BABEŞ-BOLYAI UNIVERSITY
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING
DEPARTMENT OF ORGANIC CHEMISTRY

MODELING OF THE LIPOPHILICITY OF BIOLOGICALLY ACTIVE COMPOUNDS USING PHYSICO-CHEMICAL AND TOPOLOGICAL PROPERTIES BY MULTIVARIATE REGRESSION METHODS

PhD Thesis Abstract

Scientific Advisor:
     Prof. Dr. MIRCEA V. DIUDEA

PhD student:
     CRISTINA ONIŞOR (POP)

Cluj-Napoca
2012

Babeş-Bolyai University
Faculty of Chemistry and Chemical Engineering
Department of Organic Chemistry

**MODELING OF THE LIPOPHILICITY OF BIOLOGICALLY ACTIVE
COMPOUNDS USING PHYSICO-CHEMICAL AND TOPOLOGICAL
PROPERTIES BY MULTIVARIATE REGRESSION METHODS**

Scientific Advisor:
Prof. Dr. Mircea V. Diudea

PhD student:
Cristina Onişor

President:
Conf. Dr. **Cornelia Majdik**

Jury:
Assistant Prof. Dr. **Mihalj Poša**, University of Novi Sad, Serbia
Professor Dr. **Maruţoiu Constantin**, Universitatea Babeş- Bolyai, Cluj-Napoca
Professor Dr. **Mihai Medeleanu**, Universitatea Politehnică, Timisoara

ACKNOWLEDGMENTS

"It's never enough for me to return for what I've been given!"

TABLE OF CONTENTS

THEORETICAL PART

GENERAL INTRODUCTION

The lipophilicity of molecules, as a complex physicochemical property, seems to be the main factor governing the transport and the distribution of drug molecules in biological systems and as a direct consequence it is the most important property in classical quantitative structure-activity relationship (QSAR) studies [1].

Chapter 1
LIPOPHILICITY

Lipophilicity, a physicochemical parameter of compounds, is an important factor in QSAR studies and also a property of a molecule which depends on and can be changed by modifications in molecular structure. Transport phenomena in vivo and through membranes proved to be dependent on lipophilic contributions. Therefore, the first important parameter in understanding the activity of unknown compounds is lipophilicity. The parameter of the lipophilicity, partition coefficient (log P), is commonly used in drug designing and it is a numeric characteristic of lipophilicity of the examined substance, potential drug [4].

From the magnitude of the Log P of a compound, one can infer its ease of transport through the cell membrane and other related events [5].

The 1-octanol/water partition coefficient (log P) is the most widely used parameter in medicinal chemistry and can give accurate predictions of activity in a complex biological system such as membranes, keeping in mind the obvious limitation that the activity of the drug depends on its lipophilic nature [7].

1.1. Analytical Methods for the Determination of Lipophilicity

The same basic intermolecular actions determine the behavior of chemical compounds in both biological and chromatographic environments. As a consequence, the chromatographic approach has been quite successful in duplicating Log P data derived by traditional "shake-flask" technique or other procedures [5].

1.1.1. Extractive Method

The logarithm of the partition coefficient of a chemical in the *n*-octanol-water system (P_{ow}), usually measured by the "shake flask" method, is widely used because of its simplicity and because there is same similarity between *n*-octanol and biological membranes:

$$P = C_0 / C_w \quad \text{or} \quad \log P = \log C_0 - \log C_w \quad (1.1)$$

where C_0 and C_w represent molar concentrations of the partitioned compound in organic and aqueous phase, respectively.

1.1.2. Chromatographic Methods

The alternative "indirect" techniques for experimentally measurement of log P are the chromatographic methods, RPHPLC and RPTLC. These methods are based on the assumed linear relationship between log P as measured with shake flask and the logarithm of chromatographic capacity factor data (log K and R_M, respectively, for RPHPLC and RPTLC.

1.1.2.1. Lipophilicity Determination by Reveresed-Phase Thin Layer Cromatography

The use of RPTLC is based on the assumed linear relationship between the molecular parameter and Log P:

$$R_M = \log (1/R_F - 1) \tag{1.2}$$
$$R_M = R_{Mo} + bC \tag{1.3}$$

where R_M values were calculated using Eq. 30 and C is the concentration of the organic component of the mobile phase.

1.1.2.2. Lipophilicity determination by Reveresed-Phase High-Performance Liquid Cromatography

The most popular lipophilicity indices measured by RP-HPLC are expressed through the Soczewinski-Snyder equation:

$$\log k = \log k_w - SC , \tag{1.4}$$

where

$$\log k = \log(\frac{t_r - t_o}{t_o}) , \tag{1.5}$$

The isocratic chromatographic hydrophobicity index, φ_0, represents the volume fraction of organic solvent in the mobile phase for which the amount of solute in the mobile phase is equal to that in the stationary phase, and in RP-HPLC parameter φ_0 is calculated as the ratio of the intercept and slope [23, 24]:

$$\varphi_0 = \frac{\log k_w}{S} \tag{1.6}$$

The φ_0 estimation by RP-TLC is given by [25, 26]:

$$\varphi_0 = \frac{R_{M_0}}{b} \tag{1.7}$$

Moreover, the last decade came with more revolutionary lipophilicity indices by employing the Principal Component Analysis (PCA) in this analytical area. We used the lipophilicity scale obtained by applying Principal Component Analysis (PCA) directly to the matrix retention data (k, log k, R_F and R_M) resulted for all compounds and combinations of methanol-water and so the eigenvalues of the covariance matrix were obtained [25, 27].

Chapter 2

COMPUTATIONAL CALCULATION FOR DETERMINATION OF LIPOPHILICITY

Introduction

The log P_{ow} is obtained by summing the fragment values and the correction terms for intramolecular interactions.

Calculation methods are based on the theoretical fragmentation of the molecule into suitable substructures for which reliable log P_{ow} increments are known. The log P_{ow} is obtained by summing the fragment values and the correction terms for intramolecular interactions.

In general, the reliability of calculation methods decreases as the complexity of the compound under study increases. In the case of simple molecules of low molecular weight and with one or two functional groups, a deviation of 0.1 to 0.3 log P_{ow} units between the results of the different fragmentation methods and the measured value can be expected.

2.1.   Atomic and Molecular Approaches

- *Substructure approaches* have in common that molecules are cut into atoms (*atom contribution methods*) or groups (*fragmental methods*);
- *Whole molecule approaches* inspect the entire molecule; they use for instance *molecular lipophilicity potentials (MLP), topological indices* or *molecular properties* to quantify log *P*.

2.1.1. Substructure Approaches

Thus, fragmental methods apply correction rules coupled with molecular connectivity.

A. Fragmental methods
B. Atom contribution methods

2.1.2.   Whole Molecule Approaches

Whole molecule approaches use: a) *molecular lipophilicity potentials (MLP)*, b) *topological indices* or c) *molecular properties* such as charge densities, surface area, volume and electrostatic potential to quantify log *P* [1].

2.2.   Computational Programs for Lipophilicity Determination

- *HyperChem Pro 6*
- *XLOGP 2.0*
- *CLOGP*
- *miLogP 1.2*
- *ALOGPS 2.1*
- *IA logP*
- *KOWWIN*

Chapter 3

COMPUTATIONAL CHEMISTRY

Introduction

This field of science combines mathematics and chemistry and is called computational chemistry. It deals with finding relationships called Quantitative Structure Property-Relationships (QSPRs), Quantitative Structure Activity-Relationships (QSARs) or Quantitative Structure Retention-Relationships (QSRRs) [1].

The conception that there exists a close relationship between bulk properties of compounds and the molecular structure of those compounds is quite rooted in chemistry. This idea allows one to provide a clear connection between the macroscopic and the microscopic properties of matter, and thus has been firmly established as one of the central foundations of chemistry. Therefore, it is the basic tenet of chemistry to attempt to identify these assumed relationships between molecular structure and physical chemistry properties and then to quantify them.

QSAR/QSPR/QSRR methods are unquestionably of great relevance in modern chemistry and biochemistry.

3.1. Quantitative Structure-Activity Relationships (QSARs)

Structure-activity research is based on the fact that the physicochemical parameters of a compound mainly determine its pharmacokinetic profile by influencing the amount and rate of compound absorption.

3.2. Quantitative Structure-Property Relationships (QSPRs)

The QSPR studies have become an integral part of rational drug design and formulation development.

3.3. Quantitative Structure-Retention Relationships (QSRRs)

QSRR in LC is the method to identify the properties of the solutes that control their retention by correlating the retention factors (k) measured at certain separation conditions in LC and solutes properties (Pi, descriptors).

3.4. Tools and Techniques of QSRs

3.4.1. Multiple Linear Regression Analysis (MLR)

One of the most widely used examples of estimation in statistics is provided by linear regression.

3.4.2. Principal Component Analysis (PCA)

PCA finds linear combinations of the original measurement variables that describe the significant variation in the data.

2D and 3D scores plot

2D and 3D loading plots

3.4.3. Genetic Algorithms (GA)

Genetic Algorithms use a population of individuals as a model to search for the globally optimum solution to a problem.

CONCLUSIONS

This parameter can be either determined experimentally or calculated. Because experimental measurements are time consuming and difficult, computational methods are very valuable tools for calculation of LogP's for large sets of compounds in QSAR studies, particularly at the screening stage.

ORIGINAL CONTRIBUTIONS

Chapter 4
MODELING AND PREDICTION (CORRECTION) OF PARTITION COEFFICIENTS OF CARBOXYLIC ACIDS BY MULTIVARIATE REGRESSION METHODS

4.1. General Overview

Carboxylic acids are hydrocarbon derivatives containing a carboxyl (COOH) moiety. Since carboxylic acids have a carbonyl group and an alcohol group they share some basic physico-chemical properties with aldehydes, ketones and alcohols.

4.1.1. Nomenclature of Carboxylic Acids

Most simple carboxylic acids, rather than being called by their IUPAC names, are more often referred to by common names that are older than their systematic names. Most simple carboxylic acids were originally isolated from biological sources; because their structural formulas were often unknown at the time of isolation they were given names that were generally derived from the names of the sources. The acids containing an odd number of carbon atoms greater than nine generally do not have common names. The reason is that long-chain carboxylic acids were originally isolated from fats (which are carboxylic esters), and generally these fats contain carboxylic acids with only an even number of carbon atoms (because the process by which living organisms synthesize such fatty acids puts the molecules together in two-carbon pieces) [6-9].

Table 4.1. Names of Saturated Monocarboxylic Acids.

No.	n	IUPAC Name	Common Name	Log Kow
1	0	Methanoic Acid	Formic Acid	-0.54
2	1	Etanoic Acid	Acetic Acid	-0.17
3	2	Propanoic Acid	Propionic Acid	0.33
4	3	Butanoic Acid	Butiric Acid	0.79
5	4	Pentanoic Acid	Valeric Acid	1.39
6	5	Hexanoic Acid	Caproic Acid	1.84
7	6	Heptanoic Acid	Enantic Acid	2.42
8	7	Octanoic Acid	Caprilic Acid	3.05
9	8	Nonanoic Acid	Pelargonic Acid	3.42
10	9	Decanoic Acid	Capric Acid	4.09
11	10	n-Hendecanoic Acid	n-Undecanoic Acid	-
12	11	Lauric Acid	n-Dodecanoic Acid	4.2
13	12	Tridecylic Acid	n-Tridecanoic Acid	5.56
14	13	Myristic Acid	n-Tetradecanoic Acid	6.11
15	14	Pentadecanoic Acid	n-Pentadecyclic Acid	-
16	15	Palmitic Acid	n-Hexadecanoic Acid	7.17
17	16	Margaric Acid	n-Heptadecanoic Acid	7.69

18	17	Stearic Acid	n-Octadecanoic Acid	8.23
19	18	Nondecylic Acid	n-Nonadecanoic Acid	8.75
20	19	Arachidic Acid	n-Eicosanoic Acid	9.28
21	20	Henicosanoic acid	n-Heneicosanoic Acid	9.81
22	21	Behenic Acid	n-Docosanoic Acid	10.34
23	22	Tricosanoic acid	n-Tricosanoic acid	10.87
24	23	Lignoceric Acid	n-Tetracosanoic Acid	11.4
25	24	Pentacosanoic Acid	n-Pentacosanoic Acid	-
26	25	Cerotic acid	n-Hexacosanoic acid	12.47
27	26	Heptacosanoic Acid	n-Heptacosanoic Acid	-
28	27	Montanic acid	n-Octacosanoic acid	-
29	28	Nonacosanoic Acid	n-Nonacosanoic Acid	-
30	29	Melissic acid	n-Triacontanoic acid	11.61

$$(CH_3)n$$
$$HO-C=O$$
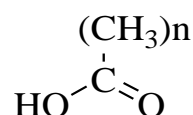
4.1.2. Classification of Carboxylic Acids

Monocarboxylic acids represent an important minor component of cells. Long-chain acids (called fatty acids) are important substrates for cell energetics; in addition they play a regulatory role in the organism. It was shown by different methods that the neutral form of fatty acids can easily penetrate through membranes, and an important step of permeation is the acid binding to the membranes. It may be proposed therefore that the mechanism of acid permeation across the membrane depends on the length of the hydrocarbon chain of the acid.

Acids with a longer hydrocarbon chain length have specific transport properties which may be attributed to the ability to form micelles [2-4].

4.1.3. Sources of Monocarboxylic Acids

Acrylic acid   Fatty acids   Docosahexaenoic acid   Eicosapentaenoic acid
    Amino acids
Keto acids   Pyruvic acid   Acetoacetic acid [10, 11]

4.1.4. Importance and Applications of Carboxylic Acids

Important carboxylic-acid derivatives include esters, anhydrides, amides, halides (see halogen), and salts (see soap).

Carboxylic acids occur widely in nature. The fatty acids are components of glycerides, which in turn are components of fat. Hydroxyl acids, such as lactic acid (found in sour-milk products) and citric acid (found in citrus fruits), and many keto acids are important metabolic products that exist in most living cells. Proteins are made up of amino acids, which also contain carboxyl groups [2, 12].

Carboxylic acid derivatives have varied applications, they are important in many areas of organic chemistry such as synthesis, separation and identification of organic compounds.

4.2. Lipophilicity of Carboxylic Acids

Lipophilicity of organic chemicals has also been identified as an important parameter to predict adsorption in soils and sediments.

However, most of them determined a single Kow value (also called P), reflecting the lipophilicity of the neutral species only. It is essential to determine Kow values over the full range of pH that occurs in the environment in order to get an appropriate predictor [21-23].

In many cases, this molecular parameter strongly correlates with the biological activity of chemicals, as well as with other important physicochemical properties [24–28].

The experimental octanol/water partition coefficient (log P) has been considered one of the most important descriptors in the QSAR analysis of toxicity, because of the possible importance of hydrophobic interactions in determining the genotoxic activity of compounds.

The measured partition coefficients of a subset of the monocarboxylic acids were determined by the shake-flask method which is the classical and most reliable method of log P determination. It consists of dissolving some of the solute in question in a volume of octanol and water, then measuring the concentration of the solute in each solvent.

The studied carboxylic acids with their corresponding lipophilicity/hydrophobicity values (Log Kow) determined experimentally (Table 4.1) were taken from the literature and are the most reported in the literature [39, 40].

4.3. Experimental Part
Computed Descriptors

A data set consisting of 30 monocarboxylic acids was characterized by 1223 theoretical descriptors calculated using the software Dragon 5.4 (Dragon 2005) [41] and 17 descriptors calculated using SciQSAR module (Alchemy software) [42] allowed the calculation of several descriptors such as constitutional, topological, geometrical, molecular and quantum mechanical. The model significance obtained in this work, with the exclusion of redundant and noisy information, was analyzed by MobyDigs v.1.0 software (Todeschini 2004) [44] that calculated the regression models by using (GA) genetic algorithms to perform variable selection.

4.4. Results and Discussion

For the validation of the developed models, the data sets was split into a training set and external prediction set containing 22 and 8 compounds, respectively.

In this study forward stepwise multiple linear regression analysis (MLR) has been applied to the modeling of partition coefficient (lipophilicity) of monocarboxylic acids by means of 1240 different computed descriptors.

Different validation procedures including linear regression of original and predicted data and also the graphing representation of partition coefficients corresponding for monocarboxylic acids confirm the high quality of models and their predictive capability.

The powerful predictive ability of the models allowed the estimation of unknown partition coefficients for some monocarboxylic acids and a correct prediction of partition coefficients for Lauric Acid (Cpd. 12) and Melissic Acid (Cpd. 30).
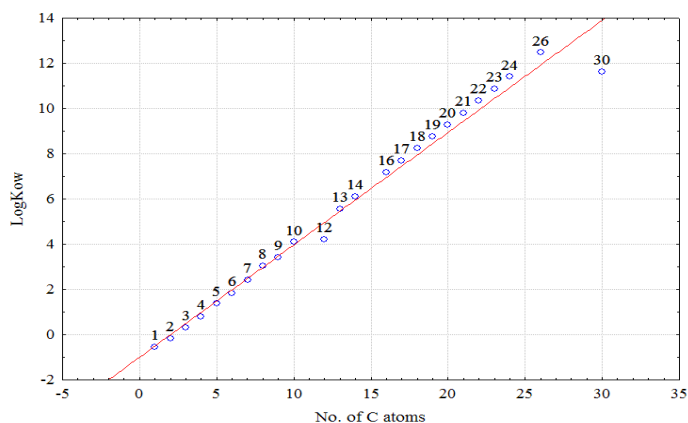
Figure 4.1. Graphical Representation of Log Kow Values with Outliers
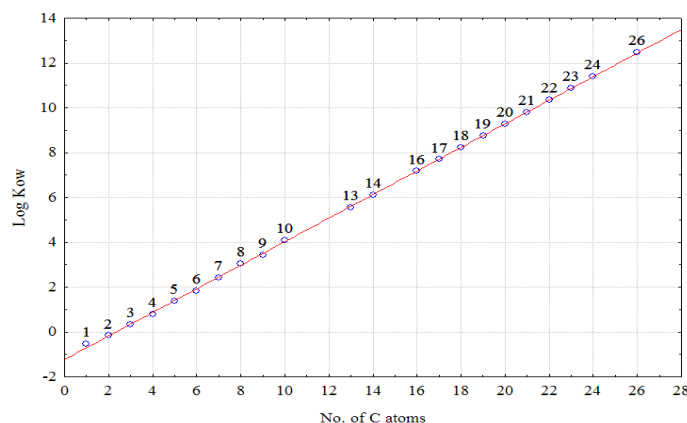
Figure 4.2. Graphical Representation of Log Kow Values without Outliers

By applying forward stepwise regression analysis, high-quality regression equations were obtained and the Log Kow partition coefficients of monocarboxylic acids were estimated in both cases: including the experimental values for Lauric Acid (Cpd. 12) and Melissic Acid (Cpd. 30) in the model (Eqs. 4.1, 4.3, 4.5 and 4.7) and without these values in the model (Eqs. 4.2, 4.4, 4.6 and 4.8). The contribution of the different groups of molecular descriptors of monocarboxylic acids partition coefficients is given by the equations as follows:

Table 4.2. Monovariable Regression Equations Based on Dragon Descriptors.

0D descriptors	Equation	No.
With outliers	Log Kow = -5.506 + 0.331Ss	(4.1)
Without outliers	Log Kow = -1.568 + 0.329Sv	(4.2)
2D descriptors	**Equation**	**No.**
With outliers	Log Kow = -1.489 + 0.497$MPC01$	(4.3)
Without outliers	Log Kow = -2.061 + 1.053$X1sol$	(4.4)
3D descriptors	**Equation**	**No.**
With outliers	Log Kow = -0.784 + 7,254$Mor14p$	(4.5)
Without outliers	Log Kow = -0.457 + 0.396$RDF025v$	(4.6)
Other descriptors	**Equation**	**No.**
With outliers	Log Kow = -3.968 + 0.775$PCWTe$	(4.7)
Without outliers	Log Kow = -1.627 + 0.114AMR	(4.8)

Selection of a regression model is commonly based on finding the fit with the highest R^2 value given the data. The Q^2 is proposed as a criterion for detecting influential observations or outliers.

The statistical details of the QSAR models given above speak for its good statistical quality which indicates that these equations represent a satisfactory model for some of the calculated descriptors and have definite physical meaning corresponding to different intermolecular interactions. All the R^2 values are above 0.9, which suggest that a good percentage of the total variance in lipophilicity is accounted by the model.

The statistical significance of each model is evaluated by the determination coefficient R^2, leave-one-out crossvalidation coefficient Q^2, predictive error sum of squares PRESS, Fisher test F and standard error s (Tables 4.4-4.5).

The high Q^2 and R^2 values are considered as a proof of high predictive ability and robustness of the obtained models. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the model. The standard error measures the model error.

Table 4.3. Monovariable Regression Equations Based on SciQSAR Software.

SciQSAR descriptors	Equation	No.
With outliers	Log Kow = -0.973 + 0.993$^1\chi v$	(4.9)
Without outliers	Log Kow = -2.061 + 1.053$^1\chi$	(4.10)

The predicted values for Lauric Acid (Cpd. 12) and Melissic Acid (Cpd. 30) were much higher in both cases. All the predicted values are presented in Table 4.6-4.8.

The results suggest that topological, electronic and molecular aspect of the molecule seem to be dominant in the retention mechanism and as a consequence they seem to control the lipophilicity.

Table 4.4. Statistical Quality Parameters of Multiple Regression Models
Based on Dragon Descriptors.

Type of descriptor	Crt.	Q^2	R^2	s	PRESS	F
0D	With outliers	0.9726	0.9813	0.583	10.985	1158
0D	Without outliers	0.9998	0.9998	0.056	0.079	117100
2D	With outliers	0.9725	0.9906	0.583	11.001	1157
2D	Without outliers	0.9999	0.9999	0.042	0.043	204500
3D	With outliers	0.9798	0.9857	0.511	8.068	1511
3D	Without outliers	0.9999	0.9999	0.002	0.050	172220
Other	With outliers	0.9739	0.9824	0.567	10.447	1225
Other	Without outliers	0.9998	0.9998	0.056	0.081	115091

Table 4.5. Statistical Quality Parameters of Multiple Regression Models
Based on SciQSAR and ChemDraw Descriptors

Type of descriptor	Crt.	Q^2	R^2	s	PRESS	F
SciQSAR	With outliers	0.9726	0.9814	0.582	10.968	1160
SciQSAR	Without outliers	0.9999	0.9999	0.042	0.043	204500

Table 4.6. Predicted Log Kow Values With Experimental Values for Lauric Acid and Melissic Acid in the Regression Model.

Cpds	Dragon descriptors type			
	0D	2D	3D	other
1	-0.54	-0.54	-0.54	-0.54
2	-0.17	-0.17	-0.17	-0.17
3	0.33	0.33	0.33	0.33
4	0.79	0.79	0.79	0.79
5	1.39	1.39	1.39	1.39
6	1.84	1.84	1.84	1.84
7	2.42	2.42	2.42	2.42
8	3.05	3.05	3.05	3.05
9	3.42	3.42	3.42	3.42
10	4.09	4.09	4.09	4.09
11	**4.48**	**4.47**	**4.42**	**4.32**
12	*4.97*	*4.97*	*4.76*	*4.79*
13	5.56	5.56	5.56	5.56
14	6.11	6.11	6.11	6.11
15	**6.46**	**6.46**	**6.75**	**6.55**
16	7.17	7.17	7.17	7.17
17	7.69	7.69	7.69	7.69
18	8.23	8.23	8.23	8.23
19	8.75	8.75	8.75	8.75
20	9.28	9.28	9.28	9.28
21	9.81	9.81	9.81	9.81
22	10.34	10.34	10.34	10.34
23	10.87	10.87	10.87	10.87
24	11.40	11.40	11.40	11.40
25	**11.42**	**11.43**	**11.24**	**11.43**
26	12.47	12.47	12.47	12.47
27	**12.42**	**12.42**	**12.14**	**12.41**
28	**12.91**	**12.93**	**12.59**	**12.89**
29	**13.41**	**13.41**	**13.06**	**13.38**
30	*13.91*	*13.91*	*13.49*	*13.87*

Table 4.7. Predicted Log Kow Values Without Experimental Values for Lauric Acid and Melissic Acid in the Regression Model.

Cpds	Dragon descriptors type			
	0D	2D	3D	other
1	-0.54	-0.54	-0.54	-0.54
2	-0.17	-0.17	-0.17	-0.17
3	0.33	0.33	0.33	0.33
4	0.79	0.79	0.79	0.79
5	1.39	1.39	1.39	1.39
6	1.84	1.84	1.84	1.84
7	2.42	2.42	2.42	2.42
8	3.05	3.05	3.05	3.05
9	3.42	3.42	3.42	3.42
10	4.09	4.09	4.09	4.09
11	**4.55**	**4.54**	**4.51**	**4.55**
12	*5.08*	*5.07*	*5.04*	*5.08*
13	5.56	5.56	5.56	5.56
14	6.11	6.11	6.11	6.11
15	**6.65**	**6.65**	**6.66**	**6.65**
16	7.17	7.17	7.17	7.17
17	7.69	7.69	7.69	7.69
18	8.23	8.23	8.23	8.23
19	8.75	8.75	8.75	8.75
20	9.28	9.28	9.28	9.28
21	9.81	9.81	9.81	9.81
22	10.34	10.34	10.34	10.34
23	10.87	10.87	10.87	10.87
24	11.40	11.40	11.40	11.40
25	**11.91**	**11.92**	**11.92**	**11.91**
26	12.47	12.47	12.47	12.47
27	**12.96**	**12.97**	**12.98**	**12.96**
28	**13.49**	**13.49**	**13.51**	**13.49**
29	**14.01**	**14.02**	**14.03**	**14.01**
30	*14.54*	*14.55*	*14.56*	*14.54*

Table 4.8. Predicted Log Kow Values With and
Without Experimental Values for Lauric Acid
and Melissic Acid in the Regression Model.

| Cpds | SciQSAR descriptors | |
	With outliers	Without outliers
1	-0.54	-0.54
2	-0.17	-0.17
3	0.33	0.33
4	0.79	0.79
5	1.39	1.39
6	1.84	1.84
7	2.42	2.42
8	3.05	3.05
9	3.42	3.42
10	4.09	4.09
11	**4.48**	**4.54**
12	*4.97*	*5.07*
13	5.56	5.56
14	6.11	6.11
15	**6.46**	**6.65**
16	7.17	7.17
17	7.69	7.69
18	8.23	8.23
19	8.75	8.75
20	9.28	9.28
21	9.81	9.81
22	10.34	10.34
23	10.87	10.87
24	11.40	11.40
25	**11.42**	**11.92**
26	12.47	12.47
27	**12.42**	**12.97**
28	**12.91**	**13.49**
29	**13.41**	**14.02**
30	*13.91*	*14.55*

4.5. Conclusions

The high predictive ability of both QSPR models enabled the correction of Lauric Acid and Melissic Acid partition coefficients. The powerful predictive ability of the models allowed the estimation of unknown partition coefficients for some monocarboxylic acids and a correct prediction of partition coefficients for Lauric Acid and Melissic Acid. The predicted values for Lauric Acid and Melissic Acid were much higher in both cases.

The results suggested also that the most significant descriptors encounter volume, polarizability, topological and molecular aspects of the molecule.

Chapter 5
BILE ACIDS AND THEIR DERIVATIVES

Introduction
The most abundant BAs present in the human body are: cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA), lithocholic acid (LCA), ursodeoxycholic acid (UDCA) and the derivatives of 5β-cholan-24-oic acid. These compounds are found in the human body primarily in their conjugated forms to glycine and taurine *via* amidation on carbon 24 [5].

The BA structure accounts for this phenomenon: a strongly hydrophobic steroid bulk mainly determines the lipophilic character of these molecules and the ionization of the C-24 carboxy group plays a minor role. According to these findings, both ionized and protonated species must be considered to partition in the 1-octanol [7, 8].

5.1. Structure and Classification of Bile Acids and Their Derivatives
Bile acids are steroid compounds with complex physico–chemical properties according to the number, position and orientation of the hydroxyl groups, and the type of conjugation to form glyco-, tauro-, oxo-, and diacetoxy- bile acid derivatives [11] (Table 5.1).

5.2. Properties of Bile Acids and Their Derivatives
Elimination of excess cholesterol; solubilize cholesterol in the bile; emulsifying agents; hormones; pivotal role; amphipatic properties.

5.3. Applications and Importance of Bile Acids
Their blood concentrations are important prognostic and diagnostic indicators of hepatobiliary and intestinal dysfunction. They are widely used for example as drugs to dissolve cholesterol gallstone or as promoters of drug absorption by nonparenteral routes; in each of these cases their interaction with biological membranes or the lipid environment is of crucial importance [15-18].

5.4. Modeling of Chromatographic Lipophilicity of Bile Acids and Their Derivatives Estimated by Reversed-Phase Thin Layer Chromatography

Reversed-phase chromatographic indices, log k_w as alternative measure, are usually calibrated towards the octanol–water system and conditions are chosen so that the best parallelism with Log P is achieved [1, 2]. The development of sophisticated HPLC technology has provided new possibilities for rapid and precise bile acid separation and measurement [3, 4].

5.4.1. Experimental Part
RP-HPTLC has been used for compounds lipophilicity estimation. A large number of computed and experimental indices have been investigated, characterized and compared.

Table 5.1. Abreviation of bile acids and their derivatives.

No.	Name of compounds	Abbreviations	Position and orientation of hydroxyls, diacetoxy- and oxo- groups
1	Lithocholic acid	LC	3α
2	Deoxycholic acid	DC	$3\alpha,12\alpha$
3	Chenodeoxycholic acid	CDC	$3\alpha,7\alpha$
4	Cholic acid	C	$3\alpha,7\alpha,12\alpha$
5	Ursodeoxycholic acid	UDC	$3\alpha,7\beta$
6	Hyocholic acid	HC	$3\alpha,6\alpha,7\alpha$
7	Hyodeoxycholic acid	HDC	$3\alpha,6\alpha$
8	Glycochenodeoxycholic acid sodium salt	GCDC	Glyco conjugate of CDC
9	Taurodeoxycholic acid sodium salt	TDC	Tauro conjugate of DC
10	Glycocholic acid sodium salt	GC	Glyco conjugate of C
11	Glycodeoxycholic acid sodium salt	GDC	Glyco conjugate of DC
12	Taurolithocholic acid sodium salt	TLC	Tauro conjugate of LC
13	Taurochenodeoxycholic acid sodium salt	TCDC	Tauro conjugate of CDC
14	Glycolithocholic acid sodium salt	GLC	Glyco conjugate of LC
15	Taurocholic acid sodium salt	TC	Tauro conjugate of C
16	$3\alpha,7\alpha$-Dihydroxy-12-oxo-5β-cholanic acid	12-oxo C	$3\alpha,7\alpha,12$-oxo
17	$3\alpha,12\alpha$-Dihydroxy-7-oxo-5β-cholanic acid	7-oxo C	$3\alpha,12\alpha,7$-oxo
18	3α-Hydroxy-7,12-dioxo-5β-cholanic acid	7,12-dioxo C	$3\alpha,7,12$-oxo
19	12α-Hydroxy-3,7-dioxo-5β-cholanic acid	3,7-dioxo C	$12\alpha,3,7$-oxo
20	3,7,12-Trioxo-5β-cholanic acid	3,7,12-trioxo C	3,7,12-oxo
21	3α-Hydroxy-12-keto-5β-cholanoic acid	3-OH,12-oxo C	3-OH,12-oxo
22	3,12-Diketo-5β-cholanoic acid	3,12-dioxo C	3,12-dioxo
23	Methyl ester of 3α-Acetoxy-12-keto-5β-cholanoic acid	Methyl ester of 3-acetoxy,12-oxo C	3α-OAc,12-oxo
24	3α-Hydroxy-7-keto-5β-cholanoic acid	3-OH,7-oxo C	$3\alpha,7$-oxo
25	Methyl ester of 3$\alpha,7\alpha$-Diacetoxy-12-keto-5β-cholanoic acid	Methyl ester of 3$\alpha,12\alpha$-diacetoxy,12-oxo C	$3\alpha,7\alpha$-OAc,12-oxo
26	Methyl ester of 3$\alpha,12\alpha$-Diacetoxy-7-keto-5β-cholanoic acid	Methyl ester of 3$\alpha,12\alpha$-diacetoxy,7-oxo C	$3\alpha,12\alpha$-OAc,7-oxo
27	Methyl ester of 3,6-Diketo-5β-cholanoic acid	Methyl ester of 3,6-oxo C	3,6-oxo

These results, followed by those obtained by HPLC will permit to determine the lipophilicity indices of 27 bile acids and their derivatives and to investigate the molecular mechanism of retention in order to find an objective manner of quantitative comparison of retention properties of different chemically bonded stationary phases.

5.4.1.2. Chromatography

Chromatographic measurements were carried out on 10x10 cm HPTLC different stationary phases: 1) HPTLC Silicagel 60RP-18$_{F254s}$; 2) HPTLC Silicagel 60RP-18W$_{F254s}$; 3) HPTLC Silicagel 60CN$_{F254s}$, Methanol-water mixtures were used as mobile phase, the volume fraction of organic solvent in the mobile phase ranged as in Table 5.2, in each case the addition step being of 5%.

Table 5.2. Mobile Phase System Corresponding to Different Investigated Stationary Phases.

Stationary Phase	Mobile Phase MeOH (%) (v/v)	Addition Step (%)
RP-18$_{F254s}$	70-90	
RP-18W$_{F254s}$	50-70	5
CN$_{F254s}$	45-65	

After being developed, the dried plates were sprayed with manganese chloride in sulfuric acid solution, and placed in the oven at 100-120^{0}C. In each case sharp spots without tailing tendency were obtained.

5.4.2. Computed Molecular Descriptors

The Log P values were calculated by Chem3D Ultra 10 [25] (LogP$_{CD}$, PartCoeff$_{CD}$), four are given by the Dragon 5.4 [26] (MLOGP-Moriguchi method, MLOGP2-Squared Moriguchi method, ALOGP-Ghose-Crippen method, ALOGP2-Squared Ghose-Crippen method), and two of them are calculated by Alchemy software (LogP$_{SciQSAR}$, LogPc$_{SciLogP}$) [27]. The ALOGPS 2.1-vcclab internet module allows the calculation of another nine LogP values (ALOGPs, AClogP, AB/LogP, COSMOFraq, miLogP, ALOGP, MLOGP, KOWWIN, XLOGP2, XLOGP3, AverageLogP) [28]. Bile acids and their derivatives were also characterized by 1267 theoretical descriptors calculated using Dragon 5.4 software [23, 34, 35]. The Dragon descriptors employed in this study can be arranged in the following groups: *descriptors 2D*: 2D autocorrelations; *descriptors 3D*: RDF, 3D-MORSE, GETAWAY, WHIM, geometrical properties and Randić molecular profiles; *others descriptors*: functional groups, atom-centered fragments, molecular properties, charge descriptors, and constitutional properties.

5.4.3. Results and Discussion

The chromatographic parameters of lipophilicity (R$_{M0}$, *b*, φ$_0$, and scores corresponding to the first principal component on R$_F$ and R$_M$ values) were estimated and compared to calculated partition coefficients.

Highly significant correlations were obtained between different experimental indices of lipophilicity (R$_{Mo}$, *b*, φ$_0$ and scores corresponding to the first principal component) estimated on CN$_{F254s}$ and RP-18$_{F254s}$ and some computed log P values which combine electronic and topological aspects.

The highest R_{M0} values were obtained on RP-18$_{F254s}$ and CN$_{F254}$ and it can be observed that the number of hydroxyl, keto- and diacetoxy- groups, as well as their position and orientation determine the chromatographic behavior of bile acids and their derivatives.

The most statistical significant QSPR models were obtained in the case of CN$_{F254s}$ and RP-18$_{F254s}$ stationary phase. On the basis of presented correlations, it may be appreciated that the lipophilicity indices determined on RP-18$_{F254s}$ and CN$_{F254s}$ stationary phases might be the best choice for the lipophilicity prediction of bile acids and their derivatives and the most important aspect of the high linear correlation was considered, by some authors, as a possibility of finding congeneric classes within large groups of compounds [14, 37, 38].

The scatterplot of scores (R_M values) shows interesting results on CN$_{F254s}$, compared to RP-18$_{F254s}$ (Figure 5.6), and it appears clearly that the studied compounds form practically five different congeneric classes (from left to right) in a good agreement with their chemical structure.
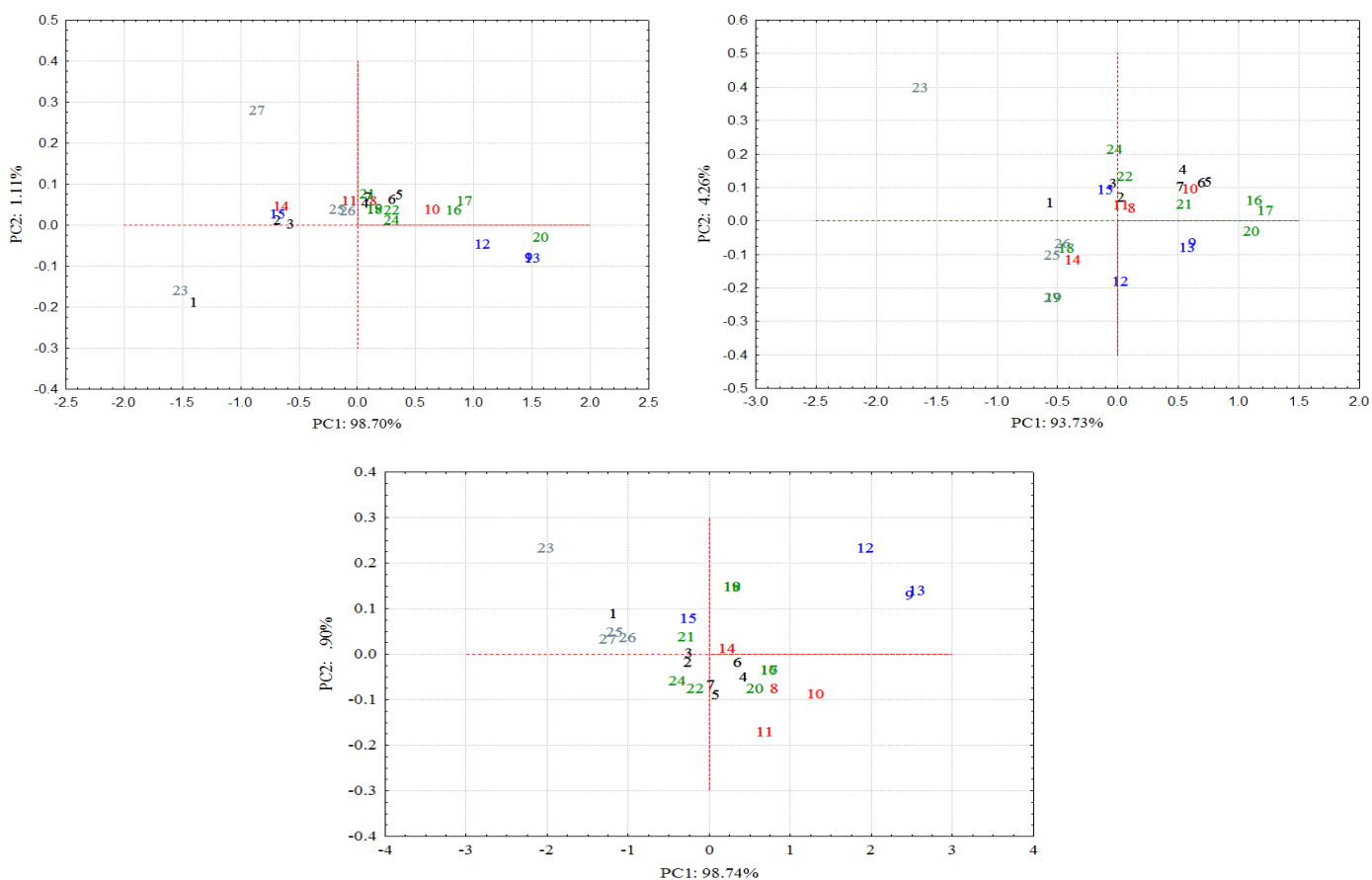
Figure 5.6. Lipophilicity Charts Corresponding to R_{M0} Values: (a) HPTLC Silicagel 60RP-18$_{F254s}$; (b) HPTLC Silicagel 60RP-18W$_{F254s}$; (c) HPTLC Silicagel 60CN$_{F254s}$

5.4.4. Concluding Remarks

The contribution of 2D and 3D descriptors which are related to atomic mass, together with reactivity parameters such as polarizability and electronegativity seem to control the chromatographic mechanism (lipophilicity) on all stationary phases.

5.5. Modeling of Chromatographic Lipophilicity of Bile Acids and Their Derivatives Estimated by Reversed-Phase High-Performance Liquid Chromatography

5.5.1. Experimental Part

The chromatography was performed on an Agilent 1100 Series LC system which was connected to an 1100 MSD mass spectrometer. The chromatographic behavior of the compounds was studied on: C18 (LiChroCART, LiChrosphere RP-18e, 125x4mm, 5µm), C8 (Zorbax, Eclipse XDB-C8, 150x4.6mm, 5µm) and CN (SAULENTECHNIK KNAUER LiChrosphere 100CN, 250x4mm, 5µm) columns.

The mobile phases used were methanol-water mixtures in different portion as in Table 5.18.

Table 5.18. Mobile Phase System Corresponding to Different Investigated Colums

Column	Mobile Phase MeOH (%) (v/v)		Addition Step (%)	
C18	30-35	38-40	1.25	0.50
C8	80-84		1.00	
CN	61-65		1.00	

5.5.2. Results and Discussion

The log *k* values of bile acids decrease linearly as the methanol volume fraction increases in all cases of columns.

The highest log k values were obtained on C8 and C18 columns compared to CN column. Highly significant correlations were obtained between different experimental indices of lipophilicity and computed Log P values, C8 column seems to be more suited for the estimation of lipophilicity.

The profiles of retention indices (log *k*) illustrate regular changes of retention factors for all three types of columns (Figure 5.9). These systematic regularities of retention observed might indicate that the same mechanism (lipophilic interactions) is dominant in all cases and no secondary mechanisms were highlighted.

In addition, the results obtained in this study applying PCA may be used in interpreting the molecular mechanism of interactions between eluents and columns with different polarities and to explain the chromatographic behavior of compounds.
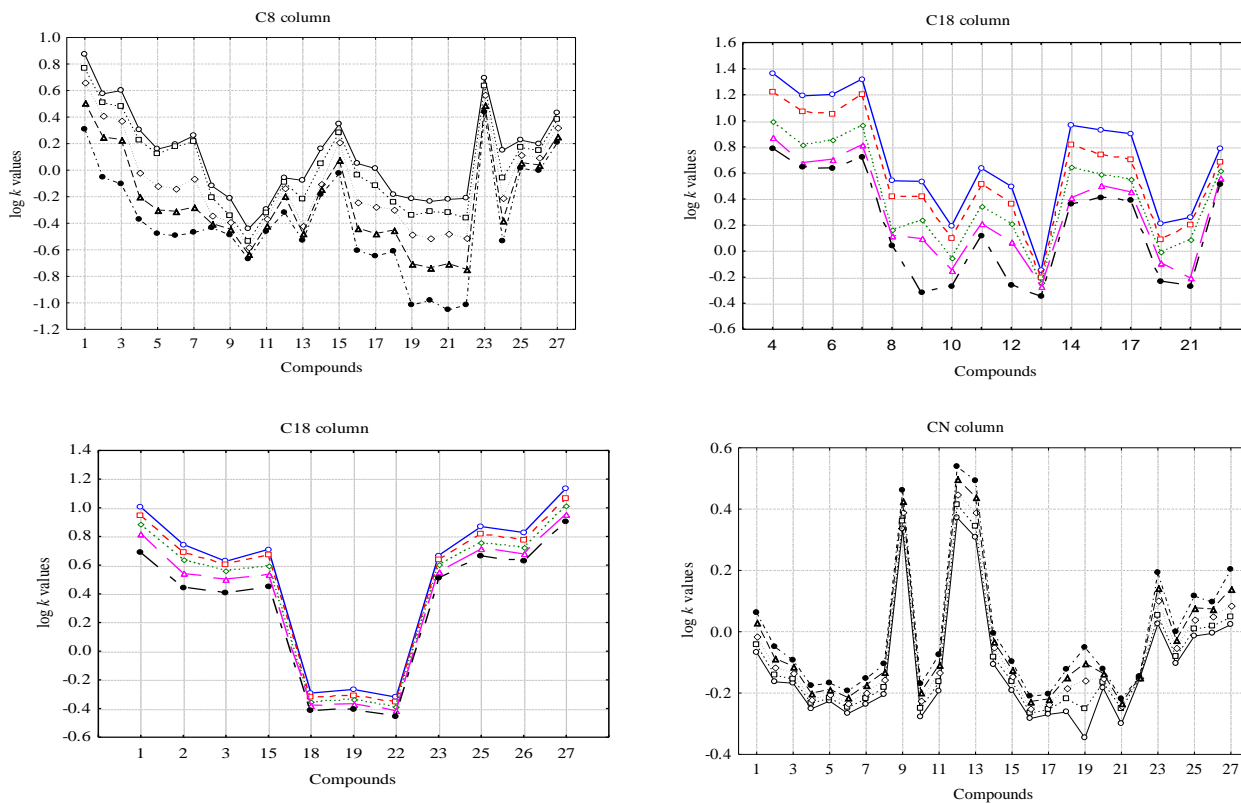
Figure 5.9. The Relationship Between Retention (log *k*) and Mobile Phase Composition.

Table 5.28. Chromatographic Indices Estimated from Retention Data.

Cpd.	C18 column					
	$\log k_{\text{w}}$	S	φ_{o}	$PC1_k$	$PC1_{\text{Log }k}$	R
1	6.821	-0.076	-89.398	-7.941	-1.004	0.9797
2	6.477	-0.075	-86.132	-0.032	-0.425	0.9877
3	4.763	-0.054	-88.031	1.650	-0.263	0.9689
4	4.938	-0.060	-82.567	-20.578	-1.360	0.9870
5	4.734	-0.059	-79.828	-10.399	-0.984	0.9728
6	4.728	-0.059	-80.132	-10.735	-1.008	0.9886
7	5.158	-0.064	-80.721	-18.113	-1.266	0.9906
8	3.635	-0.052	-70.037	6.235	0.407	0.9646
9	5.459	-0.081	-67.481	6.322	0.550	0.9672
10	2.991	-0.047	-64.195	8.800	1.068	0.9979
11	3.838	-0.053	-71.882	4.924	0.167	0.9953
12	4.876	-0.072	-67.438	6.740	0.590	0.9778
13	0.955	-0.018	-51.897	9.907	1.540	0.9900
14	4.854	-0.065	-74.901	-1.477	-0.445	0.9872
15	5.683	-0.065	-87.160	0.479	-0.383	0.9929
16	3.932	-0.051	-77.548	-0.415	-0.432	0.9829
17	3.908	-0.051	-76.770	0.421	-0.355	0.9779
18	2.023	-0.030	-66.546	8.639	1.729	0.9969
19	2.295	-0.034	-68.113	8.602	1.697	0.9987
20	2.754	-0.042	-64.946	8.691	0.998	0.9962
21	3.839	-0.059	-65.287	8.339	0.952	0.9675
22	2.386	-0.027	-88.691	8.716	1.809	0.9894
23	2.138	-0.032	-65.994	0.706	-0.385	0.9995
24	3.687	-0.040	-93.104	1.337	-0.421	0.9949
25	4.765	-0.051	-92.891	-3.753	-0.765	0.9997
26	4.554	-0.049	-92.935	-2.611	-0.681	0.9998
27	5.422	-0.057	-95.959	-14.455	-1.329	0.9977

Table 5.29. Chromatographic Indices Estimated from Retention Data.

Cpd.	C8 column					
	$\log k_{\text{w}}$	S	φ_{o}	$PC1_k$	$PC1_{\text{Log }k}$	R
1	11.988	-0.139	-86.494	-8.449	-1.661	0.9888
2	12.745	-0.151	-84.235	-3.179	-1.022	0.9548
3	13.934	-0.166	-83.886	-3.079	-0.969	0.9663
4	14.563	-0.178	-81.904	-0.124	-0.236	0.9917
5	13.777	-0.170	-81.280	0.571	0.012	0.9842
6	15.151	-0.186	-81.368	0.424	-0.004	0.9778
7	15.925	-0.195	-81.667	0.112	-0.109	0.9847
8	6.491	-0.083	-78.393	1.521	0.392	0.9763
9	4.957	-0.065	-76.263	1.742	0.555	0.9712
10	3.912	-0.055	-71.512	2.113	0.992	0.9832
11	3.129	-0.043	-73.276	1.777	0.545	0.9938
12	4.958	-0.062	-79.462	1.131	0.073	0.9603
13	9.207	-0.117	-79.029	1.548	0.494	0.9677
14	7.165	-0.088	-81.514	0.521	-0.181	0.9662
15	7.963	-0.095	-83.908	-0.956	-0.676	0.9900
16	13.793	-0.171	-80.520	1.125	0.305	0.9935
17	13.456	-0.168	-80.192	1.298	0.408	0.9971
18	8.366	-0.106	-78.625	1.633	0.523	0.9696
19	15.547	-0.196	-79.198	1.879	0.974	0.9810
20	15.126	-0.191	-79.109	1.893	0.981	0.9860
21	16.255	-0.205	-79.291	1.872	0.982	0.9735
22	15.767	-0.199	-79.154	1.901	1.012	0.9899
23	5.949	-0.066	-90.540	-5.795	-1.544	0.9988
24	13.634	-0.169	-80.768	0.921	0.198	0.9984
25	4.534	-0.054	-84.117	-0.256	-0.547	0.9992
26	4.293	-0.051	-83.852	-0.099	-0.499	0.9993
27	5.029	-0.057	-87.611	-2.048	-0.997	0.9985

Table 5.30. Chromatographic Indices Estimated from Retention Data.

Cpd.	CN column					
	$\log k_w$	S	φ_o	$PC1_k$	$PC1_{\log k}$	R
1	2.093	-0.033	-62.838	-0.012	-0.129	0.9942
2	1.666	-0.028	-59.092	0.473	0.106	0.9948
3	1.074	-0.019	-55.953	0.567	0.155	0.9970
4	0.943	-0.018	-51.266	0.857	0.343	0.9942
5	0.718	-0.015	-49.164	0.808	0.306	0.9898
6	0.900	-0.018	-50.274	0.902	0.375	0.9983
7	1.151	-0.021	-53.766	0.789	0.296	0.9987
8	1.437	-0.025	-56.794	0.645	0.205	0.9998
9	2.401	-0.032	-75.503	-3.391	-1.025	0.9980
10	1.468	-0.027	-54.757	0.872	0.358	0.9999
11	1.717	-0.029	-58.395	0.562	0.158	0.9994
12	3.089	-0.042	-73.902	-4.268	-1.159	0.9986
13	3.322	-0.047	-71.439	-3.464	-1.027	0.9975
14	1.549	-0.026	-60.741	0.237	-0.019	0.9971
15	1.280	-0.023	-56.650	0.606	0.180	0.9971
16	0.908	-0.018	-49.612	0.950	0.411	0.9982
17	0.813	-0.017	-48.946	0.920	0.387	0.9997
18	2.013	-0.035	-57.673	0.744	0.274	0.9963
19	4.472	-0.074	-60.520	0.666	0.258	0.9889
20	0.808	-0.015	-53.125	0.642	0.198	0.9988
21	0.886	-0.018	-49.217	0.955	0.415	0.9214
22	-0.053	-0.002	35.533	0.648	0.196	0.9862
23	2.776	-0.042	-65.476	-0.679	-0.376	0.9965
24	1.591	-0.026	-60.946	0.224	-0.024	0.9994
25	2.116	-0.033	-64.319	-0.296	-0.245	0.9951
26	1.682	-0.026	-64.938	-0.289	-0.247	0.9977
27	2.949	-0.045	-65.246	-0.667	-0.369	0.9849

Legend: $\log k_w$ -isocratic k value for pure water; S -solvent strength of organic modifier; φ_o - hydrophobicity index; $PC1_k$, $PC1_{\log k}$ -scores corresponding to first principal component on k and $\log k$ values

The log k values of bile acids (BA) decrease linearly in all cases of RP-HPLC as the methanol concentration increases. The highest log k values were obtained on C8 and C18 columns compared to CN column, and it can be observed that the number of hydroxyl, keto- and diacetoxy groups, as well as their position and orientation determines the chromatographic behavior of BA and their derivatives.

The chromatographic behavior of the investigated compounds is in a very good agreement with their polarity (Table 5.28-5.30) as can be easily observed from the profiles of retention indices presented in Figure 5.9. By carefully examining the patterns the similarity and differences between the bonded phases investigated can be clearly observed.

Studies within the bile salts demonstrate that HPLC mobility, which correlates with hydrophobicity, was markedly influenced by both position and orientation, in addition to number of hydroxyl and oxo-functions, in that mobility decreased in the order C>CDC>DC>LC. The cholic acid molecule has the largest planar polarity since its β side of the molecule is separated from the hydrophilic α side. For its mono- and diketo- derivatives planar polarity decreases because the β side of the molecule becomes more hydrophilic (less hydrophobic) due to shift of the oxo group toward steroid skeleton mean plane. Moreover, partial inversion of polarity occurs for DC acid because β side of the molecule becomes more polar because of displacement of the oxygen atoms at C3, C7 and C12 oxo group, and α side becomes less polar (more hydrophobic) due to appearance of the hydrophobic island on the α side of the steroid skeleton.

The eigenvalues obtained by applying PCA show that the first principal component accounts for 99.65% (k) and 97.10% (log k) of the total variance in the case of C18, 97.37% (k) and 94.29% (log k) for C8, and 99.72% (k) and 99.20% (log k) for CN column, respectively.

Among the Log P values, the most similar to the experimental partition coefficients were those obtained on C8 and C18 columns (higher than 0.7) (Table 5.32). The highest compatibility of experimental log k_w values was found with:

- CSLogP, LogPcSciLogP, miLogP, AB/LogP and MolLogP on C8 (values between 0.80-0.90)

- miLogP, LogPcSciLogP, LogPCD, VirtualLogP, XLOGP2, AvLogP and cLogP on C18 (values between 0.80-0.85) and as for CN column worse corelations were obtained.

Among theoretical values of partition coefficients, CSLogP and miLogP correlate better with log k_w on C8 and C18 columns.

The hydrophobicity index, φ_0, appeares to be the best solution for the lipophilicity scale resulted from retention data, in all cases the values being > 0.85. Comparison of these calculation procedures reveal that the most appropriate Log P values for bile acids chromatography are the ones which combine additive atomic contributions, atom-type electrotopological-state (E-state), neural network modeling indices and group contributions.

It is interesting to observe that the congeneric series of compounds form practically five different congeneric classes (Figure 5.10 (a-c) in a good agreement with their chemical structure: diacetoxy- (23, 25-27), oxo-derivatives (16-22, 24), primary and secondary bile acids (1-7), and finally the glyco- (8, 10, 11, 14) and tauro- (9, 12 13, 15) conjugates. The position of each compound within the graphs is also in a good agreement with the position and orientation of hydroxyls and the presence of polar groups –COOH, –SO$_3^-$, -C=O and -OCOCH$_3$, respectively (Figure 5.10).

The best models yield a determination coefficient over 0.70 in the case of C8 (Eq. 5.16-5.20) and C18 (Eq. 5.21-5.25) columns (Tables 5.33), which seem to be adequate for the estimation and characterization of bile acids lipophilicity as follows:

Table 5.33. Predictive Models Obtained for C8 and C18 Columns.

Column	Equation	Eq. No.
C8	$\log k_w = 95.591 - 366.036\textbf{\textit{PW3}} - 13.409\textbf{\textit{Mor32m}} + 93.994\textbf{\textit{HATS6e}}$	(5.16)
	$S = -1.179 + 4.514\textbf{\textit{PW3}} + 0.161\textbf{\textit{Mor32m}} - 1.159\textbf{\textit{HATS6e}}$	(5.17)
	$\varphi_0 = -438.882 + 0.504D/Dr06 + 0.827RDF130e + 507.235REIG$	(5.18)
	$PC1_k = -42.595 - 6.605ASP - 2.824MLOGP2 - 17.193BLTA96$	(5.19)
	$PC1_{\log k} = 67.201 - 101.496X0Av + 12.012MATS4e - 0.135Tp$	(5.20)
C18	$\log k_w = 10.747 - 4.934\textbf{\textit{EEig15x}} + \textbf{0.229}RDF075p + 7.040\textbf{\textit{Mor20}}$	(5.21)
	$S = 1.429 + 0.046\textbf{\textit{EEig13d}} - 0.123ESpm09u - 0.038\textbf{\textit{Mor20u}}$	(5.22)
	$\varphi_0 = -262.220 - 1.626\textbf{\textit{RDF045p}} - 26.937\textbf{\textit{Mor20u}} - 35.679\textbf{\textit{Mor05v}}$	(5.23)
	$PC1_k = -70.192 + 0.438TI1 - 32.019\textbf{\textit{Mor20p}} - 6.396H\text{-}047$	(5.24)
	$PC1_{\log k} = -7.523 - 0.869\textbf{\textit{Mor05m}} - 3.295\textbf{\textit{Mor20v}} + 1.664Hypertens\text{-}80$	(5.25)

The regression equations (Eq.5.26-5.30) obtained for CN column present lower determination coefficients (Table 5.34):

Table 5.34. Predictive Models Obtained for CN Column.

Column	Equation	Eq.No.
CN	$\log k_w = -151.228 - 1.038MAXDP + 154.910PCR - 3.271\textbf{\textit{Infective-80}}$	(5.26)
	$S = 0.811 - 0.055\textbf{\textit{EEig05d}} + 0.209VEA1 + 0.046\textbf{\textit{Infective-80}}$	(5.27)
	$\varphi_0 = -95.764 - 1.605RDF100u + 1.876RDF050m + 1025.198R1v+$	(5.28)
	$PC1_k = 35.460 - 17.857\textbf{\textit{MATS4e}} - 11.623\textbf{\textit{EEig03d}} + 38.407G1p$	(5.29)
	$PC1_{\log k} = -45.996 + 46.595Me + 12.029\textbf{\textit{MATS1v}} - 0.577Mor26m$	(5.30)

Predictive models with three of the most contributing descriptors were ascertained with statistical parameters in order to establish the correlation of lipophilicity indices of the studied compounds with their structural and physicochemical properties (Table 5.35).

Table 5.35. Statistical Quality Parameters of Multiple Regression Models.

Column	Variable	Q^2	R^2	s	PRESS	F
C8	$\log k_w$	0.9236	0.9402	1.204	42.555	120.5
	S	0.9100	0.9297	0.016	0.008	101.5
	φ_o	0.7636	0.8311	1.749	98.516	37.7
	$PC1_k$	0.7653	0.8158	1.160	39.411	33.9
	$PC1_{\log k}$	0.6527	0.7535	0.410	5.449	23.4
C18	$\log k_w$	0.7580	0.8230	0.632	12.557	35.6
	S	0.6691	0.7649	0.008	0.002	24.9
	φ_o	0.7084	0.7669	5.895	999.991	25.2
	$PC1_k$	0.6621	0.7394	4.724	665.536	21.8
	$PC1_{\log k}$	0.7528	0.8170	0.454	6.4	34.2
CN	$\log k_w$	0.6354	0.7284	0.544	9.127	20.6
	S	0.5723	0.7292	0.008	0.002	20.6
	φ_o	0.1436	0.3405	16.949	8578.822	4.0
	$PC1_k$	0.8800	0.9073	0.460	6.135	75.1
	$PC1_{\log k}$	0.8679	0.9131	0.141	0.697	80.6

Legend: determination coefficient, R^2; leave-one-out crossvalidation coefficient, Q^2; standard error, s; predictive error sum of squares, PRESS; Fisher test, F.

On the basis of resulted correlations, it may be appreciated that the lipophilicity indices determined on C8 and C18 columns might be the best choice for the lipophilicity prediction of bile acids and their derivatives. If we prove the high linear correlation then the retention factor can be used to conduct predictive analytics.

The contribution of 2D and 3D descriptors which are related to atomic mass and volumes, together with reactivity parameters such as polarizability and electronegativity seem to control the chromatographic mechanism (lipophilicity) on all columns.

Table 5.35. Statistical Quality Parameters of Multiple Regression Models.

Column	Variable	Q^2	R^2	s	PRESS	F
C8	$\log k_w$	0.9236	0.9402	1.204	42.555	120.5
	S	0.9100	0.9297	0.016	0.008	101.5
	φ_o	0.7636	0.8311	1.749	98.516	37.7
	$PC1_k$	0.7653	0.8158	1.160	39.411	33.9
	$PC1_{\log k}$	0.6527	0.7535	0.410	5.449	23.4
C18	$\log k_w$	0.7580	0.8230	0.632	12.557	35.6
	S	0.6691	0.7649	0.008	0.002	24.9
	φ_o	0.7084	0.7669	5.895	999.991	25.2
	$PC1_k$	0.6621	0.7394	4.724	665.536	21.8
	$PC1_{\log k}$	0.7528	0.8170	0.454	6.4	34.2
CN	$\log k_w$	0.6354	0.7284	0.544	9.127	20.6
	S	0.5723	0.7292	0.008	0.002	20.6
	φ_o	0.1436	0.3405	16.949	8578.822	4.0
	$PC1_k$	0.8800	0.9073	0.460	6.135	75.1
	$PC1_{\log k}$	0.8679	0.9131	0.141	0.697	80.6

Figure 5.10. Lipophilicity Charts Corresponding to log *k* Values C8, C18 and CN Column.

Chapter 6

MODELING OF CHROMATOGRAPHIC INDICES OF QUATERNARY AMMONIUM AND NITRONE DERIVATIVES AND THEIR THIAZOLIC SALTS

Introduction

Drug substances with quaternary ammonium structure are present in antispasmodic drugs which have affinity for the nervous system, thus it regulates the flow of blood in the arteries, expels urine from the urinary bladder, and it also acts upon the gastrointestinal tract and genital system. These were the reasons for which these compounds were biologically investigated on an isolated guinea pig ileum, using acetylcholine as contractile agent for an antispasmodic mechanism. All compounds proved a slow antispasmodic effect together with the increase of acetylcholine.

As regarding the thiazolic quaternary ammonium and nitrone salts the influence upon the antimicrobial activity was investigated, the nature of the base increasing the activity, being favorable to pyridine; the presence of the iodine or the methyl group on the benzene nuclei favors the antimicrobial activity [19-22].

One can conclude that the presence of ionic link that can be considered a pharmacophore group is favorable to compounds having an average or good antimicrobial effect. In conclusion, according to our anticipation, the presence of ionic linkage is important to a good antimicrobial activity. The compounds exerting the highest activity upon gram-positive microorganisms have a pyridine rest in the thiazol 4 positions [23].

The goal of this study was to estimate the lipophilicity and to investigate the molecular mechanism of retention and to find an objective manner of quantitative comparison of chemically bonded stationary phases for high performance thin layer chromatography (HPTLC) in terms of their (dis)similarities for 15 carefully designed, structurally diverse quaternary ammonium and nitrone derivatives and their thiazolic salts with distinctly distinguished functional groups.

6.1. Experimental Part

6.1.1. Chemicals and Reagents

The 15 quaternary ammonium and nitrone derivatives and their thiazolic salts (Table 6.1) were synthesized in the Department of Chemistry, Section of Organic Chemistry, Faculty of Farmacy, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca (Romania).

Table 6.1. Chemical Structures of Quaternary Ammonium and Nitrone Derivatives and Their Thiazolic Salts

Cpd.	R	R', R''
1	H_3C—(4-methylphenyl)—	Cl^- —CH_2—N^+(pyridinium)
2	Cl—(4-chlorophenyl)—	I^- —CH_2—N^+(pyridinium)
3	H_3C—(4-methylphenyl)—	I^- —CH_2—N^+(pyridinium)
4	H_3C—(4-methylphenyl)—	I^- —CH_2—N^+(2-methylpyridinium), CH_3
5	H_3C—(4-methylphenyl)—	—H_2C—N^+(quinolinium) I^-
6	H_3C—(4-methylphenyl)—	I^- —H_2C—N^+(diazabicyclo), N, N
7	H_3C—(4-methylphenyl)—$N(CH_3)$—C(=O)—CH_3	I^- —H_2C—N^+(diazabicyclo), N, N

8		
9		
10		R''= Br
11		
12		
13		R''= Br
14		
15		

6.1.2. Chromatography

Chromatographic measurements were carried out on 10x10 cm HPTLC different stationary phases as it follows:

1) HPTLC Silicagel 60RP-18$_{F254s}$;
2) HPTLC Silicagel 60RP-18W$_{F254s}$;
3) HPTLC Silicagel 60CN$_{F254s}$;
4) HPTLC Silicagel 60NH$_{2F254s}$ all being provided by Merck (Darmstadt, Germany).

The mobile phases used were methanol-water mixtures of various compositions. The concentration of organic solvent in the mobile phase ranged as according to Table 6.2, in each case the addition step being 10%.

Table 6.2. Mobile Phase System Corresponding to Different Investigated Stationary Phases.

Stationary Phase	Mobile Phase MeOH (%) (v/v)	Addition Step (%)
RP-18$_{F254s}$	30-45	
RP-18W$_{F254s}$	25-45	10
CN$_{F254s}$	25-45	
NH$_{2F254s}$	5-25	

The compounds investigated were separately dissolved in methanol (1 mg/ml) and 2 μL of each solution were spotted on the plates. Chromatograms were developed by the ascending technique at room temperature with previous saturation of the chamber with solvent for 15 minutes, the developing distance being 8 cm in each case.

The plates were dried at room temperature, and grey, brown or yellow spots appeared on a colorless background at 254 nm, and white or fluorescent spots appeared on a violet background at 365 nm.

6.2. Molecular Descriptors

One experimental data set consisting of 15 thiazolic quaternary ammonium salts was characterized by 1263 theoretical descriptors calculated using the software Dragon 5.4 [24], and 15 descriptors calculated using ChemDraw Ultra Plus 9.0 [25]; due to the difficulties encountered in the case of ionic molecules, only these two software programs allowed the drawing and calculation of several descriptors, which included parameters of all types such as constitutional, topological, geometrical, and quantum mechanical.

The model significance obtained in this work, with the exclusion of redundant and noisy information, was analyzed by MobyDigs v.1.0 software [27] that calculated the regression models by using (GA) genetic algorithms to perform variable selection.

The Dragon descriptors employed in this study can be arranged in the following groups: descriptors 2D: 2D autocorrelations (529 descriptors); descriptors 3D: RDF (130 descriptors), 3D-MORSE (160 descriptors), GETAWAY (194 descriptors), WHIM (99 descriptors), geometrical properties (41 descriptors) and Randić molecular profiles (41 descriptors); others descriptors: functional groups (6 descriptors), atom-centered fragments (13 descriptors), molecular properties (12 descriptors), charge descriptors (14

descriptors), and constitutional properties (24 descriptors). The ChemDraw descriptors can be arranged in topological properties (7 descriptors) and fizico-chemical properties (8 descriptors) [28, 29].

6.3. Results and Discussion

the chromatographic indices estimated from retention data, those obtained on the RP-C18 and CN stationary phase being somewhat higher than those obtained on the RP-C18W plates, and much higher than NH_2.

The chromatographic behavior of the investigated compounds on the bonded phases used in this study is more or less similar and in a very good agreement with their polarity as can be easily observed from the profiles of retention indices presented in Figure 6.1-6.2. The patterns illustrate good regularities of retention factors on RP-C18 and CN stationary phase and these findings might indicate that the same mechanism (lipophilic interactions) is dominant in both cases and that CN phase carries similar properties to RP-C18 phase (see Figure 6.1).

It may be concluded that the structure of the quaternary ammonium and nitrone derivatives and their thiazolic salts as well as the methanol concentration in the mobile phase have a larger influence on the interactions of the compounds with the RP-C18 and CN stationary phase than with the RP-C18W, respectively NH_2 stationary phase. In the case of NH_2 phase, hydrogen bonding seems to be dominant.

Their retention results from the combination of the ionic interactions and hydrophobic contribution to the retention, which strongly depend on the nature and size of R' and R'' groups on the thiazolic ring. The more polar compound 11 which has the largest molecular size has low R_{Mo} values on all stationary phases. It can be observed that the contribution of two diverse halogen atoms in compounds 2 and 10 with respect to compounds 1 and 14 increases R_{Mo} values, this behavior being dictated by the increasing size of the halide orbitals which are interacting with the carbon orbital in the bond. However, the contribution of the heteroatom in compound 10 beside the presence of Br and I halogens is also dependent on its volume as can be observed from the smaller contribution difference in compound 14, in which Br atom misses.

The statistics obtained illustrated that the linear equations fits in a very good way the chromatographic data, the linear correlation coefficients being between 0.95 and 0.99 in the majority of cases (lowest regression coefficients were on RP-C18W).

The correlation between different lipophilicity indices is presented in Table 6.7. As expected, a highly significant correlation was found between intercept (R_{Mo}) and slope (*b*) for all stationary phases as compared with the correlation between these indices and φ_0. The scores corresponding to the first principal component corresponding to R_F and R_M values are better correlated with R_{Mo} and *b*. The highest correlation was obtained for NH_2 bonded phase.

The best models yield a determination coefficient over 0.95% in the case of isocratic hydrophobicity index, φ_o (Eq. 6.1-6.4) and the scores corresponding to the first principal component resulted from R_F values (Eq. 6.5-6.18), which appeared to be the best solutions for the lipophilicity scale on RP-C18, RP-C18W, CN and NH_2 stationary phases in this order as follows (Table 6.9):

Table 6.9. Predictive Models Obtained for Isocratic Hydrophobicity Index, φ_o and the Scores Corresponding to the First Principal Component on R_F Values.

Stat. Phase	Equation	No.
RP-18$_{F254s}$	φ_0 = -81.218 − 1.367*RDF025u* + 205.716**G3u** + 19.933*H4e*	(6.1)
RP-18W$_{F254s}$	φ_0 = 161.480 − 309.838***MATS3p*** + 8.061***RDF075m*** − 39.966***RDF060p***	(6.2)
CN$_{F254s}$	φ_0 = -341.647 + 75.764*ATS1v* − 32.595*EEig13x* − 181.423***qnmax***	(6.3)
NH$_{2F254s}$	φ_0 = 660.902 - 170.625*ATS8m* − 1296.946***Gs*** + 371.939***H5m***	(6.4)
RP-18$_{F254s}$	PC1$_{RF}$ = -1.209 − 0.552***MATS6m*** + 7.288***G3u*** − 1.118*R6v*	(6.5)
RP-18W$_{F254s}$	PC1$_{RF}$ = 0.705 − 0.865*BIC1* + 0.318*DISPe* − 1.273*G3u*	(6.6)
CN$_{F254s}$	PC1$_{RF}$ = -0.101 + 0.006*RDF080p* − 0.079***R3u*** + 3.027*R2e+*	(6.7)
NH$_{2F254s}$	PC1$_{RF}$ = -4.481 + 0.939***EEig06r*** + 14.129***Gs*** − 2.938***H6p***	(6.8)

Higher determination coefficients were obtained for R_{Mo}, molecular lipophilicity (over 0.96) in case of CN and NH$_2$ (Eq. 6.9-6.12) (Table 6.9):

Table 6.9. Predictive Models Obtained for R_{M0} Values.

Stat. Phase	Equation	No.
RP-18$_{F254s}$	R_{Mo} = -27.032 + 4.752*HNar* + 15.230*BELp2* + 2.849*Mor17m*	(6.9)
RP-18W$_{F254s}$	R_{Mo} = 2.876 − 0.254***RDF105v*** + 8.480*HATS6u* -39.393***R2u+***	(6.10)
CN$_{F254s}$	R_{Mo} = -3.891 + 3.224***GATS5p*** + 2.307*Mor22p* − 8.528***qnmax***	(6.11)
NH$_{2F254s}$	R_{Mo} = 5.826 − 4.393*MATS8m* − 1.763***GATS4e*** − 14.612***G3u***	(6.12)

As regarding the scores corresponding to the first principal component resulted from R_M values (Eq. 6.13-6.16) and slope, *b* (Eq. 6.17-6.20) same order of determination coefficients were found on RP-C18 and RP-C18W stationary phases (Table 6.10):

Table 6.10. Predictive Models Obtained for *b* and the Scores Corresponding to the First Principal Component on R_M Values.

Stat. Phase	Equation	No.
RP-18$_{F254s}$	*b* = 0.100 + 0.042***MATS6e*** − 0 .021*GATS4m* − 0.568***G3u***	(6.13)
RP-18W$_{F254s}$	*b* = 0.158 + 0.060***MATS7e*** − 0.072*GATS5p* − 0.242*HATS6e*	(6.14)
CN$_{F254s}$	*b* = 0.479 − 0.794*X1A* − 0 049***GATS6v*** − 0.133*SPP*	(6.15)
NH$_{2F254s}$	*b* = -0.011 + 0.097*Mor31e* − 1.158***G1u*** + 0.513***G3u***	(6.16)
RP-18$_{F254s}$	PC1$_{RM}$ = -0.491 + 0.512*ATS7m* − 15.151***G3u*** + 10.028*G3s*	(6.17)
RP-18W$_{F254s}$	PC1$_{RM}$ = -3.907 + 0.284***RDF105p*** + 0.576*Mor12e* + 47.109***R2u+***	(6.18)
CN$_{F254s}$	PC1$_{RM}$ = 5.135 − 1.963***GATS5e*** + 0.726*GGI4* − 1.953***R3u***	(6.19)
NH$_{2F254s}$	PC1$_{RM}$ = -8.533 + 1.809***EEig06r*** + 26.680***Gs*** − 5.647***H6p***	(6.20)

The most statistical significant QSRR models were obtained in the case of CN (R^2 = 98.90%) and NH$_2$ (R^2 = 96.66%) stationary phases.

Table 6.11. Statistical Quality Parameters of Multiple Regression Models.

Stationary phase	Lipophilicity Index	Q^2	R^2	s	PRESS
RP-18$_{F254s}$	R_{M0}	0.8218	0.8994	0.046	0.438
	b	0.7887	0.9109	0.004	0.001
	φ_0	0.9347	0.9653	1.613	54.000
	$PC1_{RF}$	0.8855	0.9541	0.043	0.052
	$PC1_{RM}$	0.8508	0.9121	0.139	0.316
RP-18W$_{F254s}$	R_{M0}	0.8218	0.8994	0.046	0.438
	b	0.7887	0.9109	0.004	0.001
	φ_0	0.9347	0.9653	1.613	54.000
	$PC1_{RF}$	0.8855	0.9541	0.052	0.043
	$PC1_{RM}$	0.8508	0.9121	0.139	0.361
CN$_{F254s}$	R_{M0}	0.9077	0.9592	0.193	0.923
	b	0.9424	0.909	0.004	0.000
	φ_0	0.9799	90.890	14.418	4187.000
	$PC1_{RF}$	0.9176	90.480	0.009	0.001
	$PC1_{RM}$	0.8485	0.0905	0.916	0.715
NH$_{2F254s}$	R_{M0}	0.9478	0.9666	0.100	0.172
	b	0.8427	0.9177	0.004	0.000
	φ_0	0.8991	0.9543	2.090	106.000
	$PC1_{RF}$	0.8994	0.9090	0.002	0.009
	$PC1_{RM}$	0.9313	0.9523	0.098	0.151

Our study demonstrated that 2D and 3D descriptors related to atomic mass, symmetry together with reactivity parameters such as polarizability and electronegativity seem to control the lipophilicity on all stationary phases; the maximal negative charge of the molecule on CN phase, and topological aspects of the molecule for NH_2 are decisive for retention.
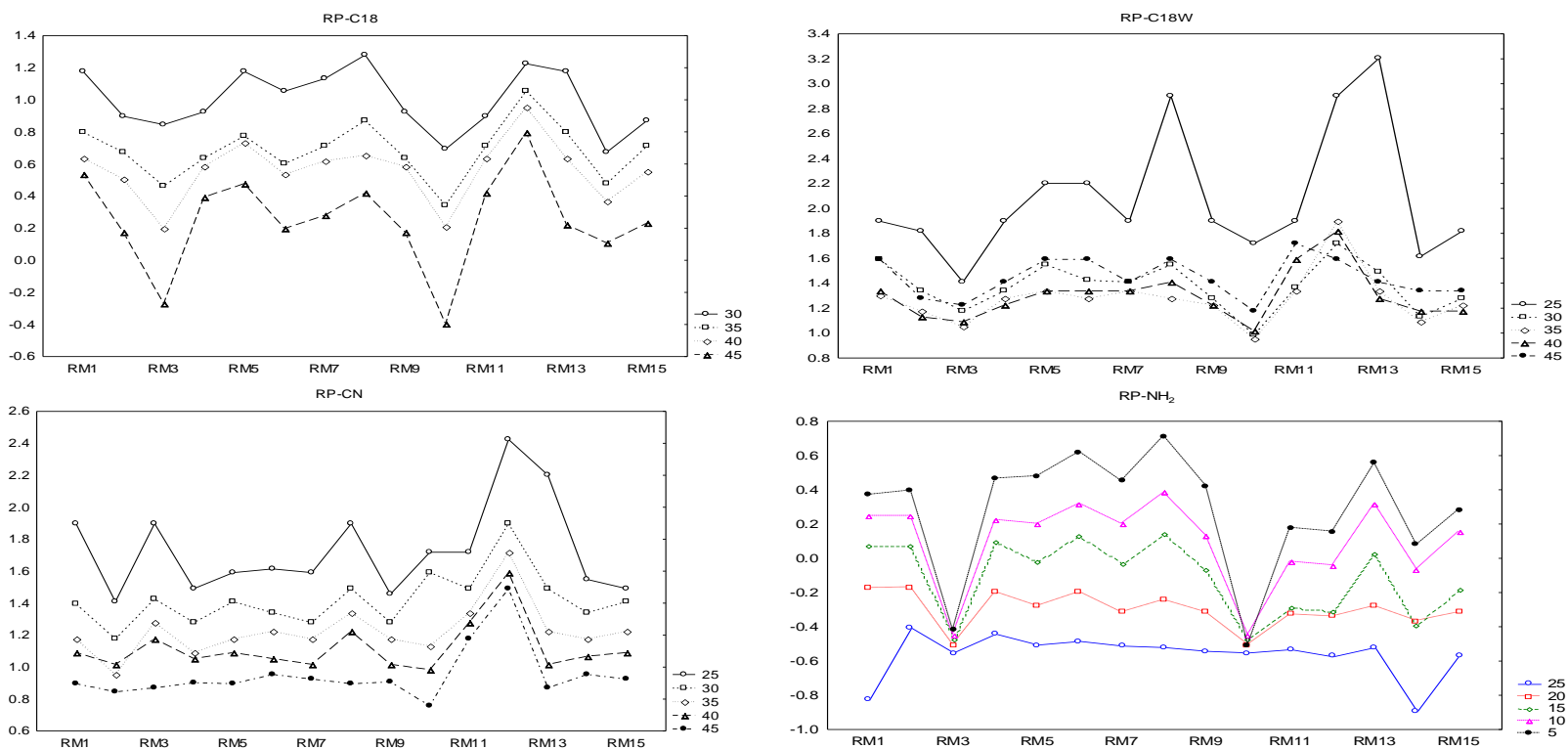
Figure 6.1. Profiles of R_M Values for All Fractions of Methanol on: RP-C18, RP-C18W, CN and NH$_2$ Stationary Phases.

Table 6.7. Correlation Matrix.

Var.	C18					C18W					CN					NH$_2$				
	R_{Mo}	b	φ_0	PC1$_{RF}$	PC1$_{RM}$	R_{Mo}	b	φ_0	PC1$_{RF}$	PC1$_{RM}$	R_{Mo}	b	φ_0	PC1$_{RF}$	PC1$_{RM}$	R_{Mo}	b	φ_0	PC1$_{RF}$	PC1$_{RM}$
R_{Mo}	1.00	-0.92	0.56	0.37	-0.24	1.00	-0.98	0.44	0.58	-0.94	1.00	-0.94	0.69	-0.37	-0.76	1.00	-0.98	-0.85	0.99	0.99
b		1.00	-0.81	-0.70	0.60		1.00	-0.52	-0.43	0.85		1.00	-0.88	0.04	0.50		1.00	0.86	-0.93	-0.93
φ_0			1.00	0.86	-0.86			1.00	-0.09	-0.20			1.00	0.39	-0.07			1.00	-0.82	-0.82
PC1$_{RF}$				1.00	-0.98				1.00	-0.80				1.00	0.87				1.00	1.00
PC1$_{RM}$					1.00					1.00					1.00					1.00

Chapter 7
MODELING OF MOLECULAR LIPOPHILICITY INDICES OF SOME FORMYL- AND ACETYLPYRIDINE-3-THIOSEMICARBAZONE DERIVATIVES

Introduction

In the last decades, considerable attention has been focused on thiosemicarbazones and on first row of transition metal complexes with such ligands due to their interesting biological activities [15, 16] e.g. anticarcinogenic, antibacterial, anti-HIV anticancer, fungicides, antiviral , antifungal , antitumour [17], anti-inflammatory, antiparasitic, antituberculosis [18], and antileukemic properties [19]. Of all the thiosemicarbazones studied so far, the 2-formyl- pyridinethiosemicarbazone (HPATS) probably got the most attention due to its marked antitumour properties [20].

Moreover palladium(II) complexes with nitrogen containing ligands are the subject of intensive biological evaluation in the search for less toxic and more selective anticancer therapies [21-23].
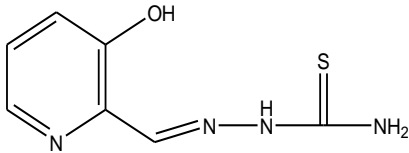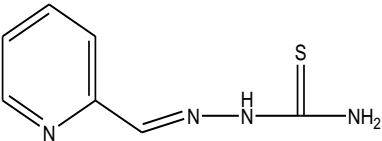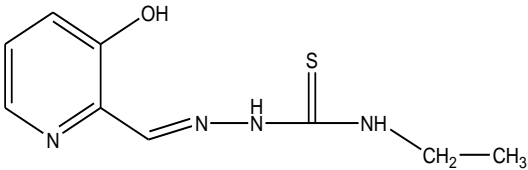
It had been reported that zinc complexes of thiosemicarbazones are antioxidant and have an effect in vitro on cell proliferation and differentiation [24].

7.1. A Comparative Study of Molecular Lipophilicity Indices of Some Formyl- and Acetylpyridine-3-Thiosemicarbazone Derivatives and Calculated Log *P* Values

7.1.1.1. Chemicals and Reagents

The formyl- and acetylpyridine-3-thiosemicarbazone derivatives and their palladium and zinc complexes are listed in Table 7.1.

Table 7.1. Structure of Formyl- and Acetylpyridine-3-Thiosemicarbazone Derivatives.

Compound	Chemical formula	Chemical structure
1	$C_7H_8N_4SO$	
2	$C_7H_8N_4S$	
3	$C_9H_{12}N_4SO$	

4	**C$_8$H$_{10}$N$_4$S**	
5	**C$_9$H$_{12}$N$_4$S**	
6	**C$_{13}$H$_{18}$N$_4$SO**	
7	**C$_{13}$H$_{18}$N$_4$S**	
8	**C$_{14}$H$_{20}$N$_4$S**	
9	**C$_{14}$H$_{20}$N$_4$S**	
10	**C$_{16}$H$_{18}$N$_6$SO**	
11	**C$_{16}$H$_{18}$N$_6$S**	

12	**$C_{17}H_{20}N_6S$**	

7.1.1. Experimental Part

7.1.1.1. Chemicals and Reagents

The formyl- and acetylpyridine-3-thiosemicarbazone derivatives and their palladium and zinc complexes (Table 7.1) were synthesized in the Department of Chemistry, Section of Inorganic and Analytical Chemistry, University of Ioannina, (Greece). The silica gel bounded plates were provided from Merck (Darmstadt, Germany).

7.1.1.2 Chromatography

The chromatographic behavior of the formyl- and acetylpyridine-3-thiosemicarbazone derivatives was studied on two stationary phases: RP-C18/UV$_{254s}$ (20X20 cm) and RP-C18W/UV$_{254s}$ (10x20 cm) silica gel bounded plates.

Table 7.2. Variation of Methanol in the Mobile Phase.

Stationary Phase	Mobile Phase MeOH (%) (v/v)	Addition Step (%)
RP-C18/UV$_{254s}$	50-90	10
RP-C18W/UV$_{254s}$	30-70	

The solutions (1μL) were applied manually to the origin of the plates by means of a 10μL Hamilton (Switzerland) microliter syringe. Chromatography was performed in a normal developing chamber at room temperature (~20^{0}C). Colored zones appeared on a colorless background and fluorescent blue-orange under UV lamp (λ = 254 nm).

7.1.2. Log *P* Computational Methods

The calculated log *P* values for the precursors of formyl- and acetylpyridine-3-thiosemicarbazone derivatives, can be roughly correlated with drug absorption [27], and were calculated using different Log *P* estimation computer programs: SciQSAR (LogP), SciLogP (LogPc) [28], Chem3D Ultra 8.0 (LogP, PartCoeff) and XLOGP (XLOGP) [29] (based on atom contributions), KOWWIN (KOWWIN) [30] (based on atom/fragment contributions), cLogP (cLogP) [31] (based on fragmental contributions), ALOGPS (ALOGPs, AB/LogP, miLogP, AvLogP, ALOGpS, IAlogS, AB/LogS, AvLogS) [32] and IAlogP (IAlogP) [33] (based on atom-type electrotopological-state indices and neural

network modeling). SMILES (Simplified Molecular Input Line Entry System) notation created by the structure-drawing program CambridgeSoft's (ChemDrawPro) is used as the chemical structure input. The results are shown in Table 7.3. Statistical analysis of the results was performed using the StatSoft 7.0 program [34].

7.1.3. Results and Discussion

As expected, the experimental data obtained revealed a linear relationship between retention and concentration of organic modifier in eluent. The isocratic hydrophobicity index, φ_o, was also calculated for each compound in both stationary phases, improving the model.

The profiles of retention indices (R_M values) illustrate regular changes of retention factors for both types of stationary phases (Figure 7.1). These systematic regularities of retention observed might indicate that the same mechanism (lipophilic interactions) is dominant in all cases on both stationary phases used.

The compounds enter into stronger interactions with the more nonpolar RP-C18 stationary phase and that aromatic radicals strongly influence the retention mechanism.
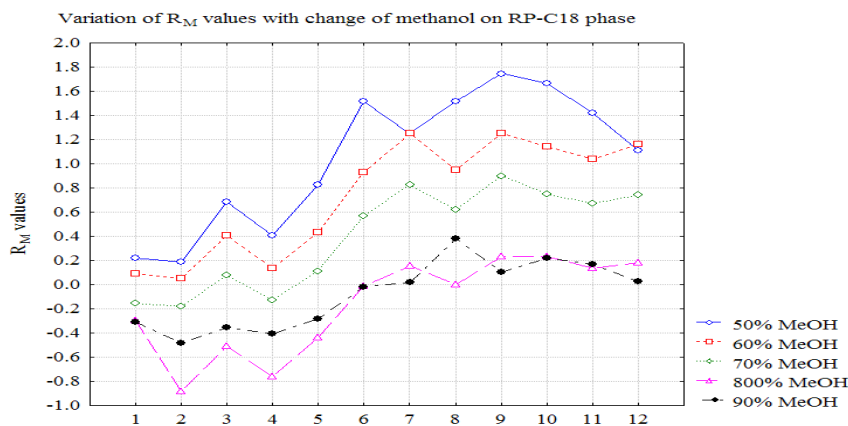
Figure 7.1. Relationship between Retention (R_M) and Mobile Phase Composition on RP-C18 Stationary Phase.
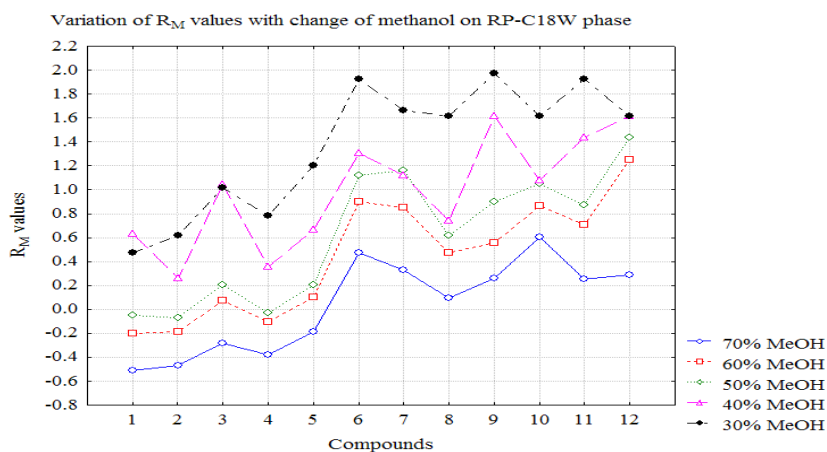
Figure 7.2. Relationship between Retention (R_M) and Mobile Phase Composition on RP-C18W Sationary Phase.

41

Table 7.6 shows that the correlation coefficients obtained on the RP-C18W stationary phase are somewhat higher than those obtained on the RP-C18 plates (see the R values) and higher retention coefficients were obtained for piperazinyl-, cyclohexyl-, and hexamethylenimine derivatives than for formyl- and acetyl derivatives. This means that the compounds enter into stronger interactions with the more nonpolar RP-C18 stationary phase and that aromatic radicals strongly influence the retention mechanism.

Also, from the variation in R_{Mo} on the RP-C18W stationary phase, it can be concluded that the structure of the formyl- and acetylpyridine-3-thiosemicarbazone derivatives as well as the methanol concentration in the mobile phase have a larger influence on the interaction of the compounds with the RP-C18W stationary phase than with the RP-C18 stationary phase according to their solubility in water.

The intercorrelations of the calculated log P values were also determined and the Log P values obtained with different softwares available on the internet are listed in Table 7.3 and these values seem to correlate strongly in between them, with slight differences (Table 7.9).

Statistically, highly significant correlations were found between lipophilicity indices, R_{Mo}, φ_o and PC1, and the calculated values by ALOGPs, AvLogP, XLOGP and LogP(ChemDraw) for RP-C18 plate and XLOGP and AvLogP for RP-C18W stationary phase. The scores corresponding to the first principal component (PC1) appeared to be the best solution for the lipophilicity scale resulted from the retention data, and the correlation matrix shows that these correlation coefficients are similar to those obtained for φ_o.

Comparison of these calculation procedures reveals ALOGPs, XLOGP and AvLogP as the most appropriates for chromatography.

7.1.4. Conclusions

In the present study, the relationship between lipophilicity parameters, which are important for cell membrane penetration as well as blood–brain barrier penetration of potential drug, studied experimentally by RP-HPTLC, and various calculated log P values of formyl- and acetylpyridine-3-thiosemicarbazone derivatives has been investigated.

Comparison of these calculation procedures reveals ALOGPs, XLOGP and AvLogP as the most appropriates for chromatography.

Table 7.3. Estimated Log*P* Values using Different Sofwares.

Cpd	LogP	LogPc	LogP (ChDraw)	PartCoeff	ALOGPs	AB/LogP	COSMOFrag	miLogP	KOWWIN	XLOGP	AvLogP	ALOGpS	AB/LogS	AvLogS
1	0.95	0.25	0.40	1.38	1.01	1.69	-0.13	0.83	1.31	0.08	0.80	-2.79	-2.77	-2.78
2	1.28	2.38	0.78	1.09	1.09	1.09	0.13	0.71	0.79	0.49	0.72	-2.81	-3.06	-2.94
3	1.27	2.77	1.25	1.51	1.69	3.05	1.34	1.58	2.27	1.02	1.82	-3.32	-2.68	-3.00
4	1.47	2.34	0.35	1.32	1.32	0.79	0.79	0.63	1.95	0.63	1.02	-3.07	-3.16	-3.11
5	1.53	1.75	1.64	1.11	1.72	2.45	1.69	1.46	1.75	1.43	1.75	-3.46	-2.69	-3.07
6	2.62	2.80	2.44	2.77	2.78	3.11	1.92	2.86	3.84	1.96	2.75	-3.62	-2.50	-3.06
7	2.46	3.69	2.83	2.48	2.83	2.51	2.22	2.74	3.32	2.37	2.67	-3.81	-2.91	-3.36
8	2.95	2.39	2.40	2.28	2.94	2.6	2.88	2.66	4.48	2.51	3.01	-3.80	-3.07	-3.44
9	2.97	3.04	2.42	3.24	3.12	3.15	3.98	2.9	4.69	2.98	3.47	-4.44	-3.49	-3.96
10	1.36	2.67	2.50	1.41	2.37	3.21	2.55	2.14	3.16	2.03	2.58	-3.44	-2.22	-2.83
11	2.93	2.54	2.89	1.12	2.43	2.61	2.80	2.02	2.64	2.43	2.49	-3.57	-2.99	-3.28
12	2.83	2.26	2.46	0.92	2.54	2.70	3.45	1.93	3.80	2.58	2.83	-3.70	-3.25	-3.47

Table 7.6. Regression Data of Studied Compounds for RP-C18 and RP-C18W Silica Gel Bounded Plates.

RP - HPLC phases	RP-C18						RP-C18W					
Cpd	R_{Mo}	b	φ_o	$PC1_{RF}$	$PC1_{RM}$	R	R_{Mo}	b	φ_o	$PC1_{RF}$	$PC1_{RM}$	R
1	0.92	-0.03	-31.89	-0.434	1.153	**0.966**	1.47	-0.06	-26.25	-0.523	1.414	**0.9304**
2	1.33	-0.05	-29.27	-0.600	1.486	0.8352	1.34	-0.05	-25.62	-0.547	1.480	**0.9878**
3	2.16	-0.06	-36.06	-0.277	0.726	**0.9442**	2.20	-0.07	-30.78	-0.241	0.646	**0.9519**
4	1.61	-0.05	-32.02	-0.495	1.227	0.8736	1.52	-0.06	-27.27	-0.464	1.273	**0.9738**
5	2.30	-0.06	-37.11	-0.217	0.568	**0.9449**	2.07	-0.07	-30.99	-0.222	0.657	**0.9712**
6	3.41	-0.08	-42.49	0.180	-0.518	**0.9712**	2.8	-0.07	-42.31	0.363	-1.018	**0.9776**
7	3.20	-0.07	-44.87	0.295	-0.715	**0.9567**	2.49	-0.06	-42.53	0.317	-0.767	**0.9534**
8	2.95	-0.06	-45.75	0.276	-0.662	0.8842	2.37	-0.07	-35.70	0.084	-0.054	**0.9317**
9	3.86	-0.09	-44.86	0.355	-1.083	**0.9859**	3.30	-0.09	-36.87	0.234	-0.820	**0.9875**
10	3.47	-0.08	-45.59	0.352	-0.949	**0.9725**	2.17	-0.04	-48.30	0.372	-0.777	**0.9516**
11	3.08	-0.07	-45.10	0.288	-0.671	**0.9698**	3.08	-0.08	-37.78	0.255	-0.779	**0.9885**
12	2.85	-0.06	-45.24	0.278	-0.562	**0.9534**	2.76	-0.06	-45.54	0.372	-1.255	0.8623

Table 7.9. Correlation Matrix for Both Stationary Phases.

Stationary phase	RP-C18/UV$_{F254s}$					RP-C18W/UV$_{F254s}$				
Computed estimators	R_{Mo}	b	φ_o	PC1$_{RF}$	PC1$_{RM}$	R_{Mo}	b	φ_o	PC1$_{RF}$	PC1$_{RM}$
LogP	**0.75**	**-0.67**	**-0.79**	**0.79**	**-0.78**	**0.86**	**-0.60**	-0.56	**0.75**	**-0.78**
LogPc	**0.70**	**-0.77**	**-0.53**	**0.53**	**-0.55**	**0.53**	**-0.24**	**-0.52**	0.59	-0.57
LogP($_{ChemDraw}$)	**0.92**	**-0.83**	**-0.96**	**0.95**	**-0.95**	**0.86**	**-0.36**	**-0.86**	**0.95**	**-0.94**
PartCoeff	**0.60**	**-0.62**	-0.43	0.48	-0.53	0.52	-0.44	-0.27	0.43	-0.42
ALOGPs	**0.95**	**-0.88**	**-0.94**	**0.94**	**-0.95**	**0.88**	-0.46	**-0.78**	**0.91**	**-0.90**
AB/LogP	**0.81**	**-0.77**	**-0.79**	**0.79**	**-0.80**	**0.78**	-0.42	**-0.73**	**0.79**	**-0.78**
COSMOFrag	**0.90**	**-0.82**	**-0.91**	**0.91**	**-0.91**	**0.89**	-0.53	**-0.74**	**0.85**	**-0.87**
miLogP	**0.92**	**-0.87**	**-0.89**	**0.90**	**-0.92**	**0.84**	-0.45	**-0.75**	**0.88**	**-0.86**
KOWWIN	**0.87**	**-0.79**	**-0.87**	**0.87**	**-0.88**	**0.81**	-0.43	**-0.71**	**0.82**	**-0.82**
XLOGP	**0.93**	**-0.85**	**-0.95**	**0.95**	**-0.95**	**0.90**	-0.50	**-0.78**	**0.90**	**-0.91**
AvLogP	**0.95**	**-0.88**	**-0.95**	**0.95**	**-0.96**	**0.91**	-0.50	**-0.79**	**0.91**	**-0.91**
ALOGpS	**-0.89**	**0.86**	**0.83**	**-0.84**	**0.86**	**-0.89**	**0.65**	**0.61**	**-0.77**	**0.79**
AB/LogS	0.01	-0.03	0.01	-0.01	0.01	-0.22	0.48	-0.25	0.09	-0.01
AvLogS	**-0.61**	**0.58**	**0.58**	**-0.59**	**0.60**	**-0.73**	**0.70**	0.29	-0.49	0.54
R_{Mo}	1.00	**-0.97**	**-0.92**	**0.93**	**-0.95**	1.00	**-0.73**	**-0.67**	**0.85**	**-0.88**
b		1.00	**0.81**	**-0.82**	**0.85**		1.00	-0.02	-0.26	0.32
φ_o			1.00	**-0.99**	**0.98**			1.00	**-0.95**	**0.93**
PC1$_{RF}$				1.00	**-1.00**				1.00	**-0.99**
PC1$_{RM}$					1.00					1.00

7.2. Development of QSAR Models of Molecular Lipophilicity of Some Formyl- and Acetylpyridine-3-Thiosemicarbazone Derivatives by Topological Descriptors

7.2.1. Experimental Part

7.2.1.1. Chemicals and Reagents
The set of compounds is presented in the previous chapter at the same chapter.

7.2.1.2. Chromatography
The chromatographic behavior of the formyl- and acetylpyridine-3-thiosemicarbazone derivatives was studied as presented in previous chapter at the same section.

7.2.2. Computed Topological Descriptors
The molecular structures of these molecules were drawn into HyperChem Program (HyperCube Inc.) [24] and optimized by using the MM+ molecular mechanics force field and then a more precise optimization is done by semiempirical AM1 method procedure.

The optimized geometries were loaded into the DRAGON Plus version 5.4 and TOPOCLUJ 3.0 software packages and in order to define the character of the compounds structure, the following descriptors were taken into consideration and used as independent variables.

We derived a total set of D= 246 topological descriptors from which 74 given by the software Dragon 5.4 [25], that included only the topological descriptors. Furthermore, 172 additional variables provided by TOPOCLUJ software package [26] were added to the pool and were calculated for every molecule.

7.2.3. Results and Discussion
In the present study, the relationship between chromatographic retention indices (R_{Mo}, b, and $PC1_{RF}$) studied experimentally by RP-HPTLC, and the calculated descriptors of the formyl- and acetylpyridine-3-thiosemicarbazone derivatives computed with Dragon and TOPOCLUJ software's has been investigated.

Good structure–retention index models show the efficiency of these indices in the structure–retention index correlations. Much higher correlation coefficients were obtained when the lipophilicity indices were estimated in dependence of the topological indices computed with TOPOCLUJ software program than with Dragon software and higher correlation coefficients were obtained for molecular lipophilicity on RP-C18 stationary phase compared to RP-C18W. The scores corresponding to the first principal component (PC1) appeared to be the best solution for the lipophilicity scale resulted from the retention data.

The shape of the molecule is an important index which should be taken into consideration because it plays a dominant role in the chromatographic behavior on both stationary phases with different polarities.

The best corresponding regression equations are shown in the Table 7.10 (Eq. 7.1-7.9) for both stationary phases and as well as for the two groups of topological descriptors used in this study.

Lower correlation coefficients were obtained for R_{Mo} and *b*, and much higher values for $PC1_{RF}$. The results suggest also that the 2- and 3-path Kier alpha-modified shape index (S2K and S3K) seem to be dominant in the retention mechanism and as a consequence they seem to control the lipophilicity in the case of RP-C18 stationary phase.

Lower correlation coefficients were obtained for RP-C18W phase and the regression equations are the following (Table 7.11) (Eq. 7.10-7.18):

Table 7.11. Regression Equations using Dragon Topological Descriptors on RP-C18W Stationary Phase.

Type of regression equation	Regression equation	Eq. no.
Multiple variable	R_{M0} = -5.435 − 0.136*SPI* + 0.631*PHI* + 10.000*PW2*	(7.10)
	b = 0.121 + 0.194*Xt* − 0.115*BLI* − 0.139*PJI2*	(7.11)
	$PC1_{RF}$ = 0.409 + 0.037*S2K* − 0.001*VAR* − 0.013*Lop*	(7.12)
Two variable	R_{M0} = -0.382 − 0.089*SPI* + 0.661*PHI*	(7.13)
	b = 0.101 − 0.118*BLI* − 0.052*PJI2*	(7.14)
	$PC1_{RF}$ = 0.434 + 0.029*S2K* − 0.012*Lop*	(7.15)
One variable	R_{M0} = -0.801 + 0.623*PHI*	(7.16)
	b = 0.029 − 0.095*BLI*	(7.17)
	$PC1_{RF}$ = 0.378 + 0.036*S2K*	(7.18)

As for RP-C_{18}W stationary phase, the descriptor that brings a slight higher contribution upon the retention mechanism is PW2-path/walk 2-Randic shape index related to the shape of the molecule.

As for the case when these lipophilicity indices where estimated in dependence of the topological descriptors computed by TOPOCLUJ 3.0 software program, the regression equations were the followings (Table 7.12 and 7.13) (Eq. 7.19-7.27 and Eq. 7.28-7.36):

Table 7.12. Regression Equations using TOPOCLUJ 3.0 Topological Indices on RP-C18 Stationary Phase.

Type of regression equation	Regression equation	Eq. no.
Multiple variable	R_{M0}= 10.454 − 3.622*C[LM[Density]]* + 0.005*VEA3* − 18.171*X[LM[Density]]*	(7.19)
	b = -0.091 + 0.00005*PDS7[LM[Density]]* 0.0002*VEA3* + 0.337*X[LM[Density]]*	(7.20)
	$PC1_{RF}$ = 0.428+ 0.0004*PDS8[LM[Density]]* − 0.0001*PDS8[LM[Mass]]* + 0.043*VAD2*	(7.21)
Two variable	R_{M0} = 11.353 − 3.918*C[LM[Density]]* − 18.405 *X[LM[Density]]*	(7.22)
	b = -0.067 − 0.0001*VEA3* + 0.209*X[LM[Density]]*	(7.23)
	$PC1_{RF}$ = 0.560 + 0.0003*PDS8[LM[Density]]* − 0.00003*PDS8[LM[Mass]]*	(7.24)
One variable	R_{M0} = 3.405 − 14.074*X[LM[Density]]*	(7.25)
	b = -0.075 + 0.208*X[LM[Density]]*	(7.26)
	$PC1_{RF}$ = 0.559 + 0.0002*PDS8[LM[Density]]*	(7.27)

Table 7.13. Regression Equations using TOPOCLUJ 3.0 Topological Indices on
RP-C18W Stationary Phase.

Type of regression equation	Regression equation	Eq. no.
Multiple variable	R_{M0} = -110.456 + 0.006*PDS8[LM[Density]]* − 0.016*PRD^2S[Sh[Detour]]* + 926.853*X[LM[Atomic radius]]*	(7.28)
	b = -0.997 − 0.0002*VRA1* + 0.000*W4[Atomic radius_Detour]* + 7.483*X[Sh[Conectivity]]*	(7.29)
	$PC1_{RF}$ = 0.395 − 0.0001*PDS6[Sh[Detour]]* + 0.0002*PDS8[LM[Density]]* − 0.001*VEA1* 0.154*VED3*	(7.30)
Two variable	R_{M0} = 2.073 + 0.006*PDS8[LM[Density]]* − 0.007*PRD^2S[Sh[Detour]]*	(7.31)
	b = -1.268 − 0.0001*VRA1* + 9.538*X[Sh[Conectivity]]*	(7.32)
	$PC1_{RF}$ = 0.544 + 0.0002*PDS8[LM[Density]]* − 0.001*VEA1*	(7.33)
One variable	R_{M0} = 1.476 + 0.002*PDS8[LM[Density]]*	(7.34)
	b = -0.747 + 5.357*X[Sh[Conectivity]]*	(7.35)
	$PC1_{RF}$ = 0.544 + 0.0002*PDS8[LM[Density]]*	(7.36)

Slight differences were observed comparing the correlation coefficients of the regression equations obtained using the two classes of software programs, but higher correlation coefficients were obtained for lipophilicity indices on RP-C_{18} stationary phase compared to the results obtained previously so it seems that the topological descriptors given by TOPOCLUJ 3.0 program are more confident in describing the retention mechanism.

It was also shown that score plots can be used to search for structural similarities within groups of compounds, since similar structures are grouped. The scatterplot of scores shows interesting results (Figure 7.5 and 7.6). Three clusters appear to be well defined and in a good agreement to the structure of compounds for both classes of descriptors computed: one of them corresponds to the compounds 1, 2, 3, 4, and 5 (formyl- and acetyl- derivatives), the second include the group of piperazinyl- derivatives (10, 11 and 12), and the third group, the hexamethylenimine- derivatives, (6, 7 and 8) with the exception of compound 9 (cyclohexyl- derivative).

Much more, the scatterplot given by TOPOCLUJ software shows a more compact classification of the compounds, compared to classification given by Dragon software, which states that these descriptors are similar in between them.

The results of these investigations also indicate that the topological descriptors computed by TOPOCLUJ 3.0 program are a most useful basis for study of QSAR/QSPR/QSRR for formyl- and acetylpyridine-3-thiosemicarbazone derivatives, which assures broad evaluations in these domains.

Table 7.10. Regression Equations using Dragon Topological Descriptors on RP-C18 Stationary Phase.

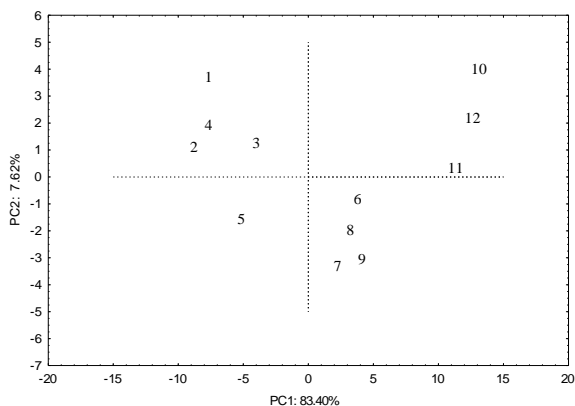| Type of regression equation	Regression equation	Eq. no.
Multiple variable	R_{M0} = -24.629 + 8.889$GNar$ + 6.683BLI + 3.174$PJI2$	(7.1)
	b = 0.047 − 0.024$S3K$ + 0.056$PW2$ − 0.509$PW5$	(7.2)
	$PC1_{RF}$ = 0.509 − 0.012STN + 0.027$S2K$ − 0.0262Lop	(7.3)
Two variable	R_{M0} = -17.177 + 6.889$GNar$ + 6.151BLI	(7.4)
	b = 0.075 − 0.024$S3K$ − 0.439$PW5$	(7.5)
	$PC1_{RF}$ = 0.498 + 0.021$S2K$ − 0.015Lop	(7.6)
One variable	R_{M0} = -13.869 + 8.270$GNar$	(7.7)
	b = 0.056 − 0.028$S3K$	(7.8)
	$PC1_{RF}$ = 0.429 + 0.028$S2K$	(7.9)

Figure 7.5. Scatterplot of Scores given by Dragon Descriptors.
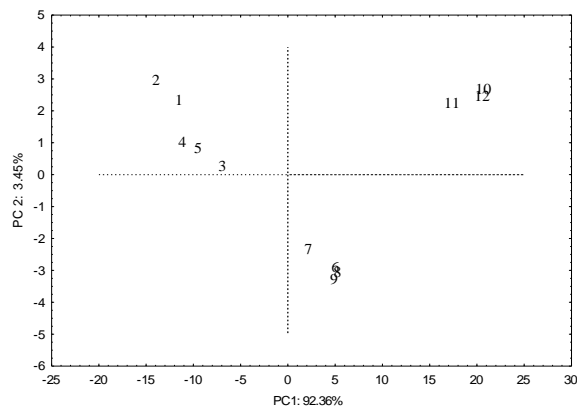
Figure 7.6. Scatterplot of Scores given by TOPOCLUJ Indices.

Chapter 8

LEL – A NEWLY DESIGNED MOLECULAR DESCRIPTOR

Introduction

A large number of structure–retention index correlations have been developed for different compounds but among these molecular descriptors, topological indices are of particular interest because they can be readily calculated directly from the molecular structures [8-12].

Using molecular graphs the chemical structure of a compound can be expressed by means of various graph matrices, polynomials, spectra, spectral moments, sequences counting distances, paths, and walks, or topological indices [13]. In general, the topological indices offer a simple way of coding molecular structure information into numerical values [14-16].

Alkanes represent an interesting class of compounds as a starting point for the application of molecular modeling procedures. Many properties of the alkanes vary function of molecular mass or branching, and alkanes can be described by using a single type of (carbon) atom. There are properties well accounted by a single molecular descriptor, e.g., octane number MON, entropy S, volume MV, refraction MR, etc. Other properties, such as, boiling point BP, heat of vaporization HV, total surface area TSA, partition coefficient Log P, density DENS, critical temperature CT, critical pressure CP and heat of formation DHF are notable exceptions, being not well modeled by any of the parameter sets.

The purpose of this work was to evaluate the relative performances of a pool of descriptors in relating the hydrocarbon molecular structures to a set of physical properties. . In this respect, the newly designed index LEL and those provided by the TOPOCLUJ software are of basic importance. The sets of studied molecules were selected among the representative and sufficiently complex structures (octane isomers and polycyclic aromatic hydrocarbons (PAH)) and the correlations used were in monovariate regression, in view of a more direct interpretation of the results.

8.1. Description of Indices

In any process of molecular modeling, either quantum or correlating one, the need for a representation of molecular structure is critical and its role is significant to find appropriate predictive models.

TIs are single number descriptors associated with a molecular graph representing a molecule, which does not depend on the numbering and pictorial representation of a molecular graph. In this section, definitions for LEL and the best scored TIs, among the indices provided by the TOPOCLUJ software, are presented.

8.1.1. LEL - An Index Built on the Laplacian Matrix

Let G=G(V,E) be a finite, undirected graph with n vertices V={1,2,...,n} and m=|E| edges. The degree of a vertex u in V is denoted by d_u. Let G have the adjacency matrix A with eigenvalues
, and Laplacian matrix L=D-A, where D is the diagonal matrix of vertex degrees, with eigenvalues

. The Laplacian-like energy, shortly LEL, of G is defined [8] as:

$$LEL = \sum_{i=1}^{n} \sqrt{\mu_i}$$

(8.1)

Similarly, among unicyclic graphs of order n, the star with an edge between two of its leaves has the minimum LEL, and the cycle C_n has the maximum LEL [10].

8.1.2. Walk Indices or Wiener-Type Indices of Higher Rank

Walks of length e, starting from the vertex $i \in V(G)$ can be counted by summing the entries in the ith row of the eth power of the adjacency matrix A:

$$^eW_i = \sum_{j \in V(G)} [A^e]_{ij}$$

(8.3)

^eW_i is called the *walk degree* (of rank e) of vertex i (or atomic walk count) [18, 19]. Local and global invariants based on walks in graph were considered for correlating with physico-chemical properties [19].

8.1.3. Indices Designed on Layer/Shell Matrices

Define the entries in the shell matrix (of pair vertex property) SM as:

$$SM_{i,k} = \underset{v|d_{i,v}=k}{\Omega} [M]_{i,v}$$

(8.13)

with the most used operation being the summation.

Shell matrix is a collection of the above defined entries:

$$SM = [SM]_{i,k}; i \in V(G); k \in [0,1,..,d(G)]$$

(8.14)

The TOPOCLUJ software package [7] is designed to calculate topological descriptors from topological matrices and/or polynomials. Several weighting schemes including group electronegativity, group mass and partial charges are enabled. Topological indices derived from the matrices: adjacency, connectivity, distance, detour, distance-path, detour-path, Cluj, their reciprocal matrices, walk-matrices, walk-operated matrices, layer- and shell-matrices were successfully used in correlating studies and graph discriminating analysis during the last decade [17,30]. The values of the best scored TIs for octane isomers are listed in Table 8.1.

Table 8.1. Topological Indices for Octanes.

Molecule	LEL	1WD	2WD	1WW	1WH	2WH	1WK	2WK	1/1WK	1/2WK	2WUCJD	PDS3[Sh[Sz]]
1	9.153	84	1848	84	13.743	48.279	10.564	29.040	0.095	0.034	1596	120.000
2	9.120	79	1628	79	14.100	51.050	10.862	31.153	0.092	0.032	1396	78.600
3	9.115	76	1512	76	14.267	52.495	10.981	32.125	0.091	0.031	1284	94.320
4	9.114	75	1476	75	14.317	52.947	11.014	32.411	0.091	0.031	1248	94.320
5	9.108	72	1360	72	14.483	54.377	11.133	33.373	0.090	0.030	1136	110.040
6	9.065	71	1316	71	14.767	56.500	11.433	35.426	0.087	0.028	1112	78.600
7	9.079	70	1280	70	14.733	56.317	11.367	35.024	0.088	0.029	1072	110.040
8	9.082		1312	71	14.650	55.560	11.300	34.454	0.088	0.029	1102	94.320

9	9.088	74	1420	74	14.467	53.939	11.167	33.343	0.090	0.030	1206	78.600
10	9.056	67	1176	67	15.033	58.878	11.633	37.107	0.086	0.027	978	110.040
11	9.074	68	1208	68	14.867	57.482	11.467	35.847	0.087	0.028	1004	125.760
12	9.073	67	1172	67	14.917	57.924	11.500	36.124	0.087	0.028	968	141.480
13	9.049	64	1072	64	15.250	60.792	11.800	38.493	0.085	0.026	880	125.760
14	9.023	63	1032	63	15.417	62.042	11.967	39.621	0.084	0.025	850	78.600
15	9.031	66	1128	66	15.167	59.771	11.767	37.911	0.085	0.026	940	141.480
16	9.020	62	1000	62	15.500	62.799	12.033	40.191	0.083	0.025	820	125.760
17	9.044	65	1096	65	15.167	59.889	11.733	37.791	0.085	0.026	906	141.480
18	8.971	58	868	58	16.000	67.000	12.500	43.750	0.080	0.023	706	125.760

The inter-correlation of indices is presented in Table 8.3 while in Table 8.4 that of the properties of the octane isomers.

Table 8.3. Intercorrelation Matrix for the Best Scored Indices in Ooctanes.

Variable	LEL	1WD	2WD	1WW	1WH	2WH	1WK	2WK	1/1WK	1/2WK	2WUCJD	PDS3 [Sh[SZ]]
LEL	1.00	0.96	0.95	0.96	-0.99	-0.98	-0.99	-0.99	0.99	0.98	0.94	**-0.34**
1WD		1.00	1.00	1.00	-0.99	-0.99	-0.98	-0.98	0.99	0.99	1.00	**-0.43**
2WD			1.00	1.00	-0.98	-0.98	-0.97	-0.97	0.98	0.99	1.00	**-0.42**
1WW				1.00	-0.99	-0.99	-0.98	-0.98	0.99	0.99	1.00	**-0.43**
1WH					1.00	1.00	1.00	1.00	-1.00	-1.00	-0.98	**0.40**
2WH						1.00	1.00	1.00	-1.00	-0.99	-0.98	**0.41**
1WK							1.00	1.00	-1.00	-0.99	-0.97	**0.38**
2WK								1.00	-1.00	-0.99	-0.97	**0.39**
1/1WK									1.00	1.00	0.98	**-0.38**
1/2WK										1.00	0.99	**-0.39**
2WUCJD											1.00	**-0.43**
PDS3[Sh[S]]												1.00

Table 8.4. Intercorrelation Matrix for the Selected Molecular Properties of Octanes.

Variable	BP	MON	HV	MV	S	TSA	AF	MR	LogP	DENS	CT	CP	DHF
BP	1.00	-0.32	0.11	0.12	0.62	0.06	**0.63**	-0.31	0.18	-0.15	**0.75**	0.08	0.33
MON		1.00	0.08	-0.38	-0.62	-0.39	**-0.66**	0.24	-0.09	0.37	0.05	0.43	0.49
HV			1.00	-0.02	0.09	-0.32	0.03	-0.09	-0.25	0.00	0.16	0.16	0.37
MV				1.00	**0.73**	0.29	**0.68**	**-0.90**	-0.03	**-1.00**	0.13	-0.04	**-0.61**
S					1.00	0.41	**0.95**	-0.68	0.07	**-0.74**	0.30	-0.28	-0.32
TSA						1.00	0.55	0.07	**0.53**	-0.25	-0.46	**-0.82**	**-0.61**
AF							1.00	-0.56	0.15	**-0.67**	0.19	-0.44	-0.39
MR								1.00	0.15	**0.92**	-0.51	-0.35	0.29
LogP									1.00	0.05	-0.07	-0.28	-0.24
DENS										1.00	-0.18	-0.01	0.58

CT							1.00	**0.71**	0.50
CP								1.00	0.50
DHF									1.00

Data for a second set of 82 polycyclic aromatic hydrocarbons are included in Table 8.5 while the correlations are given in Table 8.6.

Table 8.5. Data for Polycyclic Aromatic Hydrocarbons (PAH).

No.	Molecule	MP	BP	LogP	LEL	W	χ
1	naphtalene	81	218	3.35	13.341	109	4.966
2	1-methylnaphthalene	-22	245	3.87	14.572	140	5.377
3	2-methylnaphthalene	35	241	4	14.575	144	5.36
4	1-ethylnaphthalene	-14	259	4.39	15.837	182	5.915
5	2-ethylnaphthalene	-7	258	4.38	15.841	190	5.898
6	2-6-dimethylnaphthalene	110	262	4.31	15.808	186	5.754
7	2-7-dimethylnaphthalene	97	262	-	15.808	185	5.754
8	1-7-dimethylnaphthalene	-14	263	4.44	15.805	180	5.771
9	1-5-dimethylnaphthalene	80	269	4.31	15.802	176	5.788
10	1-2-dimethylnaphthalene	-4	271	4.31	15.803	178	5.788
11	1-3-7-trimethylnaphthalene	14	280	-	17.037	226	6.165
12	2-3-5-trimethylnaphthalene	25	285	-	17.035	224	6.182
13	2-3-6-trimethylnaphthalene	101	286	4.73	17.038	230	6.165
14	phenalene	85	-		17.919	210	6.449
15	1-phenylnaphthalene	45	334	-	21.739	412	7.949
16	2-phenylnaphthalene	104	360	-	21.744	436	7.933
17	anthracene	216	340	4.5	19.197	279	6.933
18	1-methylanthracene	86	363	-	20.426	334	7.343
19	2-methylanthracene	209	-	-	20.429	342	7.327
20	2-7-dimethylanthracene	241	370	-	21.66	413	7.72
21	2-6-dimethylanthracene	250	370	-	21.66	414	7.72
22	2-3-dimethylanthracene	252	-	-	21.658	408	7.737
23	9-10-dimethylanthracene	183	-	5.69	21.646	378	7.788
24	phenanthrene	101	338	4.52	19.194	271	6.949
25	1-methylphenanthrene	123	359	5.08	20.422	326	7.36
26	2-methylphenanthrene	56	355	5.24	20.425	334	7.343

27	3-methylphenanthrene	65	352	5.15	20.425	330	7.343
28	4-methylphenanthrene	50	-	-	20.422	322	7.36
29	9-methylphenanthrene	91	355	-	20.421	322	7.36
30	3-6-dimethylphenanthrene	141	363	-	21.656	396	7.737
31	4-5-methylenephenanthrene	116	359	-	21.195	300	7.433
32	tetracene	257	-	5.76	25.047	569	8.899
33	benzo[a]anthracene	162	-	5.91	25.043	553	8.916
34	chrysene	256	441	5.86	25.039	545	8.933
35	benzo[c]phenanthrene	68	-	-	25.038	529	8.933
36	triphenylene	199	439	5.49	25.032	513	8.949
37	pyrene	156	393	5	22.49	362	7.933
38	1-methylpyrene	70	410	-	23.717	428	8.343
39	2-methylpyrene	144	410	-	23.72	434	8.327
40	4-methylpyrene	148	410	-	23.717	424	8.343
41	2-7-dimethylpyrene	-	396	-	24.949	515	8.72
42	pentacene	271	-	-	30.894	1011	10.865
43	dibenzo[ai]anthracene	264	-	6.81	30.889	987	10.882
44	dibenzo[ah]anthracene	270	-	5.8	30.885	971	10.899
45	dibenzo[aj]anthracene	198	-	-	30.885	955	10.899
46	benzo[b]chrysene	294	-	-	30.885	971	10.899
47	dibenzo[ac]anthracene	205	-	-	30.877	907	10.916
48	pycene	-	519	-	30.881	963	10.916
49	benzo[a]pyrene	177	496	5.97	28.331	680	9.916
50	benzo[e]pyrene	179	493	-	28.325	652	9.933
51	perylene	278	-	6.25	28.326	654	9.933
52	coronene	360	-	6.5	34.906	1002	11.899
53	anthranthrene	261	-	-	31.621	839	10.899
54	benzo[ghi]perylene	283	-	6.9	31.617	815	10.916
55	dibenzo[ae]pyrene	234	-	-	34.163	1082	11.916
56	1-methylchrysene	161	-	-	26.265	620	9.343
57	6-methylchrysene	257	-	-	26.267	632	9.343
58	3-methylcholanthrene	180	-	6.75	29.54	804	10.327
59	indeno[1-2-3-cd]pyrene	163	-	-	31.599	845	10.916
60	pentaphene	263	-	-	30.889	979	10.882
61	hexaphene	308	-	-	36.734	1589	12.848
62	indano	-51	178	-	12.043	79	4.466
63	indene	-2	183	2.92	12.043	79	4.466
64	azulene	100	270	3.22	13.335	107	4.966
65	acenaphthene	96	279	3.92	16.624	166	5.949
66	acenaphthylene	93	270		16.624	166	5.949
67	fluorene	117	294	4.18	17.899	219	6.449

68	1-methylfluorene	87	318	4.97	19.128	267	6.86
69	2-methylfluorene	104	318	-	19.131	274	6.843
70	3-methylfluorene	88	316	-	19.131	272	6.843
71	4-methylfluorene	71	-	-	19.128	265	6.86
72	9-methylfluorene	47	-	-	19.125	262	6.877
73	1-2-benzofluorene	190	407	5.4	23.746	461	8.433
74	fluoranthene	111	383	5.2	22.466	364	7.949
75	2-3-benzofluorene	209	402	5.75	23.75	471	8.416
76	3-4-benzofluorene	125	406	-	23.745	453	8.433
77	benzo[ghi]fluoranthene	149	432	5.78	25.759	478	8.933
78	benzo[k]fluoranthene	217	481	-	28.313	698	9.916
79	benzo[b]fluoranthene	168	481	-	28.307	676	9.933
80	benzo[j]fluoranthene	166	480	-	28.309	678	9.933
81	ovalene	473	-	-	47.307	2106	15.865
82	quaterryllene	483	-	-	58.242	4544	19.865

Legend: melting point MP, boiling point BP, partition coefficient n-octanol/water Log P and the corresponding LEL, Wiener W and Randić indices

Table 8.6. Correlation of PAH Properties with Selected Topological Indices.

Property	LEL	W	χ
MP (n = 80)	0.857	0.748	0.855
BP (n = 53)	0.989	0.955	0.988
LogP (n = 37)	0.945	0.905	0.948

8.2. Results and Discussion

A correlating study of topological indices provided by TOPOCLUJ software package and LEL, a newly proposed index, on thirteen properties of octanes revealed good correlating ability of a dozen selected TIs, all related to the Wiener index as shown in Table 8.3.

Table 8.3. Intercorrelation Matrix for the Best Scored Indices in Ooctanes.

Variable	LEL	1WD	2WD	1WW	1WH	2WH	1WK	2WK	1/1WK	1/2WK	2WUCJD	PDS3 [Sh[SZ]]
LEL	1.00	0.96	0.95	0.96	-0.99	-0.98	-0.99	-0.99	0.99	0.98	0.94	**-0.34**
1WD		1.00	1.00	1.00	-0.99	-0.99	-0.98	-0.98	0.99	0.99	1.00	**-0.43**
2WD			1.00	1.00	-0.98	-0.98	-0.97	-0.97	0.98	0.99	1.00	**-0.42**
1WW				1.00	-0.99	-0.99	-0.98	-0.98	0.99	0.99	1.00	**-0.43**
1WH					1.00	1.00	1.00	1.00	-1.00	-1.00	-0.98	**0.40**
2WH						1.00	1.00	1.00	-1.00	-0.99	-0.98	**0.41**
1WK							1.00	1.00	-1.00	-0.99	-0.97	**0.38**
2WK								1.00	-1.00	-0.99	-0.97	**0.39**
1/1WK									1.00	1.00	0.98	**-0.38**
1/2WK										1.00	0.99	**-0.39**
2WUCJD											1.00	**-0.43**

PDS3[Sh[S]]												1.00

LEL is the best correlated with the WK indices. It describes well the properties which are well accounted by the majority of the selected molecular descriptors: octane number MON, entropy S, volume MV, or refraction MR, particularly the AF parameter, but also more difficult properties like boiling point, melting point and logP. Among the desirable attributes required by a good TI, LEL fulfils: good correlation with at least one property, good discrimination of isomers and simplicity. In addition, it is well defined mathematically and shows interesting relations in particular classes of graphs. This index and those provided by the TOPOCLUJ software as well, was proved to be of basic importance in QSAR/QSPR studies (Table 8.4 and Table 8.6).

Table 8.4. Intercorrelation Matrix for the Selected Molecular Properties of Octanes.

Variable	BP	MON	HV	MV	S	TSA	AF	MR	LogP	DENS	CT	CP	DHF
BP	1.00	-0.32	0.11	0.12	0.62	0.06	**0.63**	-0.31	0.18	-0.15	**0.75**	0.08	0.33
MON		1.00	0.08	-0.38	-0.62	-0.39	**-0.66**	0.24	-0.09	0.37	0.05	0.43	0.49
HV			1.00	-0.02	0.09	-0.32	0.03	-0.09	-0.25	0.00	0.16	0.16	0.37
MV				1.00	**0.73**	0.29	**0.68**	**-0.90**	-0.03	**-1.00**	0.13	-0.04	**-0.61**
S					1.00	0.41	**0.95**	-0.68	0.07	**-0.74**	0.30	-0.28	-0.32
TSA						1.00	0.55	0.07	**0.53**	-0.25	-0.46	**-0.82**	**-0.61**
AF							1.00	-0.56	0.15	**-0.67**	0.19	-0.44	-0.39
MR								1.00	0.15	**0.92**	-0.51	-0.35	0.29
LogP									1.00	0.05	-0.07	-0.28	-0.24
DENS										1.00	-0.18	-0.01	0.58
CT											1.00	**0.71**	0.50
CP												1.00	0.50
DHF													1.00

Table 8.6. Correlation of PAH Properties with Selected Topological Indices.

Property	LEL	W	χ
MP (n = 80)	0.857	0.748	0.855
BP (n = 53)	0.989	0.955	0.988
LogP (n = 37)	0.945	0.905	0.948

8.3. Conclusions

LEL is the best correlated with the WK indices. It describes well the properties which are well accounted by the majority of the selected molecular descriptors: octane number MON, entropy S, volume MV, or refraction MR, particularly the AF parameter, but also more difficult properties like boiling point, melting point and logP. Among the desirable attributes required by a good TI, LEL fulfils: good correlation with at least one property, good discrimination of isomers and simplicity. In addition, it is well defined mathematically and shows interesting relations in particular classes of graphs. This index and those provided by the TOPOCLUJ software as well, was proved to be of basic importance in QSAR/QSPR studies.

CONCLUDING REMARKS

- The correlation obtained between partitions coefficients and different descriptors for monocarboxylic acids is highly significant; therefore they can be used to predict the values of other members of the series.
- The powerful predictive ability of the models allowed the estimation of unknown partition coefficients for some monocarboxylic acids and a correct prediction of partition coefficients for Lauric Acid and Melissic Acid. The predicted values for Lauric Acid and Melissic Acid were much higher in both cases.
- The results obtained on RP-HPTLC, followed by those obtained by RP-HPLC will permit to determine the lipophilicity indices of 27 bile acids and their derivatives and to investigate the molecular mechanism of retention in order to find an objective manner of quantitative comparison of retention properties of different chemically bonded stationary phases.
- The goal of the study was to estimate and compare the lipophilicity of some quaternary ammonium and nitrone derivatives and their thiazolic salts and to investigate the molecular mechanism of retention in order to find an objective manner of quantitative comparison of retention properties of different chemically bonded stationary phases used in thin layer chromatography.
- Our study demonstrated that 2D and 3D descriptors related to atomic mass, symmetry together with reactivity parameters such as polarizability and electronegativity seem to control the lipophilicity on all stationary phases; the maximal negative charge of the molecule on CN phase, and topological aspects of the molecule for NH_2 are decisive for retention.
- The objective of this work was to analyze and discuss the correlations found between the chromatographic retention indices of tested compounds (R_{Mo}, b, and $PC1_{RF}$) and the calculated topological descriptors obtained through different software's of the formyl- and acetylpyridine-3-thiosemicarbazone derivatives.
- The shape of the molecule is an important index which should be taken into consideration because it plays a dominant role in the chromatographic behavior on both stationary phases with different polarities.
- A correlating study of topological indices TIs provided by TOPOCLUJ software package and LEL, a newly proposed index built up on the eigenvalues of Laplacian matrix, on thirteen properties of octanes, revealed good correlating ability of a dozen selected TIs, all related to the Wiener index, and of LEL as well. LEL describes well the properties which are well accounted by the majority of the selected molecular descriptors: octane number MON, entropy S, volume MV, or refraction MR, particularly the acentric factor AF parameter, but also more difficult properties like boiling point, melting point and logP. LEL is the best correlated with the WK (Wiener-type number, taken the reciprocal of entries in the combinatorial D_p matrix, of higher rank, calculated by TOPOCLUJ software) indices. In a second set of polycyclic aromatic hydrocarbons, LEL was proved to be as good as the Randić χ index (a connectivity index) and better than the Wiener index (a distance based index). In addition, it is well defined mathematically and shows interesting relations in particular classes of graphs, these recommending LEL as a new, powerful

topological index. The actual study proved the considered TIs are basic topological descriptors in prediction of various molecular properties, with good perspective in QSPR/QSAR studies.

REFERENCES

1. Otto, M. *Chemometrics. Statistical and Computer Application in Analytical Chemistry*, Wiley-VCH-Wienheim, New York, **1999.**

2. Roy, P.P., Leonard, J.T. and Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemom. Intell. Lab. Syst.*, **2008**, 90, 31–42.

3. Wang, R., Fu, Y. and Lai, L. A New Atom-Additive Method for Calculating Partition Coefficients, *J. Chem. Inf. Comput. Sci.*, **1997**, *37(3)*, 615-621.

4. Sârbu, C., Karajan, K. and Kevresan, S. Evaluation of the lipophilicity of bile acids and their derivatives by thin-layer chromatography and principal component analysis. *J. Cromatogr. A*, **2001**, *917(1/2)*, 361-366.

5. Karelson, M. *Molecular Descriptors in QSAR/QSPR*, Wiley & Sons, New York, **2000.**

6. Roda, A. Minutello, A. Angelotti M.A. and A. Fini, Bile acid structure-activity relationship: evaluation of bile acid lipophilicity using 1-octanol/water partition coefficient and reverse phase HPLC, *J. Lipid Res.*, **1990**, *31*, 1433-1443.

7. McCall, J.M. Liquid-lipquid partition coefficients by high-pressure liquid chromatography**,** *J. Med. Chem.,* **1975**, *18*, 549-552.

8. Rekker, R.F. and Mannhold, R. *Calculation of drug lipophilicity: The hydrophobic fragmental constant approach*, VCH Publishers, Inc. New York, **1992**.

9. Kastner, P., Klimeš, J., Zimová G. and Klimešová, V. Reversed-phase thin-layer chromatographic determination of the lipophilicity of potential antituberculotic compounds, *J. Planar Chromatogr.*, **2001**, *14*, 291-295.

10. Consonni, V. and Todeschini, R. Structure - Activity Relationships by autocorrelation descriptors and genetic algorithms. In *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (Lohdi H. and Yamanishi Y., eds.), IGI Global, Hershey, PA (USA), **2009**.

11. Pavan, M., Mauri, A. and Todeschini, R. Total ranking models by the Genetic Algorithms Variable Subset Selection (GA-VSS) approach for environmental priority settings. *Anal. Bioanal. Chem.*, **2004**, *380*, 430-444.

12. Li, Y-H., Tanno, M., Itoh, T. and Yamada H. Role of the monocarboxylic acid transport system in the intestinal absorption of an orally active b-lactam prodrug: carindacillin as a model, International Journal of Pharmaceutics, **1999**, *191*, 151–159.

13. Kah, M. and Brown, C.D. LogD: Lipophilicity for ionisable compounds, *Chemosphere*, **2008**, *72*, 1401–1408.

14. Valko, K. Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution, *J. Chromatogr. A*, **2004**, *1037*, 299–310.

15. Abraham, M.H., Ibrahim, A. and Zissimos, A.M. Determination of sets of solute descriptors from chromatographic measurements, *J. Chromatogr. A,* **2004**, *1037,* 29–47.

16. Berthod, A. and Carda-Broch, S. Determination of liquid–liquid partition coefficients by separation methods, *J. Chromatogr. A,* **2004**, *1037*, 3–14.

17. Pyka, A. and Miszczyk, M. Chromatographic evaluation of the lipophilic properties of selected pesticides, *Chromatographia,* **2005**, *61*, 37–42.

18. Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res. 2009 37(Database issue):D603-610*, http://www.hmdb.ca/metabolites, Retrived 2010-04-04.

19. Sangster J. "LOGKOW: A databank of evaluated octanol-water partition coefficients (LogP)". Sangster Research Laboratories. http://logkow.cisti.nrc.ca/logkow/. Retrieved 2010-04-04.

20. DRAGON for Windows (software for molecular descriptor calculations), Version 5.4 – 2005. http://www.talete.mi.it.

21. Alchemy 2000 software, http://www.cambridgesoft.com.

22. HyperChem(TM) Professional 7.5 for Windows, Molecular Modeling System, Hypercube, Inc. and Autodesk, Inc.

23. Todeschini, R., Moby Digs Academic version software for variable subset selection by genetic algorithms, Rel. 1.0 for Windows, Talete, Milan, 2004. http://www.talete.mi.it.

24. Devillers, J. and Balaban, A.T. *Topological indices and related descriptors in QSAR and QSPR.* Gordon and Breach Science, The Netherlands, **1999**.

25. Ghose, A.K., Pritchett, P. and Crippen, G. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. *J. Comput. Chem.,* **1988**, *9*, 80–90.

26. Poša, M. and Kuhajda, K. Hydrophobic and haemolytic potential of oxo derivatives of cholic, deoxycholic and chenodeoxycholic acids, *Steroids*, **2010**, *75*, 424-431.

27. Costescu, A., Moldovan, C. and Diudea, M.V., QSAR modeling of steroid hormones; *MATCH-Commun. Math. Comput Chem.,* **2006**, *55*(2), 315-329.

28. Ursu, O., Costescu, A., Diudea, M. And Pârv, B. QSAR modeling of antifungal activity of some heterocyclic compounds, *Croat. Chem. Acta*, **2006**, *79*(3), 483-488.

29. Moldovan, C.D., Costescu, A., Katona, G. and Diudea, M.V. A novel QSAR approach in modeling antifungal activity of some 5-or 6-methyl-2-substituded benzoxazoles/benzimidazoles against C. albicans using molecular descriptors, MATCH-*Commun. Math. Comput. Chem*., **2008**, *60*(3), 977-984.

30. Tiperciuc, B., Zaharia, V., Câmpean, R., Curticăpean, M., Costescu, A. and Diudea, M.V. A QSAR Study on Antimicrobial Activity of Some New Sulfonylhydrazinothiazoles, MATCH-*Commun. Math. Comput. Chem*., **2008**, *60*(3), 985-996.

31. Costescu, A., Moldovan, C., Katona, G. and Diudea, M.V. QSAR modeling of human catechol O-methyltransferase enzyme kinetics, *J. Math. Chem.*, **2009**, *45*(2), 287-294.

32. Moldovan, C., Costescu, A., Katona, G. and Diudea, M.V. Application to QSAR studies of 2-furylethylene derivatives, *J. Math. Chem.*, **2009**, *45*(2), 442-451.

33. ALOGPS 2.1software, http://www.vcclab.org/lab/alogps/start.html.

34. ChemSilico software, http://www.chemsilico.com.

35. OSIRIS software, http://www.organic-chemistry.org/prog/peo/osiris property explorer.

36. Drug likeness and molecular property prediction, http://www.molsoft.com.VEGA online, http://www.ddl.unimi.it.

37. Chemaxon software, http://intro.bio.umb.edu/111-112/OLLM.

38. Palage, M., Oniga, S., Parnau, A., Zaharia, V., Belegan, C., Vlase, L. And Muresan, A. Synthesis and physico-chemical characterization of some quaternary ammonium salts of 2-aryl thiazole derivatives, *Farmacia,* **2009**, *57*(5), 598-608.

39. Palage, M., Parvu, M., Oniga, S. and Muresan, A., Fungicidal-fungistatic Action of Some 2-Aril-thiazol compounds, *Farmacia*, **2007**, *LV*(2), 203-206.

15. Randić, M., Woodworth, W.L. and Graovac, A. A. Unusual Random Walks, *Int. J. Quant. Chem.,* **1983**, *24*, 435-452.

40. Diudea, M.V., Gutman, I. and Jäntschi, L. *Molecular Topology*, NOVA, New York, **2002**.

41. Diudea, M.V., Florescu, M.S. and Khadikar, P.V. Molecular Topology and Its Applications, EFICON, Bucharest, **2006**.

42. Rucker, G. and Rucker, C. Counts of all walks as atomic and molecular descriptors, *J. Chem. Inf. Comput. Sci.,* **1993**, *33*, 683-695.

43. Diudea, M.V. Walk Numbers eWM : Wiener-Type Numbers of Higher Rank, *J. Chem. Inf. Comput. Sci*., **1996**, *36*, 535-540.

44. Diudea, M.V., Topan, M. and Graovac, A. A. Layer Matrices of Walk Degrees, *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 1071 -1078.

45. Wiener, H. Structural determination of Paraffin boiling points, *J. Amer. Chem. Soc.*, **1947**, *69***,** 17-20.

46. Diudea, M.V. Indices of Reciprocal Properties or Harary Indices, *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 292-299.

47. Diudea, M.V. and Gutman, I. Wiener-Type Topological Indices, *Croat. Chem. Acta,* **1998**, *71*, 21-51.

48. Diudea, M.V. Cluj Matrix CJu: Source of various graph descriptors, *MATCH Commun. Math. Comput. Chem.,* **1997**, *35*, 169-183.

49. Diudea, M.V. Cluj matrix invariants, *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 300-305.

50. Diudea, M.V., Pârv, B. and Topan, M.I. Derived Szeged and Cluj Indices, *J. Serb. Chem. Soc*. **1997**, *62*, 267-276

51. Janežič, D., Nikolić, S. and Trinajstić, N. *Graph Theoretical Matrices in Chemistry*, MCM, Kragujevac, **2007**.

52. Diudea, M.V. Layer matrices in molecular graphs, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1064-1071.

53. Diudea, M.V. and Ursu, O. Layer matrices and distance property descriptors, *Indian J. Chem.*, 42*A*, **2003**, 1283-1294.

54. Katona, G. and Panea, T. Modeling Physical-chemical properties by topological indices *Acta Univ. Cibiniensis*, **2005**, 8(*2)*, 33-45.