to develop models that relate the structures of 90 refrigerants compounds to their boiling point temperatures. The optimal QSPR model was developed based on a 4-4–1 artificial neural network (ANN) architecture using molecular descriptors calculated from molecular structure alone. The root mean square errors (RMSE) in normal boiling points predictions were 4.46°C for the training set, 3.86°C for the validation set and 4.99 °C for the prediction set.

In the third study (chapter 4) relationships between the molecular structure and the critical micelle concentration (CMC) of cationic surfactants were investigated using a quantitative structure-micellization relationship (QSMR) approach. The CMC of a set of

29 tetra-alkyl ammonium and 15 alkylpyridinium salts was related to molecular structure descriptors using an ordinary least squares regression (OLS) method. Among different models obtained, three equations were selected as the best and their specifications are given. The results obtained for the simultaneous modeling of tetra-alkyl ammonium and alkylpyridinium salts indicate that geometric characteristics such as the hydrophobic chain length (L_C), hydrophobic volume (V_H), area of hydrophilic portion (A_{HP}) and radius of the hydrated counter ions (R_{HCI}) play a major role in micelle formation. Root mean square error of prediction (RMSEP) and average relative error ($\overline{RE}\%$) of prediction set for simultaneous model were about 0.0938 and 2.1124%, respectively.

The fourth study (chapter 5) involves a QSPR study was performed to develop models those relate the structures of 150 drug organic compounds to their n-octanol–water partition coefficients ($\log P_{o/w}$). A genetic algorithm was also applied as a variable selection tools in QSPR analysis. The models were constructed based on 110 training compounds, and predictive ability was tested on 40 compounds reserved for that purpose. Modeling of logarithm of $\log P_{o/w}$ of these compounds as a function of the theoretically derived descriptors was established MLR and ANN. The neural network employed here is a connected back-propagation model with a 4-4-1 architecture. Four descriptors for these compounds molecular volume (MV) (Geometrical), hydrophilic-lipophilic balance (HLB) (Constitutional), hydrogen bond forming ability (HB) (Electronic) and polar surface area (PSA) (Electrostatic) are taken as inputs for the models. The root mean square error of prediction (RMSEP) and square correlation coefficient (R^2) for MLR and ANN models were 0.2158, 0.9864 and 0.1838, 0.9876 for the prediction set $\log P_{o/w}$, respectively.

The fifth study (chapter 6) involves quantitative structure-retention relationship (QSRR) analysis is a useful technique capable of relating chromatographic retention time to the chemical structure of a solute. A QSRR study has been carried out on the reversed-phase high-performance liquid chromatography (RP-HPLC) retention times ($\log t_R$'s) of 62 diverse drugs (painkillers) by using molecular descriptors. MLR is utilized to construct the linear QSRR model. The applied MLR is based on a variety of theoretical molecular descriptors selected by the stepwise variable subset selection procedure. Stepwise regression was employed to develop a regression equation based on 50 training compounds, and predictive ability was tested on 12 compounds reserved for that purpose. The geometry of all drugs were optimized by the semi-empirical method AM1 and used to calculate different molecular descriptors. The regression equation included three parameters that consisted of *n*-octanol–water partition coefficient (*logP*), molecular surface area (SM) and hydrophilic-lipophilic balance (HLB) of the drug molecules, all of which could be related to retention time property. The results indicate that a strong correlation exists between the $logt_R$ and mentioned descriptors for drug compounds.

In the sixth study (chapter 7) a QSPR study was performed to develop a model that relates the structures of 150 drug organic compounds to their aqueous solubility ($\log S_w$). Molecular descriptors derived solely from 3D structure were used to represent molecular structures. A subset of the calculated descriptors selected using stepwise regression that used in the QSPR model development. Stepwise regression was employed to develop a regression equation based on 110 training compounds, and predictive ability was tested on 40 compounds reserved for that purpose. The final regression equation included three parameters that consisted of octanol/water partition coefficient ($\log P$), molecular volume (MV) and hydrogen bond forming ability (HB), of the drug molecules, all of which could be related to solubility property. The prediction results are in good agreement with the experimental values. The root mean square error of prediction (RMSEP) and square correlation coefficient (R^2) of prediction of $\log S_w$ were 0.0959 and 0.9954, respectively.

The seventh study (chapter 8) involves QSPR models for the stability constants of 58 complexes of 1,4,7,10,13-pentaoxacyclopentadecane ethers (15C5) were established with the CODESSA program. Experimental stability constants ($\log K$) for a diverse set of 58 complexes of 15C5 structures are correlated with computed structural descriptors using CODESSA. Stability constants for complexes of 15C5 ethers with potassium cation (K^+) have been determined at 25 °C in methanol solution. The best multilinear regression method (BMLR) encoded in CODESSA software was used to select significant descriptors for building multilinear QSPR model and the predictive power of model is estimated with the leave-one-out (LOO) cross-validation method. The proposed model can be used for the prediction of the stability constants of 15C5 complexes. The best QSPR model with five descriptors has $R^2 = 0.9452$, $s^2 = 0.0110$, and $F = 67.0312$.

TABLE OF CONTENTS

Content	Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Chapter2

Chapter3

Chapter4

Chapter5