Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
R. Hmamouch / Mor. J. Chem. 4N°4 (2016) 1061-1075

Density Functional Theory Based Quantitative Structure-Activity Relationship Study of Cycloguanil Derivatives Acting as Plasmodium falciparum.

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

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

** Corresponding author: l majdoulinelarif@yahoo.com*
Received 21 Sept 2016,
Revised 10 Dec 2016,
Accepted 22 Dec 2016

Abstract

This work presents a study of quantitative structure-activity relationship (QSAR) on the cycloguanil derivatives which are reported as growth inhibitors of clone of Plasmodium falciparum (T9/94 RC17) which houses A16V+S108T mutant dihydrofolate reductase (DHFR) enzyme. A set of 24 molecule-derived cycloguanil was modeled using the Gauss View software (03) using DFT B3LYP 6,6-31G-31G (d) as a base function. The obtained descriptions are purely electronic. The set constitute the inhibitory activity and the calculated electronic descriptors were statistically processed with principal component analysis (PCA), multiple linear regression (MLR), multiple nonlinear regressions (MNLR) and artificial neural network (ANN). The results obtained by the artificial neural network (ANN) show that the expected activities are in good agreement with the experimental results, with equal correlation coefficient R = 0, 912.To determine the architecture of this network, we varied the number of hidden layers, the number of neurons in the hidden layers, the transfer functions and the pairs of transfer functions. The best results were obtained with a network architecture [3-3-1], activation functions (Tansig-Purelin) and a learning algorithm of Levenberg-Marquardt.

Keywords: Quantitative structure–activity relation ; inhibitory activity ; PCA ; MLR ; MNLR ; (ANN) ; Levenberg-Marquardt

1. Introduction

Malaria is the most prevalent infectious disease in tropical and sub-tropical regions of the world. It is mainly due to a parasite of the genus Plasmodium, spread by the bite of certain species of Anopheles mosquitoes. In fact, P. falciparum is the most dangerous agent that is responsible for the majority of mortality associated with this disease [1]. Cycloguanil antimalarial molecule is used as a medicine containing the role of inhibiting dihydrofolate reductase function (DHFR) enzyme by disrupting the DNA formation that ultimately results in the death of the parasite cells. In response to the emergence of chloroquine-resistant strains, many regions of the world made this drug. Despite this significant loss of efficiency, there are several reports in the literature on the QSAR models developed for the purpose of understanding of the mechanism of drug resistance to cycloguanil P. falciparum [2]. Maitarad et al. Made 3D-QSAR/CoMFA and 3D-QSAR/CoMSIA studies using similar cycloguanil against wild-type and mutant enzymes quadruple PfDHFR [3]. A 3D-QSAR study was done on the basis of the reported activities against cycloguanil derivatives A16V+S108T PfDHFR mutant enzyme.

This ledus to generate a quantitative structure-activity relationship study using the studied compounds for which activities (pIC_{50}) growth inhibition against P. falciparum (T9/94 RC17) hosting A16V + S108T mutant DHFR enzyme are reported in the literature [4, 5]. We relied on a table containing 24 molecules collected randomly from the work of Legesse Adane et al [6]. Using a GaussView (03) [7, 8], we inserted electronic parameters such as activation energy (E_a), the total energy (E_T), the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), (μ) the dipole moment (λ_{max}) the maximum absorption and the factor of oscillation (f_{SO}).

To do this, we applied linear regression, nonlinear regression, principal component analysis (PCA) and neural networks. This showed that the model for the prediction is of LM learning algorithm of PMC type and [3-3-1] architecture.

2. Materials and methods

2.1. Chemicals

Previous studies [4, 5] have introduced a number of derivatives of cycloguanil compounds that were evaluated for their inhibitory activities against the 16V+S108T mutant enzyme, these activities are defined as the constant inhibitor concentration required for 50% growth inhibition (pIC_{50}). A set of data derived from 24 cycloguanil was used as input data for the current study. And for this purpose, we have introduced training and testing selection that contains 20 and 4 compounds produced randomly. Hence, pIC_{50} values were used as dependent variables for the results of the RLM, RLNM and ANN analysis.

The following figure shows the chemical structures of the compounds studied and their corresponding experimental pIC_{50} activities.

Figure 1: Structures and biological activities (pIC_{50}-Obs) of drift studied cycloguanil

The experimental activity of the studied compounds was collected from previous works of Legesse Adane and Prasad V. Bharatam [6]. All chiral molecules are enantiomerically pure and in the configuration S (Sinister). The range of data on activity ranges from 5,44 to 8,40.

2.2. Calculation and selection of descriptors

We utilized the Gauss View (03) molecular modeling software to represent the molecules. Hence, we used the «The functional theory of density» (DFT) method to obtain the final geometry. It is known from literature that this method has become very popular in recent years due to its capacity to reach similar accuracy to other methods in less time and lower computing cost. In accordance with the DFT results, the energy of the fundamental state of a polyelectronic system can be expressed by the total electron density, which is used instead of the wave function to compute the energy composing the fundamental basis of DFT [9], by pressing B3LYP, 6-31G* as a base function [10]. The optimized molecular geometries were transferred by the same Gauss View (03) computer software as a file (notepad) in which we extract the values corresponding to the chosen descriptors.

2.3. Methods of Data Analysis

2.3.1. Principal component analysis (PCA)

The principal component analysis (PCA) lets us transform a set of variables correlated to a new set of variables, called principal components. They are fewer but independent. Using these new variables, the dimensionality of the system is reduced with a minimum loss of information [11]. The obtained matrix of coordinates allows us to analyze the dispersion of individuals in the new defined space [12, 13, 14, and 25]. Thus, two samples that are very close graphically carry similar information, those which contribute the least are close to the origin, and those contributing the most are close to the large values (positive or negative).It follows that two highly correlated variables will be

close. Therefore, we can say that the PCA is an unsupervised analysis because it considers all the variables independently. It is useful in the identification of key variables as well as groups of variables correlated. Beyond its use in the reduction of dimensionality of the problem, it serves to emphasize the most characteristic information from a set of data both in terms of variables and in terms of descriptors. With this type of analysis, it is possible to graphically categorize samples into classes or to highlight correlated or not significant variables in a set.

2.3.2. Multiple linear regressions

The multiple linear regression (MLR) was performed with the XLSTAT (2013) software. It is the simplest and most commonly used software for the development of predictive models [15]. It rests on the assumption that there is a linear relationship between a quantitative variable "y" from several explanatory variables {x_1, x_2, x_3, x_4 ... x_p}, which are taken electronic descriptors by software Gauss View (03). Or "p" is the number of variables.

$$y = b_0 + b_1x_1 + \cdots + b_px_p$$

2.3.3. Validation

**** Coefficients and standard statistical tests***

The quality of fit was assessed by:

- ✓ *Correlation coefficient (R):*

$$R = \sqrt{1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \sum_i \frac{\hat{y}_i}{n})^2}}$$

Where y_i and $\hat{y}_i$ are the observed and calculated values of the dependent variables.

n: number of the considered data points.

These coefficients determine the variance of the target activity that is explained by the model QSAR, which is to say by the regression of the target activity based on the initial activity. These activity coefficients are not affected by the selected unit of measurement and interpret:

* A good correlation between the target activity and initial activity if r is close to 1.

* A non-linear correlation between the target activity and initial activity if r is close to 0.

- ✓ ***Standard Deviation (S):***

$$S = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{(n-k)}}$$ *where k: number of restrictions to the degrees of freedom.*

It measures the change in the target activity that is not explained by the QSAR model. In particular, the smaller the standard deviation is the best the correlation will be. Its value is always function of the unit of measurement of the target activity and also takes into account experimental errors which explain that a too small value has no meaning.

Cross-validation techniques have been applied for the evaluation of the internal prediction model.

**Internal validation*

Cross-validation is the most commonly used method to determine the stability of the predictive model and to test the influence of each sample on the final model. In fact, there are at least three cross-validation techniques: "Test set validation" or "holdout method", "k-fold cross-validation" and "leave-one-out cross validation" (LOOCV).

This process involves extracting a number n of molecules of the original set of k molecules and building a new model with the remaining n-k molecules using the chosen descriptors (only the regression constants change).This new model

is then used for the prediction phase on the n withdrawn molecules. This process is then repeated to withdraw and predict the values of all the molecules of the training set [16, 17].

External validation

To test that reliably predictive power, the use of a set of external validation, not used for the development of the model is required. As long as the original data set is large enough, the latter can be easily divided into two: a driving game in which the model is developed and a set of validation used to characterize its predictive power.

2.3.4. Artificial neural networks (ANNs)

The ANN analysis was performed with the use of the Matlab (2014) software, neural mounting tool (nntool) toolbox, on all cycloguanil derived data. Artificial neural network is a non-linear empirical model [18, 19] that is used in the prediction of biological activity, while its application is booming in many disciplines. It is, among others, an interesting alternative to traditional statistics for the data processing. In this work, we explained some key concepts of how RNA and especially the multi-layer preceptor work.

2.3.4.1. Architecture of neural networks

Typically, a neural network is defined by the architecture which is characterized by the transfer function, and how the interconnection is made between neurons.

There are several transfer functions, the choice is made depending on the problem to solve. They are also chosen because of their ease of implementation and that of their derivative which is involved in the optimization algorithms.

In our case, the selected network is a multilayer one. This choice is made for the ease and speed of construction in addition to the fact that our problem has a limited number of input variables [20, 21].

2.3.4.2. Multilayer perceptron **(MLP)**

The **MLP** is a layer propagation network model (Figure 2). The neurons are organized in layers: an input layer, an output layer and in between one or more intermediate layers, which are also called hidden layers.

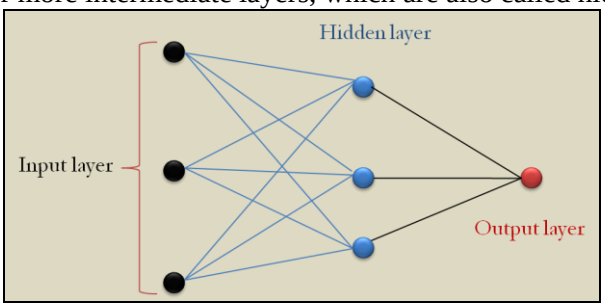

Figure 2: Multilayer Perceptron [3-3-1].

While in theory an **MLP** can have multiple layers, in practice a single hidden layer is sufficient (Hornik, 1991). To establish an **MLP** it is necessary to choose the transfer functions, to identify the relevant inputs, the number of neurons in the hidden layer, choose the algorithm then optimize and test the network.

• ***Transfer functions***

Neural networks are used for approximation of non-linear models. Nonlinearity is introduced by the transfer functions used especially for the nodes of the hidden layer. The transfer of the output layer is linear function while in theory, any nonlinear function can be used. Those selected are generally those easy to calculate and to derive.

According to Dawson and Wilby (2001), the logic sigmoid transfer function (logsig) is the most widely used. It is defined as:

$f(x) = \frac{1}{1+e^{-x}}$ Bounded between 0 and 1

In this work, we mainly use the linear transfer function (purlin), and function of hyperbolic tangent sigmoid transfer (tansig). Who are the most used in the modeling.

- ✓ **Purlin** (Linear transfer function): purelin is a neural transfer function. Transfer functions calculate a layer's output from its net input.
- ✓ **Tansig** (Hyperbolic tangent sigmoid transfer function): tansig is a neural transfer function. Transfer functions calculate a layer's output from its net input.

3. Results and discussion

In this study, we focused on a series of 24 cycloguanil derivatives to determine a quantitative relationship between structure and biological activity pIC_{50}. In this section, we will use the same approach as we have already used in previous works [12, 14]. Table 1 shows the values of the calculated parameters obtained from optimized structures by optimized DFT/B_3LYP 6-31G (d).

Table 1: Values of the obtained parameters by DFT/B3LYP 6-31G (d) optimization of studied compounds:

Molec.	pIC_{50}	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	Ea	λ_{max}	$f_{(so)}$
1	5,61	-1162,669	-5,179	-0,343	4,836	1,713	4,05	306,12	0,409
2	6,50	-1084,044	-6,819	-0,025	6,795	4,272	2,765	448,44	0,156
3	6,46	-1123,355	-6,668	0,046	6,714	5,991	3,965	312,71	0,466
4	6,31	-1162,669	-5,475	0,051	5,526	6,018	4,012	309,05	0,505
5	6,64	-1201,983	-3,388	0,058	3,446	6,105	4,018	308,57	0,506
6	6,60	-1241,296	-7,239	0,061	7,300	6,110	4,01	309,16	0,507
7	5,55	-1196,819	-3,307	-1,351	1,956	6,142	3,91	317,10	0,439
8	7,36	-1354,399	-6,630	-0,208	6,422	4,517	3,956	313,41	0,401
9	5,44	-742,386	-6,401	0,191	6,592	2,253	3,924	315,95	0,460
10	6,33	-703,072	-6,407	0,232	6,640	3,074	4,010	309,21	0,440
11	6,29	-742,384	-6,337	0,237	6,574	3,039	4,009	309,24	0,427
12	6,82	-781,698	-5,595	0,243	5,838	3,121	4,013	308,98	0,424
13	5,46	-781,696	-5,658	0,316	5,974	2,859	4,008	309,33	0,415
14	7,41	-934,119	-6,448	0,009	6,458	2,295	3,957	313,33	0,378
15	6,35	-703,068	-6,587	0,187	6,775	2,522	3,971	312,23	0,443
16	6,00	-802,301	-6,549	-0,042	6,506	3,645	3,944	314,40	0,466
17	6,45	-624,442	-3,328	0,360	3,688	3,997	3,971	312,23	0,443
18	6,51	-723,675	-7,300	0,121	7,421	5,143	3,921	316,21	0,456
19	6,53	-1162,665	-6,006	-0,122	5,884	4,210	3,957	313,34	0,450
20	7,55	-1123,351	-6,604	-0,089	6,515	5,185	3,909	317,16	0,463
21	7,62	-1315,125	-5,727	-0,263	5,463	4,749	3,988	310,92	0,383
22	7,54	-1315,125	-5,753	-0,282	5,471	5,859	4,055	305,74	0,437
23	8,40	-1359,222	-5,829	-0,076	5,753	5,249	4,006	309,53	0,410
24	8,40	-1162,669	-5,179	-0,343	4,836	1,713	4,050	306,12	0,409

3.1. Statistical analysis (implementation of the PCA)

The graphical representation of cycloguanil molecules and the study of the electronic properties have shown that a large number of chemical and electronic parameters were significant; largely, there is a link between these parameters. So it seems worth trying to process statistical data using a multivariate analysis method such as principal component analysis (PCA).

3.2. Study of the eigenvalues

The purpose of drawing a bar graph representing the total inertia is to get the maximum inertia preserved with minimal factors.

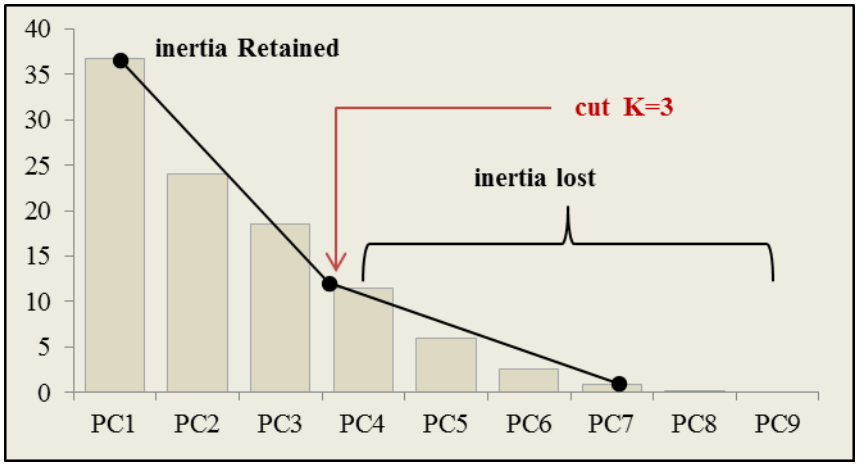

Figure 3: Total inertia diagram

We observe a significant drop from the 4th axis (PC4) (from 11,44% to 0,85% of inertia) as this one retains 3rd PC3 axis representing almost 79,28% of the total inertia c. That is to say, we can explain 79,28% of the information of the table.

3.3. Correlation Matrix

Table 2: Correlation matrix (Pearson (n)) between different obtained descriptors

Variables	**pIC_{50}**	**E_T**	**E_{HOMO}**	**E_{LUMO}**	**ΔE**	**μ**	**E_a**	**λ_{max}**	**$f_{(so)}$**
pIC_{50}	**1**								
E_T	**-0,578**	**1**							
E_{HOMO}	0,026	-0,041	**1**						
E_{LUMO}	**-0,535**	**0,791**	0,04	**1**					
ΔE	-0,136	0,205	**-0,978**	0,167	**1**				
μ	0,236	**-0,51**	-0,013	-0,082	-0,004	**1**			
E_a	0,148	-0,311	0,432	-0,251	-0,478	-0,071	**1**		
λ_{max}	-0,145	0,31	-0,432	0,246	0,477	0,071	-1	**1**	
$f_{(so)}$	-0,364	0,033	0,03	0,289	0,03	**0,565**	-0,12	0,119	**1**

- ✓ E_T the total energy is positively correlated with the energy LUMO (r = 0,791 p<0,05) at a significant level and negatively correlated with the dipole moment μ (r = 0,791 p <0,05) at an insignificant level.
- ✓ The LUMO energy is negatively correlated with the energy ΔE (r = 0,978 p <0,05) at a very significant level.

✓ The dipole moment μ is positively correlated with $f_{(so)}$ (r = 0,791 p <0,05) at an insignificant level.

Bold values are different from 0 at a significant level for p<0,05

At a very significant level for p<0,01

At a highly significant level for p<0,001

3.4. Construction of projected point clouds

The correlation circle shows the projection of the electronic variables on the factorial PC1; PC2 shows that the axis PC1 characterizes molecules having a high activation energy (E_a), and high wavelength (λ_{max}).

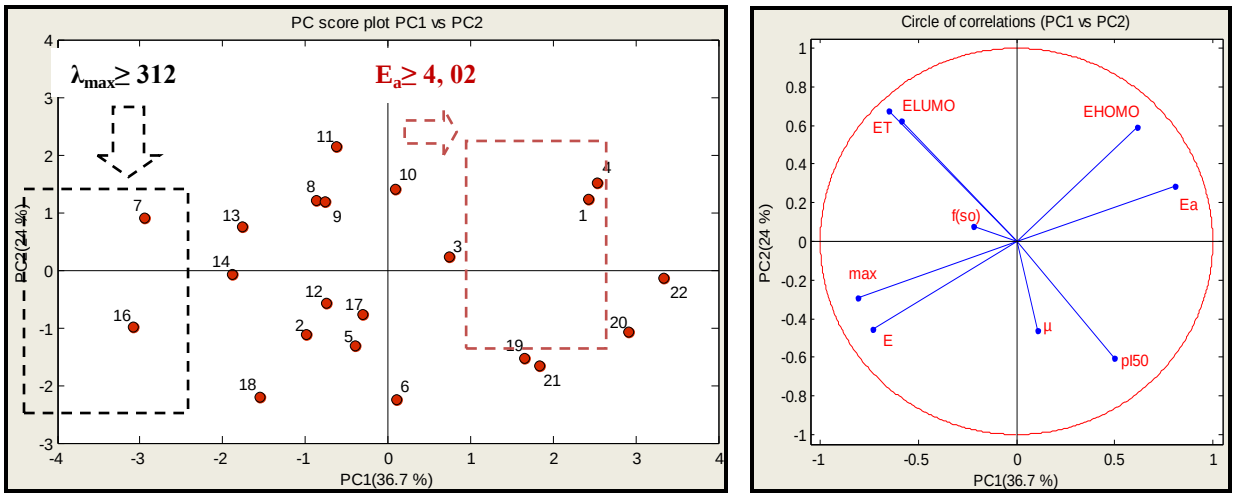

Figure 4: Screening of individuals and variables on factorial Plans (PC1, PC2).

So we can say that the PC1 axis opposes chemical molecules [1; 4; 20; 22] having a high activation energy (E_a), to chemical molecules [7; 13; 14; 16; 8] with a high wavelength (λ_{max}).

As against the contribution of the PC2 axis, this study characterizes the total energy (E_T) with a cosine equal 0,455 is not significant.

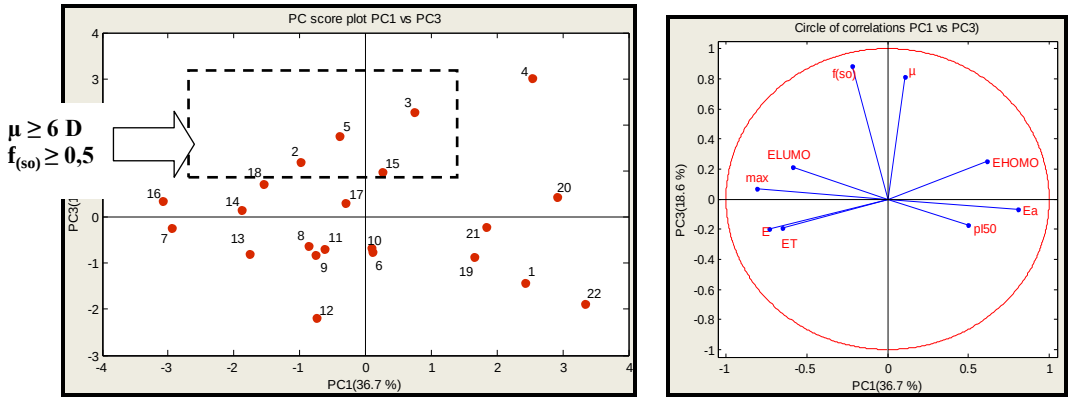

Figure 5: Screening of individuals and variables on factorial Plans (PC1, PC3).

The axis PC3 characterizes molecules [3; 4; 5] having strong dipole moment (μ) and oscillation factor ($f_{(so)}$).

On the other hand, the PC1-PC2-PC3 projection (Fig.6) (79,272% of the total variance) also shows that we can distinguish three groups of molecules with special structure convenience.

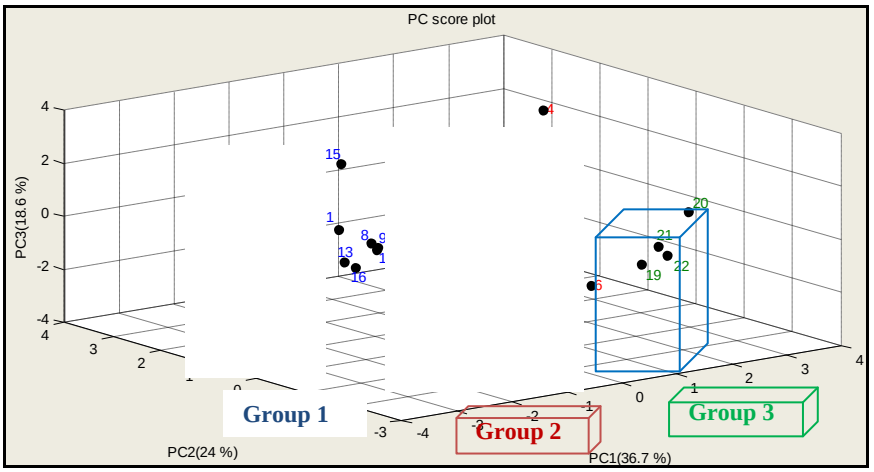

Figure 6: Cartesian diagram according to PC1, PC2 and PC3: Separation between the three groups.

- ✓ **Group 1** (G1): Cycloguanil containing compounds with aliphatic groups in position (C6) and position (C4'-C5 ').
- ✓ **Group 2** (G2): Containing cycloguanil compounds having aliphatic moieties in position (C6) and substituent such as Cl, H at position (C4'-C5 ') p-chlorophenyl.
- ✓ **Group 3** (G2): containing cycloguanil compounds having aliphatic moieties in position (C6) and substituent such as H, Cl position (C4'-C5 ') meta-chlorophenyl.
- ✓ Taking account of the following structure:

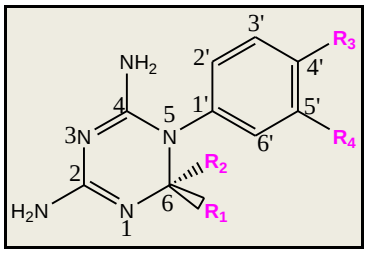

Figure 7: cycloguanil structure

3.5. Multiple linear regressions

The descriptors obtained from molecular structures optimized at DFT / B3LYP 6-31G (d), were used for the development of QSPR models to predict the biological activity of compounds 24 cycloguanil.

The multiple linear regression technique with the correlation coefficient (R), Mean Squared Error (MSE) and Fisher test (F) was used to extract the best performing models. Initially, a QSPR model was developed on 20 molecules in the database (driving set) and 4 randomly chosen molecules for the external validation. Descriptors corresponding to these 20 compounds have been incorporated into our data analysis approach. To develop a quantitative model MLR technique was used by XLSTAT (2014) software. The best model is an equation with three descriptors correlated well with experiment (R = 0.841) and more robust with (R_{cv}= 0,729).

$$pIC_{50} = 18{,}153 - 1{,}585 \times 10^{-3} \times E_T - 2{,}608 \times 10^{-2} \times \lambda_{max} - 10{,}710 \times f_{(so)} \quad (1)$$

N = 20 R= 0,841 F = 12, 89 MSE = 0,173

Where E_T is the total energy, wavelength λ_{max} and $f_{(so)}$ oscillation factor.

The predicted activity (pIC_{50} calculated from equation 1 in the optimal model of multiple linear regressions) and the values observed are given in table 7. Descriptors proposed in this equation were used as input parameters for the non-linear regression and the artificial neural network.

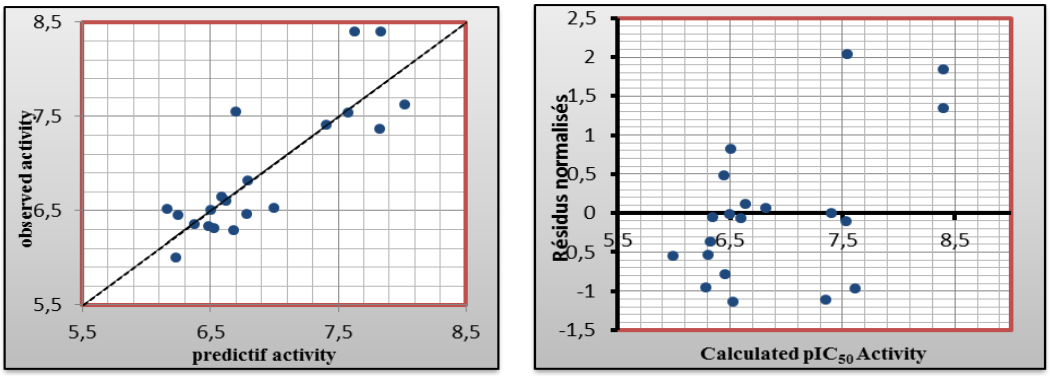

Figure 8: Relationship between the estimated values of pCI_{50}, their predictions and their residues established by (MLR).

3.6. Multiple non-linear regressions

We also used a non-linear regression model to improve the structure-activity relationship and to evaluate the effect of the substituent. We applied the proposed descriptors by multiple linear regressions for 20 molecules in the whole formation and we used the correlation coefficient (R) and the Mean squared error (MSE) to select the best performance of regression. We used a pre-programmed function XLSTAT of a second degree polynomial type.

The resulting equation was:

$$pIC_{50} = 30{,}448 - 3{,}873 \times 10^{-3} \times E_T - 0{,}176 \times \lambda_{max} + 36{,}16 \times f_{(so)} - 1{,}071 \times 10^{-6} \times E_T^2 + 2{,}383 \times 10^{-4} \times \lambda_{max}^2 - 53{,}007 \times f_{so}^2$$

N = 20 R= 0,848 MSE = 0,205

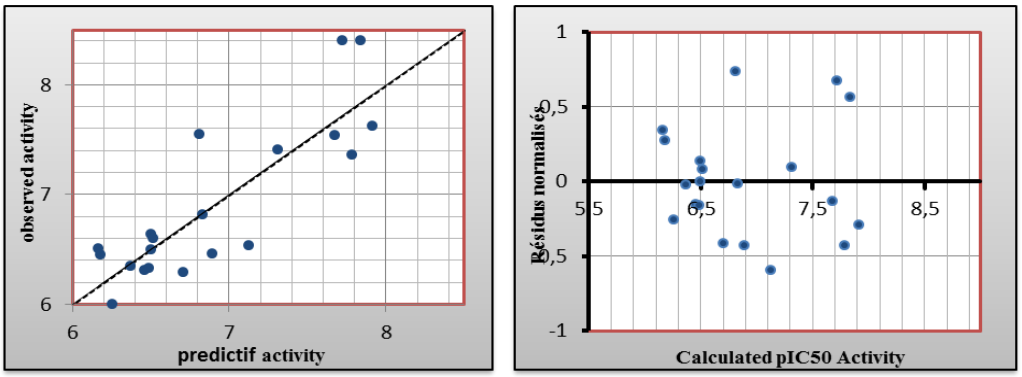

Figure 9: Relationship between the estimated values of pCI_{50}, their predictions and their residues established by (NMLR)

The predicted activities calculated from the equation (2) and the values observed are given in table 7. The real predictive power of a quantitative model of structure-activity relationship is its ability to accurately predict the activities of the compounds in an external test set (compounds not used in the developed model). The activities of the left 4 molecules derived from the training set by the multiple linear and nonlinear regressions. A Comparison of pIC_{50}-test values and pIC_{50}-obs shows a good forecast for the four compounds:

***Multiple linear regressions*:**

N = 4 R_{test}= 0,915 R^2_{test} = 0,837

Multiple non-linear regressions:

N = 4 R_{test}= 0,93 R^2_{test} = 0,862

Table 3: the observed activities predicted pIC_{50} and residuals in MLR and MNLR for 4 test compounds (test set).

No	Obs	MLR		MNLR	
		Pred-test	**Resid-test**	**Pred-test**	**Resid-test**
1	5,61	7,630	-2,020	7,720	-2,110
8	5,55	7,078	-1,528	7,195	-1,645
16	5,44	6,162	-0,722	6,167	-0,727
20	5,46	6,877	-1,417	6,897	-1,437

3.7. Artificial neural networks

In order to increase the probability to characterize the compounds, artificial neural networks can be used to generate predictive models of quantitative structure-activity relationship between a set of molecular descriptors obtained from the multiple linear regression and the observed activities. The determination of the type of architecture PMC-time neural network asks a question about the choice of the number of hidden layer, the number of hidden neurons, the number of iterations and transfer functions. For this we randomly divided our database into three parts: 60% for the training and 20% for the test and 20% for validation.

- ***Choice of number of hidden layers:***

Table 4: presents the calculations of R and MSE for one, two, three and four hidden layers.

Number of hidden layers	MSE	R
1	**0,013**	**0,707**
2	0,25	0,584
3	0,79	0,327

An increase in the number of hidden layers increases the load calculations without any performance report. We can therefore ensure that the use of a single hidden layer is preferable for the PMC model type.

- ***Choice of the number of hidden neuron:***

A parameter, ρ, was proposed for determining the number of hidden neurons, which play a major role in determining the best architecture of artificial neural network [22, 23], defined as follows:

$$\rho = \frac{\text{Number of data points in the training set}}{\text{Sum of the number of connections in the artificial neural}}$$

In order to avoid over-fitting or under-fitting, it is recommended that $1.8 < \rho < 2.3$ [24]. So with three hidden neurons, the output layer represents the calculated activity values (pIC_{50}). The architecture of the ANN used in this work **[3-3-1]** is depicted in figure 10. All calculations of NN are done on Matlab 2014.

- ***Choice of transfer functions and the number of iterations:***

In this study, we used the Levenberg-Marquardt algorithm (LM) which has high performance qualified learning.

In this case, we changed the number of neurons in the hidden layer and the pairs of transfer functions. The performance was evaluated by the mean squared error (MSE) and the correlation coefficient (R).

The following table shows the best performance found for various combinations of transfer couples.

Table 5: Transfer functions torques according to their performance.

Appellations	*Hidden Layer function*	*Output Layer function*	*R*	*MSE*	*Number of Iterations*
T-T	Tansig	Tansig	0,130	1,360	10
T-L	Tansig	Logsig	0,805	0,841	9
T-P	**Tansig**	**Purlin**	**0,912**	**0,094**	**6**
L-L	Logsig	Logsig	0,250	2,570	11
L-T	Logsig	Tansig	0,623	3,310	7
L-P	Logsig	Purlin	0,515	0,217	8
P-P	Purlin	Purlin	0,556	0,479	12
P-L	Purlin	Logsig	0,507	1,120	9
P-T	Purlin	Tansig	0,556	0,186	8

From the results obtained, the transfer couple functions (Tansig-Purlin) gave a correlation coefficient R = 0,912 and a mean square error MSE = 0,094, with an architecture of [3-3-1]. With this configuration we get to a better performance of the LM learning algorithm. This performance was met after 6 iterations. We can say from these results that the most powerful model in predicting the biological activity of cycloguanil compounds is the one used as a transfer function, the tansig function in the hidden layer and purlin function in the layer output while using a learning algorithm LM, PMC configuration deviation [3-3-1] and containing three layers (Figure 10):

- ✓ 3 neurons in the grafted layer, representing the independent electronic variables;
- ✓ 3 neurons in the hidden layer;
- ✓ One neuron of the output layer, representing the biological activity (pIC_{50}) of cycloguanil compounds.

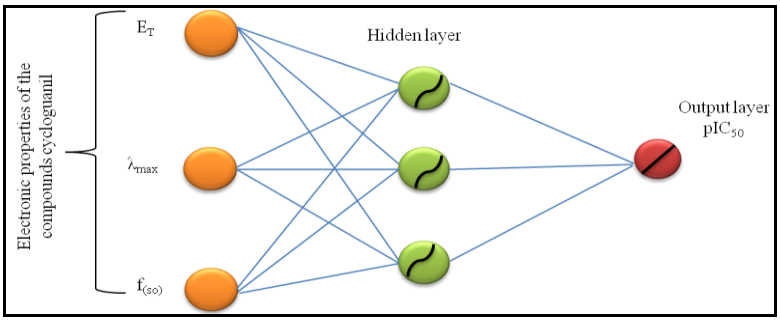

Figure10: The architecture of a PMC to 3 input variables, 3 neurons in the hidden layer and one neuron to the output layer.

The ANN calculated activity model were developed using the properties of several studied compounds. The correlation between ANN calculated and experimental activity values were very significant as indicated by R and R^2 values.

N = 20 R = 0,912 R^2 = 0,831 MSE = 0,094

These values show that the relationship between the estimated values of pIC_{50} and their residues established by artificial neural networks. They are illustrated in figure 11. The statistic of the three steps of the calculation by the ANN: Training, validation and test are illustrated in table 6. In this part, we investigated the best linear QSAR regression equations established in this study. Based on these findings, a comparison of the quality of MLR and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives

better results than MLR. ANN was able to establish a satisfactory relationship between the electronic descriptors and the activity of the studied compound.

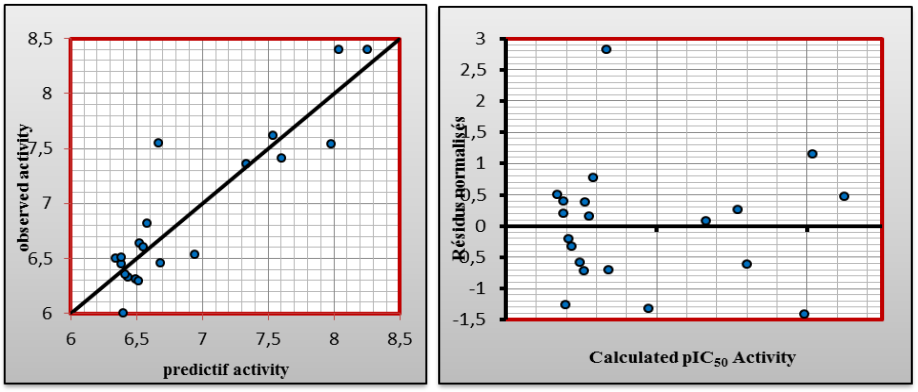

Figure 11: The relationship between the estimated values of pCI_{50}, their predictions and their residues established by (RNA)

Table 6: Values obtained by ANN

	Samples	**RMSE**	**R**	**R^2**
Training	12	0,038	0,9461	0,8951
Validation	4	0,012	0,983	0,9663
Test	4	0,025	0,972	0,9448

4. Conclusion

In this work, we studied QSAR to predict the biological activity of several cycloguanil compounds which were evaluated for their inhibitory activities against A16V + S108T mutant enzyme responsible for malaria, the results show that the relationship between the anti-malaria activity pIC_{50} on other electronic parameters cycloguanil molecules is non-linear. More precisely, we can say that the artificial neural network had significantly better predictive ability than the other two models, with more predictive power.

The Levenberg-Marquardt algorithm has exhibited better performance in terms of statistical indicators, as well as the network architecture is [3-3-1], with a non-linear activation function for the hidden type tansig layer and Purelin for the output layer, with a very good prediction of antimalarial activity.

We have established meaningful relationships between several electronic descriptors and inhibitory activity against A16V + S108T mutant enzyme with a powerful cross-validation, the model proposed in this activity precise descriptors $\mathbf{E_T}$, $\boldsymbol{\lambda_{max}}$, $\mathbf{f_{(so)}}$, are highly relevant.

Comparison of key statistical terms as R or R^2 different models obtained using different statistical tools and various electronic descriptors is shown in table 7.

Table 7: observed values and calculated values of pCI_{50} according to different methods

N° of		MLR		NMLR		ANN		CV	
Compound	**pIC_{50}(obs)**	**pred**	**Resid**	**pred**	**Resid**	**pred**	**Resid**	**pred**	**Resid**
1	6,50	6,505	-0,005	6,502	-0,002	6,290	0,158	6,022	-3,523
2	6,46	6,788	-0,328	6,891	-0,431	6,659	-0,222	6,869	-0,409
3	6,31	6,532	-0,222	6,462	-0,152	6,452	-0,181	6,59	-0,280
4	6,64	6,591	0,049	6,502	0,138	6,486	0,117	6,651	-0,011
5	6,60	6,627	-0,027	6,517	0,083	6,519	0,047	6,659	-0,059
6	7,36	7,827	-0,467	7,787	-0,427	7,367	0,027	7,932	-0,572
7	6,33	6,486	-0,156	6,488	-0,158	6,391	-0,105	6,596	-0,266
8	6,29	6,689	-0,399	6,708	-0,418	6,481	-0,227	6,798	-0,508
9	6,82	6,795	0,025	6,835	-0,015	6,546	0,242	6,792	0,028
10	7,41	7,414	-0,004	7,312	0,098	7,66	-0,192	7,410	0,000
11	6,35	6,376	-0,026	6,371	-0,021	6,372	-0,067	6,591	-0,241
12	6,00	6,230	-0,230	6,259	-0,259	6,35	-0,397	6,867	-0,867
13	6,45	6,251	0,199	6,178	0,272	6,338	0,064	6,109	0,341
14	6,51	6,167	0,343	6,164	0,346	6,338	0,124	5,997	0,513
15	6,53	7,004	-0,474	7,127	-0,597	6,946	-0,416	6,924	-0,394
16	7,55	6,704	0,846	6,813	0,737	6,643	0,883	6,564	0,986
17	7,62	8,024	-0,404	7,915	-0,295	7,591	0,080	8,558	-0,938
18	7,54	7,584	-0,044	7,675	-0,135	8,073	-0,443	8,291	-0,751
19	8,40	7,840	0,560	7,839	0,561	8,368	0,146	7,528	0,872
20	8,40	7,636	0,764	7,723	0,677	8,133	0,362	7,361	1,039

Acknowledgements

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

References

[1] O.A. Santos-Filho, A. J. Hopfinger, A search for sources of drug resistance by the 4D QSAR analysis of a set of antimalarial dihydrofolate reductase inhibitors, *J. Comput.-Aided Mol. Des*. 15 (2001) 1–12.

[2] D. Hecht, M. Cheung, G.B. Fogel, QSAR using evolved neural networks for the inhibition of mutant Pf DHFR by pyrimethamine derivatives, *Biosystems* 92 (2008) 10–15.

[3] P. Maitarad, S. Hannongbua, S. Kamchonwongpaisan, J. Vanichtanankul, T. Vilaivan, T. Yuthavong, Interactions between cycloguanil derivatives and wild typeand resistance-associated mutant P. falciparum dihydrofolatereductases, *J. Comput.-Aided. Mol. Des*. 23 (2009) 241–252.

[4] Y. Yuthavong, T. Vilainvan, N. Chareonsethakul, S. Kamchonwongpaisan, W. Sirawaraporn, R. Quarrell, G. Lowe, Development of a lead inhibitor for the A16V + S108T mutant of DHFR from cycloguanil-resistant strain (T9/94) of P. falciparum, *J. Med. Chem*. 43 (2000) 2738–2744.

[5] S. Kamchonwongpaisan, R. Quarrell, N. Charoensetakul, R. Ponsinet, T. Vilaivan, J. Vanichtanankul, B. Tarnchompoo, W. Sirawaraporn, G. Lowe, Y. Yuthavong, Inhibitors of multiple mutants of P. falciparum DHFR and their antimalarial activities, *J. Med. Chem*. 47 (2004) 673–680.

[6] L. Adane, 3D-QSAR analysis of cycloguanil derivatives as inhibitors of A16V + S108T mutant Plasmodium falciparum dihydrofolate reductase enzyme, *Journal of Molecular Graphics publishing and Modelling* 28 (2009) 357-367.
[7] K. Laarej; M. Bouachrine; S. Radi; S. Kertit and B. Hammouti, *E-Journal of Chemistry*, (2010) 7(2), 419-424.
[8] H. Zarrok; H. Oudda; A. Zarrouk; R. Salghi; B. Hammouti; M. Bouachrine, *Der PharmaChemica*, (2011) 3 (6): 576-590.
[9] C. Adamo, V. Barone, *Chem. Phys. Lett.*, 330 (2000) 152-160. , Gaussian 03, Revision B.01, M J. Frisch; and al., Gaussian Inc., 2003, Pittsburgh, PA.
[10] AD. Becke, *J. Chem. Phys.*, 1993, 98, 1372, C. Lee; W. Yang; RG. Parr, *Phys. Rev*., 1988, B. 37, 785-789., C. Lee; W. Yang; RG. Parr, *Phys. Rev*., B. 37 (1988) 785-789.
[11] STATITCF Software, Technical Institute of cereals and fodder, (1987) Paris, France.
[12] R. Hmamouchi; M. Larif; A. Adad; M. Bouachrine; T. Lakhlifi, Int. *J. Adv. Res. Comp. Sci. Soft. Eng.*, (2014) 4 (2), 241-251.
[13] R. Hmamouchi; M. Larif; A. Adad; M. Bouachrine; T Lakhlifi, *Journal of Computational Methods in Molecular Design*, (2014) 4 (3):61-71.
[14] R. Hmamouchi; A. I. Taghki,; M. Larif; A. Adad; A. Abdellaoui; M. Bouachrine and T. Lakhlifi, *Journal of Chemical and Pharmaceutical Research*, 5(9) (2013) 198-202.
[15] M. Lejeune, Statistiques : la théorie et ses applications, Springer-Verlag, Paris, 2004.
[16] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model*., (2002) 20, 269-276.
[17] A. Tropsha, P. Gramatica, K.V. Gomba, QSAR Comb. Sci., (2003) 22, 69-77.
[18] D. Mantzaris, G. Anastassopoulos, "Intelligent prediction of vesicoureteral reflux disease,*"WSEAS Trans. Syst*, vol. 4, (2005) pp. 1440-1449.
[19] S. Baboo, I. Shereef, "An efficient weather forecasting system using artificial neural network".
[20] I. Manssouri, M. Manssouri, B. El Kihel, "Fault Detection by K-NN algorithm and MLP neuronal networks in distillation column," *Journal of information, Intelligence and knowledge*, vol. 3, (1992) pp.72-75.
[21] R. Nayak, L. Jain, B. Ting, "Artificial neural networks in biomedical engineering: a review," *Proc. 1st Asian-Pacific Congr. Comput.Mech*.; (2001) pp. 887-892.
[22] W.G. Richards, Application of neural networks: quan-titative structure-activity relationships of the derivatives of 2,4diamino (substituted-benzyl) pyrimidines as DHFR inhibitors, *J. Med. Chem*. 35 (1992) 3201–3207.
[23] T.A. Andrea, H. Kalayeh, Applications of neural networksin quantitative structure-activity relationships of dihydrofolatereductase inhibitors, *J. Med. Chem*. 34 (1991) 2824–2836.
[24] M. Elhallaoui, Modélisatrice moléculaire et étude QSAR d'antagonistes non compétitifs du récepteur NMDA par les méthodes statistiques et le réseau de neurones (Doctoral thesis), Fez, 2002.
[25] R. Hmamouchi, M. Larif, S. Chtita, A. Adad, M. Bouachrine, T. Lakhlifi, *Journal of Taibah University for Science*, Vol. 10, (2016) pp. 451-461.