***ScientiaAmazonia*, v. 3, n.3, 89-96, 2014**
Revista on-line http://www.scientia.ufam.edu.br
Set-Dez ISSN:2238.1910

QSAR study of quinoline metanols with antimalarial activity against *Plasmodium falciparum*

Carla dos Santos Duarte [1], Alberto dos Santos Marques[2] and Yuji Takahata[3]

Submetido 05/10/2014 – Aceito 28/12/2014 – Publicado on-line 30/12/2014

Resumo

Um modelo QSAR foi formado com um total de vinte e cinco derivados da quinolina methanol substituídos na posição-4, com atividade anti-malária que inibe o crescimento da sepa W2 do *Plasmodium falciparum.* O programa DRAGON foi usado para produzir os descritores. O programa MOBYDIGS foi usado para construir um modelo QSAR composto de três descritores, RDF065m, Mor15e, G1m. O modelo QSAR satisfaz todos os critérios requeridos para a sua validação, indicando que o modelo é robusto com boa habilidade de previsão. A atividade biológica de vinte e três novos compostos foram modelados com cálculos usando o modelo QSAR. Os resultados previram que cinco dos novos compostos modelados têm maiores atividades biológicas do que qualquer um dos observados. Sugere-se a síntese desses compostos.

Palavras-chave: Malária, Relacionamento quantitativo da estrutura e atividade, modelagem molecular.

Abstract

A QSAR model was formed with total of twenty five 4-position quinoline methanol antimalarials that inhibit the growth of W2 strain of the *Plasmodium falciparum*. The DRAGON software was used to produce descriptors. The MobyDigs software was used to build a QSAR model, which is composed of three descriptors, RDF065m, Mor15e, G1m. The QSAR model satisfies all the criteria required for validation, indicating that the model is robust and good predictive ability. Biological activity of twenty three newly modeled compounds were calculated using the QSAR model. Five newly modeled compounds are predicted to have higher biological activities than any one of the observed. They are suggested for synthesis.

Keywords: Malaria, quantitative structure activity relationship, molecular modeling,

[1] Universidade Federal do Amazonas, Departamento de Bioquímica, Av. Rodrigo Otávio, 6.200 - Coroado I, 69077-000, Manaus, AM.
[2] Universidade Federal do Amazonas, Instituto de Ciencias Exatas, Departamento de Química, Laboratório de Tecnologia com Moléculas Bioativas, Av. Rodrigo Otávio, 6.200 - Coroado I, 69077-000, Manaus, AM. E-mail: marquesalbertods@gmail.com
[3] Universidade do Estado do Amazonas-UEA, Av. Carvalho Real, n°1777- Cachoeirinha, Manaus, AM.; Universidade Estadual de Campinas - Unicamp, Cidade Universitária "Zeferino Vaz", Campinas, SP.

ScientiaAmazonia, v. 3, n.3, 89-96, 2014
Revista on-line http://www.scientia.ufam.edu.br
Set-Dez ISSN:2238.1910

1. Introduction

Malaria is one of the oldest diseases known by man (CAMARGO, 2003). The protozoa of the generous *Plasmodium* disease cause this disease. Its transmission to humans occurs by infected female Anopheles. In 2008, four species of human malaria were described, *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. A study published in 2012 indicates that malaria is responsible for killing 1.2 million people per year worldwide, with nearly double the previously estimated by the OMS (CHAN, 2011, MURRAY, et al., 2012). In Brazil, more than 500 000 are affected each year and of these, 99.9 % of the cases are concentrated in the Amazon region, because it is a region favorable to the development of *Plasmodium* (SAMBO et al., 2000). *Plasmodium falciparum* is the most virulent species of the plasmodium and is the cause of more serious form of malaria, which in most cases may lead to death (SANTOS, 2009). An aggravating factor is the resistance to antimalarial drugs. It is justified by the principle of evolution of species (FRANCA, 2008). This makes the search for new drugs for known targets, as well as the search for new targets for chemotherapy. It becomes an ongoing challenge for the scientific community.

Taking into consideration the severe problem in cases of resistance of *Plasmodium falciparum* to existing antimalarials (MOORE et al: 1961, MOCKENHAUPT et al. 1995.). Milner and collaborators (2010) reported the construction of a library of next generation of quinoline methanol (QM) obtained from mefloquine, for the purpose of early identification of compounds that have biological properties compatible with the profile of the target products for carrying out the chemoprophylaxis of malaria.

Quantitative Structure Activity Relationship (QSAR) is a powerful technique used to study a set of compounds. It is useful to predict the unknown biological activity of a newly modeled compound. It helps to propose new compounds with competitive biological activities, before performing their synthesis in the laboratory.

The first objective of this work is to build a QSAR model for the 25 compounds of the antimalaria of quinoline methanol type, with antimalarial activity against W2 strain of *Plasmodium falciparum* determined experimentally (MILNER et al., 2010). They are classified into five groups, A, B, C, D, and E (Figure 1). The second objective is to propose new compounds with competitive biological activities. Here instead of using the original names of compounds (MILNER et al., 2010) that begin with the letters WR, we will use the numbers **1, 2, 3**, ..., 25, (Table 1).

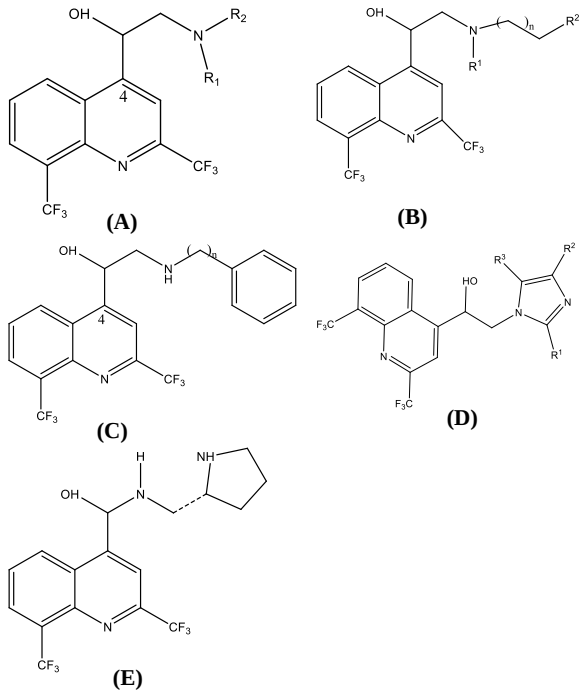

Figure 1 - Five groups (A) - (E) quinoline methanol with substituents in position 4.

2. Materials and methods

Models and computational details:

Initially we built a database with the 25 methanol quinoline compounds substituted in position 4, adapted from the literature (MILNER et al., 2010), with quantitative values of IC_{90} (ng / mL) in vitro against W2 strain of *Plasmodium falciparum*. These values were converted to concentration, mol / L, and biological activity of each compound was calculated as log $1/IC_{90}$ (mol / L) . (Table 1).

Each molecule in the ground state was optimized in vacuum with the semi- empirical AM1 method (DEWAR at al.,1985), assuming that the effect of solvent on the molecular geometry of substituted quinoline methanol is small due to the rigidity of the central part of the molecules. Here the use of semi- empirical method is suitable due to the size and the large number of molecules to be optimized.

ScientiaAmazonia, v. 3, n.3, 89-96, 2014
Revista on-line http://www.scientia.ufam.edu.br
Set-Dez ISSN:2238.1910

Table 1 - The IC_{90} values (ng / ml) in experimental in vitro against W2 strains of P. *falciparum* of 25 compounds divided into five groups of compounds of quinoline methanol (QM) from literature,[10] A (Fig. 1A), B (Fig. 1B), C (Fig. 1C), D (Fig.1D), E (fig.1E) . The values of biological activity, log $1/IC_{90}$ were calculated using IC_{90} (mol/L).The original name of each compound begins with the letters WR

Compounds					IC_{90} (ng/mL)	IC_{90} (mol/L)	log $1/IC_{90}$
Group A; Figure 1A			R_1	R_2			
				H	470	1.449E-06	5.84
				Me	17	5.025E-08	7.30
1	WR308314		H	Me	20	5.677E-08	7.25
2	WR308245		H	Et	9	2.366E-08	7.63
3	WR308246		Me	n-Pr	1	2.730E-09	8.56
4	WR308254		Et	i-Pr	5	1.365E-08	7.86
				n-Pr	5	1.224E-08	7.91
				n-Bu	2	5.258E-09	8.28
5	WR183544		H				
6	WR308257		H				
7	WR308277		n-Pr	n-Bu	2	4.582E-09	8.34
8	WR177000		H				
9	WR176990		n-Bu				
10	WR308607		H	i-Bu	2	5.258E-09	8.28
11	WR183545		H	t-Bu	19	4.995E-08	7.30
12	WR308442		H	n-hex	124	3.036E-07	6.52
	Group B; Figure 1B	**n**	**R_1**	**R_2**			
13	WR308258	1	H	OH	260	7.339E-07	6.13
14	WR308412	1	H	OMe	15	4.073E-08	/7.39
15	WR308622	2	H	OMe	12	3.139E-08	7.50
16	WR308278	2	H	SMe	10	2.510E-08	7.60
17	WR308411	0	H	CF3	500	1.275E-06	5.89
18	WR308396	1	H	NHBn	6	1.353E-08	7.87
	Group C; Figure 1C	n					
19	WR308251	0			484	1.209E-06	5.92
20	WR308252	1			19	4.585E-08	7.34
21	WR308253	2			19	4.435E-08	7.35
	Group D; Figure 1D	**R_1**	**R_2**	**R_3**			
22	WR308437	Et	Me	H	69	1.546E-07	6.81
23	WR308623	n-Pr	H	H	360	8.064E-07	6.09
24	WR308764	H	Benzeno	Benzeno	120	2.156E-07	6.67
	Group E; Figure 1E						
25	WR308621				46	1.129E-07	6.95

The DRAGON software (DRAGON, 2006) was used to calculate the descriptors. The DRAGON offers a package of twenty blocks of descriptors, ranging from 0D to 3D descriptors. Initially, it was calculated for a total of 1,664 descriptors and those with inter correlation values greater than 0.9 were eliminated, which reduced the number of molecular descriptors of each molecule to 1226.

The 25 compounds were divided into two sets: training set and test set. The test set consists of four compounds **2**, **4**, **8** and **13**, the remaining twenty-one compounds belong to the training set. The MobyDigs program (MOBYDIGS, 2009) was used for the selection of three descriptors and the construction of a QSAR model for the training set. The QSAR model was also used to assess the predictability of the biological activity of the test set.

We employ the three descriptors in the QSAR model, because it requires at least five compounds for each variable included in the model (GAUDIO et al. 2001). The genetic algorithm (GA) (GOLDBERG et al., 1989) was used for selection of the best combinations of variables. The QSAR model was evaluated to the following aspects;

ScientiaAmazonia, v. 3, n.3, 89-96, 2014
Revista on-line http://www.scientia.ufam.edu.br
Set-Dez ISSN:2238.1910

degree of fitness, degree of significance, degree of predictability and others. Many authors use q^2 (square of correlation coefficient of the cross-validation) to evaluate degree of predictability and q^2 values above 0.5 is considered as the predictive model (GOLBRAIKH et al., 2002). To evaluate the predictability of a set of external compounds, q^2_{ext} (the square of the correlation coefficient of the external cross-validation) is widely used (CHIRICO et al., 2011).

Todeschini et al. (2004) proposed a set of validation criteria severe than q^2, using RQK adjustment function, which was also adopted in this work . For the QSAR model to be considered valid and robust, the QUICK rule must be obeyed: 1. Collinearity between K_{xy} - K_x > DK variables -, otherwise there is high collinearity, where DK means the difference (D) of the values of K; 2. Capacity of foresight $q^2 > q°$ (Reference $q^0 > 0.5$), otherwise the model must be reject; 3. Capacity of the model prediction $q^2 - q^2_{ASYM} > dq$, otherwise the predictive capacity of the model is questionable, where q^2_{ASYM} is q^2 asymptotic and dq means the difference (d) of the values of q^2 ; 4. Redundancy in $r^P > t^P$ variables, otherwise exists redundancy in the explanatory variables; 5. overfitting $r^N > t^N$, otherwise there is overfitting due to noise variables . Unless all these criteria are not simultaneously satisfied, the model is rejected. Since all the parameters of the five criteria, such as q^2_{ASYM}, r^P and others, are defined and explained in detail in the literature (CHIRICO et al., 2011), we do not repeat here anymore .

3. Results and discussion

The best QSAR model obtained with the 21 compounds of the training set is represented by

Equation 1 (Eq.1), where Y = log $1/IC_{90}$ (mol), biological activity;

Y = 24.19487 (± 4.69343) -0.16668 (± 0.06613) RDF065m -1.4386 (± 0.60577) Mor15e -78.68588 (23.81521 ±) G1m (1)

(n = 21, r^2 = 0.8035 s = 0.394, F = 23.2, q^2 = 0.7335 ; q^2_{boot} = 0.6166, ; q^2_{ext} = 0.8031 ; SDEP = 0.413 ; SDEC = 0.355 ; Kxx = 23.24, KXY = 31.34, a (r^2) = 0.066, a ($q^{2)}$ = -0.481) .

The statistical parameters of Eq. 1 are:

i) n is the number of compounds included in the model;
ii) r^2 is the square of the correlation coefficient;
iii) s is the standard deviation and F is the significance parameter or Fisher's F test (95 % confidence);
iv) q^2 is the square of the correlation coefficient of cross validation;
v) q^2_{boot} is the square of the correlation coefficient of cross validation" bootstrap" method;
vi) q^2_{ext} is the square of the correlation coefficient of the external cross-validation;
vii) SDEP is standard deviation error of prediction;
viii) SDEC is the standard deviation error of calculation;
ix) Kxx is the total correlation in the model predictors (x) and Kxy is the total correlation in the set given by the model predictors X plus the response Y.
x) a(r^2) and a(q^2) are the parameters derived from Y- randomization.

The degree of predictability of the model can be assessed by q^2_{ext} value. The q^2_{ext} = 0.8031 value indicates that the quality of the prediction is good. The q^2_{boot} value = 0.6166 is close to the value of q^2 = 0.7335, indicating that predictability is reasonable. According to the literature (ERIKSSON, et al., 2003), if the values of the parameters derived from Y- randomization are met, a(r^2) < 0.3 and a(q^2) < 0.05, the QSAR model can be considered robust. The two parameters obtained in Eq.1 has the values a(r^2) = 0.066, and a(q^2) = -0.481, respectively. As the values a($r^{2)}$ and a($q^{2)}$ calculated satisfy the criteria in the literature, our QSAR model (Eq.1) is robust .

Table 2 shows the validation of QSAR, with QUIK rule using RQK parameters. The calculated RQK parameters satisfy all the five criteria listed in the table. This demonstrates that the QSAR model (Eq.1) satisfies a validation more severe than q^2.

The analysis of the quality of the model can be made with the aid of Figure 2 and Table 3 also. It is observed that most of the compounds is close to the fitted line. Table 3 shows the experimental (Yexp) and calculated (Ycalc) values of activity of the 21 compounds of the training set and the values of the predicted activity (Ypred) of the 4 test compounds, as well as the corresponding errors and Hat, obtained by using the QSAR model (Eq. 1).

ScientiaAmazonia, v. 3, n.3, 89-96, 2014
Revista on-line http://www.scientia.ufam.edu.br
Set-Dez ISSN:2238.1910

Table 2 - The validation of QSAR model (equation 1) with the QUIK rule using RQK parameters

Criteria for Approval				
1. Collinearity among variables $K_{xy} - K_x > DK$	K_{xy} 31.34	K_x 23.24	$K_{xy} - K_x$ 8,1	DK 0.081
2. Capacity of Prediction $q^2 > q^o$ (Referência $q^o > 0,5$)	q^2 0,77	q^o >0,5		
3. Predictive capacity of the model $q^2 - q^2_{ASYM} > dq$	q^2 0,7699	q^2_{ASYM} 0.77075	$q^2 - q^2_{ASYM}$ -0,00085	dq - 0.005
4. Redundancy in the variables $r^P > t^P$	r^P 0,494	t^P 0,1		
5.Overfitting (overfitting) $r^N > t^N$	r^N -0,101	t^N -0,311		

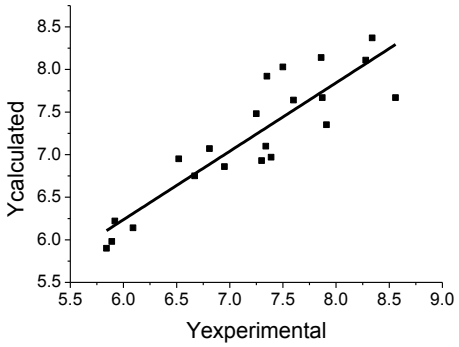

Figure 2- Experimental biological activity (Yexperimental) versus calculated activity (Ycalculated).

Table 3 - The values of experimental activity (Yexp), calculated (Ycalc), predicted (Ypred), the calculated error (Calc. Err.), The predicted error (Err. Pted.) and Hat, for the training and test sets .

Criteria for Approval				
1. Collinearity among variables $K_{xy} - K_x > DK$	K_{xy} 31.34	K_x 23.24	$K_{xy} - K_x$ 8,1	DK 0.081
2. Capacity of Prediction $q^2 > q^o$ (Referência $q^o > 0,5$)	q^2 0,77	q^o >0,5		
3. Predictive capacity of the model $q^2 - q^2_{ASYM} > dq$	q^2 0,7699	q^2_{ASYM} 0.77075	$q^2 - q^2_{ASYM}$ -0,00085	dq - 0.005
4. Redundancy in the variables $r^P > t^P$	r^P 0,494	t^P 0,1		
5.Overfitting (overfitting) $r^N > t^N$	r^N -0,101	t^N -0,311		

The Williams plot (Figure 3) is to assess the applicability domain (AD) of each of the training set (GRAMATICA, 2007). All the compounds are within the limit of AD, inside of the 3σ range . All compounds are below the limit of the Hat value indicated by the dotted line between 0.5 and 0.6 .

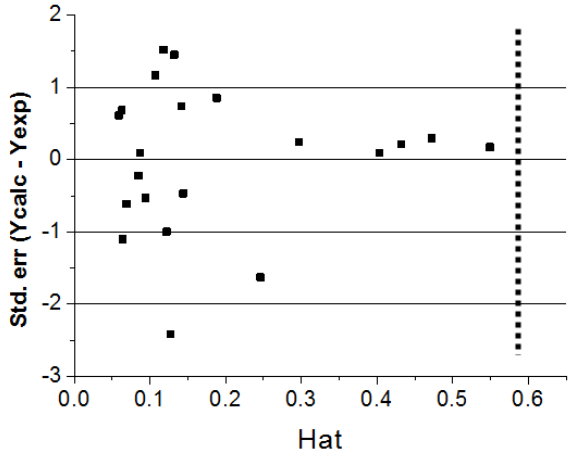

Figure 3 - The William plot. The dotted vertical line represents the limit of the applicability domain.

The RDF065m descriptor in the QSAR model (Eq. 1) corresponds to Radial Distribution Function -6.5 weighted by atomic mass. The Mor15e descriptor represents the signal 3D Morse - signal 15 weighted by atomic Sanderson electronegativity. The G1m descriptor represents the first component symmetry directional WHIM index weighted by atomic mass. The WHIM (Weighted Holistic Invariant Molecular descriptors) are geometric descriptors based on statistical indices and calculated projections of the atoms along the principal axes. The three descriptors are constructed to capture relevant information of the molecule. They are important to build the QSAR model (Eq.1) that is robust and high predictability. All coefficients of the three descriptors have negative signs. This means that the smaller the values of the descriptors, greater will be the biological activity of the compound. To model a new compound with high activity, the new compound should have the lowest possible values of the three descriptors. The QSAR model (Eq.1) does not allow one to extract any practical information about the relationship between biological activity and molecular structure of the compound, because of the complex nature of the three selected descriptors. The principal utility of the QSAR model is that it can be used to *predict* the biological activity of a new compound, whose biological activity is not yet known. The largest

ScientiaAmazonia, v. 3, n.3, 89-96, 2014
Revista on-line http://www.scientia.ufam.edu.br
Set-Dez ISSN:2238.1910

experimental activity in Table 1 is Y_{exp} = log $1/IC_{90}$ = 8.56 of the compound 5 (WR183544). Our next goal is to model new compounds whose activities may be greater than 8.56. Twenty three new compounds was modeled and listed in Table 4 (**n1 - n23**. They all belong to group A (Figure 1A) which shows the highest experimental activities (Table 1). The model was treated as follows: (i) The molecular geometry of each compound was optimized by the semiempirical method AM1. (ii) The values of the three descriptors RDF065m, Mor15e, G1m, of the each optimized molecule were calculated with the Dragon program, and they are listed in Table 4. Finally, the biological activity of the new molecule was calculated by the QSAR model (Eq. 1) replacing the three descriptors in the equation. The calculated value is the value of the predicted activity (Y_{pred}), which is listed in Table 4. Five new compounds listed present higher activities than 8.56, which is the largest observed activity. They are **n19** (Y_{pred} = 8.74), **n11** (Y_{pred} = 8.73), **n20** (Y_{pred} = 8.65), **n8** (Y_{pred} = 8.64), **n23** (Y_{pred} = 8.63) which are suggest for synthesis.

Table 4 - List of 23 new compounds. All compounds belong to the Group A (Figure 1A). The predicted biological activity (Ypred), was calculated with the QSAR model (Eq. 1) for each new compound, using the three selected descriptors, RDF065m, Mor15e, and G1m.

new compounds	R1	R2	Yprev	RDF065m	Mor15e	G1m
n1	Me	Et	8,03	22,199	0,075	0,157
n2	Me	n-Pr	8,09	23,449	-0,003	0,155
n3	Me	n-Bu	8,47	23,987	-0,219	0,153
n4	Me	n-Pent	7,79	24,183	-0,152	0,16
n5	Et	n-Pr	8,39	22,177	0,049	0,153
n6	Et	n-Bu	7,88	23,469	-0,129	0,16
n7	Et	n-Pent	7,44	23,569	-0,167	0,166
n8*	n-Pr	n-Bu	8,64*	25,249	-0,263	0,149
n9	n-Pr	n-Pent	8,05	25,391	-0,255	0,156
n10	n-Bu	n-Pent	8,10	27,878	-0,246	0,15
n11*	n-Pent	n-Pent	8,73*	28,222	-0,394	0,144
n12	H	cyclohexane	8,12	21,038	-0,123	0,162
n13	H	cyclopropane	6,33	19,801	0,662	0,173
n14	H	cyclobutane	7,96	19,799	0,457	0,156
n15	H	cyclopentane	7,77	22,649	0,367	0,154
n16	n-Pr	cyclobutane	7,00	23,142	1,067	0,15
n17	Cyclobutane	cyclobutane	7,88	23,754	-0,11	0,159
n18	n-Pr	cyclohexane	8,15	25,173	-0,239	0,155
n19*	Cyclohexane	cyclohexane	8,74*	29,475	-0,879	0,15
n20*	H	-C=C-C-C-C	8,65*	18,529	0,233	0,154
n21	H	-C≡C-C-C-C	8,44	19,468	0,161	0,156
n22	H	-C(O)-N-C-C-C	6,56	22,862	1,185	0,154
n23*	H	-C(O)-N-cicloexano	8,63*	20,394	0,308	0,149

* Predicted activity is greater than 8.56, which is the highest activity observed.

4. **Conclusions**

It was possible to construct a robust QSAR model (Eq.1) and the good predictability for 25 methanol quinoline compounds substituted in position 4, that have biological activity against W2 strain of *Plasmodium falciparum*. The biological activities of 23 newly modeled compounds were calculated using the QSAR model obtained. Among the 23 compounds, there are five with predicted activities major larger than 8.56. Our finding suggest the synthesis these five compounds.

Acknowledgements

CSD, thanks FAPEAM for award of a master degree fellowship . ADSM, thanks CNPq for financial support coming from the edicts (553292/2005-6, CT / AMAZONAS / MCT / CNPq) and (MCT / CNPq / FNDCT No. 19/2009, 556853 / 2009-1), and the Research fellowship from Muraki, Amazon State, Brazil. YT, thanks FAPEAM for the Senior Visiting Researcher (PVS) fellowship.

Disclosure

This article is unpublished and not being considered for any other publication. The author(s) and reviewers did not report any conflict of interest during their evaluation. Therefore, the Journal Scientia Amazonia owns the copyright and has the approval and permission of authors for publication this article electronically.

References

MALÁRIA, available in: **http://biomedicinaunic.blogspot.com/2010/09/malaria.html. .** Access at 03/16/2011.

CAMARGO, E. P.; Malária, Maleita, Paludismo, **Ciência e Cultura,** v. 55(1), p.26-29, 2003.

CHAN, M. **World malária report 2011**, available in: http://www.who.int/malaria/world_malaria_report_2011/9789241564403_eng.pdf>. Access at: 03/02/2012.

MURRAY, C. J.; ROSENFELD, L. C.; LIM, S. S.; ANDREWS, K. G.; FOREMAN, K. J. ; HARING, D.; FULLMAN, N.; NAGHAVI, M.; LOZANDO, R.; LOPEZ, D.; Global malaria mortality between 1980 and 2010: a systematic analysis. **Lancet**. V.379, p.413-431, 2012.

SAMBO, M. R.; BORGES, A.; BORGES, F.; Malária cerebral e síndroma neurológica pós-malaria. **Revista Spmi, Portugal**. v.7(3), p. 170-175, 2000.

SANTOS, C. B. R.; GOMES, T. L. D.; MACEDO, W. J. C.; MACIEL, A. A.; PINHEIRO, J. C.; Desenvolvimento de fármacos derivados da diidroartemisinina contra a malária usando métodos de química quântica e quimiométricos available in: **<http://www.abq.org.br/cbq/2009/trabalhos/3/3-21-651.htm.>** Access at 12/10/2010.

FRANCA, T. C. C.; SANTOS, M. G.; FIQUEROA-VILLAR, J. D.; Malária:aspectos históricos e quimioterapia. **Química Nova**, v. 31(5), p.1271-8, 2008.

MOORE, D. V.; LAINER, J. E.; Observations on two Plasmodium falciparum infections with abnormal response to chloroquine. **American Journal of Tropical Medicine and Hygiene** v.10, p.5-9, 1961.

MOCKENAUPT, F. P.; Mefloquine resistance in Plasmodiumfalciparum. **Parasitology Today** v.11, p.248-253, 1995.

MILNER, E.; MCCALMONT, W.; BHONSLE, J.; CARIDHA, D.; CARROLL, D. W.; GARDNER, S.; GERENA, L.; GETTAYACAMIN, M.; LANTERY, C.; LUONG, T.; MELENDEZ, V.; MOON, J. K.; RONCAL, N. E.; SOUZA, J. C.; TANGTEUNG, A.; WIPF, P.; DOW, G. S.; Structure-Activity Relationships amongst 4-Position Quinoline Methanol Antimalarials that Inhibit the Growth of Drug Sensitive and Resistant Strains of Plasmodium Falciparum. **Bioorganic & Medicinal Chemistry Letters.** v.20, p.1347-1351, 2010.

DEWAR, M. J. S.; ZOEBISCH, E. G.; HEALY, E. F.; STEWART, J. J. P.; Development and use of quantum mechanical molecular model. 76. AM1. A new general propose quantum mechanical molecular model. **Journal of the American Chemical Societ,** v.107, p.3902-3909, 1985.

DRAGON FOR WINDOWS AND LINUX, TALETE SRL, Milano, Italy, 2006.

MOBYDIGS, Version 1.1, TALET SRL, Milano, Italy, 2009.

GAUDIO, A. C.; ZANDONATE, E.; Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. **Quimica Nova.** V.5, p.658-671, 2001.

GOLDBERG, D. E.; **Genetic Algorithms in Search, Optimization and Machine Learning**, ADDISON-WESLEY: Indianapolis, 1989.

GOLBRAIKH, A.; TROPSHA, A.; Beware of q2! **Journal of Molecular Graphics and Modelling,** v.20, p.269-76, 2002.

CHIRICO, N.; Gramatica, P.; Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. **Journal of Chemical Information**

and Modeling, V.51(9), p.2320-2335, 2011. DOI: 10.1021/ci200211n

TODESCHINI, R.; CONSONNI, V.; MAURI, A.; PAVAN, M.; Detecting "bad" regression models: multicriteria fitness functions in regression analysis**.** **Analytica Chimica Acta**, v. 515(1), p.199-208, 2004.

ERIkSSON, L.; JAWORSkA, J.; WORTH, P. A.; CRONIN, M. T. D.; MCDOWELL, R. M.; GRAMATICA, P.; Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. **Environmental Health Perspectives**. v.111, p.1361-1375, 2003.

GRAMATICA, P.; Principles of QSAR models validations internal and external. **QSAR & Combinatorial Science**, v.26(5), p.694-701, 2007.