QSAR Study of conductors polymers based on thiophene

K. Dguigui [1*], M. Mbarki [1], M. Elhallaoui [1], M. Elasri [2], M. Bouachrine [3]

[1] LCPPEM, Département de Chimie Faculté des Sciences Dhar El Mahraz, Université Sidi Mohammed ben Abdallah, B.P.1796 ATLAS. Fès. Maroc.
[2] Département de Chimie, Faculté des Sciences et Techniques Fès-Saïs. Université Sidi Mohammed Ben Abdallah, B.P.2202 Route D'Immouzer Fès. Maroc.
[3] LRMM, Département de Chimie Faculté des Sciences et Techniques. Errachidia. Université Moulay Ismail.B.P.509 Boutalamine. Errachidia. Maroc.

Received in 03 Oct 2008, Revised 24 Oct, Accepted 25 Oct 2010.
Email: khalidchimie76@hotmail.com; Tel.: +212658127928; Fax: +212535732981

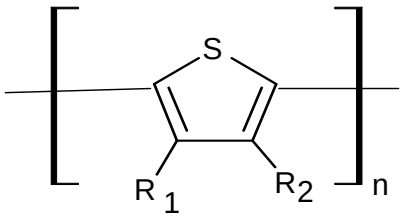
Abstract

A quantitative structure-activity relationship (QSAR) model was constructed to predict the electrical conductivity of 46 compounds of monothiophene derivative. In the present study where used to establish QSARs based on quantum mechanically derived molecular descriptors, hydrophobicity and steric effects. The structure-activity relationships were discussed by multiple linaire regression (MLR) r =0,91 and neural network (NN) r =0,95. The correlation coefficient obtained with MLR and NN is fairly good to evaluate a quantitative model. To test the performance of this model we have used the cross validation method (r =0,85).

Key words: Quantitative structure activity relationships (QSAR), Conductivity σ, multiple linear regressions (MLR), Neural Network (NN), Cross validation (CV).

1. Introduction

Over 40 years ago, conjugated polymers were envisioned as futuristic new materials that would lead to the next generation of electronic and optical devices. The development of electronic devices based on conjugated polymers, since polythiophenes (PTs) possess excellent and high conductivity when doped, **[1]**. To date PTs have been used in variety of application including: transistors, hole injection layer in polymer, electrical conductors, environmental sensors, and solar cells, **[2].**
The geometric structure of polythiophene is sketched in fig.1, it exists in a continuous network, often a simple chain, of adjacent unsaturated carbon atoms, i.e., carbon atoms in the sp^2 hybridized state, **[3].** This type of structure gives the corresponding materials proprieties of semiconductor wide band gap E_p >1.5 eV, **[4].**



Fig, 1. Chemical structure of polythiophene

QSARs can play an important role in molecular electronic, solid physics, materials chemistry, and the ability to predict electrical activity, for example electrical conductivity is valuable in any number of industries.
The main contributions to the widespread use of QSAR models come from the development of novel structural descriptors and statistical equations relating various physical, chemical, and biological proprieties to the chemical structure.
The success of the QSPR and QSAR approach can be explained by the insight offered into the structural determination of chemical proprieties and possibility to estimate the proprieties of new chemical compounds without the need to synthesize and test them.
The main hypothesis in the QSPR and QSAR approach is that all proprieties (physical, chemical,..) of a chemical substance are statistically related to its molecular structure **[5].**
In our study, we are concerned to seek the relationship between electrical conductivity of polythiophene and physico-chemical proprieties using statistical methods and neural network.

2. Methodology

2.1. Data and Software
2.1.1. Database
In this work, a set of 47 organic compounds derivatives of monothiophene has been investigated and analyzed to determine a quantitative relationship between structure and conductivity.
The conductivity is expressed in logarithmic form and its unit is S/cm.

2.1.2 Software
Statistical evaluation of the data and multivariate data analysis has been performed mainly by the software products Mystat version 12 **[6],** Statlab version 3 **[7],** Matlab version 7 **[8].**
The work has been performed on a personal computer running Microsoft Windows operating system version 7.

2.2. Molecular Descriptors
The descriptors are used to characterize the molecules to be analyzed, they can be calculated from the structure (constitution, configuration and molecular conformation) or properties (physical, chemical, biological) belonging to molecules.
The molecular descriptors were selected in such a way that they represent the features necessary to quantify the conductivity properties **[9].**
Eight descriptors were chosen to model response factors, which are molecular weight (MW), Van Der Waals surface (WS), Dipole moment (μ), Hydrogen bond Acceptor (HBA), constant Rekker (F_{rekker}), HOMO Energy (E_{HOMO}), LUMO Energy (E_{LUMO}) and chemical hardness (η).

2.3. Descriptors calculation
Molecular descriptors are formal mathematical representations of a molecule, obtained by a well-specified algorithm, and applied to a defined molecular representation or a well-specified experimental procedure: the molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment [10].

2.3.1. Molecular weight
The molecular weight is the weight of all atoms present in molecule.

2.3.2. Van Der Waals surface
Van der waals surface or van der waals envelope is the surface of the union of the spherical atomic surface defined by the van der waals redius of each component atom in the molecule **[11].**

2.3.3. Dipole moment
The dipole moment μ (or electric dipole moment) is a vectorial quantity that encodes displacement with respect to the center of gravity of positive and negative charges in a molecule defined as:

$$\mu = \int_{i=1}^{n} q_i \cdot r_i \qquad (1)$$

Where q_i are point charges located at position r_i. The SI unit for dipole moments is the coulomb meter, but they are often expressed in Debye.

The elements of the vector μ are called dipole moment components:

$$\mu_x = \int_{i=1}^{n} q_i \cdot x_i \;\; ; \;\; \mu_y = \int_{i=1}^{n} q_i \cdot y_i \;\; ; \;\; \mu_z = \int_{i=1}^{n} q_i \cdot y_z \qquad (2)$$

Where x, y, z, are the coordinates of the charges. Molecules with zero dipole moments are called nonpolar, others polar, moreover, dipole moments equal to zero indicate molecules with a center of symmetry **[12].**
In our studies, we calculated the dipole moment aside from the HF method utilizing a 6-31G basis set by the software Gaussian 03.

2.3.4. E_{HOMO} and E_{LUMO}
The E_{HOMO} and E_{LUMO} values were calculated for lowest energy conformer without further geometry optimization using the ab-initio Hartree-Fock method utilizing a 6-31G basis set.

2.3.5. Chemical hardness
The definition of the chemical hardness given by eqs:

$$\eta = \left(\frac{\partial^2 E}{\partial N^2} \right)_{\upsilon} = \left(\frac{\partial \mu}{\partial N} \right)_{\upsilon} \qquad (3)$$

It's need from the knowledge of the energy as a function of the number of electrons.
In terms of orbital energies, the chemical hardness is:

$$\eta = \varepsilon_L - \varepsilon_H \qquad (4)$$

Where ε_H and ε_L are the eigenvalues of the highest occupied molecular spin-orbital (HOMO) and the lowest unoccupied molecular spin-orbital (LUMO) **[13].**

2.3.6. Hydrogen bond accepter
Hydrogen bond is the attractive interaction of a hydrogen atom with an electronegative atom.
Katritzky, Mu et al say that all the H-bond descriptors are assigned zero if hydrogen atoms in the molecule can be donated, moreover, hydrogen bond acceptors are usually restricted to oxygen, nitrogen, and sulfur atoms.

2.3.7. Constant rekker
The log P of a molecule is calculated by summing up the fragmental contributions and applying the appropriate correction factors as:

$$\log P = b_0 + \int_{i=1}^{n} f_i . N_i + \int_{j=1}^{n} C_j . N_j \qquad (5)$$

Where f_i and N_i are the hydrophobic constant and the number of occurrences of ith fragment in the considered compound, N_j is the number of occurrences of the jth correction factor.
C_j is the value of the considered correction factor describing some special structural features; in practice, it can be calculated as: $C_j = K_{j*}\ 0.219$
Where 0.219 is the so-called "magic constant" and K_j is an integer value characterizing the jth correction factor [14].

3. Results and discussion

A QSAR study was carried for a series of 46 derivatives of polythiophenes, in order to determine a quantitative relationship between structure and conductivity.
Table 1 shows the chemical structure of the series of training set, the values of σ_{obs} and values of calculated σ by MLR and NN.

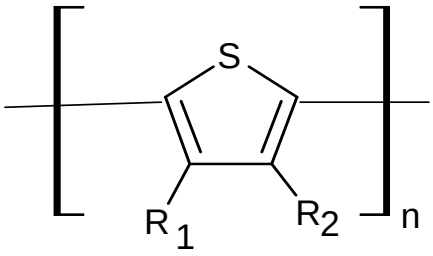


Table 1. he chemical structure of the studied compounds, the values of $\log(\sigma_{obs})$ corresponding to reference [15] and values of predicted $\log(\sigma)_{MLR}$ and $\log(\sigma)_{NN}$ are calculated using MLR and NN respectively. The values of $\log(\sigma)_{CV}$ are calculated with the cross validation method with "leave one out" procedure used to test the performance of the model.

N	R_1	R_2	$Log\sigma_{obs}$ (S/cm)	$Log\sigma_{MLR}$ (S/cm)	$Log\sigma_{NN}$ (S/cm)	$Log\sigma_{CV}$ (S/cm)
1	C_2H_5	H	2.25	2.27	1.93	1.88
2	C_3H_7	H	2.20	2.29	1.85	1.81
3	C_4H_9	H	2.17	2.12	1.77	1.74
4	C_5H_{11}	H	2.14	2.00	1.71	1.67
5	C_6H_{13}	H	1.47	1.86	1.65	1.66
6	C_7H15	H	1.65	1.74	1.59	1.59
7	C_8H_{17}	H	1.60	1.60	1.52	1.51
8	C_9H_{19}	H	1.92	1.50	1.47	1.46
9	$C_{10}H_{21}$	H	1.84	1.34	1.38	1.36
10	$C_{12}H_{25}$	H	0.60	4.37	1.69	1.70
11	$C_{14}H_{29}$	H	0.84	0.82	1.12	1.16
12	$C_{18}H_{37}$	H	0.30	0.30	0.85	1.01
13	$C_{20}H_{41}$	H	1.04	0.07	0.71	0.55
14	C_6H_5	H	2.14	2.70	2.14	1.87
15	$mCH_3O\text{-}C_6H_5$	H	1.60	-0.40	0.92	0.39
16	$mCF_3O\text{-}C_6H_5$	H	-2.74	-2.65	-2.50	-1.43
17	$C_2H_4\text{-}C_6H_5$	H	1.11	2.60	1.64	1.70
18	$C_3H_6\text{-}C_6H_5$	H	1.39	2.45	1.56	1.53
19	$C_2H_4\text{-}O\text{-}CH_2\text{-}C_6H_5$	H	1.30	0.12	0.61	0.44
20	$O\text{-}CH_3$	H	-0.88	-1.21	-1.24	-1.39
21	$O\text{-}C_2H_5$	H	-2.01	-1.27	-1.27	-1.24
22	$O\text{-}C_3H_7$	H	-2.22	-1.42	-1.42	-1.09
23	$O\text{-}(CH_2)_9\text{-}CH_3$	H	-2.52	-2.24	-2.62	-2.61
24	$O\text{-}(CH_2)_{10}\text{-}CH_3$	H	-2.52	-1.70	-2.70	-2.74
25	$O\text{-}(CH_2)_{11}\text{-}CH_3$	H	-1.22	0.84	-2.19	-1.84
26	$O\text{-}(CH_2)_{12}\text{-}CH_3$	H	-1.69	-1.48	-2.32	-1.01
27	$O\text{-}(CH_2)_{13}\text{-}CH_3$	H	-2.69	-1.86	-1.43	-1.03
28	$O\text{-}(CH_2)_{14}\text{-}CH_3$	H	-3.01	-2.18	-3.22	-3.33
29	$CH_2\text{-}O\text{-}C_2H_5$	H	-2.01	-0.86	-0.33	0.48
30	$CH_2\text{-}O\text{-}CH_3$	H	1.00	-1.00	-0.77	-2.48
31	$CH_2\text{-}O\text{-}C_7H_{15}$	H	-1.01	-1.36	-1.00	-1.15
32	$(CH_2)_2\text{-}O\text{-}C_4H_9$	H	1.74	-0.11	1.43	1.78

33	$(CH_2)_2$-O-C_6H_{13}	H	1.69	1.84	1.69	0.29
34	$(CH_2)_3$-O-C_5H_{11}	H	1.00	1.32	1.00	0.23
35	CH_2-O-CH_2-O-$(CH_2)_2$-O-CH_3	H	-3.01	-2.21	-2.78	-0.94
36	$(CH_2)_2$-O-$(CH_2)_2$-O-CH_3	H	2.39	1.16	1.72	1.24
37	$(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_2$-CH_3	H	1.00	1.01	1.71	0.68
38	H	CH_3	2.69	2.57	1.98	1.80
39	CH_3	CH_3	-0.31	0.79	1.03	0.83
40	CH_3	OCH_3	2.34	1.27	2.68	2.77
41	H	OC_4H_9	-3.01	-2.48	-3.09	-3.39
42	OC_4H_9	OC_4H_9	-5.01	-4.61	-4.97	-3.51
43	CH_3	OC_4H_9	0.30	0.67	0.22	0.32
44	CH_3	OC_8H_{17}	0.00	0.00	-0.20	0.13
45	CH_3	$OC_{12}H_{25}$	0.69	1.05	1.26	1.31
46	H	S-C_2H_5	-3.01	-3.08	-2.92	0.08

3.1. Multiple linear regressions

The electrical conductivity was first regressed with the molecular descriptors selected for the QSAR study (table 2).

Table 2. Type of molecular descriptors selected for the QSAR study

Molecular descriptors	**Type**
Van Der Waals surface (SW)	steric
Dipole moment (μ)	electronic
Energy of the highest occupied molecular orbital (E_{HOMO})	electronic
Energy of the lowest unoccupied molecular orbital (E_{LUMO})	electronic
Chemical hardness (η)	electronic
Hydrogen bond acceptor (HBA)	steric
Rekker constant (F_{rekker})	hydrophobicity
Molecular weight (MW)	Physici-chemical properties

The selection of descriptors is based on the values of p-value, the descriptors with a p-value greater than 0.05 are eliminated. Descriptors selected are a p-value less than 0,05.

The equation of the model is:

$$\text{Log}\,(\sigma)_{MLR} = 26.89 + 0.059*(PM) - 2.233*(SW) - 0.737*(\mu) + 0.731*(HBA) + 1.970*(F_{Rekker}) + 272.635*(E_{HOMO}) + 289.30*(E_{LUMO}) + 295.714*(\eta) \quad (6)$$

$$N = 47,\ r = 0.91,\ p \leq 0.05,\ SE = 0.9 \quad (7)$$

The coefficient of correlation shows that the model has a good and power explanatory.

The correlation between $\log(\sigma)_{obs}$ and $\log(\sigma)_{MLR}$ is illustrated in the following plot (r=0.91) **Figure1**.

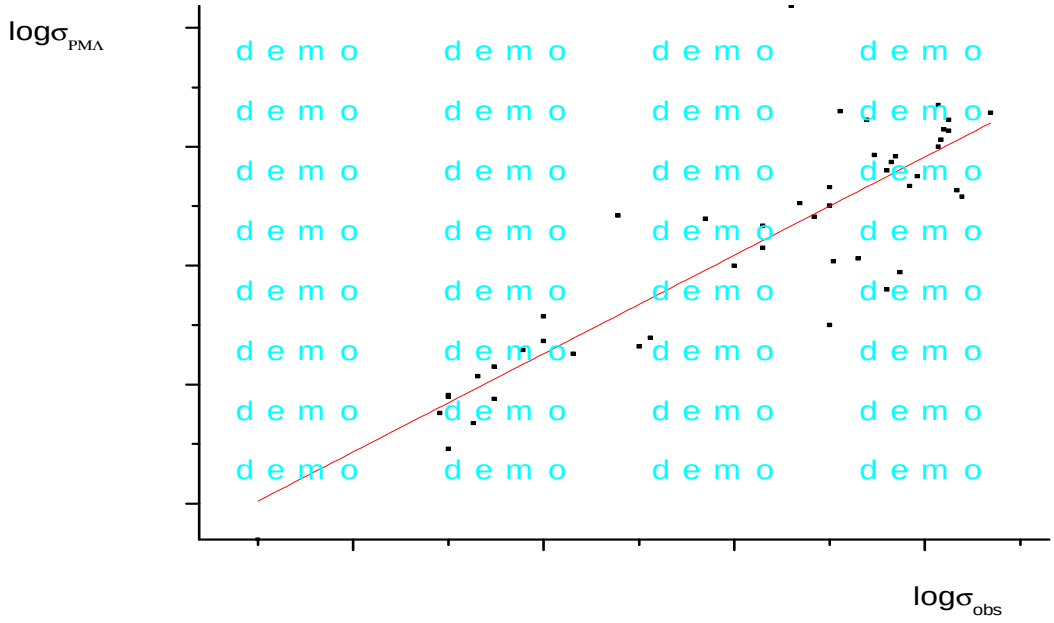


Figure1. Graphical representation of log $(\sigma)_{obs}$ versus log $(\sigma)_{MLR}$.

3.2. Neural network

Despite the good results we obtained by multiple linear regressions, it's likely that any nonlinear relationship occurs. The neural network is the proper concept to accomplish this stain.

In our study, the layer is formed by entering the eight descriptors proposed by linear regression.

The hidden layer contains two tansig neural and the output layer neural is a linear or network configuration is (8-2-1) mark, and the number of weights should be between 1.8 and 2.3.

The correlation coefficient obtained by neural network is 0.95 which enabled us to offer a good model.

The correlation between $\log(\sigma)_{obs}$ and $\log(\sigma)_{NN}$ is illustrated in the following plot **Figure2**.

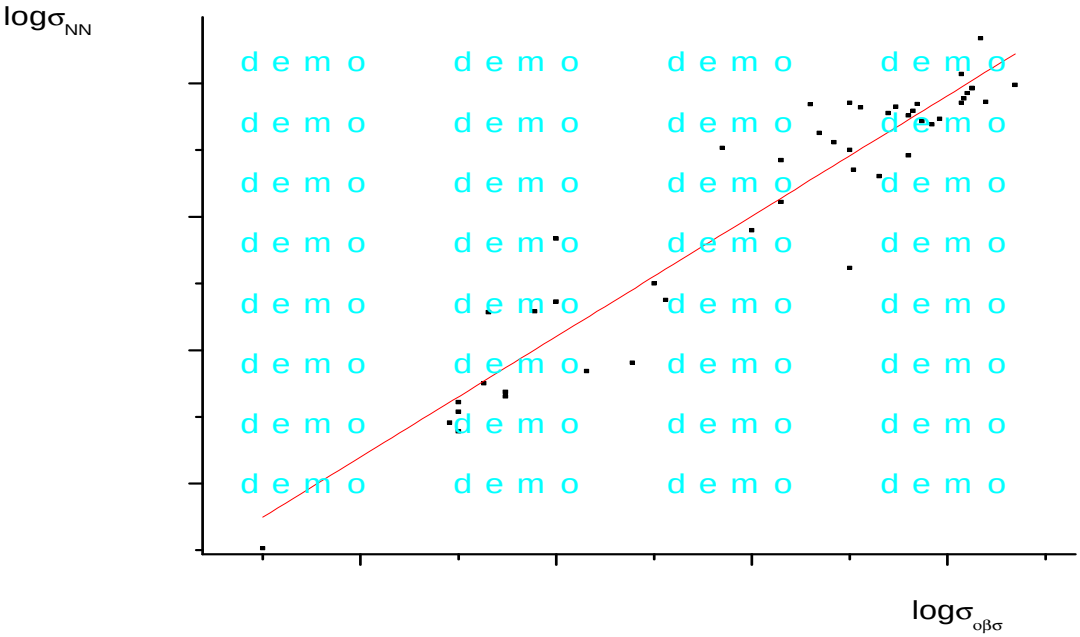


Figure2. Graphical representation of $\log(\sigma)_{obs}$ versus $\log(\sigma)_{NN}$

3.3. Cross validation (leave-one-out)

Leave-one-out croos validation (LOOCV) involves using a single observation from the original sample as the validation datum and the remaining observation as the training data. This is repeated such that each observation in the sample is used once as the validation data. This is the same as K-fold cross-validation with K being equal to the number of observations in the original sample.

In this part of work, we can test the performance of the neural network and validity our choices of descriptors by MLR. The correlation coefficient obtained by LOOCV is 0.85. The results are illustrated in the following plot **figure 3**.

N.B: for after all the results we obtained and submitted in the graphs, we note that althought two parts: high when $\log\sigma \geq 0.85$ and low when $\log\sigma \leq 0.85$.

After the regression equation we note that the coefficient of contribution of electronic descriptors is very important which means that the conductivity is directly related with the energy of the systeme (E_{HOMO}, E_{LUMO}, η).

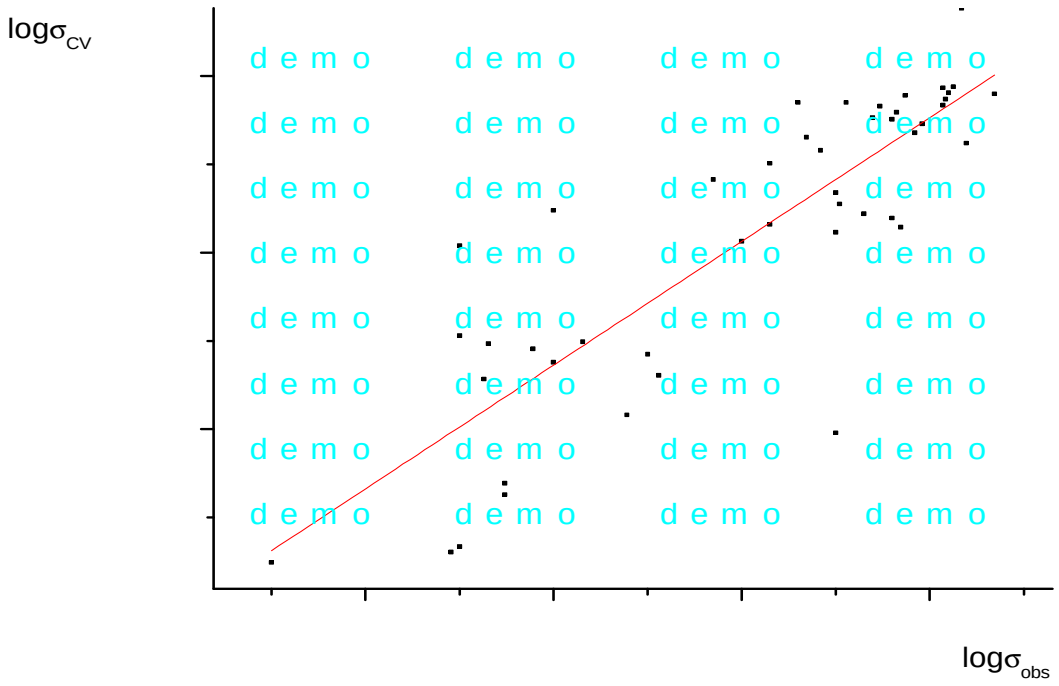


Figure 3. Graphical representation of $\log(\sigma)_{obs}$ versus $\log(\sigma)_{CV}$

4. Conclusion

The statistical analysis that we have undertaken to establish a structure-conductivity relationship for the polythiophenes, shows that the conductivity is related to the electronic descriptors which are calculated by ab-initio methods.

The reliability of this study has been tested by three different statistical methods: MLR, NN and CV.

A comparison of all the methods indicates that the NN is more reliable than others and has a high predictive power.

Finally QSAR is a broadly used tool for developing relationships between the effects (e.g.activities and properties of interest) of a series of molecules with their structural properties. It is used in many areas of science. It is a dynamic area that integrates new technologies at a staggering rate.

Acknowledgements

We thank Prof M Lahmouad for many fruitful discussions. Likewise, thank is due to the groups of theoretical chemistry, thank is due to A. Daoudi for the many help extended.

References

1. Torsi, L., Tafuri, A., Cioffi, N., Gallazzi, M.C., Sassella, A., Sabbatini, L., Zambonin, P.G. Regioregular polythiophene field-effect transistor employed as chemical sensors. *Sensors and Actuators, B: Chemical* 93 (2003) 257.
2. Andersson, M.R., Thomas, O., Mammo, W., Svensson, M., theander M., Inganas O., Substituted polythiophenes designed for optoelectronic devices and conductors, *J. Matter. Chem.* 9 (1999) 1933.
3. Salaneck, W.R., Friend, R.H., Brédas, J.L,. Electronic structure of conjugated polymers: Consequences of electron-lattice coupling. *Physics Repports* 319 (1999) 231.
4. Dora Demeter., Nouveaux Systèmes conjugués fonctionnels dérivés du thiophène. Thèse Doctorat (2008).4.
5. Ovidiu Ivanciuc., Ivanciuc T., Alexandru, Balaban, T,. QSAR models for the dermal penetration of polycyclic aromatic hydrocarbons ; *Internet Journal of Molecular Desing*, 1 (2002) 559.
6. The Mystat Inc., software Mystat, Press, L. Wilkinson, Chicago, 1982.
7. The Statlab SLP., software Statlab, Press, A.S.A. France, 1990-1997.
8. The mathworks Inc., Softwar Matlab, the Mathworks, Inc. Press, Natick, MA, 2000.
9. F. Liu., Y. Liang., Chenzhong Cao. QSPR modeling of thermal conductivity detection response factors for divers organic compound. *Chemometrics and Intelligent Laboratory Systems* 81(2006) 120.
10. Vivianna, C., Todeschini R., Milano chemometrics and QSAR research group. Bicocca, Piazza della scienza 1.120126, Milan, italy.
11. Bondi A., *J. Phys. Chem.*, 68 (1964) 441.
12. Todeschini, R., Consonni, V., Wiley- VCH verlog GmbH, Co. KGaA,. *Molecular descriptors for chimoinformatics*, (2000) P 251.
13. José, L,. Gazquez,. Perspectives one the density functional theory of chemical reactivity, *J. Mex. Chem. Soc.* 52 (2008) 3.
14. Nys, G.G,. Rekker, R.F.. The concept of hydrophobic fragmental constant (f-values). II. Extension of its applicability to the calculation of lipophilicities of aromatic and heteroaromatic structures. *Eur. J. Med. Chem.*, 9 (1974) 361.
15. Moratti S,. Handbook of conducting polymers, second Edition, New York, Skotheim A. T., Elsenbaumer R. L., Reynolds J.R. and references therein. (1998) Chap 13, p343.