Journal of Computational Methods in Molecular Design, 2014, 4 (3):72-83

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Two different antibacterial activities against *Staphylococcus aureus* and *Bacillus subtilis* of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazine derivatives. Studies by combining DFT and QSAR results

Azeddine Adad[1], Majdouline Larif[2], Rachid Hmamouchi[1], Mohammed Bouachrine[3] and Tahar Lakhlifi[1*]

[1]Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, University Moulay Ismail, Meknes, Morocco
[2]Separation Process Laboratory, Faculty of Science, University IbnTofail, Kenitra, Morocco
[3]ESTM, University Moulay Ismail, Meknes, Morocco

ABSTRACT

To establish a quantitative structure-activity relationship for antibacterial activity against Staphylococcus aureus and Bacillus subtilis, a series of seventeen 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazine derivatives molecules was submitted to a principal components analysis (PCA), to a multiple regression analysis (MRA), to a regression partial least squares (PLS), to a non-linear regression (RNLM) and to an artificial neural network (ANN). We accordingly propose a quantitative model, and we interpret the activity of the compounds relying on the multivariate statistical analysis. Density functional theory (DFT) and ab-initio molecular orbital calculations have been carried out in order to get insights into the structure, chemical reactivity and property information for the series of study compounds. The topological descriptors were computed with ACD/ChemSketch and Gaussian 03W program, respectively. This study shows that the MRA, PLS, and ANN have served also to predict activities, but when compared with the results given by the RNLM, we realized that the predictions fulfilled by this latter were more effective.

Keywords: QSAR, DFT, 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazine, *Staphylococcus aureus*, *Bacillus subtilis*.

INTRODUCTION

The development of simple synthetic routes to widely used organic compounds using readily available reagents is one of the main objectives of organic synthesis. Nitrogen heterocycles are of special interest because they constitute an important class of natural and non-natural products, many of which exhibit useful biological activities. Investigation of the 1,3-oxazine heterocycles has shown that they possess varied biological properties such as analgesic, anticonvulsant, antitubercular, antibacterial and anticancer activity. Particular attention has been paid to these compounds since the discovery of the non-nucleoside reverse transcriptase inhibitor trifluoromethyl-1,3-oxazine-2-one, which shows high activity against a variety of HIV-1 mutant strains. In addition, naphthoxazine derivatives have exhibited therapeutic potential for the treatment of Parkinson's disease [1].

Several lipopeptides have potent antibiotic activity and have been the subject of several studies on the discovery of new antibiotics. The list includes surfactin, produced by *B.subtilis*, the most powerful biosurfactant known to date. These compounds have many pharmacologicalactivities: antibacterial, antifungal, antiviral, and antimycoplasma properties; inhibition of the fibrin clot formation and hemolysis; formation of ion channels in lipid bilayer membranes; antitumor activity against Ehrlich's ascites carcinoma cells; and inhibition of the cyclic adenosine 3,5-monophosphate phosphodiesterase[2,3].

Quantitative Structure-Property/Activity Relationship (QSPR/QSAR) methods are among the most practical tools in computational physical chemistry. These methods are based on the axiom that the variance in the physicochemical properties and activities of chemical compounds is determined by the variance in their molecular structures. Thus, if experimental data are available for only some chemicals in a group, one can predict the missing from molecular descriptors calculated for the whole group and suitable mathematical model [4]. The global prediction of toxicity using QSAR has been the goal of many workers who utilized a variety of approaches. This goal is alluring, but has yet to be achieved satisfactorily. There are a number of reasons for the absence of success [5]. The deficiency of available toxicity data has clearly held back progress. This lack of success has been compounded in many studies by a poor appreciation of the insufficient heterogeneity, or chemical diversity, in the dataset. Further, while some molecular properties (such as hydrophobicity) are well described, others, including electrophilic reactivity, ionization, and hydrogen bonding, are poorly parameterized. Last, mechanisms of toxic action are not fully understood or misinterpreted, or their relevance in the modelling of toxicity is ignored [6].

We hereby report QSAR studies of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines synthesized in recent study [3]. To the best of our knowledge, this is the first report on the correlation of molecular descriptors with the antimicrobial activity of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines.

MATERIALS AND METHODS

Experimental data

Antecedent studies [3] had established a quantitative model of molecular-structure and antibacterial activitiespMIC$_{sa}$ and pMIC$_{bs}$ of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines.The following figure shows the chemicals structures of studied compounds optimized.

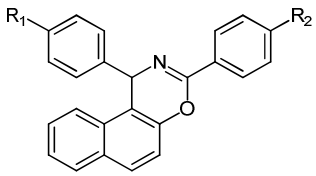

Figure 1: Chemical structures of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines

The experimental pMIC$_{sa}$ and pMIC$_{bs}$of the studied compounds have been collected from recent work [3] (Table1). The range of the **IC$_{50}$** data varies from 1.430 to 2.100 (μM).

Table 1: Molecular structures of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines used in QSAR studies

Compound	R_1	R_2	pMIC$_{sa}$	pMIC$_{bs}$
1	H	H	1.4300	2.0300
2	CH_3	CH_3	1.4600	2.0700
3	OCH_3	OCH_3	1.5000	2.1000
4	Cl	Cl	1.5100	1.8100
5	F	F	2.0700	1.7700
6	H	OCH_3	1.7700	1.7700
7	H	CH_3	1.7500	1.4500
8	H	NO_2	1.7800	1.7800
9	H	F	1.7700	1.7700
10	H	Br	1.7500	2.0500
11	CH_3	H	1.5200	1.8200
12	CH_3	OCH_3	1.4500	1.7500
13	CH_3	NO_2	1.7800	1.7800
14	CH_3	H	1.8000	1.8000
15	CH_3	F	1.4900	1.7900
16	CH_3	Br	1.4700	2.0700
17	H	Cl	1.8300	1.8300

Computational methods

An attempt has been made to correlate the activity of these compounds with various physicochemical parameters.

DFT (density functional theory) methods were used in this study. These methods have become very popular in recent years because they can reach similar precision to other methods in less time and less cost from the computational point of view. In agreement with the DFT results, energy of the fundamental state of a polyelectronic system can be expressed through the total electronic density, and in fact, the use of electronic density instead of wave function for calculating the energy constitutes the fundamental base of DFT [7-9] using the B3LYP functional [10,11] and a 6-31G (d) basis set. The B3LYP, a version of DFT method, uses Becke's three-parameter functional

(B3) and includes a mixture of HF with DFT exchange terms associated with the gradient corrected correlation functional of Lee, Yang and Parr (LYP). The geometry of all species under investigation was determined by optimizing all geometrical variables without any symmetry constraints.

The 3D structures of the molecules were generated using the Gauss View 3.0, and then, all calculations were performed using Gaussian 03W program series, Geometry optimization of thirteen compounds was carried out by B3LYP method employing 6–31G(d) basis set. ChemSketch program (Demo version 10.0) [12] was employed to calculate the others molecular descriptors.

Calculation of molecular descriptors

Calculation of descriptors using Gaussian 03W

Several quantum chemical methods and quantum-chemistry calculations have been performed in order to study the molecular structure and electronic properties [13-20], from the results of the DFT calculations, the quantum chemistry descriptors were obtained for the model building as follows: the total energy (ET (u.a)), the highest occupied molecular orbital energy (E_{HOMO} (eV)), the lowest unoccupied molecular orbital energy (E_{LUMO} (eV)), the energy difference between the LUMO and the HOMO energy (Gap (eV)), the total dipole moment of the molecule (μ(Debye)), absolute hardness (η), absolute electron negativity (χ), reactivity index (ω),total energy E_T, activation energy E_a, absorption maximum λ_{max} and factor of oscillation $f_{(SO)}$ [21].

η,χ andω were determined by the following equations:

$$\eta = \frac{(E_{LUMO} - E_{HOMO})}{2} \quad \chi = -\frac{(E_{LUMO} + E_{HOMO})}{2} \quad \omega = \frac{\chi^2}{2\eta}$$

Calculation of descriptors using ACD/ChemSketch

Advanced chemistry development's ACD/ChemSketch program [12] was used to calculate Formula Weight (PM), Molar Volume (MV (cm^3)), Molecular Weight (MW), Molar Refractivity (MR (cm^3)), Parachor (Pc (cm^3)), Density (D (g/cm^3)), Refractive Index (n), Surface Tension (γ(dyne/cm),Polarizability (αe (cm^3)) and octanol/water partition coefficientsLogP.

- ***Molecular Weight*(M):** Used as the descriptor in systems such as transport studies where diffusion is the mode of operation. It is an important variable in QSAR studies pertaining to cross resistance of various drugs in multi-drug resistant cell lines.

- ***Molar Volume* (Vm):** The molar volume calculates from additive increments. The additive atomic increments were obtained using a database of density and calculated M:

$$Vm = \frac{M}{d}$$

- ***Density*(d):**The density is calculates from M and the calculated molar volume:

$$d = \frac{M}{Vm}$$

- ***Molar Refractivity* (A):** The Lorentz-Lorenz equation relates Molecular weight, density, and refractive index:

$$A = \frac{n^2 - 1}{n^2 + 2}\frac{M}{d}.$$

- ***Parachor*(P):** The parachor is calculates from additive increments. The additive atomic increments were obtained using a database of density, surface tension, and calculated MW:

$$P = \left(\frac{M}{d}\right) S^{\frac{1}{4}}$$

- ***Refractive Index* (n):** By the Lorentz-Lorenz equation:

$$n = \sqrt{\frac{2\,A + M}{Vm - A}}$$

The refractive index calculates from the molar volume and molar refractivity, both of which are calculated as above.

- ***Surface Tension* (S):** Calculated from calculated Vm and calculated Pc:

$$S = \left(\frac{P}{Vm} \right)^4$$

- ***Polarizability*(αe):**Calculates from the Molar Refractivity as follows:

$$\alpha e = 0{,}3964308 \times A$$

- ***Partition coefficients LogP:***The partition coefficient is a ratio of concentrations of un-ionized compound between the two solutionsoctanol/water:

$$\text{Log}\,P = \text{Log}\left(\frac{[solute]_{octanol}}{[solute]_{water}}\right)$$

Statistical analysis

To explain the structure-activity relationship, these 20 descriptors are calculated for 17 molecules using the Gaussian 03W, Gauss View and ChemSketch software.

The study we conducted consists of:

- The principal component analysis (PCA) available in a software called XLSTAT.
- The multiple linear regressions (MLR) available in the XLSTAT software.
- The regression partial least squares (PLS) available in the XLSTAT software.
- The non-linear regression (RNLM) available in XLSTAT software.
- The Neural Network (RN) available in the software MATLAB Version 9.

The structures of the molecules based on 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines, (1–17) were studied by statistical methods based on the principal component analysis (PCA)using the software XLSTAT version Demo 2009 [22]. PCA is a statistical technique useful for summarizing all the information encoded in the structures of the compounds. It is also very helpful for understanding the distribution of the compounds [23]. This is an essentially descriptive statistical method which aims to present, in graphic form, the maximum of information contained in the data table 1 and table 2.

The multiple linear regression (MLR) analysis with descendent selection and elimination of variables was employed to model the structure activity relationships. It is a mathematic technique that minimizes differences between actual and predicted values. It has served also to select the descriptors used as the input parameters in the partial least squares (PLS), and the Multiples nonlinear regression (MNLR) and artificial neural network (ANN).

The (MLR), the (PLS), and the (MNLR) were generated using the software XLSTAT version Demo 2009 [12], to predict cytotoxic effects IC_{50} activities. Equations were justified by the correlation coefficient (R), mean squared error (MSE), fishers F-statistic (F), and significance level (F value).

ANN is artificial systems simulating the function of the human brain. Three components constitute a neural network: the processing elements or nodes, the topology of the connections between the nodes, and the learning rule by which new information is encoded in the network. While there are a number of different ANN models, the most frequently used type of ANN in QSAR is the three-layered feed-forward network [17]. In this type of networks, the neurons are arranged in layers (an input layer, one hidden layer and an output layer). Each neuron in any layer is fully connected with the neurons of a succeeding layer and no connections are between neurons belonging to the same layer.

Table 2: The values of the sixteen chemical descriptors

N°	E_t(eV)	E_{HOMO}(eV)	E_{LUMO}(eV)	ΔE(eV)	M	E_a(eV)	λ_{max}(nm)	f (SO)	χ	η	ω	A	Vm	P	n	S	d	αe	M	Log P
1	-28700.5883	-5.7555	-1.2353	4.5201	1.0021	3.9741	311.9800	0.0137	3.4954	2.2601	2.7030	104.8500	287.4000	745.9000	1.6500	45.3000	1.1800	41.5600	339.1623	5.2400
2	-30840.2875	-5.6964	-1.1630	4.5335	0.7204	3.9962	310.2500	0.0138	3.4297	2.2667	2.5946	113.6900	317.8000	808.1000	1.6340	41.8000	1.1500	45.0700	367.1936	6.4200
3	-34932.9463	-5.6393	-1.0819	4.5574	0.7057	4.0120	309.0300	0.0061	3.3606	2.2787	2.4780	116.4700	330.7000	846.4000	1.6220	42.9000	1.2000	46.1700	399.1834	4.0400
4	-53711.8108	-5.9772	-1.5208	4.4565	3.4095	3.8964	318.2000	0.0133	3.7490	2.2282	3.1538	114.0500	306.0000	803.6000	1.6680	47.5000	1.3300	45.2100	407.0844	5.9600
5	-34100.8808	-5.8861	-1.3458	4.5403	2.5407	4.0029	309.7400	0.0147	3.6159	2.2701	2.8798	104.5900	293.1000	746.2000	1.6320	41.9000	1.2800	41.4600	375.1435	5.4400
6	-31816.7836	-5.6940	-1.1322	4.5618	2.0953	4.0373	307.1000	0.0170	3.4131	2.2809	2.5537	110.6600	309.0000	796.2000	1.6340	44.0000	1.1900	43.8700	369.1729	4.7800
7	-29770.4420	-5.7228	-1.1896	4.5332	0.7747	3.9954	310.3200	0.0143	3.4562	2.2666	2.6351	109.2700	302.6000	777.0000	1.6410	43.4000	1.1600	43.3200	353.1780	5.8300
8	-34265.0578	-6.0191	-2.5898	3.4293	6.3936	3.0159	411.1000	0.0055	4.3045	1.7146	5.4031	110.5100	292.7000	791.4000	1.6790	53.4000	1.3100	43.8100	384.1474	4.3400
9	-41206.2009	-5.8567	-1.4051	4.4516	2.9244	3.8914	318.6100	0.0130	3.6309	2.2258	2.9615	109.4500	296.7000	774.7000	1.6590	46.4000	1.2500	43.3900	373.1233	5.7700
10	-31400.7413	-5.8110	-1.2805	4.5305	2.1922	3.9889	310.8200	0.0139	3.5457	2.2652	2.7750	104.7200	290.3000	746.1000	1.6400	43.6000	1.2300	41.5100	357.1529	5.5100
11	-98660.3526	-5.8559	-1.4114	4.4445	2.8136	3.8838	319.2400	0.0134	3.6336	2.2222	2.9707	112.4000	299.9000	789.4000	1.6720	47.9000	1.3900	44.5600	417.0728	5.9400
12	-42276.0491	-5.8292	-1.3779	4.4513	3.0472	3.8918	318.5800	0.0126	3.6036	2.2256	2.9173	109.2700	302.6000	777.0000	1.6410	43.4000	1.1600	43.3200	353.1780	5.8300
13	-32886.6291	-5.6670	-1.1069	4.5601	1.9084	4.0347	307.2900	0.0160	3.3870	2.2801	2.5156	115.0800	324.2000	827.2000	1.6280	42.3000	1.1800	45.6200	383.1885	5.3700
14	-35334.9087	-5.9911	-2.5722	3.4189	6.5390	3.0060	412.4500	0.0056	4.2816	1.7095	5.3620	114.9300	307.8000	822.4000	1.6690	50.9000	1.2900	45.5600	398.1630	4.9300
15	-29770.4338	-5.7277	-1.2089	4.5188	1.0468	3.9735	312.0300	0.0132	3.4683	2.2594	2.6621	109.2700	302.6000	777.0000	1.6410	43.4000	1.1600	43.3200	353.1780	5.8300
16	-32470.5855	-5.7838	-1.2530	4.5307	2.2970	3.9900	310.7400	0.0134	3.5184	2.2654	2.7322	109.1400	305.5000	777.2000	1.6330	41.8000	1.2100	43.2600	371.1685	6.1000
17	-99730.2008	-5.8281	-1.3847	4.4434	2.9359	3.8832	319.2800	0.0129	3.6064	2.2217	2.9271	116.8300	315.1000	820.5000	1.6630	45.9000	1.3700	46.3100	431.0885	6.5300

Table3:The correlation matrix (Pearson (n)) between different obtained descriptors

Variables	Et	E_{HOMO}	E_{LUMO}	ΔE	μ	Ea	λmax	f (SO)	χ	η	ω	A	Vm	P	n	S	d	αe	M	Log P	pMICsa	pMICbs
Et(eV)	**1**	0,250	0,028	-0,044	-0,166	-0,010	0,037	-0,040	-0,075	-0,044	0,011	-0,420	-0,092	-0,240	**-0,522**	-0,280	**-0,775**	-0,420	**-0,768**	-0,416	-0,015	0,094
E_{HOMO}		**1**	**0,823**	**0,708**	**-0,874**	**0,730**	**-0,713**	0,435	**-0,884**	**0,708**	**-0,788**	0,058	**0,507**	0,152	**-0,817**	**-0,785**	**-0,698**	0,058	-0,347	0,021	-0,288	0,224
E_{LUMO}			**1**	**0,984**	**-0,939**	**0,988**	**-0,985**	**0,729**	**-0,993**	**0,984**	**-0,998**	-0,141	0,270	-0,134	**-0,731**	**-0,885**	-0,475	-0,141	-0,281	0,348	-0,277	0,184
ΔE(eV)				**1**	**-0,892**	**0,999**	**-1,000**	**0,769**	**-0,956**	**1,000**	**-0,992**	-0,193	0,176	-0,214	**-0,651**	**-0,853**	-0,371	-0,193	-0,240	0,427	-0,254	0,159
μ					**1**	**-0,902**	**0,894**	**-0,583**	**0,954**	**-0,892**	**0,926**	0,167	-0,250	0,139	**0,732**	**0,843**	**0,593**	0,167	0,400	-0,264	0,361	-0,216
Ea(eV)						**1**	**-0,999**	**0,774**	**-0,964**	**0,999**	**-0,994**	-0,200	0,187	-0,211	**-0,681**	**-0,869**	-0,398	-0,200	-0,259	0,403	-0,239	0,162
λmax							**1**	**-0,774**	**0,958**	**-1,000**	**0,993**	0,195	-0,177	0,215	**0,657**	**0,856**	0,377	0,196	0,245	-0,424	0,248	-0,156
f (SO)								**1**	**-0,689**	**0,769**	**-0,742**	-0,343	-0,113	-0,408	-0,391	**-0,630**	-0,260	-0,343	-0,311	**0,567**	0,035	-0,160
χ									**1**	**-0,956**	**0,984**	0,103	-0,328	0,078	**0,771**	**0,891**	**0,536**	0,103	0,303	-0,290	0,287	-0,198
η										**1**	**-0,992**	-0,193	0,176	-0,214	**-0,651**	**-0,853**	-0,371	-0,193	-0,240	0,427	-0,254	0,159
ω											**1**	0,150	-0,243	0,156	**0,698**	**0,873**	0,439	0,150	0,263	-0,386	0,279	-0,174
A												**1**	**0,837**	**0,970**	0,163	0,212	0,275	**1,000**	**0,742**	-0,017	-0,105	0,009
Vm													**1**	**0,893**	-0,403	-0,303	-0,164	**0,837**	0,430	-0,069	-0,147	0,164
P														**1**	0,018	0,158	0,139	**0,970**	**0,648**	-0,214	-0,078	0,059
n															**1**	**0,912**	**0,753**	0,163	0,462	0,066	0,080	-0,266
S																**1**	**0,636**	0,212	0,412	-0,321	0,153	-0,222
d																	**1**	0,275	**0,819**	0,083	0,319	-0,089
αe																		**1**	**0,741**	-0,017	-0,105	0,008
M																			**1**	0,041	0,200	0,004
Log P																				**1**	-0,178	-0,084
pMICsa																					**1**	-0,448
pMICbs																						**1**

- **ΔE(eV)** and **η** are perfectly correlated (r = 1).
- **A** and**αe** are perfectly correlated (r = 1).
- **A**, **P** and **αe** are highly correlated (r (A, Pc) = 0.970, r (P, αe) = 0.970).
- $\mathbf{E_{LUMO}}$ and **ω** are strongly negatively correlated (r = -0.998).

According to the supervised learning adopted, the networks are taught by giving them examples of input patterns and the corresponding target outputs. Through an iterative process, the connection weights are modified until the network gives the desired results for the training set of data. A back-propagation algorithm is used to minimize the error function. This algorithm has been described previously with a simple example of application [18] and a detail of this algorithm is given elsewhere [19].

RESULTS

Data set for analysis

The QSAR analysis was performed using the IC_{50} of the 17 compounds against the *Staphylococcus aureus*, *Bacillus subtilise*($pMIC_{sa}$ and $pMIC_{bs}$). (Experimental values) as reported in [3], the values of the 17 chemical descriptors as shown in table 2.

Principal component analyses(PCA)

The totality of the twenty descriptors (variables) coding the sixteen molecules was submitted to a principal components analysis (PCA). The first three axes F1, F2 and F3 contributing respectively 51.16%, 20.59 % and 12.70 % to the total variance, the total information is estimated to a percentage of 91.3%, were sufficient to describe the information represented by the data set. Correlations between the sixteen descriptors are shown in table 3 as a correlation matrix, in figure 2 these descriptors are represented in a correlation circles.

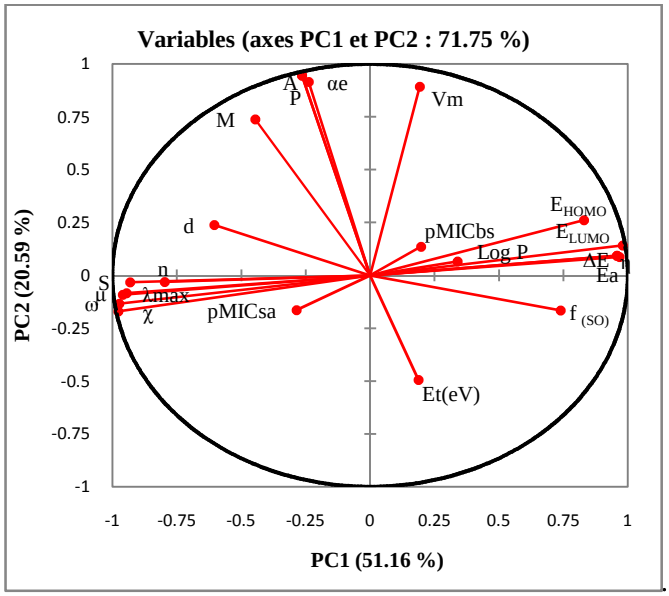

Figure2: Correlation circle

On the other hand, the projection PC1-PC2 (71,75% of the total variance) also shows that we can discern six groups of molecules with special structures propriety.

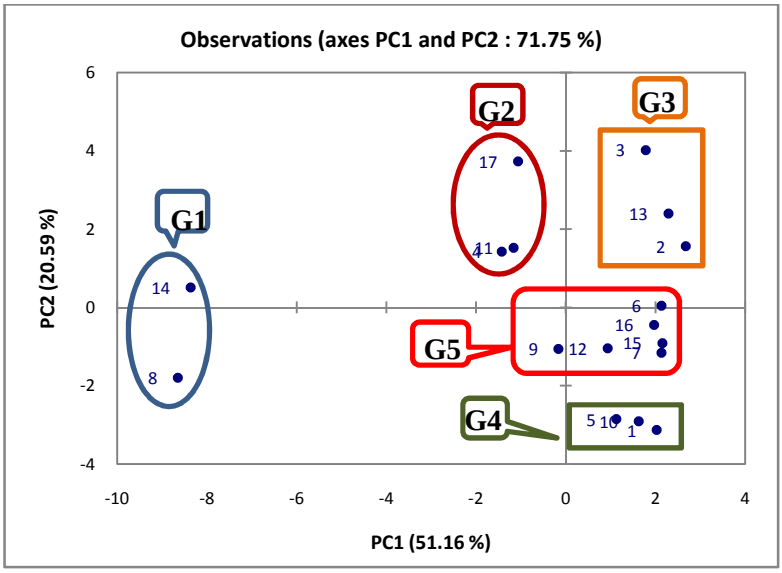

Figure3: Cartesian diagram according to PC1 and PC2: Separation between five groups.

Multiple Linear Regressions(MLR)

In order to propose a mathematical model and to evaluate quantitatively the substituent's physicochemical effects on the two activities of the totality of the set of these 17 molecules, we submitted the data matrix constituted obviously from the 20 physicochemical variables corresponding to the 17 molecules, to a progressive multiple regression analysis. This method used the coefficients R, R^2, and the F-values to select the best regression performance.
Where R is the correlation coefficient; R^2 is the coefficient of determination; MSE is the mean squared error; F is the Fisher F-statistic.

The QSAR models built using multiple linear regression (MLR) method is represented by the following equation:

pMICsa = 14,135 + 1,710 $\mathbf{E_a}$ + 2,512x $10^{-02}\boldsymbol{\lambda_{max}}$ + 60,057 $\mathbf{f_{(SO)}}$+ 5,930 x $10^{-02}\boldsymbol{\chi}$+ 10^{-02}**A**- 1,310x $10^{-02}\mathbf{V_m}$ - 0,037 **P** - 3,040 **n** + 6,057 x 10^{-02}**S** -19,856 **d** + 7,494 x 10^{-02}**M**- 0,205 **LogP**. (Equation 1)

pMICbs = 74,898 - 4,873 $\mathbf{E_a}$ - 5,810 x $10^{-02}\boldsymbol{\lambda_{max}}$ - 94,816 $\mathbf{f_{(SO)}}$ - 0,923 $\boldsymbol{\chi}$ - 0,612 **A**- 0,182 $\mathbf{V_m}$+ 0,229 **P** - 53,165 **n** - 0,182 **S** +51,480**d** - 0,167 **M** + 1,273 **LogP**. (Equation 2)

The Fisher's F test is used. Given the fact that the probability corresponding to the F value is lower than 0.05 for $pMIC_{sa}$, it means that we would be taking a lower than 0.28% risk in assuming that the null hypothesis is wrong. Therefore, we can conclude with confidence that the models do bring a significant amount of information. For $pMIC_{bs}$, the F value (F value = 0.265) is up than 0.05, the model is not significant. (Tables 4 and 5)

Table4: Analysis of variance (global model)

	Source	DDL	Sum of squares	Mean square	F	Pr> F
pMICsa	Model	12	0.343	0.029	0.540	0.816
	Error	4	0.212	0.053		
	total corrected	16	0.555			
pMICbs	**Source**	**DDL**	**Sum of squares**	**Mean square**	**F**	**Pr> F**
	Model	12	0.279	0.023	0.580	0.791
	Error	4	0.160	0.040		
	total corrected	16	0.440			

Table5: Correlation coefficient (R), Coefficient of determination (R^2), Mean squared error (MSE), Fishers F-statistic (F) and Significances level (F value)

	$pMIC_{sa}$	$pMIC_{bs}$
R^2	0,618	0.635
R	0.786	0.797
MSE	0.053	0.040
F	0.540	0.580
F value	0.816	0.791

The values of predicted activities ($pMIC_{sa}$) and ($pMIC_{bs}$) calculated from equations (1and 2), and the observed values are given in table 10. The correlations of predicted and observed are illustrated in figure 4.

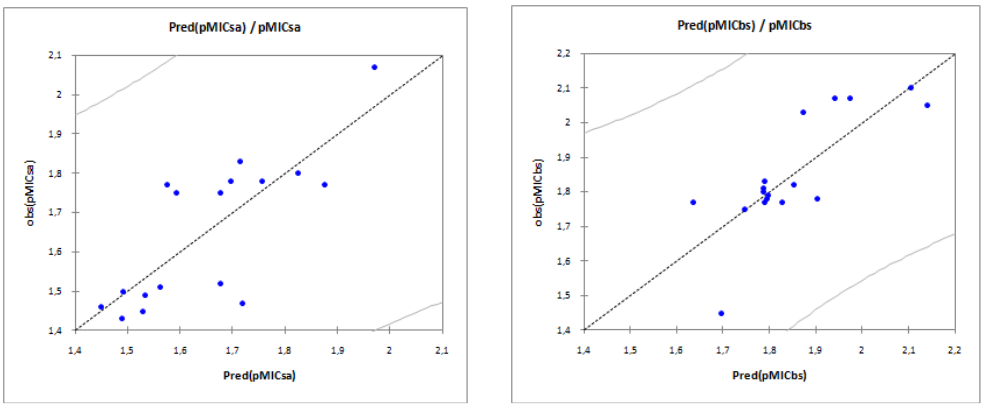

Figure 4: Correlations of observed and predicted activities calculated using MLR

The descriptors proposed in equations (1and 2) by MLR were, therefore, used as the input parameters in the partial least squares (PLS), and the Multiples nonlinear regression (MNLR) and artificial neural network (ANN).

Partial least squares PLS

Partial Least Squares regression (PLS) is an efficient and optimal for a criterion method based on covariance. It is recommended in cases where the number of variables is high, and where it is likely that the explanatory variables are correlated.

We submitted the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 17 molecules, to the partial least squares (PLS). This method used the coefficients R, R^2, and the F-values to select the best regression performance.

The QSAR models built using partial least squares (PLS) method is represented by the following equation:

pMICsa = 1,146 + 5,835 x 10^{-08} **Et** - 6,417x 10^{-02}**E_{HOMO}** - 1,423 x 10^{-02} **E_{LUMO}** - 0,017**ΔE** + 0,003 **μ** - 2,042x 10^{-02}**E_a** +1,966 x 10^{-04}**λ_{max}** - 2,535 **$f_{(SO)}$**+ 2,354 x 10^{-02} **χ**-0,035 **η** + 7,146 x 10^{-03} **ω** - 2,636 x 10^{-03} **A** - 9,974 x 10^{-04}**V_m** -2,451 x 10^{-04} **P** + 0,500 **n** + 1,908 x 10^{-03} **S**+ 0,127 **d** - 0,006 **αe** + 4,277 x10^{-04} **M** - 1,196 x 10^{-02}**Log P.** (Equation 3)

pMICbs = 2,053 - 3,019 x 10^{-08} **Et** + 0,033 **E_{HOMO}** + 7,367 x 10^{-03} **E_{LUMO}** + 9,242 x 10^{-03} **ΔE** - 1,808 x 10^{-03}**μ** + 1,057 x 10^{-02}**E_a** -1,017 x 10-04**λ_{max}** +1,311 **$f_{(SO)}$** - 1,218 x 10^{-02} **χ** + 1,848x 10^{-02} **η** - 3,697x 10^{-03} **ω** + 1,363x 10^{-03} **A** + 5,160x 10^{-04}**V_m** + 1,268x 10^{-04} **P** - 0,258 **n** -9,873x 10^{-04} **S** - 6,622 x 10^{-02} **d** + 3,448 x 10^{-03} **αe** - 2,212x 10^{-04} **M** + 6,188 x10^{-03} **Log P.** (Equation 4)

The correlation coefficient (R), coefficient of determination (R^2), Mean Squared Error (MSE) and Standard deviation (S) for the two modelsare illustrated in table 6.

Table6: Correlation coefficient (R), Coefficient of determination (R^2), Mean squared error (MSE) and Standard deviation (S).

	pMICsa	pMICbs
R^2	0.926	0.276
R	0.962	0.525
MSE	0.001	0.012
S	0.037	0.110

The values of predicted activities ($pMIC_{sa}$) and ($pMIC_{bs}$) calculated from equations (3 and 4), and the observed values are given in table 9. The correlations of predicted and observed are illustrated in figure 5.

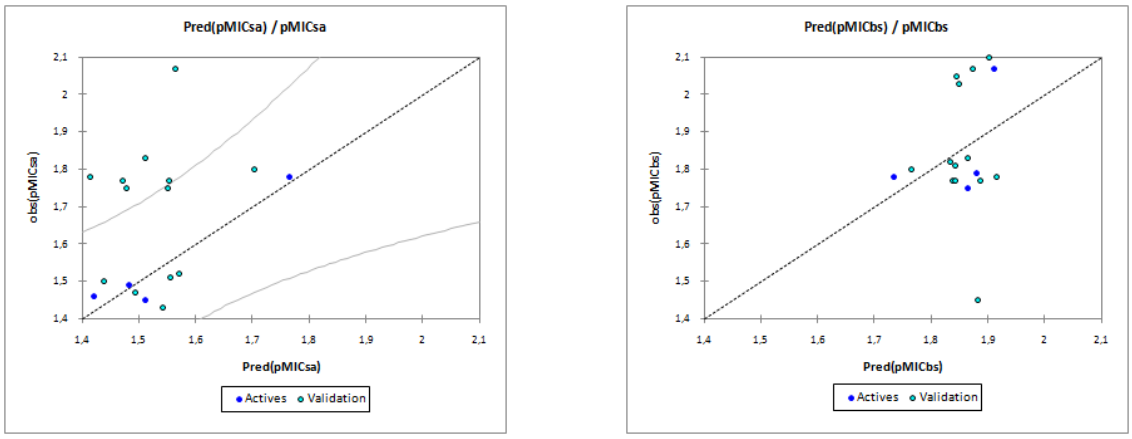

Figure5: Correlations of observed and predicted activities calculated using PLS

Squares (PLS), it is likely that any non-linear relationship took place. Nonlinear regression performed by XLSTAT software and the neural network are suitable concepts to accomplish this task.

Multiples nonlinear regression (MNLR)

We have used also the technique of nonlinear regression model to improve the structure - activity relationship to quantitatively evaluate the effect of substituent. It takes into account several parameters. This is the most common tool for the study of multidimensional data. We have applied to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 13 molecules. The coefficients R, R^2, and the F-values are used to select the best regression performance.

We used a pre-programmed function of XLSTAT following:

$$\mathbf{Y = a + (b\,X_1 + c\,X_2 + d\,X_3 + e\,X_4 \ldots) + (f\,X_1^2 + g\,X_2^2 + h\,X_3^2 + i\,X_4^2 \ldots)}$$

Where A, b, c, d,...: represent the parameters and X_1, X_2, X_3, X_4,...: represent the variables.

The resulting equations were:

pMICsa = -5146,432 - 25,979 **A** + 1,785 $\mathbf{V_m}$ + 4,252 x 10^{-02}**P** + 6096,303 **n** + 1,644 **S** -23,372 **d** + 49,012 **αe** + 0,102 **M** + 3,008 **Log P** + 2,316 x $10^{-02}$$\mathbf{A^2}$ - 7,491 x 10^{-04} $\mathbf{V_m^2}$ - 2,557 x 10^{-04} $\mathbf{P^2}$ - 1803,350 $\mathbf{n^2}$ + 3,436 x 10^{-04} $\mathbf{S^2}$ - 0,290 **Log** $\mathbf{P^2}$.(Equation 5)

pMICbs = 1872,319 + 12,309 **A** - 1,794 $\mathbf{V_m}$ + 0,465 **P** - 2295,463 **n** - 1,550 **S** + 65,330 **d**- 23,490 **αe** - 0,223 **M** - 5,133 **Log P** - 1,741 10-02 $\mathbf{A^2}$ +1,128 10-03 $\mathbf{V_m^2}$ + 1,125 10-04 $\mathbf{P^2}$ + 697,713 $\mathbf{n^2}$ - 6,248 10-03 $\mathbf{S^2}$ + 0,593 **Log** $\mathbf{P^2}$.(Equation 6)

The correlation coefficient (R), coefficient of determination (R^2), Mean Squared and Error (MSE) for the two models are illustrated in table 7.

Table7: Correlation coefficient (R), Coefficient of determination (R^2), and Mean squared error (MSE)

	$pMIC_{sa}$	$pMIC_{bs}$
R^2	0.998	0.998
R	0.999	0.999
MSE	0.0001	0.0001

The values of predicted activities calculated from equations (5and 6), and the observed values are given in table 9. The correlations of predicted and observed are illustrated in figure 6.

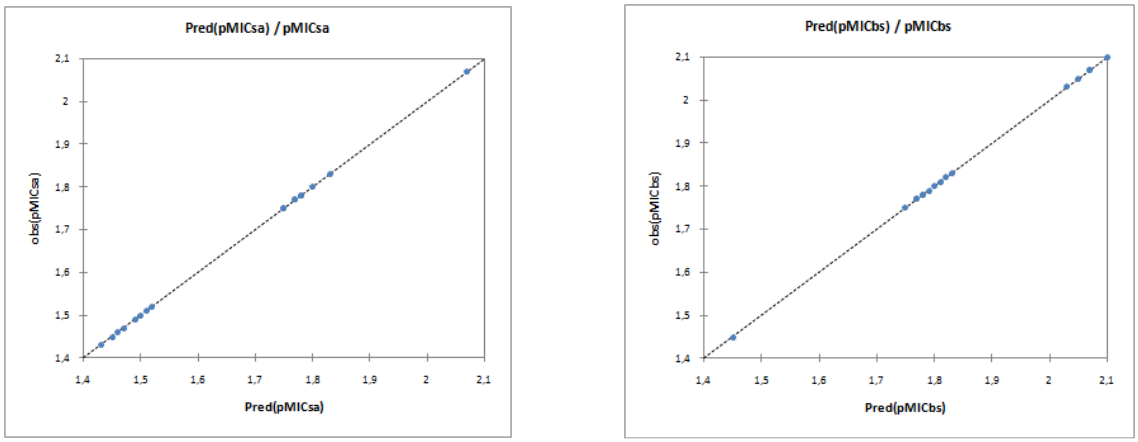

Figure6: Correlations of observed and predicted activities calculated using MNLR

Artificial neural networks (ANN)

Neural networks (ANN) can be used to generate predictive models of quantitative structure-activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR and observed activity.

The correlations coefficients and Standard Error of Estimate, obtained with the Neural network (Table 8), show that the selected descriptors by MLR are pertinent and that the model proposed to predict activity is relevant.

Table8: Correlation coefficient (R) and Coefficient of determination (R^2)

	Samples	R	R^2
Training	12	0,904	0,817
Validation	3	0,887	0,787
Test	2	0,801	0,641

The values of predicted activities and the observed values are given in table 9.

The obtained squared correlation coefficient (R^2) value confirms that the MNLR result were the best to build the quantitative structure activity relationship models.

In this part, we investigated the best linear QSAR regression equations established in this study. Based on this result, a comparison of the quality of ACP, MLR, PLS, MNLR and ANN models shows that the MNLR models have substantially better predictive capability because the MNLR approach gives better results than MLR, PLS and ANN. MNLR was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds.

The values of predicted activities calculated using ANN and the observed values are given in table 9. The correlations of predicted and observed are illustrated in figure 7.

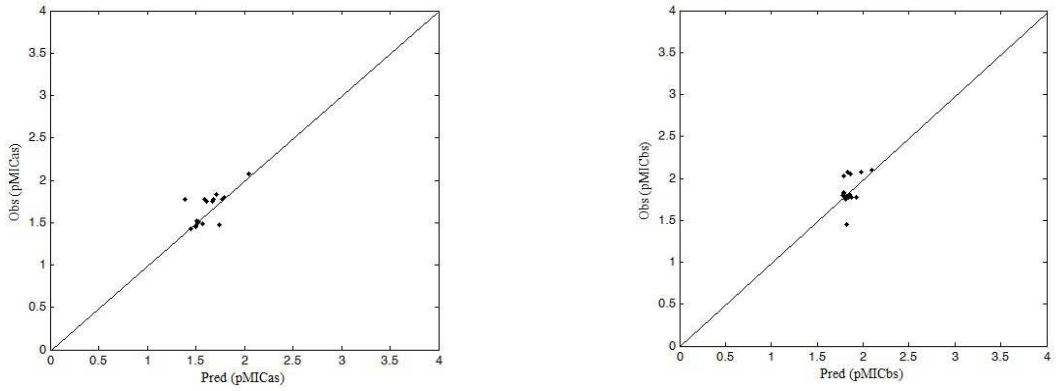

Figure7: Correlations of observed and predicted activities calculated using ANN

DISCUSSION

The principle (for the two studies) is to perform in the first time, a main component analysis (PCA), which allows us to eliminate descriptors that are highly correlated (dependent), then perform a decreasing study of MLR based on the elimination of descriptors (one by one) aberrant until a valid model (including the critical probability: p-value<0.05 for all descriptors and the model complete.

The figure 3(PCA) shows a distribution of molecules in six groups: the group 1 (G_1) which has $R_2 = NO_2$, the group 2 (G_2) variously substituted by halogens (Br and Cl), the group 3 (G_3) which substituted by - O-CH_3 and CH_3, the group 4 (G_4) fluor-substitued in R_2 and the group 5 (G_5) which contains the rest of the compounds.

Comparison of key statistical terms like R or R^2 of different models obtained by using different statistical tools and different descriptors has been shown in table 9.

Table9: Observed, predicted activities and residue according to different methods.

In vitro antimicrobial activity IC_{50} values (µM)									
	$pMIC_{sa}$								
	Observed	MLR		PLS		MNLR		ANN	
		Pred.	Resid.	Pred.	Resid.	Pred.	Resid.	Pred.	Resid.
1	1.430	1.489	-0.059	1.543	0.027	1.430	0.000	1.4531	-0.0231
2	1.460	1.449	0.011	1.420	0.035	1.460	0.000	1.5019	-0.0419
3	1.500	1.491	0.009	1.439	0.030	1.500	0.000	1.5222	-0.0222
4	1.510	1.561	-0.051	1.555	0.027	1.510	0.000	1.5346	-0.0246
5	2.070	1.970	0.100	1.564	0.026	2.070	0.000	2.0490	0.0210
6	1.770	1.876	-0.106	1.471	0.030	1.770	0.000	1.5959	0.1741
7	1.750	1.593	0.157	1.477	0.030	1.750	0.000	1.6093	0.1407
8	1.780	1.756	0.024	1.765	0.052	1.780	0.000	1.7677	0.0123
9	1.770	1.575	0.195	1.553	0.026	1.770	0.000	1.3932	0.3768
10	1.750	1.676	0.074	1.552	0.027	1.750	0.000	1.6767	0.0733
11	1.520	1.676	-0.156	1.571	0.029	1.520	0.000	1.5108	0.0092
12	1.450	1.530	-0.080	1.511	0.027	1.450	0.000	1.5033	-0.0533
13	1.780	1.698	0.082	1.415	0.034	1.780	0.000	1.6780	0.1020
14	1.800	1.824	-0.024	1.704	0.045	1.800	0.000	1.7935	0.0065
15	1.490	1.533	-0.043	1.484	0.030	1.490	0.000	1.5752	-0.0852
16	1.470	1.719	-0.249	1.494	0.029	1.470	0.000	1.7418	-0.2718
17	1.830	1.714	0.116	1.512	0.026	1.830	0.000	1.7106	0.1194
	$pMIC_{bs}$								
1	2.0300	1.872	0.158	1.848	0.080	2.030	0.000	1.7936	0.2364
2	2.0700	1.974	0.096	1.912	0.103	2.070	0.000	1.9794	0.0906
3	2.1000	2.106	-0.006	1.902	0.090	2.100	0.000	2.0962	0.0038
4	1.8100	1.788	0.022	1.842	0.080	1.810	0.000	1.8503	-0.0403
5	1.7700	1.790	-0.020	1.838	0.078	1.770	0.000	1.8433	-0.0733
6	1.7700	1.635	0.135	1.886	0.090	1.770	0.000	1.8773	-0.1073
7	1.4500	1.696	-0.246	1.882	0.090	1.450	0.000	1.8196	-0.3696
8	1.7800	1.794	-0.014	1.734	0.153	1.780	0.000	1.8078	-0.0278
9	1.7700	1.826	-0.056	1.843	0.078	1.770	0.000	1.8775	-0.1075
10	2.0500	2.141	-0.091	1.844	0.079	2.050	0.000	1.8586	0.1914
11	1.8200	1.854	-0.034	1.834	0.085	1.820	0.000	1.7906	0.0294
12	1.7500	1.746	0.004	1.865	0.081	1.750	0.000	1.8121	-0.0621
13	1.7800	1.903	-0.123	1.915	0.101	1.780	0.000	1.9225	-0.1425
14	1.8000	1.786	0.014	1.765	0.133	1.800	0.000	1.7765	0.0235
15	1.7900	1.796	-0.006	1.879	0.088	1.790	0.000	1.8191	-0.0291
16	2.0700	1.941	0.129	1.874	0.085	2.070	0.000	1.8332	0.2368
17	1.8300	1.791	0.039	1.865	0.079	1.830	0.000	1.7908	0.0392

CONCLUSION

In this work we have investigated the QSAR regression to predict toxicity of several compounds based on 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines.

The studies of the quality of the MLR, PLS, RNLM and ANN models have shown that:

- The PLS method gave low coefficients of determination (R^2), thus it was had no efficiency in predicting the values of activities.
- The nonlinear regression and the neural network ANN results have substantially better predictive capability than the other methods.
- With ANN approach, we have established a relationship between several descriptors and antimicrobial activity values (IC_{50}) of 1,3-disubstituted-1H-naphtho[1,2-e][1,3]oxazines against *Staphylococcus aureus* and *Bacillus subtilis* (($pMIC_{sa}$) and ($pMIC_{bs}$))in satisfactory manners.

Finally, we can conclude that studied descriptors, which are sufficiently rich in chemical, electronic and topological information to encode the structural feature may be used with other descriptors for the development of predictive QSAR models.

Acknowledgment
We are grateful to the "Association Marocaine des ChimistesThéoriciens" (AMCT) for its pertinent help concerning the programs.

REFERENCES

[1] Z Turgut; EPelit; A Köycü, *Molecules*, **2007**, 12, 345-352.
[2] P André; V. Fernandes et al., *Brazil. J. Microbiol.*, **2007**, 38, 704-709.
[3]V Verma et al., *Arab. J. Chem.*, **2012**, in press.
[4] E Zvinanashe; TT Du; T Griff, *Chemosphere*, **2009**, 75, 1531-1538.
[5] MJ Zhu; F Ge; RL Zhu, *Chemosphere*, **2010**, 80, 46-52.
[6] A Adad; R Hmamouchi; AI Taghki; AAbdellaoui; M Bouachrine, T Lakhlifi. *J. Chem. Pharm. Res.*,**2013**,5 (7), 28-41.
[7] C Adamo; V Barone, *Chem. Phys. Lett.*, **2000**, 330, 152–160.
[8] M Parac; S Grimme, *J. Phys. Chem.*, **2003**, A106, 6844–6850.
[9] M Parac; S Grimme, All calculations were done with Gaussian 03W software, *J. Phys. Chem.* A, **2003**, 106, 6844-6850.
[10] AD Becke,*J. Chem. Phys*.,**1993**, 98, 1372.
[11] C Lee; W Yang; R G Parr, *Phys. Rev.*, **1988**, B37, 785-789.
[12] Advanced Chemistry Development Inc., Toronto, Canada **2009**.
[13] K Laarej; M Bouachrine; S Radi; S Kertit; B Hammouti, *e-J. Chem.*, **2010**, 7(2), 419-424.
[14] H Zarrok; H Oudda; A Zarrouk; R Salghi; B Hammouti; M Bouachrine, *Der Pharma Chemica*, **2011**,3 (6), 576-590.
[15] M Larif; A Adad; R Hmamouchi; A I Taghki; A Soulaymani; A Elmidaoui; M Bouachrine; T Lakhlifi, *Arab. J. Chem.*, **2013**, in press.
[16] A Adad; M Larif; R Hmamouchi; A IdrissiTaghki; M Bouachrine; T Lakhlifi, *J. Chem. Acta*, **2013**, 2, 105-118.
[17] G Hea; L Fenga; H Chena, *International Symposium on Safety Science and Engineering in China, Proc.Engin.*, **2012**, 43, 204–209.
[18] D Cherquaoui; D Villemin, *J. Chem. Soc. Far. Trans.*, **1994**, 90, 97-102.
[19] JA Freeman; DM Skapura, *J. Op. Res. Soc.*, **1991**, 43(11), 1106.
[20] A Ousaa; B Elidrissi; M Ghamali; S Chtita; M Bouachrine; T Lakhlifi, *JCMMD*,**2014,** in press.
[21] U Sakar; R Parthasarathi; V Subramanian ; PK Chattaraji, *Internet Electronic J. Mol. Design*, **2004**, 1-24.
[22] XLSTAT.software (XLSTAT Company). http://www.xlstat.com., **2009.**
[23] S Chtita; M Larif; M Ghamali; A Adad; R Hmamouchi; M Bouachrine, T Lakhlifi, *IJIRSET*, **2013**, 2 (11), 6586-6601.