Research Article

Organic Compounds Based on (E)-N-Aryl-2-ethene-sulfonamide as Microtubule Targeted Agents in Prostate Cancer: QSAR Study

El Ghalia Hadaji,¹ Mohamed Bourass,¹
Abdelkarim Ouammou,¹ and Mohammed Bouachrine²

¹Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdallah University, Fez, Morocco

²Research Team Materials, Environment & Modeling, ESTM, University Moulay Ismail, Meknes, Morocco

Correspondence should be addressed to Mohamed Bourass; mohamedbourass87@gmail.com

Received 25 October 2016; Revised 8 January 2017; Accepted 5 February 2017; Published 18 April 2017

Academic Editor: Jan Skov Pedersen

Copyright © 2017 El Ghalia Hadaji et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(E)-N-Aryl-2-ethene-sulfonamide and its derivatives are potent anticancer agents; these compounds inhibit cancer cells proliferation. A study of quantitative structure-activity relationship (QSAR) has been applied on 40 compounds based on (E)-N-Aryl-2-ethene-sulfonamide, in order to predict their anticancer biological activity. The principal components analysis is used for minimizing the base matrix and the multiple linear regression (MLR) and multiple nonlinear regression have been used to design the relationships between the molecular descriptor and anticancer properties of the sulfonamide derivatives. The validation of the models MLR and MNLR has been done by dividing the dataset into training and test set, the external validation of multiple correlation coefficients was $R_{pIC50} = 0.81$ for MLR and $R_{pIC50} = 0.91$ for MNLR. The artificial neural network (ANN) showed a correlation coefficient close to 0.96, which concluded that this latter model is more effective and much better than the other models. This obtained model (ANN) has been confirmed by two methods of LOO cross-validation and scrambling (or Y-randomization). The high correlation between experimental and predicted activity values was observed, indicating the validation and the good quality of the derived QSAR model.

1. Introduction

Cancer is a major public health problem. The number of new cancer cases in 2012 was estimated to be at 14, 1 million and 8, 2 million deaths. It was estimated in 2008 that more than 70% of the fatalities due to cancer originated from developing countries. The frequency of cancers may go up by 50%, with 15 million new cases per year in 2020 [1, 2].

It is previewed that in 2030 the number of the fatalities due to cancer will increase up to 13.1 million. Even though its incidence is increasing in most regions of the world, the incidence rate is the highest in the more developed regions, but its fatality is relatively higher in the developing countries because of the inaccessibility to treatment and the absence of preliminary detection [3, 4].

One in two men and one in three women are affected by cancer. The distribution of cancers by age shows the trend since the beginning of registration; namely, there are a high

number of cancer cases among women, taking into account the fact that the onset of cancer in women is at the age of 39 years, while for men it is at the age of 49 years. The number of cancer cases decreases from the age of 65 years in women whereas it begins to increase in men in this age [5].

Scientific advances of recent years now make it possible to decipher the genetic code of cancer and to understand how this disease is related to the mechanisms of life itself. For this reason, the cancer disease will probably never be completely eradicated, and basic research is redirected to the pharmaceutical industry that provides a plethora of increasingly targeted molecules, ever more efficient and always more expensive, so as to combine them in the manufacture of medicines.

Many sulfonamide derivatives have been reported to finally show in the pharmaceutical industry significant anti-tumor properties. The sulfonamides were the first widely used medications and they are systematically used as preventive and chemotherapeutic agents against various

FIGURE 1: The general structure of sulfonamides.

diseases. Over 30 medicines containing this feature are used clinically, including the antihypertensive bosentan, antibacterial, antiprotozoal, antifungal, anti-inflammatory, and nonpeptide antagonists of vasopressin receptors [6–8].

Sulfonamides are compounds which have a general structure represented in Figure 1. After the discovery of sulfanilamide, thousands of chemical changes were studied and the best therapeutic results were obtained from compounds where hydrogen ring (SO_2NH_2) has been replaced by heterocyclic [9, 10]. To date, more than twenty thousand sulfanilamide derivatives were synthesized. These syntheses have resulted in the discovery of new compounds having pharmacological properties that vary in said main structure; R , R_1 may be hydrogen, alkyl, aryl, or heteroaryl, and so forth [11, 12].

Chemically, for its anticancer activity, sulfonamide exists in a variety of pharmacological targets. Drug discovery is a long and complex process. This discipline can occur at different levels of the process of drug discovery. Among the techniques of chemoinformatics, we can mention QSAR techniques of finding a correlation between biological activity measured for a panel of compounds and some molecular descriptors. Quantitative structure-activity relationship (QSAR) methodology is an essential tool in medicinal chemistry. The modern QSAR appeared since the year 1960 but the first correlation investigations of biological activity with physicochemical properties began nearly 60 years before the important work of Overton and Meyer linking aquatic toxicity for partitioning of lipids in water. In 1962 came the seminal work of Corwin Hansch and colleagues, which arose great interest in predicting biological activities. Since 2011, the QSAR studies have begun to grow, with more than 1,400 publications per year [13]. QSAR techniques are based on the concept postulating that similar structures have similar properties and that the more the molecules are different, the harder it is to correlate the physicochemical properties and biological activity, whereas the opposite is easier [14]. The application of the quantitative structure-activity relationship (QSAR) technique for the purpose of molecular modeling and drug design has provided a potential approach in the field of computational chemistry. This tool deals with the determination of the quantitative correlation between molecular structures and their activities employing various chemometric tools. The prime importance of the QSAR technique lies in its ability to determine the essential structural requirements of the molecules for exhibiting definite responses and to predict the activity of untested molecules followed by the design of virtual libraries [11]. On the other

hand, QSAR studies were reported to pick out important structural features answerable for the anticancer activity [12]. The quantitative structure-activity relationships (QSAR) are assuredly a significant factor in coval drug design. Consequently, it is quite evident why a great number of users of QSAR [13, 14] are located in industrial research units. So, classical QSAR and 3D-QSAR are highly active areas of research in the drug design [15, 16]. The basis for different quantitative structure-activity relationship (QSAR) methods is the “description” of the molecular structures by means of numbers. Right now, there are a large number of molecular descriptors that can be used in QSAR studies [17–19].

Our objective is to highlight a fundamental and original research on molecules sulfonamide basic core, in order to develop the relationship between the structure and the activity of the active chemical substance and its derivatives.

In this study, principal component analysis (PCA), multiple linear regression (MLR) analysis, multiple nonlinear regression (MNLR) analysis, artificial neural network (ANN) calculations, Crosse validation, and scrambling or Y-randomization are applied to a series of (E)-N-Aryl-2-ethene-sulfonamide inhibitors in order to set up a 3D-QSAR model reliable to predict anticancer activity.

2. Materials and Methods

2.1. Experimental Data. In the present study, we chose 40 substitutes of (E)-N-Aryl-2-ethene-sulfonamide of which the anticancer activities are reported in the literature by Shiri et al. [20]. On the other hand and for the 3D-QSAR study, the reported values of IC_{50} have been converted into pIC_{50} by taking negative logarithm ($\text{pIC}_{50} = \log_{10} \text{IC}_{50}$) and subsequently used as the dependent variable for the 3D-QSAR model development. Figure 1 presents the basic structure of the flavonoids and Table 1 shows the studied compounds and their corresponding experimental activities and pIC_{50} .

2.2. Computational Methods. The quantum method DFT (density functional theory) was used in this study to predict the various physicochemical properties, in order to search for the best correlation between themselves. 3D structures of the molecules were generated using the Gauss View 3.0, and all quantum calculations of all compounds have been carried out using the Gaussian 03. The optimization geometry and the physicochemical properties of the 40 compounds were predicted by employing the B3LYP function coupled with 6–31 G basis set [21, 22].

2.3. Calculation of the Molecular Descriptors. Before every modelization, it is necessary to calculate or to measure a big amount of different descriptors because the mechanisms which determine molecules’ activity or one of its properties are frequently bad-known. Thus, we must select among variable cells the ones that are the most pertinent to modelization. This selection is done using the regression linear multiple methods. The optimized molecules were used to calculate a certain number of electronic descriptors: dipolar moment

TABLE 1: Study compounds and radicals.

Compound	R	R ₁	pIC50
1	H	H	5.00
2	4-Cl	H	4.70
3	4-F	4-Br	5.00
4	4-F	4-OCH ₃	5.00
5	4-OCH ₃	4-OCH ₃	5.30
6	4-OCH ₃	2,4-(OCH ₃) ₂	4.82
7	4-OCH ₃	2,6-(OCH ₃) ₂	6.43
8	4-OCH ₃	2,4,6-(OCH ₃) ₃	6.7
9	4-OCH ₃	3,4,5-(OCH ₃) ₃	4.46
10	2,4,6-(OCH ₃) ₃	4-OCH ₃	5.12
11	4-OCH ₃	2,6-(OCH ₃) ₂ , 4-OH	5.00
12	4-OCH ₃	2,4,6-F ₃	4.12
13	3-F, 4-OCH ₃	2,4,6-(OCH ₃) ₃	7.52
16	3-NH ₂ , 4-OCH ₃	2,4,6-(OCH ₃) ₃	4.12
17	3-NO ₂ , 4-OCH ₃	3,4,5-(OCH ₃) ₃	4.46
18	3-NH ₂ , 4-OCH ₃	3,4,5-(OCH ₃) ₃	4.00
22	3-NO ₂ , 4-F	2,4,6-(OCH ₃) ₃	5.00
23	3-NH ₂ , 4-F	2,4,6-(OCH ₃) ₃	5.00
24	3,5-(NO ₂) ₂ , 4-OCH ₃	2,4,6-(OCH ₃) ₃	5.00
25	3,5-(NH ₂) ₂ , 4-OCH ₃	2,4,6-(OCH ₃) ₃	5.60
26	3-F, 4-OCH ₃	4-OCH ₃	5.30
27	3-F, 4-OCH ₃	2,3,4,5,6-F ₅	5.00
28	3-NO ₂ , 4-OCH ₃	2,3,4,5,6-F ₅	4.00
29	3-NH ₂ , 4-OCH ₃	2,3,4,5,6-F ₅	4.12
30	2,3,4,5,6-F ₅	3-NO ₂ , 4-OCH ₃	4.00
31	2,3,4,5,6-F ₅	3-NH ₂ , 4-OCH ₃	4.46
32	2,3,4,5,6-F ₅	2,3,4,5,6-F ₅	4.46
Compound	R ₂	pIC50	
33	CH ₂	6.46
34	CH(CH ₃)	7.00	
35	C(CH ₃) ₂	6.70	
36	CH(C ₆ H ₅)	5.60	
37	CH(C ₆ H ₄ 4-F)	5.60	
38	CH(C ₆ H ₄ 4-Br)	5.12	
Compound	R ₂	pIC50	
39	CH ₂	6.40
40	CH(CH ₃)	7.40	
41	C(CH ₃) ₂	7.15	
42	CH(C ₆ H ₅)	7.12	
43	CH(C ₆ H ₄ 4-F)	7.12	
44	CH(C ₆ H ₄ 4-Cl)	7.12	
45	CH(C ₆ H ₄ 4-Br)	6.60	

(DM), orbital borders energy (E_{HOMO} , E_{LUMO}), total energy E_{totale} , and repulsion energy (RE) [23].

ChemBio Office (2015) was used to calculate the following parameters: molecular weight (MW), partition coefficient

(log P), the hydrogen bond acceptor (HA), and the hydrogen bond donor (HD) [24–26].

ChemSketch program was used to calculate the following parameters: molar volume (MV (cm³)), molar refractivity

(MR (cm³)), parachor (Pc (cm³)), density (g/cm³), refractive index, surface tension (dyne/cm), and polarizability (cm³) [27, 28].

2.4. Statistical Analysis. To present the structure-activity relationship for 40 studied molecules, 16 descriptors are calculated using the Gaussian 03, chemoffice 2012, and chemsketch. The procedure of the study is as follows.

The principal component analysis (PCA) [29] was generated using the software XLSTAT, version 2015 [30], to predict anticancer activities pIC50. It is a statistical method based on minimizing all the information encoded in the structures of the compounds. It is also very helpful to understand the distribution of the compounds. This is an essentially descriptive statistical method which aims to present, in a graphic form, the maximum of information contained in the data listed in Tables 1 and 2.

The multiple linear regression statistic techniques are used to study the relationship between one dependent variable and several independent variables. It is a mathematical technique that minimizes the difference between actual and predicted values. It also serves to select the descriptors that are used as input parameters in multiple nonlinear regressions (MNLR) and the multiple linear regression.

We also used the technique of nonlinear regression model to improve the structure-activity relationship to quantify the substituting effect. We applied the data matrix formed clearly from the descriptors proposed by MLR corresponding to the 39 molecules (training set). The coefficients R and R^2 and F -values are used to select the best regression performance. We used a preprogrammed function of XLSTAT as follows:

$$Y = a + (bX1 + cX2 + dX3 + eX4 \dots) + (fX12 + gX22 + hX32 + iX42 \dots), \quad (1)$$

where $a, b, c, d, \dots$ represent the parameters and $X1, X2, X3, X4, \dots$ represent the variables.

All the feedforward NN used in this paper are three-layer networks; the first (input) layer comprises six neurons, representing the pertinent descriptors obtained in MLR technique [31, 32], whereas there are neither theoretical nor empirical rules to determine the number of quiet layers or the number of neuron layers; one hidden layer seems to be sufficient in the most chemical application of ANN. Some authors [33, 34] have offered a parameter ρ , leading to determining the number of hidden neurons, which plays a major role in determining the best ANN architecture. It is defined as follows:

$$\rho = (\text{number of data points in the training set} / \text{sum of the number of connections in the NN}).$$

Therefore, in order to avoid overfitting or underfitting, it is recommended to take into consideration the ρ value: $1.8 < \rho < 2.3$ [35]. Thus, the ANN used in this work is formed by two hidden neurons, and the output layer represents the calculated activities values log(IC50) [35, 36].

The cross-validation technique is one of the most famous ways for the selection of regression models based on

FIGURE 2: The principal components and their variances.

minimizing the cross-validation (CV) criterion of Stone among [37] an appropriate class of model candidates. This may be particularly motivated when augury (or, similarly, estimation of the unknown regression function) is the aim of the statistical analysis [38].

The additional procedure named scrambling must be used to validate the model and also to verify that this obtained model cannot change. In some cases, the built models have lower performance on the test set. These cases can be explained by the quality of the obtained descriptors of the aforementioned models. To assess the share of luck in models built, a technique called scrambling or Y-randomization is used [39].

3. Results and Discussion

3.1. Dataset for Analysis. QSAR study was carried out on a series of 40 substitutes of (E)-N-Aryl-2-ethene-sulfonamide, in order to determine a quantitative relationship between the structure and the antiviral activities. The values of the 16 descriptors are shown in Table 2. The results obtained for 3D-QSAR using MLR, MNLR, ANN, CV, and Y-randomization are represented in Tables 3 and 4.

3.2. Principal Component Analysis. The totality of the 16 descriptors coding the 40 molecules is submitted to a principal components analysis (PCA) [32]. 16 principal components were obtained (Figure 2).

The first three principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variances are 43.24%, 21.12%, and 13.00% for the axes $F1$, $F2$, and $F3$, respectively. The total information is estimated to a percentage of 77.36%. Table 5 shows the correlation matrix (Pearson (n)) therefore obtained between different descriptors.

The Cartesian diagram analyses of projections are according to the plane $F1$ - $F2$ (60.57%) of the total variance of the studied molecules.

TABLE 2: Structures of dataset used for 3D-QSAR analysis of a series of novel N-phenylbenzamide and N-phenylacetophenone.

MD	Rep E	MW	HA	HD	logP	polar	dent	Surften	parc	MR	R Ind	MV	E _{HOMO}	E _{LUMO}	E _{TOTAL}
1	6.55	1313.26	259.32	2	1	3.33	28.98	1.3	57.6	547.8	73.1	198.7	-6.423	-2.244	-31352.99
2	6.68	1517.44	293.76	2	1	3.95	30.89	1.39	58.8	583.7	77.93	210.7	-6.559	-2.430	-43945.49
3	4.77	1846.46	256.22	3	1	4.29	32.08	1.63	58.2	605.4	80.93	219.1	-6.676	-2.565	-104516.20
4	7.19	1661.17	307.34	4	1	3.4	31.54	1.35	52.7	611	79.58	226.9	-6.488	-2.271	-37208.38
5	6.01	1786.57	319.37	4	1	3.14	34.02	1.29	51.5	661.1	85.83	246.7	-5.988	-2.045	-37626.78
6	7.03	2037.67	349.4	5	1	3.03	36.55	1.29	49.3	717.8	92.19	270.7	-5.673	-1.894	-40763.80
7	5.72	2250.6	349.4	5	1	3.03	36.55	1.29	49.3	717.8	92.19	270.7	-5.872	-1.910	-40763.80
8	5.51	2522.44	379.43	6	1	2.94	39.07	1.29	47.6	774.5	98.56	294.7	-5.659	-2.033	-43900.55
9	4.70	2384.79	379.43	6	1	2.68	39.07	1.29	47.6	774.5	98.56	294.7	-5.978	-2.053	-43900.28
10	6.69	2452.79	379.43	6	1	2.94	39.07	1.29	47.6	774.5	98.56	294.7	-5.470	-1.798	-43900.55
11	8.98	2298.19	365.43	6	2	2.63	37.15	1.36	54.9	732.8	93.72	269.2	-5.402	-2.326	-42824.28
12	9.82	2162.67	343.32	6	1	3.72	31.63	1.46	49.9	625.9	79.8	235.4	-6.272	-2.858	-42643.44
13	2.65	2109.79	397.42	7	1	3.1	39.11	1.33	46.7	781.6	98.76	299	-4.767	-1.168	-32198.01
16	9.91	2606.21	394.42	7	2	2.12	40.5	1.33	52.6	800.3	102.18	297	-4.978	-1.455	-45416.87
17	9.48	2825.23	424.42	7	1	2.12	40.96	1.32	44.9	831.8	103.32	321.1	-6.155	-3.169	-49501.11
18	7.81	2553.51	394.44	7	2	1.86	40.5	1.33	52.6	800.3	102.18	297	-5.206	-1.977	-45416.6
22	8.06	2772.74	412.39	7	1	2.9	38.6	1.36	44.6	781.7	97.38	302.3	-6.131	-3.722	-49082.99
23	11.44	2459.85	382.41	7	2	2.38	38.02	1.38	53.7	750.8	95.92	277.2	-5.46	-1.743	-44998.47
24	9.03	3522.25	469.42	8	1	2.08	43.05	1.42	50	876.4	108.61	329.5	-5.987	-3.689	-55102.22
25	12.16	3073.97	409.46	8	3	1.3	41.94	1.37	58	826.2	105.79	299.3	-5.069	-1.899	-46931.54
26	4.88	1946.17	337.37	5	1	3.29	34.07	1.34	50.2	668.3	85.94	250.9	-6.144	-2.099	-40339.65
27	4.03	2455.9	379.29	9	1	4.2	31.77	1.6	46.3	647.2	80.14	248	-6.468	-2.983	-50799.05
28	6.77	2692.35	424.29	9	1	3.48	33.79	1.56	40.3	681.9	85.23	270.5	-6.368	-3.391	-53681.81
29	6.22	2450.84	394.32	9	2	3.22	33.16	1.6	53.6	665.9	83.64	246.1	-5.476	-2.776	-49597.29
30	7.01	2727.73	424.29	9	1	3.48	33.79	1.56	40.3	681.9	85.23	270.5	-6.645	-3.957	-53681.81
31	9.76	2426.24	394.32	9	2	3.22	33.16	1.6	53.6	665.9	83.64	246.1	-5.662	-2.666	-49597.29
32	4.34	2860.76	439.23	12	1	4.94	29.42	1.82	43.6	619.1	74.22	240.8	-7.524	-3.451	-58534.89
33	9.27	2333.89	376.43	5	2	2.6	39.19	1.35	57.5	767.7	98.86	278.7	-5.244	-2.011	-43324.61
34	6.32	2675.05	390.45	5	2	3.12	41.01	1.32	54.9	804.8	103.47	295.6	-5.170	-2.032	-44401.43
35	5.85	2707.64	404.48	5	2	3.24	42.86	1.3	53.2	842.3	108.13	311.7	-5.069	-2.050	-45478.52
36	5.83	3193.46	452.53	5	2	4.16	48.98	1.33	57.9	938.5	123.55	340.1	-5.170	-2.004	-49654.01
37	3.93	3308.33	470.52	6	2	4.32	49.02	1.37	56.8	945.6	123.66	344.3	-5.278	-2.101	-52372.36
38	6.62	3797.02	531.42	5	2	4.96	52.04	1.49	59.3	989	131.27	356.3	-5.364	-2.008	-120098.58
39	7.03	2254.82	362.4	5	3	2.29	37.27	1.43	67	724.9	94.02	253.3	-5.309	-2.069	-42248.06
40	9.09	2392.44	376.43	5	3	2.81	39.1	1.39	63.2	762.1	98.63	270.2	-5.458	-2.036	-43324.88
41	5.28	2537.25	390.45	5	3	2.93	40.94	1.36	60.7	799.1	103.29	286.4	-5.167	-1.998	-44401.7
42	6.32	3059.09	438.5	5	3	3.85	47.06	1.39	65.5	895.8	118.71	314.8	-5.605	-2.095	-48576.91
43	10.00	3197.74	456.49	6	3	4.01	47.1	1.43	64.1	902.9	118.83	319	-5.444	-2.046	-51295.81
44	10.25	3320.18	472.94	5	3	4.47	48.97	1.45	66	931.6	123.54	326.7	-5.476	-2.064	-61169.40
45	10.00	3628.51	517.394	5	3	4.64	50.12	1.56	66.7	946.3	126.43	331	-5.458	-2.054	-119022.04

TABLE 3: Comparison between the observed and predicted activities of statistically significant models obtained by 3D models training set.

Compound	pIC50	pIC50 MLR	pIC50 ANN	pIC50 CV	COMP	pIC50	pIC50 Y-R	Compound	pIC50	pIC50 RNLM
1	5.000	4.819	4.9718	5.03	3	5	4.78	1	5.000	5.016
2	4.700	4.931	4.8771	4.96	10	6.7	5.64	2	4.700	4.965
3	5.000	4.435	4.8861	5.13	9	4.46	4.51	3	5.000	4.221
4	5.000	5.296	5.1855	5.09	8	5.12	5.16	4	5.000	5.156
5	5.300	5.582	5.3373	5.33	11	5	4.66	5	5.300	5.526
7	6.430	5.696	5.6839	5.92	13	7.52	6.61	6	4.820	5.647
8	6.700	5.433	5.9661	5.34	16	4.12	4.36	7	6.430	5.620
9	4.460	5.191	4.6112	4.79	17	4.46	4.52	9	4.460	4.943
10	5.120	5.658	6.1726	5.71	18	4	4.06	10	5.120	5.616
11	5.000	5.449	4.9940	4.66	23	5	4.82	12	4.120	4.503
13	7.520	6.132	7.3923	6.61	27	5	5.02	13	7.520	6.849
16	4.120	5.580	4.1307	4.36	28	4	3.72	17	4.460	3.758
17	4.460	4.259	4.3915	4.52	29	4.12	4.22	18	4.000	4.323
18	4.000	4.858	3.7573	4.04	30	4	3.99	22	5.000	4.539
23	5.000	5.332	5.2075	4.52	31	4.46	4.57	23	5.000	5.199
25	5.600	4.798	5.7078	5.33	33	6.46	6.49	26	5.300	5.373
26	5.300	5.543	5.4727	5.34	34	7	6.99	27	5.000	4.389
27	5.000	4.667	5.0062	5.02	35	6.7	7.19	28	4.000	4.513
28	4.000	4.489	4.2127	3.72	36	5.6	5.59	29	4.120	4.160
29	4.120	4.293	4.0668	4.3	42	7.12	6.97	30	4.000	4.622
30	4.000	3.948	3.9026	3.99	45	6.6	6.49	31	4.460	4.249
31	4.460	4.398	4.5080	4.64	1	5	5.02	32	4.460	4.435
32	4.460	4.256	4.4046	4.28	4	5	5.09	34	7.000	6.324
33	6.460	5.744	6.5096	6.54	5	5.3	5.45	35	6.700	6.696
34	7.000	6.657	6.7798	6.99	7	6.43	6.04	36	5.600	5.766
35	6.700	7.084	6.9139	6.84	25	5.6	5.33	37	5.600	5.432
36	5.600	6.161	5.5838	5.53	26	5.3	5.34	38	5.120	5.451
38	5.120	6.258	5.1458	5	2	4.7	4.93	39	6.400	6.301
40	7.400	6.802	7.0403	7.58	38	5.12	5	40	7.400	7.544
41	7.150	7.429	7.0397	7.15	40	7.4	6.95	42	7.120	6.970
42	7.120	6.272	7.0239	7.2	41	7.15	7.58	43	7.120	6.984
45	6.600	6.452	7.0172	6.46	32	4.46	4.28	45	6.600	6.840

The Pearson correlation coefficients are summarized in Table 4. The obtained matrix provides information on the negative or positive correlation between variables. The principal component analysis (PCA) was conducted to identify the link between the different variables. The correlation coefficients between the 11 descriptors are shown in Table 3 and in Figure 3 these descriptors are represented in a correlation circle.

3.3. Multiple Linear Regressions (MLR). In order to select the predominant descriptors that will affect the inhibitory activities of these compounds, correlation analysis was performed with statistical software XLSTAT 13 taking every calculated descriptor as an independent variable and pIC50 as a dependent variable. Based on the correlation analysis, the aforementioned stepwise multiple linear regression techniques were used to establish the QSAR model. However, this

method uses the coefficients R , R^2 , R^2_{adjusted} , MSE, MAE, and F -values in order 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, MAE is the mean absolute error, and F is the Fisher F -statistic. E_{LUMO} , the hydrogen bond acceptor (HA), the hydrogen bond donor (HD), ($\log P$), and R index are the descriptors dependent on the anticancer activity of a sulfonamide.

The QSAR model obtained by using the multiple linear regression (MLR) method is represented by the following equations:

$$\begin{aligned} \text{pIC50} = & 84,233 - 0,703 * \text{HA} + 2,024 * \text{HD} + 0,857 \\ & * \log P - 48,758 * R \text{ Index} + 0,956 \\ & * E_{\text{LUMO}}. \end{aligned} \quad (2)$$

TABLE 4: Comparison between the observed and predicted activities of statistically significant models obtained by 3D models test set.

Compound	pIC50	pIC50 MLR test	Compound	pIC50	pIC50 MNLR test
6	4.820	5.712	8	6.700	5.226
12	4.120	4.826	11	5.000	4.971
22	5.000	4.350	16	4.120	5.634
24	5.000	2.195	24	5.000	2.436
37	5.600	5.942	25	5.600	5.187
39	6.400	5.643	33	6.460	5.428
43	7.120	6.290	41	7.150	8.165
44	7.120	6.735	44	7.120	7.199

FIGURE 3: Circles correlations between descriptors.

The statistical characteristics of the obtained equation are as follows:

$$N = 32.$$

$$R = 0.81.$$

$$R^2 = 0.652.$$

$$R^2_{\text{adjusted}} = 0.585.$$

$$F = 9.742.$$

$$\text{MSE} = 0.418.$$

$$\text{MAE} = 0.528.$$

$$P < 0.0001.$$

As indicated in the abovementioned equation, the most significant descriptors affecting anticancer activity for inhibiting cancer cells proliferation are electronic descriptors (E_{HOMO}) and steric descriptors (R index, HA, HD, and $\log P$), in which these descriptors are used to build the QSAR model for the of (E)-N-Aryl-2-ethene-sulfonamide.

The obtained model exhibits a high correlation coefficient ($R = 0.81$) with electronic and steric descriptors. On the other

hand, the results revealed that the relationship between the anticancer activities and the parameters of the molecules was not linear ($R < 1$) for the compounds based on (E)-N-Aryl-2-ethene-sulfonamide.

For our 40 compounds, the correlation between experimental activity and the calculated one based on this model is quite significant (Figure 4) as indicated by statistical values.

Where n is the number of compounds, r is the correlation coefficient, F is the Fisher F -statistic, and MSE is a mean square error.

Several statistical parameters such as regression coefficient (r), square correlation coefficient (R^2), adjusted square correlation coefficient (R^2_{adj}), standard error of estimate (S), Fischer's value (F), and significance level ($P < 0.005$) are used to check the credibility of the developed models.

Generally, a good QSAR model has the characteristics of a large F , small MSE and MAE, very small P value, and R^2 , R^2_{adjusted} close to one. Equation (2) has verified these criteria; therefore, it is statistically acceptable. Fisher's test F is used. Given that the probability corresponding to the F value was lower than 0.05 for anticancer activity, there is a

TABLE 5: The correlation matrix (Pearson (n)) between different obtained descriptors.

Desc	MD	Rep E	MW	HA	HD	log P	polarb	density	surften	parac	MR	R Ind	MV	E_{HOMO}	E_{LUMO}	E_{T}
MD	1															
Rep E	0.259	1														
MW	0.199	0.958	1													
HA	0.066	0.384	0.442	1												
HD	0.435	0.476	0.438	-0.087	1											
log P	-0.375	0.212	0.246	-0.095	-0.026	1										
polarb	0.209	0.813	0.779	-0.122	0.609	0.124	1									
density	-0.037	0.228	0.232	0.551	0.026	0.532	-0.241	1								
surften	0.294	0.196	0.126	-0.542	0.817	0.174	0.492	-0.058	1							
parac	0.212	0.856	0.828	-0.002	0.569	0.063	0.989	-0.226	0.391	1						
MR	0.209	0.813	0.779	-0.122	0.609	0.124	1.000	-0.241	0.492	0.989	1					
R Ind	0.206	0.096	0.023	-0.714	0.676	0.257	0.470	-0.154	0.959	0.347	0.470	1				
MV	0.156	0.864	0.851	0.126	0.409	0.012	0.937	-0.235	0.174	0.974	0.937	0.136	1			
E_{HOMO}	0.222	0.320	0.333	-0.156	0.622	-0.360	0.652	-0.503	0.443	0.647	0.653	0.386	0.587	1		
E_{LUMO}	-0.022	-0.132	-0.111	-0.495	0.388	-0.136	0.351	-0.553	0.488	0.285	0.351	0.539	0.182	0.697	1	
E_{T}	-0.361	-0.164	-0.058	-0.156	-0.272	0.359	-0.094	0.037	-0.105	-0.104	-0.094	-0.056	-0.084	-0.183	-0.108	1

Polarizability and parachoc are perfectly correlated ($r = 0.989$); polarizability and MR are perfectly correlated ($r = 1$); parachoc and MR are perfectly correlated ($r = 0.989$). Both variables are redundant; parachoc and MV are perfectly correlated ($r = 0.974$). Therefore, it is necessary to remove all the correlated variables such as parachoc and MR.

FIGURE 4: Predicted antiviral activities of pIC50 by MLR in comparison with experimental values (training set in blue and test set in red).

lower than <0.01% risk in assuming that the null hypothesis is false. Therefore, we can conclude with confidence that the model will correctly predict the anticancer activity of a given compound.

As shown in (2), the descriptors having negative influence in the anticancer activity (pIC50) of the given compound are the *R* Index and the HA, and the parameters having positive influence in the activities are the HD and E_{LUMO} . The HA has a negative sign in the model, which suggests that increased activity can be achieved by decreasing the number of heteroatoms (nitrogen or oxygen atoms). The HD has a positive sign in the model, which suggests that increased activity can be achieved by increasing the heteroatom with one or more hydrogen atoms. On the other hand, we remark that E_{LUMO} has a positive sign in the model, which suggests that the substitution using a stronger donating of electrons such as phenyl and ethyl increases anticancer activity of a given compound.

The true predictive power of our model is to test its ability to perfectly predict compounds from an external test set (compounds that were not used for the developed model), by using the 16 compounds that remained from dataset used for building the quantitative model MLR, and the obtained results are tabulated in Table 3. Through these results, we remark that the correlations between the experimental plots and data predicted from multiple regression derived QSAR values are much closer, which shows that the developed models can be successfully applied to predict the inhibition of other derivatives of sulfonamide. Figure 2 shows a very regular distribution of anticancer activity values based on the observed values depending on the experimental values.

3.4. Multiple Nonlinear Regression (MNLr). The basic descriptors correspond to the MLR; 32 compounds were applied to the data matrix. The coefficients *R*, R^2 , mean absolute error (MAE), and the mean squared error (MSE) are used to select the best performance of the regression.

The resulting equations are as follows:

$$\begin{aligned} \text{pIC50} = & 1,086 - 0.925 * \text{HA} - 0,164 * \text{HD} + 2,062 \\ & * \log P + 52,785 * R \text{ Index} + 3,673 * E_{\text{LUMO}} \\ & + 2,449 \end{aligned} \quad (3)$$

$$10^{-2} * \text{HA}^2 + 0,626 * \text{HD}^2 - 0,194 * \log P^2 - 29,65$$

$$* R \text{ Index}^2 + 0,526 * E_{\text{LUMO}}^2,$$

$$N = 32,$$

$$R = 91,$$

$$R^2 = 0.841,$$

$$\text{MSE} = 0.398,$$

$$\text{MAE} = 0.476.$$

The correlation coefficient obtained in (3) is very interesting (0.91) which shows that the anticancer activity obtained from this model is close to those obtained experimentally, comparing the results obtained by the MLR method (Figure 5).

3.5. Artificial Neural Networks. In order to increase the probability of a good characterization of the studied compounds, neural network (ANN) has been used to generate the predictive models of quantitative structure-activity relationships (QSAR) between a set of molecular descriptors obtained from MLR and the activity observed experimentally. The correlation of the observed activities with the ANN calculated ones is illustrated in Figure 6. The correlation coefficient $R = 0.94$ and Standard Deviation (SD) = 0.211, obtained with the neural network, show that the selected descriptors by LMR are pertinent and that the model proposed to predict the activity is relevant:

$$N = 32.$$

$$R = 0.96.$$

$$\text{MSE} = 0.0959.$$

$$\text{MAE} = 0.204.$$

$$\text{SD} = 0,302.$$

$$P < 0.0001.$$

3.6. Validation. We use the procedure "Leave-One-Out" which removes successively a molecule of learning the game containing 32 molecules. This procedure is repeated 32 times in order to predict the properties of all molecules.

The consistency and reliability of the MLR and ANN model are validated using the cross-validation technique with a good correlation obtained with cross-validation $R_{\text{cv}} = 0.95$. So the predictive power of this model is very significant (Figure 7). Consider the following:

$$N = 32.$$

$$R = 0.95.$$

$$\text{MSE} = 0.133.$$

FIGURE 5: Graphical representation of calculated and observed pIC50 with MNLR (training set in blue and test set in red).

FIGURE 6: Correlation between pIC50 obs and pIC50 predicted by ANN.

$$\text{MAE} = 0.231.$$

$$\text{SD} = 0.371.$$

$$P < 0.0001.$$

The most important result of this investigation is that in vitro anticancer activity may be predicted using QSAR methods. It affirms that artificial neural networks were the best to build the quantitative structure-activity relationship models. So, the model proposed in this study indicates high predictive power ($R_{\text{ANN}} = 0.94$).

In this part, QSAR model includes some molecular descriptors, and the regression quality indicates that these descriptors provide valuable information and have a significant role in the assessment of the activity of (E)-N-Aryl-2-ethene-sulfonamide. The artificial neural network (ANN) techniques, considering the relevant descriptors obtained from MLR, showed that a good agreement between the observed and the predicted values was prime and the obtained model has been confirmed by cross-validation method (LOO).

3.7. Scrambling or Y-Randomization. Y-randomization is widely used in QSAR studies to ensure the robustness of the

FIGURE 7: Correlation between pIC50 Obs and pIC50 predicted by CV.

models obtained. Scrambling validates the QSAR model by comparing the performance of the original model to that of models built for permuted (randomly shuffled) responses based on the original descriptor pool and the original procedure used to build the model [40].

The aim of this procedure is to randomly mix the properties of the model in the original game and recreate cross-validation models. Consider the following:

$$N = 32.$$

$$R = 0.95.$$

$$\text{MSE} = 0.1043.$$

$$\text{MAE} = 0.214.$$

$$\text{SD} = 0.305.$$

$$P < 0.0001.$$

The correlation coefficient value of the mixture molecules is the same as that obtained by applying the full model (Figure 8). These results demonstrate the absence of dependence between descriptors ejection and inclusion in model; also the closest measurement points of the target point do not

TABLE 6: Violations of Lipinski rules.

Compound	MW	HA	HD	log (P)	Number of violations
1	259.32	2	1	3.332	0
2	293.76	2	1	3.954	0
3	256.22	3	1	4.289	0
4	307.34	4	1	3.396	0
5	319.37	4	1	3.137	0
6	349.4	5	1	3.03	0
7	349.4	5	1	3.03	0
8	379.43	6	1	2.94	0
9	379.43	6	1	2.68	0
10	379.43	6	1	2.94	0
11	365.43	6	2	2.63	0
12	343.32	6	1	3.72	0
13	397.42	7	1	3.1	0
16	394.42	7	2	2.12	0
17	424.42	7	1	2.12	0
18	394.44	7	2	1.86	0
22	412.39	7	1	2.9	0
23	382.41	7	2	2.38	0
24	469.42	8	1	2.08	0
25	409.46	8	3	1.3	0
26	337.37	5	1	3.29	0
27	379.29	9	1	4.2	0
28	424.29	9	1	3.48	0
29	394.32	9	2	3.22	0
30	424.29	9	1	3.48	0
31	394.32	9	2	3.22	0
32	439.23	12	1	4.94	1
33	376.43	5	2	2.6	0
34	390.45	5	2	3.12	0
35	404.48	5	2	3.24	0
36	452.53	5	2	4.16	0
37	470.52	6	2	4.32	0
38	531.42	5	2	4.96	1
39	362.4	5	3	2.29	0
40	376.43	5	3	2.81	0
41	390.45	5	3	2.93	0
42	438.5	5	3	3.85	0
43	456.49	6	3	4.01	0
44	472.94	5	3	4.47	0
45	517.394	5	3	4.64	1

hide other experimental data and are not involved almost exclusively in the estimate, and the data used in this validation is regularly distributed in space so the resulting model can be extrapolated for the entire series.

3.8. Five Rules of Lipinski. Each prospective drug must comply with several basic criteria, such as its low cost of production and being soluble and stable, but must also comply with the scales associated with the pharmacological

properties of absorption, distribution, metabolism, excretion, and toxicity. The list of aspects to be included is very long but it is commonly accepted that the molecules validating at least two conditions of the Lipinski rule are potential candidates. These rules are as follows:

- (1) The molecular weight is less than or equal to 500 Da.
- (2) It has 5 or less hydrogen bond donors (sum of OH and NH).

FIGURE 8: Correlation between pIC50 Obs and pIC50 predicted by Y-randomization.

- (3) It has 10 or fewer hydrogen bond acceptors (sum of O and N).
- (4) Its $\log P$ value is less than or equal to 5.

$\log P$ is equal to the logarithm of the ratio of the concentrations of the test substance in octanol and in water: $\log P = \log(\text{Coct}/\text{Ceau})$. This value allows us to understand the hydrophilic or hydrophobic (lipophilic) molecules of our compounds in octanol and water solvent, where $\log P = \log(\text{Coct}/\text{Ceau})$.

Various statistical studies have identified the best $\log P$ values for a compound (typically a drug) absorbed by the human body as follows:

- (i) Brain penetration: 2.0
- (ii) Oral absorption: 1.8
- (iii) Sublingual absorption: 5
- (iv) Percutaneous absorption: 2.6.

Molecules that violate many of these rules can have problems with bioavailability. Therefore, this rule establishes some relevant structural parameters for the theoretical prediction of the oral bioavailability profile and is widely used in the design of new drugs.

The results of the calculation (Table 6) show that all compounds satisfy the rules of Lipinski, suggesting that these compounds theoretically do not have problems with oral bioavailability except molecule 32 which has 11 acceptor sites and molecules 39 and 45 whose molecular weight is up to 500 Da [40, 41].

4. Conclusion

Theoretical calculations were performed on a series of sulfonamide chemical information presented in the form of mathematical equation. The multiple linear regression (MLR) and multiple nonlinear regression analyses were used to concept a quantitative structure propriety relation model

of (E)-N-Aryl-2-ethene-sulfonamide analogs compounds for their properties and anticancer activity.

The most important result of this investigation is that in vitro anticancer activity may be predicted using QSAR methods. It affirms that artificial neural networks results were the best to build the quantitative structure-activity relationship models. So, the model proposed in this study indicates high predictive power ($R_{\text{ANN}} = 0.94$).

In this part, QSAR model includes some molecular descriptors, and regression quality indicates that these descriptors provide valuable information and play a significant role in the assessment of the activity of (E)-N-Aryl-2-ethene-sulfonamide analogs. Furthermore, we can conclude that the studied descriptors, which are sufficiently rich in chemical, electronic, and topological information to encode the structural features, may be used with other descriptors for the development of predictive QSAR models.

Conflicts of Interest

The authors declare that they have no conflicts of interests.

Acknowledgments

The authors are grateful to the "Association Marocaine des Chimistes Théoriciens" (AMCT) for its pertinent help concerning the programs.

References

- [1] E. Esteve, D. Bazin, C. Jouanneau et al., "How to assess the role of Pt and Zn in the nephrotoxicity of Pt anti-cancer drugs?: an investigation combining μ XRF and statistical analysis. Part II: clinical application," *Comptes Rendus Chimie*, vol. 19, no. 11-12, pp. 1580–1585, 2015.
- [2] A. K. Morgans, A. C. M. Van Bommel, C. Stowell et al., "Development of a standardized set of patient-centered outcomes for advanced prostate cancer: an international effort for a unified approach," *European Urology*, vol. 68, no. 5, pp. 891–898, 2015.
- [3] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA: A Cancer Journal for Clinicians*, vol. 63, no. 1, pp. 11–30, 2013.
- [4] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA Cancer Journal for Clinicians*, vol. 65, no. 1, pp. 5–29, 2015.
- [5] A. J. Price, R. C. Travis, P. N. Appleby et al., "Circulating folate and vitamin B12 and risk of prostate cancer: a collaborative analysis of individual participant data from six cohorts including 6875 cases and 8104 controls," *European Urology*, vol. 70, no. 6, pp. 941–951, 2016.
- [6] A. Kolaczek, I. Fusiarz, J. Lawecka, and D. Branowska, "Biological activity and synthesis of sulfonamide derivatives: a brief review," *Chemik*, vol. 68, no. 7, pp. 620–628, 2014.
- [7] C. T. Supuran, A. Casini, and A. Scozzafava, "Protease inhibitors of the sulfonamide type: anticancer, antiinflammatory, and antiviral agents," *Medicinal Research Reviews*, vol. 23, no. 5, pp. 535–558, 2003.
- [8] A. Scozzafava, T. Owa, A. Mastrolorenzo, and C. T. Supuran, "Anticancer and antiviral sulfonamides," *Current Medicinal Chemistry*, vol. 10, no. 11, pp. 925–953, 2003.
- [9] N. N. Al-Mohammed, Y. Alias, Z. Abdullah, R. M. Shakir, E. M. Taha, and A. A. Hamid, "Synthesis and antibacterial evaluation

- of some novel imidazole and benzimidazole sulfonamides," *Molecules*, vol. 18, no. 10, pp. 11978–11995, 2013.
- [10] N. Özbek, H. Katircioğlu, N. Karacan, and T. Baykal, "Synthesis, characterization and antimicrobial activity of new aliphatic sulfonamide," *Bioorganic and Medicinal Chemistry*, vol. 15, no. 15, pp. 5105–5109, 2007.
 - [11] L. De Luca and G. Giacomelli, "An easy microwave-assisted synthesis of sulfonamides directly from sulfonic acids," *Journal of Organic Chemistry*, vol. 73, no. 10, pp. 3967–3969, 2008.
 - [12] S. S. Stokes, R. Albert, E. T. Buurman et al., "Inhibitors of the acetyltransferase domain of N-acetylglucosamine-1-phosphate-uridylyltransferase/glucosamine-1-phosphate-acetyltransferase (GlmU). Part 2: optimization of physical properties leading to antibacterial aryl sulfonamides," *Bioorganic and Medicinal Chemistry Letters*, vol. 22, no. 23, pp. 7019–7023, 2012.
 - [13] J. C. Dearden, "The history and development of Quantitative Structure-Activity Relationships (QSARs)," *International Journal of Quantitative Structure-Property Relationships*, vol. 1, no. 1, pp. 1–44, 2016.
 - [14] N. Adhikari, M. K. Maiti, and T. Jha, "Exploring structural requirements of 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines as antiamebic agents using comparative QSAR modelling," *Bioorganic and Medicinal Chemistry Letters*, vol. 20, no. 14, pp. 4021–4026, 2010.
 - [15] C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, and A. J. Hopfinger, "4D-QSAR: perspectives in drug design," *Molecules*, vol. 15, no. 5, pp. 3281–3294, 2010.
 - [16] M. Elhallaoui, M. Elasri, F. Ouazzani, A. Mechaqrane, and T. Lakhli, "Quantitative structure-activity relationships of noncompetitive antagonists of the NMDA receptor: a study of a series of MK801 derivative molecules using statistical methods and neural network," *International Journal of Molecular Sciences*, vol. 4, no. 5, pp. 249–262, 2003.
 - [17] Y. Yamaguchi, S. Yokoyama, and S. Mashiko, "Strong coupling of the single excitations in the Q-like bands of phenylene-linked free-base and zinc bacteriochlorin dimers: a time-dependent density functional theory study," *Journal of Chemical Physics*, vol. 116, no. 15, pp. 6541–6548, 2002.
 - [18] E. Hadaji, M. Bourass, A. Ouammou, and M. Bouachrine, "3D-QSAR models to predict the antiviral activities of a series of novel N-phenylbenzamide and N-phenylacetophenone compounds based on density functional theory using statistical methods," *Moroccan Journal of Chemistry*, vol. 4, no. 1, pp. 204–214, 2016.
 - [19] R. Hmamouchi, A. I. Taghki, M. Larif et al., "Combining DFT and QSAR result for predicting the biological activity of the phenylsuccinimide derivatives," *Journal of Chemical and Pharmaceutical Research*, vol. 5, no. 9, pp. 198–209, 2013.
 - [20] F. Shiri, S. M. Bakhshayesh, and J. B. Ghasemi, "Computer-aided molecular design of (E)-N-Aryl-2-ethene-sulfonamide analogues as microtubule targeted agents in prostate cancer," *Arabian Journal of Chemistry*, 2015.
 - [21] M. Parac and S. Grimme, "All calculations were done by GAUSSIAN 03 W software," *The Journal of Physical Chemistry A*, vol. 106, pp. 6844–6850, 2003.
 - [22] L. Becker, K. Hinrichs, and U. Finke, "New algorithm for computing joins with grid files," in *Proceedings of the IEEE 9th International Conference on Data Engineering*, pp. 190–197, Vienna, Austria, April 1993.
 - [23] S. Chtita, R. Hmamouchi, M. Larif, M. Ghamali, M. Bouachrine, and T. Lakhli, "QSPR studies of 9-anilinoacridine derivatives for their DNA drug binding properties based on density functional theory using statistical methods: model, validation and influencing factors," *Journal of Taibah University for Science*, vol. 10, no. 6, pp. 868–876, 2016.
 - [24] ACD/Labs Extension for ChemDraw Version 8.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2015, <http://www.acdlabs.com>.
 - [25] ChemBioOffice, PerkinElmer Informatics, 2010, <http://www.cambridgesoft.com>.
 - [26] N. L. Allinger, "Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms," *Journal of the American Chemical Society*, vol. 99, no. 25, pp. 8127–8134, 1977.
 - [27] ACDLABS 10, Advanced Chemistry Development, Inc., Toronto, Canada, 2015, <http://www.acdlabs.com/>.
 - [28] MarvinSketch 5.11.4, Chem Axon, 2012, <http://www.chemaxon.com>.
 - [29] M. Larif, A. Adad, R. Hmamouchi et al., "Biological activities of triazine derivatives. Combining DFT and QSAR results," *Arabian Journal of Chemistry*, 2013.
 - [30] XLSTAT 2015 software (XLSTAT Company), <http://www.xlstat.com>.
 - [31] T. A. Andrea and H. Kalayeh, "Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors," *Journal of Medicinal Chemistry*, vol. 34, no. 9, pp. 2824–2836, 1991.
 - [32] P. Y. Lee and C. Y. Chen, "Toxicity and quantitative structure-activity relationships of benzoic acids to *Pseudokirchneriella subcapitata*," *Journal of Hazardous Materials*, vol. 165, no. 1–3, pp. 156–161, 2009.
 - [33] S. Chtita, M. Larif, M. Ghamali, M. Bouachrine, and T. Lakhli, "DFT-based QSAR studies of MK801 derivatives for non competitive antagonists of NMDA using electronic and topological descriptors," *Journal of Taibah University for Chemistry*, vol. 9, no. 2, pp. 143–154, 2014.
 - [34] S. Chtita, M. Larif, M. Ghamali, M. Bouachrine, and T. Lakhli, "Quantitative structure-activity relationship studies of dibenzo[a,d]cycloalkenimine derivatives for non-competitive antagonists of N-methyl-D-aspartate based on density functional theory with electronic and topological descriptors," *Journal of Taibah University for Science*, vol. 9, no. 2, pp. 143–154, 2015.
 - [35] M. Larif, S. Chtita, A. Adad, R. Hmamouchi, M. Bouachrine, and T. Lakhli, "Predicting biological activity of chalcone (1, 3-diphenyl-2-propen-1-one) derivatives cytotoxicity against HT-29 human colon adenocarcinoma cell lines by DFT-QSAR models," *Journal of Computational Methods in Molecular Design*, vol. 4, no. 4, pp. 121–130, 2014.
 - [36] M. Larif, S. Chtita, A. Adad, R. Hmamouchi, M. Bouachrine, and T. Lakhli, "Predicting biological activity of anticancer molecules 3-ary 1-4-hydroxyquinolin-2-(1H)-one by DFT-QSAR models," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 12, pp. 32–42, 2013.
 - [37] B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 316–331, 1983.
 - [38] D. W. Osten, "Selection of optimal regression models via cross-validation," *Journal of Chemometrics*, vol. 2, no. 1, pp. 39–48, 1988.
 - [39] D. Voet and J.-G. Voet, *Biochimie*, De Boeck & Larcier, Bruxelles, Belgium, 2nd edition, 2005.

- [40] A. Monge, *Création et Utilisation de Chimiothèques Optimisées pour la Recherche in silico de Nouveaux Composés Bioactifs*, Université d'Orléans, 2006 (Français).
- [41] M. H. Abraham, J. M. R. Gola, R. Kumarsingh, J. E. Cometto-Muniz, and W. S. Cain, "Connection between chromatographic data and biological data," *Journal of Chromatography B: Biomedical Sciences and Applications*, vol. 745, no. 1, pp. 103–115, 2000.