DFT-Based QSAR Studies of Some Structurally Diverse Azaaurones Derivatives as Potential Antimalarials

Hanine HADNI¹, Menana El Hallaoui²

^{1, 2} Engineering Materials, Modeling and Environmental Laboratory, Faculty of Science, University Sidi Mohammed Ben Abdellah, Department of chemistry, B.P. 1796, Atlas, Fez, Morocco.

Abstract: A quantitative structure-activity relationship (QSAR) was carried out to analyze antimalarial activity of 34 compounds, designed as analogues of the naturally occurring aurones, using multiple linear regression (MLR), and artificial neural networks (ANN). An appropriate set of molecular descriptors were calculated to represent the molecular structures of compounds, such as constitutional, topological, Chemical and quantum-chemical descriptors. The DFT-B3LYP method, with the basis set 6-31G, was used to calculate some quantum chemical descriptors. The correlation coefficients calculated by MLR and ANN are $r_{MLR} = 0.88$ and $r_{ANN} = 0.98$ respectively. This model is statistically significant and shows a very good stability towards data variation in leave-one-out (LOO) cross-validation $r_{ev} = 0.92$. Therefore the QSAR model proposed in this paper shows satisfying predictive ability and it can be used for designing similar group of antimalarial compounds.

Keywords: antimalarial activity, QSAR, multiple linear regression, artificial neural networks, cross-validation

1. Introduction

Malaria is considered to be among the most significant world-wide public health troubles, especially, in tropical and subtropical regions [1,2]. Malaria is caused by protozoa of the genus Plasmodium, and there are five species that infect humans (P. falciparum, P. vivax, P. malariae, P. ovale and P. knowlesi) [3]. Although several attempts have been made to produce a vaccine, drugs are the only therapeutic alternative presented; however, the resistance to traditional therapies has increased the morbidity and mortality of malaria [4], making the search for new antimalarial drugs that employ new molecular targets extremely urgent [5, 6].

In the meanwhile, Aurones (2-benzylidenebenzofuran-3(2H)-ones) show antimalarial activity, they are secondary metabolite belonging to the flavonoids family (Figure 1), and structural isomers of flavones [7]. Compared to other flavonoids subclasses, somehow, aurones are much less studied.

In this work we attempt to use quantitative structure activity relationship (QSAR) tools to analyze antiplasmodial activity of 34 1-Azaaurones derived from the naturally occurring aurones (table 1) and to establish a predictive QSAR model [8]. We accordingly propose a quantitative model, and we try to interpret the activity of the compounds relying on the multivariate statistical analyses. The multiple regression analysis (MLR) has served to predict activities and to select the descriptors used as the input parameters for a back propagation network (ANN). To test the performance of this model we have used the cross validation method.



Figure 1: Structures of investigated aurones derivatives.

2. Material and Methods

2.1. Experimental Data

The experimental antimalarial activity IC_{50} (μ M) of 1-Azaaurones derived from the naturally occurring aurones are collected from recent publications [9]. The observations are converted into minus logarithm scale logIC₅₀ and are included in table 1.

Table 1: Studied compounds and their observed activities antimalarial logIC ₅₀							
Compounds	Structure of compound Nam of compound						
1	HO OH O	(Z)-2-Benzylidene-4,6-dihydroxybenzofuran-3(2H)-one	1,975				
2		(Z)-2-Benzylidene-4,6-dimethoxybenzofuran-3(2H)-one	1,780				
3	HO	(Z)-4,6-Dihydroxy-2-(4-methylbenzylidene)benzofuran- 3(2H)-one	1,802				
4	H ³ C~0	(Z)-4,6-Dihydroxy-2-(2-ethylbenzylidene)benzofuran- 3(2H)-one	1,322				
5	но	(Z)-4,6-Dihydroxy-2-(4-ethylbenzylidene)benzofuran- 3(2H)-one	2,055				
6		(Z)-2-(4-Ethylbenzylidene)-6-hydroxybenzofuran-3(2H)- One	1,447				
7		(Z)-2-(4-tert-Butylbenzylidene)-4,6- dimethoxybenzofuran-3(2H)-one	1,124				
8		(Z)-2-(4-Butylbenzylidene)-4,6-dimethoxybenzofuran- 3(2H)-one	1,072				
9		(Z)-2-(4-Bromobenzylidene)-4,6-dimethoxybenzofuran- 3(2H)-one	1,697				

10		(Z)-2-(4-Fluorobenzylidene)-4,6-dimethoxybenzofuran- 3(2H)-one	1,938
11	HO HO	(Z)-2-(4-Hydroxybenzylidene)-6-hydroxybenzofuran- 3(2H)-one	2,114
12	je co	(Z)-2-(4-Methoxybenzylidene)-6-hydroxybenzofuran- 3(2H)-one	1,041
13		(Z)-4,6-Dimethoxy-2-(pyridin-4- ylmethylene)benzofuran-3(2H)-one	1,929
14		(Z)-2-(4-Bromobenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	1,697
15		(Z)-2-(4-Chlorobenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	1,230
16		(Z)-2-(2-Chlorobenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,996
17		(Z)-2-(2,4-Dichlorobenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,924
18		(Z)-2-(2-Chloro-6-fluorobenzylidene)-4,6-dimethoxy- 2,3-dihydro-1H-indol-3-one	0,954
19		(Z)-2-(4-Ethylbenzylidene)-4,6-dimethoxy-2,3-dihydro- 1H-indol-3-one	0,000

20	(Z)-2-(2-Ethylbenzylidene)-4,6-dimethoxy-2,3-dihydro- 1H-indol-3-one	0,107
21	(Z)-2-(2,6-Dimethylbenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,959
22	(Z)-2-(2,4-Dimethylbenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,556
23	(Z)-2-(2,4,5-Trimethylbenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,748
24	(Z)-2-(2,3,5,6-Tetramethylbenzylidene)-4,6-dimethoxy- 2,3-dihydro-1H-indol-3-one	0,949
25	(Z)-2-(4-Isopropylbenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,643
26	(Z)-2-(4-tert-Buthylbenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,857
27	(Z)-2-(4-Buthylbenzylidene)-4,6-dimethoxy-2,3-dihydro- 1H-indol-3-one	0,613

28	(Z)-2-(4-Ethynylbenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	1,127
29	(Z)-2-(2,4-Dimethoxybenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,699
30	(Z)-2-(2,4,6-Trimethoxybenzylidene)-4,6-dimethoxy- 2,3-dihydro-1H-indol-3-one	0,279
31	(Z)-2-(3,4,5-Trimethoxybenzylidene)-4,6-dimethoxy- 2,3-dihydro-1H-indol-3-one	0,279
32	(Z)-2-(4-Thiomethylbenzylidene)-4,6- dimethoxybenzofuran-3(2H)-one	0,826
33	(Z)-2-(4-Morpholinobenzylidene)-4,6-dimethoxy-2,3- dihydro-1H-indol-3-one	0,949
34	(Z)-2-(4-(Dimethylamino)benzylidene)-4,6-dimethoxy- 2,3-dihydro-1H-indol-3-one	0,568

2.2 Calculation of molecular descriptors

3-D modeling and calculations were performed using the Gaussian 03 quantum chemistry package [10], for the whole of molecules. To save computational time, initial geometry

optimizations were carried out with molecular mechanics (MM) method using the MM+ force fields. The lowest energy conformations of the molecules obtained by the MM method were further optimized by the DFT method [11], by employing Becke's three-parameter hybrid functional

(B3LYP) [12], with a 6-31G basis set. These methods have become very popular in recent years because they can reach similar precision to other methods in less time and less cost from the computational point of view. In the other hand, Advanced Chemistry Development's ACD/ChemSketch program [14] and Chembio office software are used to calculate constitutional, topological and lipophilic descriptors. The totality of descriptors used in this work are listed in table 1

	• electron affinity (I)				
	I= -εHumo				
	• ionization potential (A)				
	A = -sLumo				
	 Electronegativity (v) 				
Descriptors calculated	I+A				
using Gaussian 03 W	$\chi = \frac{1}{2}$				
[15]	• Hardnass (n)				
	• Hardness (II) $I = A$				
	$\eta = \frac{1}{2}$				
	• Softness (S)				
	$S = 1/(2\pi)$				
	S = 1/(21)				
	• Electrophinicity (6) $(1 - x^2/(2\pi))$				
	$\omega = \chi /(2\eta)$				
	• Dipole moment (Dp				
	• total energy (E)				
	• Repulsion energy (EI)				
	• Molecular Weight (PM)				
	 Molecular Weight (FW) Moler Volume (MV): 				
	• Woran volume (WVV).				
	MV =				
	D				
	• Density (D):				
	D-MW				
	MV MV				
	• <i>Molar Ref</i> ractivity (MR):				
	$n^2 - 1 MW$				
Descriptors calculated	$MR = \frac{1}{2}$				
using ACD/ChemSketch	$n^2 + 1 D$				
	• Parachor (Pr)				
	$Pr = \left(\frac{MV}{M}\right) q^{\frac{1}{2}}$				
	$D^{J} \mathcal{B}_{4}$				
	• Refractive Index (n				
	2 MR + MW				
	$n = \frac{2MR + MW}{2}$				
	$\sqrt{MV - MR}$				
	• Surface Tension (y)				
	(Pr)				
	$\gamma = \left(\frac{1}{MU}\right)^4$				
	• Polarizability (go)				
	- 10 a = 0.306/308 MP				
	uc = 0.3704300 MIK				
Descriptors calculated	• Logp				
using chembio office	- Or				

2.3. Methods

Multiple Linear Regression (MLR)

The multiple linear regression statistic technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values. The multiple linear regression model (MLR) [15] was generated using the software xlstat, to predict antimalarial activities $logIC_{50}$. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

Artificial Neural Networks (ANN):

All the feed-forward ANN used in this paper are three-layer networks, the input layer contains five neurones, representing the relevant descriptors obtained in MLR technique (Parachor (Pr), Density (D), Softness (S), Molecular Weight (MW) and Dipole moment (Dp)). Although there are neither theoretical nor empirical rules to determinate the number of hidden layers or the number of neurone layers, one hidden layer seems to be sufficient in the most chemical application of ANN. Some authors [16, 17] have proposed a parameter ρ , leading to determine the number of hidden neurons, which plays a major role in determining the best ANN architecture. It is defined as follows:

 $\rho =$ (Number of data points in the training set / Sum of the number of connections in the ANN).

Therefore, in order to avoid overfitting or underfitting, it is recommended to take into account the ρ value; $1.8 < \rho < 2.3$ [18]. Thus, the ANN used in this work is formed by two hidden neurones, and the output layer represents the calculated activities values logIC₅₀. So, the final ANN architecture is (5-2-1), it is depicted in figure 2. All calculations of ANN are done on Matlab 7 using our program written in C language.



Figure 2: Schematic representation of the architecture network (5-2-1) used in this work

Cross-validation

Since a high-correlation coefficient only indicates how well the equations fit the data, cross-validation procedure was carried out in order to explore the reliability of the proposed models. In this aspect, the well-known "leave-one-out" (LOO) approach was used in which a number of models were developed with one sample ignored each time. Then, the ignored data were predicted by each model and the differences between predicted and observed activity values were evaluated. The LOO cross-validation coefficient Q^2 that is given by Eq. 1 was used as an indicator of the predictive performance and stability of a model. In general, LOO cross-validated coefficient Q^2 being higher than 0.5 can be considered as a statistical proof of the high-predictive ability [19].

$$Q^{2} = 1 - \frac{\sum_{i=1}^{n} (Yexp - Ypred)^{2}}{\sum_{i=1}^{n} (Yexp - \bar{Y})^{2}}$$

3. Results and Discussion

The selected descriptors values, and predicted activities values obtained by MLR, ANN and CV methods, are summarized in table 3.

Wherein Yexp and Ypred are the observed and predicted values for the dependent variables, respectively, and \overline{Y} is the average observed value.

Compounds	S	Dp	MW	Pr	D	$LogIc_{50}(obs)$	$LogIC_{50}(calc)$		<i>c</i>)
							MLR	ANN	CV
1	14.874	5.066	254.243	506.100	1.489	1.975	2.267	1.997	1.966
2	14.451	4.822	282.297	592.900	1.272	1.780	1.360	1.709	1.710
3	15.101	5.783	268.274	544.400	1.434	1.802	1.866	1.918	1.943
4	14.671	4.372	310.351	671.300	1.218	1.322	1.156	1.191	1.331
5	15.080	4.828	282.297	584.500	1.386	2.055	1.784	2.158	1.858
6	14.861	4.711	266.298	569.300	1.298	1.447	1.494	1.546	1.807
7	14.688	4.891	338.405	747.700	1.172	1.124	0.893	1.037	0.968
8	12.742	4.885	338.405	751.400	1.176	1.072	1.318	0.983	1.805
9	14.671	5.219	361.194	644.000	1.517	1.697	1.898	1.718	2.081
10	14.490	5.087	300.288	600.300	1.329	1.938	1.448	1.938	1.662
11	15.302	5.562	254.243	506.100	1.489	2.114	2.115	1.943	1.903
12	15.244	6.354	312.324	651.500	1.270	1.041	0.930	0.865	0.831
13	15.111	5.570	283.285	760.500	1.281	1.929	1.947	1.889	2.109
14	16.651	4.373	360.209	653.900	1.482	1.697	1.494	1.666	1.482
15	16.609	4.155	315.758	638.000	1.322	1.230	1.122	1.104	1.113
16	15.708	6.800	315.758	638.000	1.322	0.996	0.946	0.996	1.197
17	15.900	5.922	350.202	675.100	1.396	0.924	1.201	0.932	1.258
18	16.356	5.355	333.748	645.300	1.373	0.954	1.110	0.954	0.938
19	15.962	4.655	309.367	679.200	1.191	0.000	0.770	0.202	0.422
20	16.428	5.711	309.367	679.200	1.191	0.107	0.535	0.136	0.170
21	14.808	5.009	309.367	677.400	1.192	0.959	0.948	1.116	0.928
22	15.973	4.399	309.367	677.400	1.192	0.556	0.799	0.370	0.370
23	16.010	4.618	323.394	715.700	1.173	0.748	0.688	0.580	0.494
24	14.881	4.741	337.421	753.900	1.155	0.949	0.827	1.028	0.843
25	15.957	4.680	323.394	717.200	1.167	0.643	0.667	0.593	0.641
26	15.960	4.838	337.421	757.300	1.151	0.857	0.595	0.829	0.720
27	15.980	4.729	337.421	759.300	1.152	0.613	0.620	0.858	0.618
28	16.560	5.350	305.335	658.600	1.270	1.127	0.914	1.148	1.055
29	16.316	4.656	341.366	718.100	1.241	0.699	0.787	0.833	0.598
30	16.819	5.773	371.392	776.700	1.242	0.279	0.485	0.303	0.530
31	16.162	6.883	371.392	776.700	1.242	0.279	0.467	0.279	0.334
32	15.983	6.441	327.406	691.500	1.290	0.826	0.896	0.938	0.767
33	17.623	6.780	366.419	776.500	1.262	0.949	0.348	0.900	0.986
34	17.450	4.925	324.381	705.500	1.224	0.568	0.566	0.586	0.586

3.1 Multiple Linear Regression

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

LogIC50 = -1.35-0.2*S - 0.13*Dp - 1.1E-02*MW + 4.67E - 03*Pr + 5.16*D

Where n is the number of compounds, r is the correlation coefficient, r^2 is the correlation Squared, Sd is the standard deviation, F is the Fisher F-statistic. The relevant descriptors involved in the MLR model are Parachor (Pr), Density (D), Softness (S), Molecular Weight (MW) and Dipole moment (Dp).The corresponding normalized coefficients are presented in Figure 3. The correlation of the observed activities with the MLR calculated ones is illustrated in figure 4.

International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438



Figure 3: Modeling characterization by the normalized coefficients



Figure 4: Correlation of observed and predicted activities calculated using MLR.

3.2. Neural networks

In order to increase the probability of good characterization of studied compounds, artificial neural networks (ANN) are used to generate predictive model of quantitative structureactivity relationships (QSAR) between a set of molecular descriptors obtained by MLR method, and observed activities.

As it is shown in figure 5, a good correlation between observed antimalarial values and ANN predicted activities is obtained (n = 34, r = 0.98, $r^2 = 0.965$, Sd = 0.1)



Figure 5: Correlation of observed and predicted activities calculated using ANN

The correlation coefficient r = 0.98 and Standard Error of Estimate s = 0.1, obtained with the Neural network, show that the selected descriptors by MLR are pertinent and that the model proposed to predict anti-malarial activity is relevant.

3.3. Cross-validation

The QSAR model proposed to predict the activity of new compounds should be validated before its use. In this paper we validated our model with cross validation method.

 $N=34 R=0.925 R^2=0.855 S=0.22$

The correlation of the observed activities with the CV calculated ones is illustrated in figure 6:



Figure 6: Correlation of observed and predicted activities calculated using CV

A good correlation was obtained with cross validation rcv=0.925. So the predictive power of this model is very significant. The most important result of this investigation is that in vitro antimalarial activity could be predicted using QSAR methods, and that the selected descriptors are pertinent.

4. Conclusion

In this study, a QSAR model based on a set of 34 compounds designed as analogues of the naturally occurring aurones, was investigated in the goal to predict antimalarial

Volume 4 Issue 3, March 2015

<u>www.ijsr.net</u>

activity, using MLR and ANN methods. Our results could be summarized as follow:

- The MLR correlation coefficient ($r_{MLR} = 0.887$) is adequate , but when compared with the results given by the ANN ($r_{ANN} = 0.98$), we notice that the predictions fulfilled by this latter were more effective.
- The model proposed in this study is statistically significant and shows a high predictive power ($r_{cv} = 0.92$).
- The descriptors selected to represent the majority of classes of descriptors proposed to build the QSAR model show that the antimalaria activity is closely dependent of the parachor, the density, the softness, the molecular weight and the dipole moment of aurones derivatives.

References

- [1] Poinar, G. Syst. Paras. 2005, 61,47–52.
- [2] WHO, World Malaria Report, Available from http://www.who.int/malaria/worldmalariareport2011/en/ index.html, (2011).
- [3] Murray, M.C.; Perkins, M.E.; James, A.B. Medic. Chem. 1996, 31, 141–150.
- [4] Adjuik, M.; Agnamey, P.; Babiker, A.; Baptista, J.; Borrmann, S.; Brasseur, P.; Carnevale, P.; Cisse, M.; Collins, R.; Alessandro, U.; Day, N.; Boom, W.; Doherty, T.; Dorsey, G.; Garner, P.; Gikunda, S.; Gil, V.; Greenwood, B.; Guthmann, J. P.; Henry, M. C.; Kamya, M. R.; Kremsner, P. G.; Konate, E.; Krishna, S.; Lalloo, D.; Lange, P.; Loolpapit, M.; Malenga, G.; Marquino, W.; Marsh, K.; Milligan, P.; Molyneux, M.; Mugittu, K.; Niangue, J.; Nosten, F.; Ntoumi, F.; Obonyo, C.; Ochieng, F.; Olliaro, P.; Oloo, A. J.; Osorio, L.; Pinoges, L.; Priotto, G.; Rosenthal, P. J.; Ruebush, T.; Simpson, J.; Sirima, S.; Some, E.; Taylor, W.; Kuile, F.; Tiono, A.; Von Seidlein, L.; Watkins, B.; White, N.; Lancet 2004, 363, 9.
- [5] Guglielmo, S.; Bertinaria, M.; Rolando, B.; Crosetti, M.; Fruttero, R.; Yardley, V.; Croft, S. L.; Gasco, A; Eur. J. Med. Chem. 2009, 44, 5071.
- [6] Van Baelen, G.; Hostyn, S.; Dhooghe, L.; Tapolcsanyi, P.; Matyus, P.; Lemiere, G.; Dommisse, R.; Kaiser, M.; Brun, R.; Cos, P.; Maes, L.; Hajos, G.; Riedl, Z.; Nagy, I.; Maes, B. U. W.; Pieters, L.; Bioorg. Med. Chem. 2009, 17, 7209.
- [7] Boumendjel, A.; Curr. Med. Chem. 2003, 10, 2621.
- [8] Hansch, C.; Leo, A.; Ameri. Chem. Soci. Washington DC. 1995, 65.
- [9] Souard F., Okombi S., Beney C., Chevalley S., Valentin A., Boumendjel A., Bioorg. Med. Chem., 18 (2010) 5724–5731.
- [10] Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Montgomery, J.A.; Vreven Jr., T.; Kudin, K.N.; Burant, J.C.; Millam, J.M.; Iyengar, S.S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H.P.; Cross, J.B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.;

Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A.G.; Clifford, S.; Cioslowski, J.; St efanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.L.; Fox, D.J.; Keith, T.; Al- Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; Pople, J.A.; GAUSSIAN 03, Revision B.05, Gaussian, Inc., Pittsburgh, PA, (2003).

- [11] Parr, R.; Yang, W.; Density Functional Theory of Atoms and Molecules, Oxford University Press, New York, 1989.
- [12] Becke, A.D. J. Chem. Phys. 1993, 98, 5648.
- [13] Koopmans, T. Physica. 1933, 1, 104–113.
- [14] ACD/ChemSketch Version 4.5 for Microsoft Windows User's Guide.
- [15] Larif, M.; Adad, A.; Hmammouchi, R.; Taghki, AI.; Soulaymani, A.; Elmidaoui, A.; Bouachrine, M.; Lakhlifi, T. Arab. Jour. of Chem. 2013.
- [16] So, S.; Richards, G. J Med Chem. 1992, 35, 3207.
- [17] Andrea, TA. ; Kalayeh, H. J Med Chem. 1991, 34, 2824.
- [18] Ellhalaoui, M. PhD, University of Sidi Mohammed Ben Abdellah, Morocco, 2002.
- [19] Golbraikh, A.; Tropsha, A. J. Mol. Graph. Model., 2002, 20, 269.

Volume 4 Issue 3, March 2015