Moroccan Journal of Chemistry
ISSN : 2351-812X

Mor. J. Chem., 2023, Volume 11, Issue 4, Page 1086-1099
https://revues.imist.ma/index.php/morjchem
https://doi.org/10.48317/IMIST.PRSM/morjchem-v11i04.40744

Structure-toxicity relationships for phenols and anilines towards Chlorella vulgaris using quantum chemical descriptors and statistical methods

Ousaa A[1*], Aouidate A[1], Ghamali M[2], Chtita S[3], Idrissi Taourati A[2], Bouachrine M[2], Lakhlifi T[2]

[1] *Laboratory of Applied Chemistry and Environment, Faculty of Sciences- University ibn zohr,Agadir,Morocco.*
[2]*Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, University Moulay Ismail, Meknes, Morocco,*
[3]*Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, B.P 7955, Casablanca, Morocco,*

**Corresponding author, Email address: abdellahousaa@gmail.com*

Received *17 May 2023,*
Revised *02 Aug 2023,*
Accepted *06 Aug 2023*

Citation*: Ousaa A., Ghamali M., Chtita S., Idrissi Taourati A., Aouidate A., Bouachrine M., Lakhlifi T. (2023) Structure-toxicity relationships for phenols and anilines towards Chlorella vulgaris using quantum chemical descriptors and statistical methods, Mor. J. Chem., 11(4), 1086-1099*

Abstract: The study aimed to develop quantitative structure-toxicity relationship (QSTR) models to understand how chemical structure relates to the toxicity of 67 phenols and anilines. The chemical structures of these compounds were characterized using electronic and physico-chemical descriptors and DFT calculations were performed to obtain insights into the chemical structure and property information of the compounds. The study compared the predictive abilities of multiple linear regression (MLR), multiple nonlinear regression (MNLR), and artificial neural network (ANN) models for predicting the toxicity of the compounds. The ANN model was found to be the most effective compared the other two models. The developed models were able to accurately predict the toxicity of the compounds for four different toxicity endpoints, as demonstrated by leave-one-out cross validation, external validation, Y-randomized validation, and application domain analysis. The study suggests that the proposed descriptors could be useful in predicting the toxicity of phenols and anilines towards Chlorella vulgaris.

Keywords: *QSTR model, DFT study, phenols and anilines, toxicity, Chlorella vulgaris*

1. Introduction

Phenol, aniline and their derivatives have been utilized in a variety of ways in the chemical industry for many years. These have had various functions such as being used as solvents, propellants, additives, cooling agents, insecticides and herbicides. Additionally, they have found use in the production of materials such as dyes, pharmaceuticals, polymers, and synthetic resins, due to their versatility and usefulness (Aruoja *et al.,* 2011; Azzaoui *et al.*, 2015; Bakire *et al.,* 2018). Many of these chemicals were released into the environment and accumulated in nearly all-natural environments, especially in aquatic systems, Ecologically, algae play a significant role in the aquatic ecosystem as a dominant producer, providing energy and oxygen at higher levels. Adverse effects of these toxic chemicals to algae may reduce the primary productivity of the ecosystem and further disrupt the food web so it is beneficial to study seriously their potential hazard to aquatic organism.

The process of experimentation is a direct method for acquiring the toxicity data of organic compounds, however, it has its limitations. These limitations include the need for an enormous number of trial organisms, which can be quite expensive and time-consuming. In addition, there is often a variation in measured values between different researchers. Thus, obtaining toxicity data for all organic compounds through experimentation alone would be a challenging task, especially considering the constant development of new compounds and the accompanying difficulties. As such, it is necessary to utilize theoretical research to address the shortcomings of experimentation and accurately predict the toxicity data of compounds in a rapid manner .

Computational science and theoretical chemistry have rapidly evolved, making it possible to obtain quantum chemical parameters of organic compounds rapidly and precisely. The quantitative structure-toxicity relationship (QSTR) approach utilizes structural parameters of compounds and appropriate mathematical models to predict bioactivity such as toxicity, mutagenicity, and carcinogenicity.

Currently, QSTR studies have a plethora of molecular descriptors at their disposal (Rajarshi *et al.,* 2012, Zhu *et al.,* 2010). After validation, the results obtained can be utilized to predict the activities of untested compounds. The use of density functional theory (DFT) has been substantiated by comparative QSTR studies, which have demonstrated that descriptors generated using the DFT method can enhance the accuracy of results and yield more dependable QSTR outcomes (Pansal & Singh, 2018; Bouyad *et al.,* 2018). The objective of this investigation is to construct predictive QSTR models for the acute toxic effects of Phenol, aniline, and their derivatives on Chlorella vulgaris. To achieve this, statistical tools such as principal components analysis (PCA), multiple linear regression (MLR), multiple non-linear regression (MNLR), and artificial neural network (ANN) methods will be utilized.

2. Material and Methods

2.1. Data sources

Acute toxicity data of 67 phenols and anilines towards *Chlorella vulgaris* were taken from a literature (Zhu *et al.,* 2010) **IC_{50}** (**IC_{20}**) here means the millimolar concentration causing 50% (20%) inhibition of growth about 67 phenols and anilines towards *Chlorella vulgaris.* The bigger the value of **–log IC_{50} (pIC_{50})** and **pIC_{20}**, the higher is toxicity of compounds, and vice versa. For the proper validation of our data set with a QSTR model, the 67 substituted phenols were divided into training and test sets. A total of 60 molecules were placed in the training set to build the QSTR models, whereas the remaining 10 molecules composed the test set. The division was carried out by random selection. The following table shows the studied compounds and the corresponding experimental toxicties pIC_{50} and pIC_{20} (**table 1**).

2.2. Molecular descriptors

The computation of electronic descriptors was performed using the Gaussian 03W program To compute electronic descriptors, the Gaussian 03W program (Tugcuet *et al.,* 2018) was utilized. All 67 phenols and anilines that were modeled theoretically had their geometries optimized using the DFT method with the B3LYP functional and the 6-31G (d) basis set. Relevant structural descriptors such as the highest occupied molecular orbital energy (E_{HOMO}) in electron volts (eV), the lowest unoccupied molecular orbital energy (E_{LUMO}) in eV, the energy gap (ΔE) in eV, the dipole moment (μ) in Debye, and the total energy (E_T) in eV were selected from the quantum computation outcomes.

For the calculation of other molecular descriptors, the ChemSketch program (Lafridi *et al.,* 2020) was employed. These descriptors include the molar volume (MV) in cubic centimeters (cm3), the molecular weight (MW) in grams per mole (g/mol), the molar refractivity (MR) in cm3, the parachor (Pc) in cm3,

the density (D) in grams per cubic centimeter (g/cm3), the refractive index (n), the surface tension (γ) in dyne per centimeter (Dyne/cm), and the polarizability (α) in cm3. Additionally, the number of atoms (NA) and the number of electrons (NE) were considered as descriptors. To improve the quality of toxicity estimates for these compounds, it is recommended to include additional molecular descriptors that reflect specific interactions, such as the octanol/water partition coefficient (log P).

Table 1: phenols and anilines derivatives and their observed toxicities against *Chlorella vulgaris*

N°	Name (IUPAC)	pIC_{20}	pIC_{50}	N°	Name (IUPAC)	pIC_{20}	pIC_{50}
1*	2-Methylphenol	0.0313	−0.0846	**35**	4-Chloro-2-nitrophenol	2.2617	1.8485
2	2,3-Dimethylphenol	0.6072	0.3872	**36***	4-Chloro-3-nitrophenol	1.5405	1.3667
3	2,4-Dimethylphenol	0.5745	0.4424	**37**	2,6-Dichloro-4-nitrophenol	1.0765	0.9582
4	2,5-Dimethylphenol	0.5745	0.3326	**38***	3-Aminophenol	−0.2951	−0.6845
5	2,6-Dimethylphenol	0.3502	0.1377	**39**	2-Amino-4-methylphenol	0.9149	0.4378
6	3,4-Dimethylphenol	0.7724	0.5779	**40**	2-Amino-4-chlorophenol	1.4822	1.0296
7	3,5-Dimethylphenol	0.6757	0.5147	**41**	2-Nitroaniline	0.5611	0.3021
8	4-Ethylphenol	0.3971	0.2357	**42**	3-Nitroaniline	0.3963	0.1505
9	4-Methoxyphenol	−0.0257	−0.2948	**43**	2,4-Dinitroaniline	1.5467	1.2512
10	3,5-Dimethoxyphenol	0.0744	−0.1893	**44**	3,5-Dinitroaniline	1.6957	1.5046
11	2,3,5-Trimethylphenol	0.7512	0.5101	**45**	2-Methyl-3-nitroaniline	0.1553	0.0308
12	2,4,6-Trimethylphenol	0.4969	0.3561	**46**	4-Methyl-3-nitroaniline	0.7187	0.1757
13	Hydroxyhydroquinone	0.4980	0.1566	**47**	4-Chloro-2-nitroaniline	1.1968	0.8624
14	Methoxyhydroquinone	1.3884	1.2075	**48**	4-Chloro-3-nitroaniline	0.9747	0.6943
15	2,5-Dichlorohydroquinone	1.8199	1.6074	**49**	6-Chloro-2,4-dinitroaniline	1.9160	1.6815
16	5-Methylresorcinol	−0.5101	−0.6842	**50**	4-Hydroxy-3-methoxy-benzonitrile	0.3512	0.0012
17	2-Chloro-4-methylphenol	0.6637	0.5093	**51**	Phenol	0.4677	0.6003
18	2-Chloro-5-methylphenol	1.2366	0.8930	**52**	2-Chlorophenol	0.3656	0.1731
19*	4-Chloro-2-methylphenol	1.0606	0.8550	**53**	4-Chlorophenol	0.7129	0.4568
20	4-Chloro-3-methylphenol	1.3716	1.1718	**54**	2,4-Dichlorophenol	1.5401	1.2437
21	4-Chloro-3,5-dimethylphenol	1.4474	1.1534	**55**	2,6-Dichlorophenol	1.3146	0.8798
22	2-Nitrophenol	1.6591	1.1218	**56**	3,4-Dichlorophenol	1.6937	1.4718
23	3-Nitrophenol	0.9627	0.7029	**57**	3,5-Dichlorophenol	2.0081	1.6681
24	4-Nitrophenol	1.4118	1.2317	**58**	2,3,6-Trichlorophenol	1.8041	1.5101
25	2,4-Dinitrophenol	1.1488	1.0452	**59***	2,4,5-Trichlorophenol	1.9732	1.6722
26	2,5-Dinitrophenol	1.9795	1.8057	**60**	2,3,4,6-Tetrachlorophenol	1.6249	1.4462
27	3,4-Dinitrophenol	0.8382	0.6242	**61**	2,3,5,6-Tetrachlorophenol	1.6932	1.4308
28	3-Methyl-2-nitrophenol	1.0596	0.6532	**62**	Pentachlorophenol	1.7442	1.4523
29	3-Methyl-4-nitrophenol	1.1851	1.0288	**63***	Hydroquinone	0.2615	0.0178
30	4-Methyl-3-nitrophenol	0.7049	0.5671	**64**	4-Chlorocatechol	1.4611	1.1347
31	5-Methyl-2-nitrophenol	1.3170	1.1461	**65**	Resorcinol	0.1596	0.4928
32*	2-Methyl-4,6-dinitrophenol	1.6906	1.3966	**66**	4-Chlororesorcinol	0.4945	0.2652
33	2,6-Dimethyl-4-nitrophenol	1.4336	1.2099	**67**	4,6-Dichlororesorcinol	1.1922	1.0250
34	2-Chloro-4-nitrophenol	1.5767	1.4656				

* Test set

2.3. Statistical analysis

The structures of 67 phenols and anilines towards *Chlorella vulgaris* were studied by statistical methods based on the principal component analysis (PCA) (Camargo *et al.,* 2022) using the software XLSTAT version 2016 (Larif *et al.,* 2018). PCA is a statistical method useful to summarize all the information encoded in the structures of the compounds. It is also very helpful for understanding the distribution and classification of the data set (Kyaw *et al.,* 2009). This is an important descriptive statistical method which aims to present, in graphic forms.

The relationship between structure and toxicity was modeled using Multiple linear regression (MLR) analysis employing backward selection. This mathematical technique is useful as it minimizes the differences between actual and predicted values and selects the descriptors to be used as input

parameters in the multiple non-linear regression (MNLR) and artificial neural network (ANN) models (Hammoudan *et al.,* 2022; El Masaoudy *et al.,* 2023).

The MLR and MNLR were performed using the software XLSTAT version 2016 (Ousaa *et al.,* 2018), to predict toxic effects pIC_{20} and pIC_{50}. Equations were justified by the determination coefficient (R^2), mean squared error (MSE), Fisher's criterion (F) and significance level (P) (Camargo *et al.,* 2022). The ANN is an artificial system that is simulating the function of the human brain. Three components form 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 many different ANN types, the most commonly used in QSAR is the three-layered feed forward network (Aouidate *et al.,* 2018). In this type of network, 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 other neurons of a next layer and no connections are between neurons belonging to the same layer.

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 procedure, 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 (Ousaa *et al.,* 2018) and a detail of this algorithm is given elsewhere. The ANN analysis was performed using Matlab software version 2009a Neural Fitting tool (nftool) toolbox (Ghamali *et al.,* 2017).

In a QSTR study, the evaluation of the proposed models for stability, predictability, and generalization ability is an essential step. To validate the prediction ability of a QSTR model, two main methods are available: internal and external validations. Cross-validation is one of the most common methods used for internal validation. In this study, the internal predictive ability of each model was evaluated using leave-one-out cross-validation (R^2_{cv}). A good R^2_{cv} often indicates robustness and high internal predictive capacity of a QSTR model. However, recent studies (Shavalieva *et al.,* 2022; Gramatica *et al.,* 2014;) indicate that there is no clear correlation between the value of R^2_{cv} and the actual predictive capacity of a QSTR model, suggesting that R^2_{cv} alone remains inadequate as a reliable estimate of the model's predictive ability for all new chemicals. To determine both the generalizability of QSTR models for new chemicals and the true predictive ability of the models, statistical external validation is utilized during the model development step by properly employing a prediction set for validation.

3. Results

3.1. QSTR models and analysis

QSTR analysis was performed using the pIC20 and pIC50 of 67 phenols and anilines towards Chlorella vulgaris as reported in (Zhu *et al.,* 2010). From the results of the density functional theory DFT (B3LYP/6-31G (d)) calculations, fquantum chemistry descriptors obtained and the other molecular descriptors calculated by the ChemSketch program (Ousaa *et al.,* 2018) for building the model. to explain the Quantitative structure–toxicity relationship (QSTR), In this study 16 descriptors are calculated for the 67 molecules. For the correct validation of our data set with a QSAR model, the 67 compounds data were divided into training and test sets. A total of 60 molecules were placed in the training set to build the QSTR models, whereas the remaining 7 molecules composed the test set. The division was performed by random selection. The principle objective is to perform in the first time, a principal component analysis (PCA), which allows us to eliminate descriptors that are highly correlated (dependent), then an MLR analysis was performed on the remaining descriptors using the backward method until a valid model.

3.2. Principal component analysis

The set of descriptors coding the 67 phenols and anilines, electronic and physico-chemical descriptors are submitted to PCA analysis (Chtita et al.,2020). The first three principal axes are sufficient to encode the information provided by the data matrix. Indeed, the percentages of variance are 44.66%, 29.29% and 11.96% for the axes F1, F2 and F3, respectively. The total information is estimated to a percentage of 85.92%.

The principal component analysis (PCA) (Chtita et al.,2019) was conducted to identify the link between the different descriptors. Bold values are different from 0 at a significance level of p= 0.05. Correlations between the fourteen descriptors are shown as a correlation matrix. The obtained matrix provides information on the positive or negative correlation between descriptors. In general, the co-linearity (r>0.5) was observed between most of the variables, and between the variables and pIC_{50}. Additionally, to decrease the redundancy presented in our data matrix, the descriptors that are highly correlated ($R \geq 0.95$), were removed.

3.3. Multiple linear regression of the variable toxicity (MLR)

To generate the quantitative relationships between toxicity pIC_{50} and selected descriptors, our data set were subjected to the MLR and MNLR. Only variables with significant coefficients were retained.

In our study, we conducted multiple linear regression (MLR) analysis to establish quantitative relationships between toxicity (pIC20 and pIC50) and the selected descriptors. Only variables with significant coefficients were retained in the final MLR models. The objective of the analysis was to identify descriptors that showed statistically significant correlations with toxicity, as these descriptors are crucial in predicting the toxicity levels of the compounds. These models allowed us to establish quantitative relationships and gain insights into the potential toxicity of the compounds. The MLR analysis provided a valuable tool for capturing and quantifying the relationships between the selected descriptors and toxicity. It enhanced our understanding of toxicity patterns and facilitated making predictions based on the available dataset. During our analysis, we attempted to develop relationships with the indicator variables of toxicity, namely pIC_{20} and pIC_{50}. However, the best relationship obtained by this method is only one corresponding to the linear combination of two descriptors selected, the energy ELUMO and the octanol/water partition coefficient (log P) for pIC50 in **Eqn.1**, but for pIC20 is only one corresponding to the linear combination of two the total energy ET and the octanol/water partition coefficient (log P) in **Eqn.2**.

The resulting equations is:

$$\mathbf{pIC_{50}} = -1.275 - 0.170 \times \mathbf{E_{LUMO}} + 0.675 \times \mathbf{log\ \mathit{P}} \qquad \textbf{Eqn.1}$$

$$\mathbf{N = 60 \quad R^2 = 0.801 \quad R^2_{CV1} = 0.777 \quad MSE = 0.120 \quad F = 114.457 \quad p\text{-value} < 0.0001}$$

$$\mathbf{pIC_{20}} = -1.176 + 7.221\ 10^{-06}\ \mathbf{E_T} + 1.58\ \mathbf{log\ \mathit{P}} \qquad \textbf{Eqn.2}$$

$$\mathbf{N = 60 \quad R^2 = 0.821 \quad R^2_{CV2} = 0.793 \quad MSE = 0.064 \quad F = 47.672 \quad p\text{-value} < 0.0001}$$

For our 60 compounds, the correlation between experimental toxicity and calculated on based on these models is quite significant as indicated by statistical values. In the equations, **N** is the number of compounds, $\mathbf{R^2}$ is the determination coefficient, **MSE** is the mean squared error, **F** is the Fisher's criterion and **P** is the significance level.

A higher correlation coefficient and lower mean squared error indicate that the model is more reliable. A **P** that is smaller than 0.05 exhibits that the regression equation is statistically significant. The QSTR model expressed by **Eqn.1** and **Eqn.2** is cross-validated by its noticeable R^2cv value (R^2cv_1 =0.777

and R^2_{CV2} =0.793) obtained by the leave-one-out (LOO) method. A value of R^2cv is greater than 0.5 is the important criterion for qualifying a QSTR model as valid (Tiwari et al.,2022, Liu et al.,2022). The correlation coefficients between descriptors in the model were calculated by variance inflation factor (**VIF**) as shown in **table 2**. The **VIF** was defined as $1/(1-R^2)$, where R was the multiple correlation coefficients for one independent variable against all the other descriptors in the model (Rohand et al.,2021, Bakal et al., 2022). If **VIF** greater than 5, it mean that models were unstable and must be rejected, models with a **VIF** values between 1 and 4 can be accepted. As can be seen from **table 2**, the **VIF** values of the two descriptors are all smaller than 5.0, resulting that there is no-collinearity between the selected descriptors and the obtained model has good stability.

Table 2: The variance inflation factors (VIF) of descriptors in QSAR model

	pIC_{20}		pIC_{50}	
	E_{LUMO}	log *P*	E_T	log *P*
Tolerance	0.967	0.958	0.988	0.979
VIF	1.013	1.022	1.004	1.011

The elaborated QSTR model reveals that the toxicity of 60 phenols and anilines towards *Chlorella vulgaris* may be explained by the two selected descriptors in Eq (1). The negative correlation of the energy E_{LUMO} with the pIC_{50} shows that an increase in the values of this factor indicates a decrease in the value of the pIC_{50}, whereas a positive correlation of the octanol/water partition coefficient (log *P*) with the pIC_{50} reveals an increase in the value of the pIC_{50}.

3.4. Multiple nonlinear regression of the variable toxicity (MNLR)

The nonlinear regression method was also used to improve the structure toxicity in a quantitative way, taking into account several parameters. We have applied it to table containing 60 molecules associated with sixteen variables. We used a pre-programmed function of XLSTAT following:

$$\mathbf{Y = a + (b\ X1 + c\ X2 + d\ X3 + e\ X4\ \ldots)}$$

Where a, b, c, d...: represent the parameters and X1, X2, X3, X4,...: represent the variables.
The resulting equation is:

$$\mathbf{pIC_{50}} = -1{,}195 - 0{,}151\times\mathbf{E_{LUMO}} + 0{,}595\times\mathbf{log}\ \boldsymbol{P} + 8{,}116\ 10^{-3}\times\mathbf{E^2_{LUMO}} + 1{,}690\ \ 10^{-2}\times\mathbf{log^2}\boldsymbol{P} \qquad \textbf{Eqn.3}$$

$$\mathbf{pIC_{20}} = -8.376 + 6.221\ 10^{-06}\ \mathbf{E_T} + 0.601\ \mathbf{log}\ \boldsymbol{P}\ + 3.489\ 10^{-11}\ \mathbf{E^2_T} + 0.401\ 10^{-2}\ \mathbf{log^2}\ \boldsymbol{P} \qquad \textbf{Eqn.4}$$

The obtained parameters describing the topological and the electronic aspects of the studied molecules are:

$$\mathbf{N = 60 \quad R^2 = 0.802 \quad R^2_{CV3} = 0.751 \quad MSE = 0.124}$$

$$\mathbf{N = 60 \quad R^2 = 0.810 \quad R^2_{CV4} = 0.713 \quad MSE = 0.066}$$

The toxicity values pIC_{50} predicted by this model are almost similar to that observed.
The obtained coefficients of determination in **Eqn.3** and **Eqn.4** is quite very interesting. The QSTR models expressed by **Eqn.3** and **Eqn.4** is cross-validated by its appreciable R^2cv values obtained using the leave-one-out (LOO) method. A value of R^2cv is greater than 0.5 is the important criterion for qualifying a QSTR model as valid (Ousaa et al.,2018, Ghamali et 2017). To optimize the error standard deviation and to improve our model, we involve in the next part artificial neural networks (ANN).

3.5. *Artificial neural networks ANN*

In order to increase the probability of good characterization of studied compounds, neural networks (ANN) can be used to establish predictive models of quantitative structure–toxicity relationships (QSTR) between a set of molecular descriptors obtained from the MLR and observed toxicity. The ANN calculated toxicity model was developed using the parameters of the studied compounds. The correlation between ANN calculated and experimental toxicity values are very significant as illustrated in figure 1.

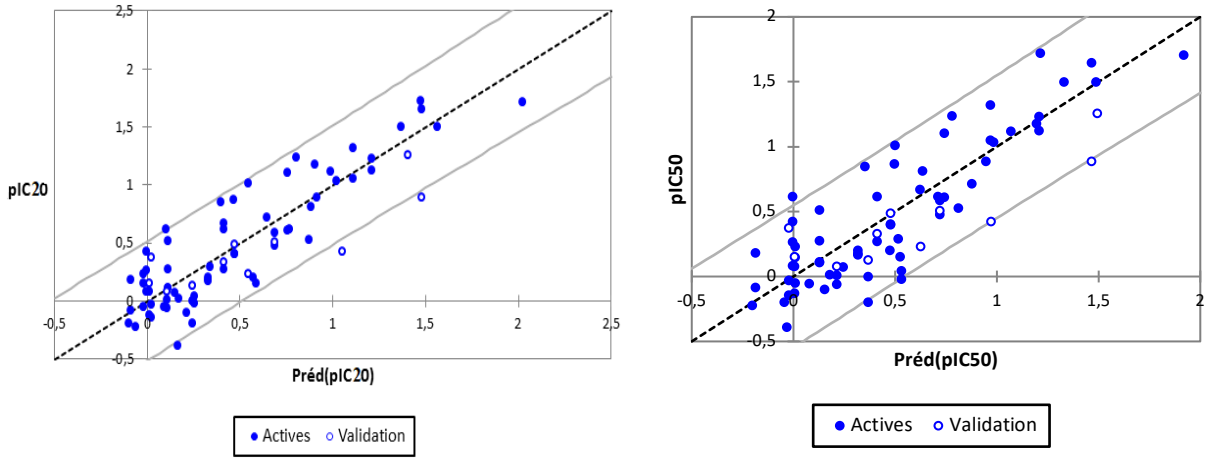

Figure 1: Correlations of observed and predicted activities calculated using ANN

pIC_{20} : **N = 60** $R^2 = 0.833$ $R^2_{CV5} = 0.764$ **MSE = 0.100**

pIC_{50} : **N = 60** $R^2 = 0.804$ $R^2_{CV6} = 0.772$ **MSE = 0.100**

The obtained determination coefficient (R^2) values is 0.833 and 0.804 for this data set of the phenols and anilines derivatives. This confirms that the artificial neural network (ANN) results are the best to predict the quantitative structure-activity relationship model. Furthermore, the R^2_{cv} values shows that the ANN model is the high predictive power.

3.6. Y-randomized analysis

To validate the robustness of the models, utilized Y-randomized analysis, as proposed by (Chtita et al.,2020, Ousaa et al 2018), to ensure the reliability of the models. This analysis involved scrambling the order of all experimental values and developing a new model based on this randomized data. The purpose was to determine if the original models exhibited a significant correlation by chance alone. For a model to be considered robust, it should have low R^2 and R^2_{cv} LOO values when applied to the randomized data. The Y-randomized analysis was repeated 400 times, and the average R^2 values (denoted as R^2 aver) for pIC_{50} and pIC_{20} were found to be 0.131,1,44, 151 and 0.136, respectively for RLM and RNLM. These values were significantly lower than the original model values. Consequently, the results indicate that there was no chance correlation during the modeling process, thereby confirming the stability and reliability of the original models (Lafridi el al 2020, Chtita et al 2019).

3.7. External validation

To evaluate the predictive capacity of the MLR, MNLR, and ANN models, it is necessary to employ a group of compounds that were not used to establish the QSTR model during the training phase. In this study, the QSTR models developed for 60 phenols and anilines were used to predict the toxicity of

seven additional compounds (test set). **Table 3** presents the primary performance measures for the three models. The results indicate that the ANN model exhibited more robust statistical indicators compared to the other models.

Table 3: Performance comparison between models obtained by MLR, RNLM and ANN

Model	Training set			Test set		
	R^2	R^2cv	MSE	R^2	R^2ext	MSE
pIC_{20}						
MLR	0.821	0.793	0.064	0.801	0.708	0.132
MNLR	0.810	0.713	0.066	0.802	0.716	0.115
ANN	0.832	0.794	0.124	0.812	0.726	0.115
pIC_{50}						
MLR	0.801	0.777	0.120	0.801	0.708	0.132
MNLR	0.802	0.751	0.124	0.802	0.716	0.115
ANN	0.824	0.794	0.100	0.824	0.773	0.109

After evaluating the optimal linear QSTR regression equations established in this research, it was observed that the predictive capability of the ANN model is better than that of the MLR and MNLR models. The ANN model provides superior results, indicating a highly satisfactory correlation between molecular descriptors and the toxicity of the compounds examined To demonstrate the accuracy, generalizability and predictability of the proposed models, we compared key statistical parameters such as the R or R^2 obtained using different statistical tools and descriptors. These results are presented in **Table 4**.

Table 4: Observed values and calculated values of pIC_{20} and pIC_{50} according to different methods

N°	pIC_{20} (obs.)	pIC20 (calc.)			N°	pIC_{50} (obs.)	pIC_{50} (calc.)		
		MLR	NMLR	ANN			MLR	NMLR	ANN
1*	0.0313	0.0350	0.0303	−0.033	**1***	−0.0846	−0.1046	0.0341	−0.2549
2	0.6072	-0.6101	0.6002	0.5859	**2**	0.3872	0.3672	0.3870	0.3973
3	0.5745	-0.5901	0.5705	0.4111	**3**	0.4424	0.4024	0.4423	0.2023
4	0.5745	0.5650	0.5715	0.5157	**4**	0.3326	0.3626	0.3324	0.4847
5	0.3502	0.4541	0.3412	0.5046	**5**	0.1377	0.1577	0.1372	0.2786
6	0.7724	0.6910	0.7314	0.4910	**6**	0.5779	0.5579	0.5777	0.3143
7	0.6757	0.6143	0.6227	0.2647	**7**	0.5147	0.5047	0.5146	0.0464
8	0.3971	0.3652	0.3831	0.4353	**8**	0.2357	0.2357	0.2354	0.2472
9	−0.0257	0.0118	0.0257	0.3112	**9**	−0.2948	−0.2848	−0.2949	0.0618
10	0.0744	0.1040	0.0734	0.1241	**10**	−0.1893	−0.1793	−0.1890	−0.0821
11	0.7512	0.7107	0.7502	0.8106	**11**	0.5101	0.5001	0.5103	0.6183
12	0.4969	0.6828	0.4939	0.7029	**12**	0.3561	0.3461	0.3561	0.4850
13	0.4980	0.4041	0.4970	0.4740	**13**	0.1566	0.1466	0.1561	0.1733
14	1.3884	1.3092	1.3814	1.2591	**14**	1.2075	1.2175	1.2074	1.0018
15	1.8199	1.7194	1.8159	1.6793	**15**	1.6074	1.5974	1.6072	1.4023
16	−0.5101	−0.397	−0.4121	−0.197	**16**	−0.684	−0.674	0.6841	−0.439
17	0.6637	0.5866	0.6677	0.7862	**17**	0.5093	0.4893	0.5093	0.5580
18	1.2366	0.8989	1.1326	0.7986	**18**	0.8930	0.8830	0.8931	0.5793
19*	1.0606	1.0790	1.0666	1.1796	**19***	0.8550	0.8350	0.8550	0.9408
20	1.3716	1.2513	1.4726	1.3514	**20**	1.1718	1.1418	1.1719	1.1294
21*	1.4474	1.3639	1.3434	1.3438	**21***	1.1534	1.1434	1.1536	1.1326
22	1.6591	1.1613	1.6581	1.1214	**22**	1.1218	1.1318	1.1213	0.8706
23	0.9627	1.2010	0.8637	1.1610	**23**	0.7029	0.7029	0.7021	0.9103

24	1.4118	1.3137	1.4128	1.2736	**24**	1.2317	1.2217	1.2318	1.0318
25	1.1488	1.5841	1.2478	1.6241	**25**	1.0452	1.0252	1.0454	1.4435
26	1.9795	1.8679	1.8785	1.8272	**26**	1.8057	1.8757	1.7056	1.6106
27	0.8382	1.3220	0.8392	1.4820	**27**	0.6242	0.6342	0.6241	1.2398
28	1.0596	0.9439	1.1506	0.9030	**28**	0.6532	0.6332	0.6533	0.7045
29	1.1851	1.0858	1.0861	0.9851	**29**	1.0288	1.0088	1.&286	0.8058
30	0.7049	0.9001	0.7359	0.9406	**30**	0.5671	0.5571	0.5673	0.7439
31	1.3170	1.3080	1.3180	1.2689	**31**	1.1461	1.1361	1.1469	1.0492
32*	1.6906	1.5347	1.6816	1.4940	**32***	1.3966	1.3466	1.2967	1.3078
33	1.4336	1.3049	1.5326	1.2642	**33**	1.2099	1.2199	1.2094	1.0439
34	1.5767	1.5440	1.4757	1.6044	**34**	1.4656	1.4456	1.4656	1.3555
35	2.2617	1.9978	2.3607	1.9772	**35**	1.8485	1.8585	1.7483	1.7178
36	1.5405	1.7601	1.4495	1.8061	**36**	1.3667	1.3567	1.3669	1.5455
37*	1.0765	1.2036	1.0755	1.2540	**37***	0.9582	0.9682	0.8582	1.0155
38	−0.2951	−0.380	−0.306	−0.486	**38**	−0.684	−0.674	−0.584	−0.846
39*	0.9149	0.5568	0.8959	0.4963	**39***	0.4378	0.4478	0.4377	0.1484
40	1.4822	1.6290	1.3832	1.7294	**40**	1.0296	1.0296	1.0294	1.3508
41	0.5611	0.4989	0.6621	0.5011	**41**	0.3021	0.3121	0.3020	0.2330
42	0.3963	0.3887	0.4953	0.4181	**42**	0.1505	−0.1605	−0.1503	0.1020
43	1.5467	1.4697	1.5477	1.6691	**43**	1.2512	1.2412	1.3517	1.3688
44	1.6957	1.6487	1.7967	1.6286	**44**	1.5046	1.5146	1.5042	1.2586
45	0.1553	0.4500	0.0543	0.7530	**45**	0.0308	0.0408	0.0306	0.4592
46	0.7187	0.6718	0.6177	0.5710	**46**	0.1757	0.1657	0.1756	0.2766
47	1.1968	1.2590	1.2958	1.3593	**47**	0.8624	0.8724	0.8625	1.0808
48	0.9747	1.0430	0.9737	1.0436	**48**	0.6943	0.6443	0.6942	0.7218
49*	1.9160	1.6798	1.8150	1.5792	**49***	1.6815	1.6315	1.6812	1.2792
50	0.3512	0.3112	0.3522	0.3110	**50**	0.0012	0.0112	0.0015	0.0968
51	0.4677	−0.309	0.4667	−0.529	**51**	0.6003	0.6103	0.6004	−0.7586
52	0.3656	0.3839	0.3646	0.5819	**52**	0.1731	0.1531	0.1736	0.3222
53	0.7129	1.0081	0.6119	1.0780	**53**	0.4568	0.4968	0.5569	0.8227
54	1.5401	1.5124	1.5411	1.5422	**54**	1.2437	1.2337	1.2433	1.2794
55	1.3146	1.1992	1.3126	1.1193	**55**	0.8798	0.8898	0.8790	0.8424
56	1.6937	1.7177	1.4937	1.7178	**56**	1.4718	1.4618	1.4710	1.4584
57	2.0081	1.7977	2.0001	1.7847	**57**	1.6681	1.6581	1.6682	1.5109
58	1.8041	1.6919	1.8001	1.6117	**58**	1.5101	1.5001	1.5100	1.3432
59*	1.9732	1.9011	1.8712	1.9408	**59***	1.6722	1.6622	1.6724	1.6765
60	1.6249	1.7430	1.6249	1.7930	**60**	1.4462	1.4362	1.4461	1.5459
61	1.6932	1.6719	1.6922	1.6811	**61**	1.4308	1.4208	1.4307	1.4265
62	1.7442	1.6078	1.7432	1.6072	**62**	1.4523	1.4623	1.4525	1.3800
63*	0.2615	0.3374	0.2605	0.4073	**63***	0.0178	0.0278	0.0173	0.1574
64	1.4611	1.4356	1.4601	1.4057	**64**	1.1347	1.1447	1.1348	1.1209
65	0.1596	−0.100	0.2506	−0.130	**65**	0.4928	0.4728	0.4923	−0.395
66	0.4945	0.7160	0.3935	0.7162	**66**	0.2652	0.2452	0.3651	0.4382
67	1.1922	1.0958	1.0912	1.0355	**67**	1.0250	1.0050	1.0252	0.7834

* Test set

3.8. Domain of applicability

To estimate the reliability of any QSTR model and its ability to predict new compounds, the domain of applicability must be essentially defined. The predicted compounds that fall within this domain may be considered as reliable. The applicability domain was discussed with the Williams graph in figure 2,

which the standardized residuals and the leverage values (h_i) are plotted (Rücker et al.,2007, Ousaa et al., 2023). It is based on the calculation of the leverage h_i for each molecule, for which QSAR model is used to predict its toxicity:

$$h_i = x_i (X^T X)^{-1} x_i^T \qquad (i = 1, \ldots, n) \qquad \textbf{Eqn. 5}$$

Where x_i is the row vector of the descriptors of compound i and X is the variable matrix deduced from the training set variable values. The index T refers to the matrix/vector transposed. The critical leverage h^* is, generally, fixed at 3(k+1)/N, where N is the number of training molecules, and k is the number of model descriptors. If the leverage value h of molecule is higher than the critical value (h^*) i.e., $h > h^*$, the prediction of the compound can be considered as not reliable (Nawaz et al.,2022, Chtita et al., 2021).

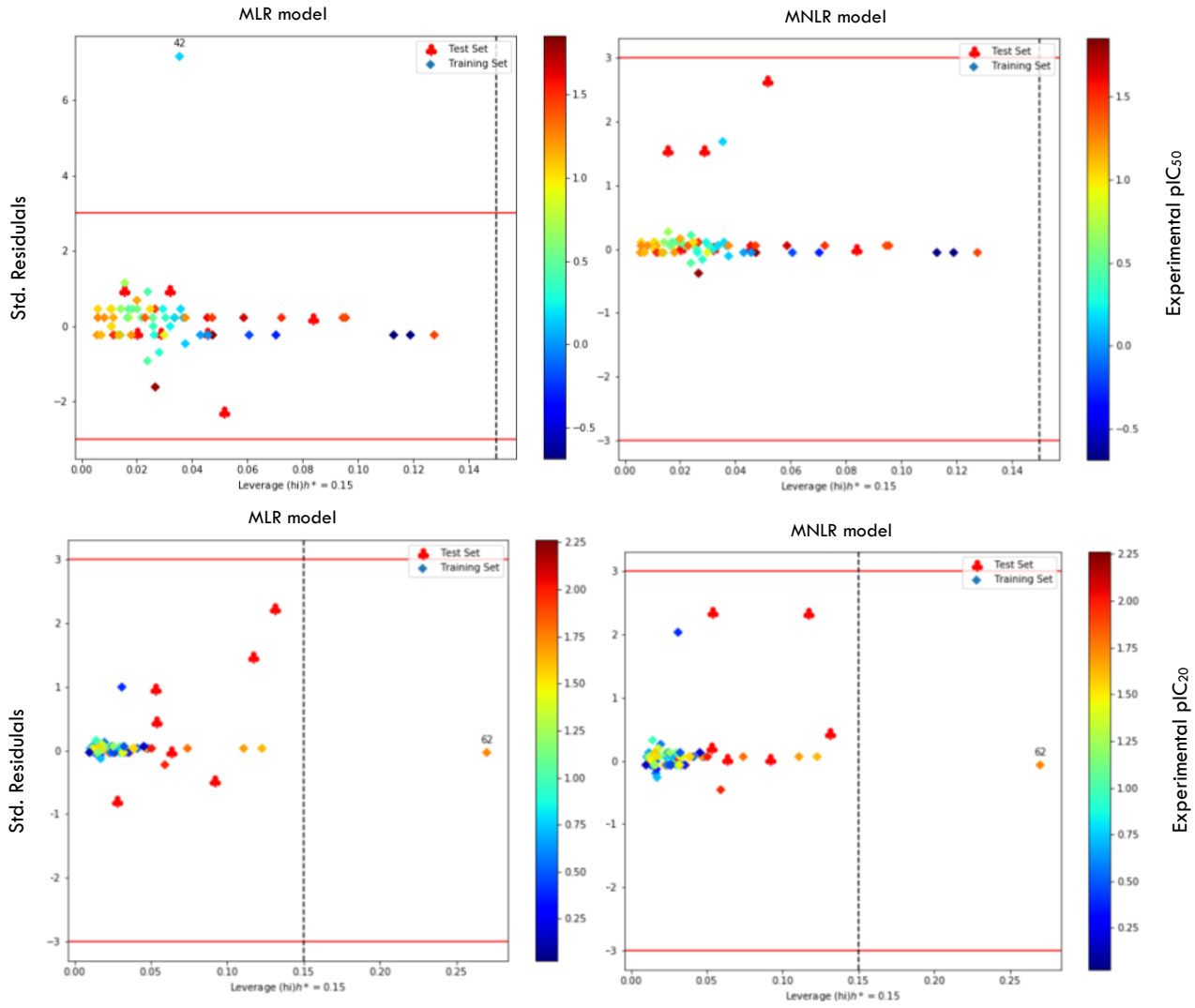

Figure 2: Williams plot for the presented MLR and MNLR model

The Williams plot for the presented MLR and MNLR models is shown in figure 2. From this plot, the leverage values (h_i) of any molecule in the training and test sets are less than the critical value (h^* = 0.15) excepting the compounds 42 and 62 as outliers. Also, the standardized residuals of all molecules in the training and test sets are less than three standard deviation units (±3σ). Thus, the predicted toxicity by the developed MLR and MNLR models is reliable.

Conclusion

This study employed three distinct techniques (MLR, MNLR, and ANN) to develop QSTR models for predicting the toxicity of phenols and anilines towards Chlorella vulgaris. The models established in this research may be employed with greater ease compared to previous models, with an enhanced degree of confidence in their prediction accuracy. Our analysis establishes a robust correlation between various descriptors and the pIC50 and pIC20 values of the phenol and aniline derivatives studied using the three techniques.

Furthermore, robust statistical analyses, including leave-one-out cross validation, Y-randomized analysis, and external validation and the applicability domains (AD) were performed to validate the models. These analyses confirmed the accuracy, stability, and reliability of the developed models. Based on these findings, it was concluded that the QSTR models built on uniform descriptors could deliver satisfactory performance in predicting multiple toxicity endpoints of chemicals towards *Chlorella vulgaris*.

These results encourage collaboration between theoretical researchers and pharmacologists, academic or industrial to protect nature and the environment, especially in aquatic systems.

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

References

Aarjane M.; Slassi S.; Ghaleb A.; Amine A. (2022) Synthesis, biological evaluation, molecular docking and *in silico* ADMET screening studies of novel isoxazoline derivatives from acridone, *Arabian Journal of Chemistry* , 14,4,103057 , https://doi.org/10.1016/j.arabjc.2021.103057

Aouidate A., Ghaleb A., Ghamali M.; Chtita S.; Ousaa A.; Choukrad M.; Sbai A.; Bouachrine M.; Lakhlifi T., (2018) QSAR study and rustic ligand-based virtual screening in a search for aminooxadiazole derivatives as PIM1 inhibitors, *Chemistry Central Journal*, 12, Article 32, https://doi.org/10.1186/s13065-018-0401-x

Aouidate A.; Ghaleb A.; Ghamali M.; Chtita S.; Ousaa A.; Choukrad M.; Sbai A.; Bouachrine M.; Lakhlifi T., (2018). 3D-QSAR and molecular docking Studie on derivatives of 7-Azaindole as inhibitors of Trk A: a strategic design in novel anticancer agents, *Letters in Drug Design & Discovery*, 15, https://doi.org/10.2174/1570180815666171229151138

Aruoja V.; Sihtmäe, M.; Dubourguier, H.; CKahru, A., (2011) Toxicity of 58 substituted anilines and phenols to algae Pseudokirchneriella subcapitata and bacteria Vibrio fischeri: comparison with published data and QSARs. *Chemosphere*, 1310–1320, https://doi.org/10.1016/j.chemosphere.2011.05.023.

Azzaoui K., Mejdoubi E., Lamhamdi A., Zaoui S., Berrabah M., Elidrissi A., Hammouti B., Fouda M.M.G., Al-Deyab S.S., (2015) Structure and properties of hydroxyapatite/Hydroxyethyl cellulose acetate composite films, *Carbohydrate Polymers* 115, 170–176, https://doi.org/10.1016/j.carbpol.2014.08.089

Bakal R.L.; Jawarkar R.D.; Manwar, J. V; Jaiswal, M.S.; Ghosh, A.; Gandhi, A.; Zaki, M.E.A.; Al-Hussain, S.; Samad, A.; Masand, V.H., (2022) Identification of Potent Aldose Reductase Inhibitors as Antidiabetic (AntiHyperglycemic) Agents Using QSAR Based Virtual Screening, Molecular Docking, MD Simulation and MMGBSA Approaches. *Saudi Pharm. J*, https://doi.org/10.1016/j.jsps.2022.04.003.

Bakire, S.; Yang, X.; Ma, G.; Wei, X.; Yu, H.; Chen, J., (2018) Developing predictive models for toxicity of organic chemicals to green algae based on mode of action. *Chemosphere*, 463–470, https://doi.org/10.1016/j.chemosphere.2017.10.028.

Belhassan A., Chtita S., Lakhlifi T., Bouachrine M., (2019) QSPR Study of the Retention/Release Property of Odorant Molecules in Water, Dairy and Pectin gels. *Mater. Today Proc*, 13, 621– 629. https://doi.org/10.1016/j.matpr.2019.04.021

Bouyad K., Kandri Rodi Y., Elmsellem H., El Ghadraoui E. H., Ouzidan Y., Abdel-Rahman I., Kusuma H.S., Warad I., Mague J.T., Essassi E.M., Hammouti B., Chetouani A. (2018) azo[4,5-b]pyridines as a new class of corrosion inhibitors for mild steel: Experimental and DFT approach, *Mor. J. Chem.* 6, 22-34

Camargo A., (2022) PCA test: testing the statistical significance of Principal Component Analysis in R. PeerJ, 10, 12967. https://doi.org/10.7717/peerj.12967.

Chen J.W.; Peijnenburg W.J.; Quan, X. (2000), The Use of PLS Algorithms and Quantum Chemical Parameters Derived from PM3 Hamiltonian in QSPR Studies on Direct Photolysis Quantum Yields of Substituted Aromatic Halides, *Chemsphere* 1319-1326.

Chtita S.; Belhassan, A.; Bakhouch, M.; Taourati, A.I.; Aouidate, A.; Belaidi, S.; Moutaabbid, M.; Belaaouad, S.; Bouachrine, M.; Lakhlifi, T., (2021) QSAR Study of Unsymmetrical Aromatic Disulfides as Potent Avian SARS-CoV Main Protease Inhibitors Using Quantum Chemical Descriptors and Statistical Methods. *Chemom. Intell. Lab. Syst*., 210, 104266, https://doi.org/10.1016/j.chemolab.2021.104266.

Chtita S.; Ghamali, M.; Ousaa, A.; A.Aouidate, A.; A. Belhassan, A.; A. Idrissi Taourati, A.; Elidrissi, B.; Bouachrine, M.; Lakhlifi, T., (2020) QSAR study of N-substituted oseltamivir derivatives as potent avian influenza virus H5N1 inhibitors using quantum chemical descriptors and statistical methods, *Royal Society of Chemistry and the Centre National de la Recherche Scientifique*, 44, 1747-1760. https://doi.org/https://doi.org/10.1039/C9NJ04909F

Chtita S.; Ghamali, M.; Ousaa, A.; Aouidate, A.; Belhassan, A.; Idrissi Taourati, A.; Hariram Masand, V.; Bouachrine, M.; Lakhlifi, T., (2019) QSAR study of anti-Human African Trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and Lipinski's descriptors, *Heliyon*., e01304. https://doi.org/10.1016/j.heliyon.2019.e01304

Daoui O.; Elkhattabi, S.; Chtita, S., (2023) Design and prediction of ADME/Tox properties of novel magnolol derivatives as anticancer agents for NSCLC using 3D-QSAR, molecular docking, MOLCAD and MM-GBSA studies. *Letters in Drug Design & Discovery*, 20, 545- 569, https://doi.org/10.2174/1570180819666220510141710.

El Masaoudy Y., Tabti K., Koubi Y., Maghat H., Lakhlifi T., Bouachrine M. (2023) In silico design of new pyrimidine-2,4-dione derivatives as promising inhibitors for HIV Reverse Transcriptase-associated RNase H using 2D-QSAR modeling and (ADME/Tox) properties, *Mor. J. Chem.,* 14(2), 300-317. https://doi.org/10.48317/IMIST.PRSM/morjchem-v11i2.35455

Ghamali M.; Chtita S.; Ousaa A.; Elidrissi B.; Bouachrine M.; Lakhlifi T. (2017) QSAR analysis of the toxicity of phenols and thiophenols using MLR and ANN, *Journal of Taibah University for Science*, 9, 1-10. https://doi.org/10.1016/j.jtusci.2016.03.002

Gramatica P. (2014) External Evaluation of QSAR Models, in Addition to Cross-Validation: Verification of Predictive Capability on Totally New Chemicals, Molecular Informatics, 33(4), 311-314, https://doi.org/10.1002/minf.201400030

Hammoudan I., Chtita S., Bakhouch M., Riffi Temsamani D. (2022) QSAR study of a series of peptidomimetic derivatives towards MERS-CoV inhibitors, *Mor. J. Chem.,* 10(3), 405-416

Jawarkar R.D.; Bakal, R.L.; Zaki, M.E.A.; Al-Hussain, S.; Ghosh, A.; Gandhi, A.; Mukerjee, N.; Samad, A.; Masand, V.H.; Lewaa, I., (2022) QSAR Based Virtual Screening Derived Identification of a Novel Hit as a SARS CoV-229E 3CLpro Inhibitor: GA-MLR QSAR Modeling Supported by Molecular Docking, Molecular Dynamics Simulation and MMGBSA Calculation Approaches. *Arab. J. Chem*., 15, 103499, https://doi.org/10.1016/j.arabjc.2021.103499.

Lafridi H.; Ousaa, A.; Zgou, H.; Bouachrine, M. (2020) QSAR modeling of antiradical properties of phenolic compounds using DFT calculations, *Moroccan Journal of Chemistry*., 8(4), 841-852. https://doi.org/10.48317/IMIST.PRSM/morjchem-v8i4.20579

Larif M.; Adad A.; Hmammouchi R.; Taghki A.I.; Soulaymani A.; Elmidaoui A.; Bouachrine M.; Lakhlifi T., (2016) Biological activities of triazine derivatives. Combining DFT and QSAR results. *Arab. J. Chem.*, 10, S946-S955. http://dx.doi.org/10.1016/j.arabjc.2012.12.033 .

Liu W.; He M.; Li Y.; Peng Z.; Wang G. (2022) A Review on Synthetic Chalcone Derivatives as Tubulin Polymerisation Inhibitors. *J. Enzyme Inhib. Med. Chem*., 37, 9–38, https://doi.org/10.1080/14756366.2021.1976772.

Llorach-Pares L.; Nonell-Canals, A.; Avila, C.; Sanchez-Martinez, M., (2022) Computer-Aided Drug Design (CADD) to De-Orphanize Marine Molecules: Finding Potential Therapeutic Agents for Neurodegenerative and Cardiovascular Diseases. *Mar. Drugs*., 20, https://doi.org/10.3390/md20010053.

Myint K.Z.; Xiang-Qun, X., (2015) Ligand Biological Activity Predictions Using Fingerprint-Based Artificial Neural Networks (FANN-QSAR), 149-164. https://doi.org/10.1007/978-1-4939-2239-0_9

Nawaz M.Z.; Attique S.A.; Ain Q.U.; Alghamdi H.A.; Bilal M.; Yan W.; Zhu D. (2022) Discovery and characterization of dual inhibitors of human Vanin-1 and Vanin-2 enzymes through molecular docking and dynamic simulation-based approach. *Int J Biol Macromol*, 213, 1088-1097, https://doi.org/10.1016/j.ijbiomac.2022.06.014.

Ousaa A.; Elidrissi, B.; Ghamali, M.; Chtita, S.; Aouidate, A.; Bouachrine, M.; Lakhlifi T. (2018) QSAR study of (5-nitroheteroaryl-1,3,4-thiadiazole-2-yl) piperazinyl derivatives to predict new similar compounds as antileishmanial agents. *Advances in Physical Chemistry*, https://doi.org/10.1155/2018/2569129

Ousaa A.; Elidrissi, B.; Ghamali, M.; Chtita, S.; Aouidate, A.; Bouachrine, M.; Lakhlifi, T., (2018) Quantitative structure-toxicity relationship studies of aromatic aldehydes to Tetrahymena pyriformis based on electronic and topological descriptors, *Journal of Materials and Environmental Sciences*, 9,1,256-266, https://doi.org/10.26872/jmes.2018.9.1.29.

Ousaa A.; Elidrissi B.; Ghamali M.; Chtita S.; Aouidate A.; Ghaleb A.; Bouachrine M.; Lakhlifi T. (2018) QSAR and docking studies on the pyrimidine derivatives analogs with antileishmanial activity. *RHAZES: Green and Applied Chemistry*, 1, 1, 50-63, https://doi.org/10.48419/IMIST.PRSM/rhazes-v1.16442

Ousaa A.; Elidrissi, B.; Ghamali, M.; Chtita, S.; Aouidate, A.; Bouachrine, M.; Lakhlifi, T. (2023) Discovery of new inhibitors of Sars-CoV: QSAR study using Density functional theory (DFT) and statistical methods, *Journal of Materials and Environmental Sciences*, 14, 326-336.

Bansal S. & Singh S. S. (2018), Synthesis of some novel 6,6'-(1,4-phenylene)bis(4-phenylpyrimidin-2-ol/-2-amine/2-thiol catalyzed by $MgFe_2O_4$ and evaluation of their antimicrobial activities Mor. J. Chem. 6 N°4, 700-707

Pervaiz M.; Munir A.; Riaz, A.; Saeed, Z.; Younas, U.; Imran, M.; Ullah, S.; Bashir, R.; Rashid, A.; Adnan, A., (2022) Amalgamation, scrutinizing, and biological evaluation of the antimicrobial aptitude of thiosemicarbazide Schiff bases derivatives metal complexes. *Inorg Chem Comms*, 141, 109459, https://doi.org/10.1016/j.inoche.2022.109459.

Rajarshi G.; Willighagen, E., (2012) A Survey of Quantitative Descriptions of Molecular Structure. *Curr Top Med Chem*, 12,18, 1946–1956, https://doi.org/10.2174/156802612804910278.

Rohand T.; Ben EL Ayouchia, H.; Achtak, H.; Ghaleb, A.; Derin, Y.; Tutar, A.; Tanemura, K. (2021) Design, synthesis, DFT calculations, molecular docking and antimicrobial activities of novel cobalt, chromium metal complexes of heterocyclic moiety-based 1,3,4- oxadiazole derivatives, *Journal of Biomolecular Structure and Dynamics*. 1-14 https://doi.org/10.1080/07391102.2021.1965031

Rücker C., Rücker, G., Meringer, M. (2007) y-Randomization and its variants in QSPR/ QSAR. *J. Chem. Inf. Model*., 47, 2345–2357, https://doi.org/10.1021/ci700157b

Shavalieva G., Papadokonstantakis S., and Gregory Peters G. (2022) Prior Knowledge for Predictive Modeling: The Case of Acute Aquatic Toxicity, *Journal of Chemical Information and Modeling*, 62(17), 4018-4031 DOI: 10.1021/acs.jcim.1c01079

Sofi M.Y.; Shafi, A.; Masoodi, K.Z. (2022) Chapter 22 - Introduction to Computer-Aided Drug Design. In; Sofi, M.Y., Shafi, A., Masoodi, K.Z.B.T.-B.; pp. 215–229. https://www.sciencedirect.com/science/article/pii/B9780323911283000021?via%3Dihub

STATITCF Software, Technical Institute of cereals and fodder, Paris, France., (1987).

Tiwari A.; Singh S. (2022) Chapter 13 - Computational Approaches in Drug Designing. In; Singh, D.B., Pathak, R.K.B.T.-B. Academic Press; 207–217 ISBN 978-0-323-89775-4, https://doi.org/10.1016/B978-0-323- 89775-4.00010-9.

Tugcu G.; Sacan M.T. (2018) A multipronged QSAR approach to predict algal low-toxiceffect concentrations of substituted phenols and anilines. *J. Hazard. Mater*, 344, 893–901, https://doi.org/10.1016/j.jhazmat.2017.11.033 .

Urbizo-Reyes U.; Liceaga, A.M.; Reddivari, L.; Kim, K.-H.; Anderson, J.M. Enzyme Kinetics., (2022) Molecular Docking, and in Silico Characterization of Canary Seed (Phalaris Canariensis L.) Peptides with ACE and Pancreatic Lipase Inhibitory Activity. J. Funct. Foods, 88, 104892, https://doi.org/10.1016/j.jff.2021.104892.

Zhu M.J.; F,Ge.; Zhu, R., (2010) A QSAR Study of the Acute Toxicity of Halogenated Phenols. *Chemosphere*, 46-52, https://doi.org/10.1016/j.proeng.2012.08.035.

(2023) ; https://revues.imist.ma/index.php/morjchem/index