Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. *75* (11) 1533–1548 (2010)
JSCS–4074
UDC 635.71:616.936:530.145:543:51–3
Original scientific paper

A study of new antimalarial artemisinins through molecular modeling and multivariate analysis

JOÃO E. V. FERREIRA[1]*, ANTONIO F. FIGUEIREDO[1], JARDEL P. BARBOSA[1], MARIA G. G. CRISTINO[1], WILLIAMS J. C. MACEDO[1], OSMARINA P. P. SILVA[1], BRUNO V. MALHEIROS[1], RAYMONY T. A. SERRA[2] and JOSE CIRIACO-PINHEIRO[1]

[1]*Laboratório de Química Teórica e Computacional, Faculdade de Química, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Avenida Augusto Correa, 01, CP 101101, CEP 66075-110, Belém, PA, Amazônia and* [2]*Centro de Ciências Biológicas e da Saúde, Universidade Federal do Maranhão, CEP 65085-580, São Luis, MA, Brasil*

(Received 26 January, revised 12 June 2010)

Abstract: Artemisinin and 18 derivatives with antimalarial activity against W-2 strains of *Plasmodium falciparum* were studied through quantum chemistry and multivariate analysis. The geometry optimization of the structures was realized with the Hartree-Fock (HF) theory and 3-21G** basis set. Maps of molecular electrostatic potential (*MEP*) and molecular docking were used to investigate the interaction between the ligands and the receptor (heme). Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were employed to select the most important descriptors related to activity. A predictive model was generated by the partial least square (PLS) method through 15 molecules and 4 used as an external validation set, which were selected in the training set, the validation parameters of which are Q^2 = 0.85 and R^2 = 0.86. The model included as molecular parameters the radial distribution function, *RDF060e*, the hydration energy, *HE*, and the distance between the O1 atom from the ligand and the iron atom from heme, d(Fe–O_1). Thus, the synthesis of new derivatives may follow the results of the *MEP* maps and the PLS analysis.

Keywords: Malaria; artemisinin; molecular docking; *MEP* maps; QSAR.

INTRODUCTION

Malaria is a very serious infectious disease caused by protozoans of the genus *Plasmodium* and is transmitted by the bite of infected female *Anopheles* mosquitoes. Every year, over one million people die owing to malaria, especially in tropical and subtropical areas. Most of the deaths are attributed to the parasite species *Plasmodium falciparum*. Many drugs have been investigated for their efficacy in the treatment of the disease, but resistant strains of *P. falciparum* to some

* Corresponding author. E-mail: joao.elias@yahoo.com.br
doi: 10.2298/JSC100126124F

of these drugs have appeared.[1–6] Hence, further discovery of new classes of more potent compounds against the disease is necessary.

However, drug design is a process involving high cost and time. Computational and quantitative structure–activity relationship (QSAR) studies have been of great value in medicinal chemistry.[7–15] Statistical tools can be used for the prediction of the biological activities of new compounds based only on the knowledge of their chemical structures, *i.e.*, not depending on experimental data, which are unknown. Such a strategy gives very useful information for the understanding of the mechanisms of the action of drugs and proposals for syntheses, in this way rationalizing drug discovery. As Doweyko[16] states "QSAR is alive and well", thus it is still of relevance today.

Some of the effective drugs used in the clinical treatment of *P. falciparum* malaria are artemisinin or *qinghaosu* and its derivatives. Artemisinin (compound **1**, Fig. 1) is a sesquiterpene containing the 1,2,4-trioxane ring system. Currently,

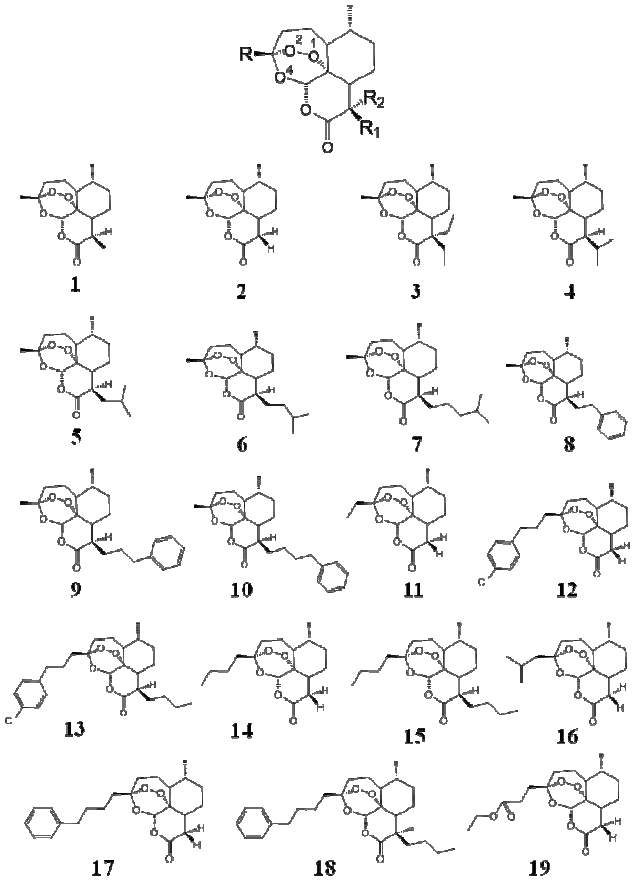

Fig. 1. Artemisinin and derivates with antimalarial activity against W-2 strains of *P. falciparum*.

semi-synthetic artemisinin derivatives play an important role in the treatment of *P. falciparum* malaria.[1,17–19] Even though the true mechanism of their biological activity against malaria has not been elucidated completely yet, various studies suggest that the trioxane ring is essential for antimalarial activity owing to the properties displayed by the endoperoxide linkage. Literature also suggests that free heme (Fig. 2) could be the molecule targeted by artemisinin in biological systems and that Fe^{2+} interacts with the peroxide when artemisinin reacts with heme.[20–23]

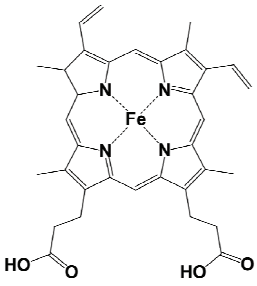

Fig. 2. Two-dimensional structure of the heme.

In this article, a quantum chemical and multivariate study of artemisinin and 18 derivatives, the training set (Fig. 1), with different antimalarial activities, tested *in vitro* against *P. falciparum*, was performed. Initially, the structures were modeled and many different molecular descriptors were computed. Maps of the molecular electrostatic potential (*MEP*) and molecular docking were employed to better understand the correlation between structure and activity, and the interaction between the ligands (artemisinin and derivatives) and the receptor (heme). Methods of multivariate analysis[24] were used to deal with such a large number of descriptors and find a predictive model. First principal component analysis (PCA) and hierarchical cluster analysis (HCA) were employed in order to choose those molecular descriptors that are most related to the biological property investigated. Then, a QSAR model was elaborated through the partial least square (PLS) method to aid future studies searching for others new drugs against malaria.

COMPUTATIONAL METHODS

The compounds studied

Initially, 19 molecules were selected from the literature[6] to build a training set (Fig. 1). They were associated to their *in vitro* bioactivity against the drug-resistant malarial strain *P. falciparum* (W-2 clone), which is cloroquine resistant but mefloquine sensitive. The samples include artemisinin (**1**) and analogs substituted in the R, R1 and R2 positions around the tetracyclic framework of the parent structure **1**. The atomic numbering adopted in this work is the same as that used in a previous work.[10] Considering that the biological data was available from different sources, the logarithm of the IC_{50} value of artemisinin over the IC_{50} value of the compounds (logarithm of relative activity, log *RA*) was used to reduce inconsistencies caused by individual experimental environments:

$$\log RA = \log (IC_{50} \text{ of artemisinin} / IC_{50} \text{ of an analog}) \quad (1)$$

where IC_{50} means the 50 % inhibitory concentration. In this work, the following classification based on the antimalarial responses was adopted: compounds with log $RA \geq 0.00$ were assumed as more potent analogs (**1**, **2**, **8**–**11**, **17** and **19**) and those with log $RA < 0.00$ were considered as less potent analogs (**3**–**7**, **12**–**16** and **18**).

Molecular modeling

The starting point in the molecular modeling step was the construction of the structures of the molecules with the aid of GaussView software.[25] To model all the compounds, it was necessary to use a quantum chemistry method and a basis set able to describe well the region around the 1,2,4-trioxane ring, which was indicated as being crucial for the antimalarial efficacy of artemisinin, as previously stated. Hence, the complete geometry optimization for all structures was performed with the Hartree-Fock (HF) method[26] and the 3-21G** basis set,[27] incorporated in the Gaussian 98 program.[28] This procedure gives good results for the geometrical parameters when compared to the experimental (crystallographic) results, as verified by Cardoso *et al.*[11]

Descriptors

After molecular modeling, various descriptors were computed for each molecule in the training set. They represent different source of chemical information (features) regarding the molecules and include geometric, electronic, quantum-chemical, physical–chemical and topological descriptors and others. They are important for the quantitative description of molecular structure and to find appropriate predictive models.[29] The descriptors computed were dipole moment, molecular weight, molecular volume, surface area, molecular refractivity, molecular polarizability, molecular hardness, molecular softness, O1–O2 bond length, atomic charge on the atoms O1, O2 and O4, logarithm of octanol–water partition coefficient (log *P*), hydration energy (HE), radial distribution functions (RDF030e and RDF060e), total energy, frontier orbital energies (HOMO, HOMO-1, LUMO, LUMO+1), maximal electrotopological negative variation (MAXDN), maximal electrotopological positive variation (MAXDP), information index on molecular size (ISIZ), energy of the interaction of the heme–ligand complex, distance between the O1 atom from the ligand and the iron atom from heme, *d*(Fe–O1), and the distance between the O2 atom from the ligand and the iron atom from heme, *d*(Fe–O2). The computation of the descriptors was performed employing the following software: Gaussian 98, e-Dragon,[30,31] Autodock 4.0,[32] Molekel,[33] and HyperChem 6.02.[34]

Molecular electrostatic potential maps

An important aspect explored in this research was the correlation of the structure–activity of the species studied here through the characteristics of the electrostatic potential in the region of the 1,2,4-trioxane-ring, since in the literature[8,11,12] it is stated that artemisinin and its derivatives with antimalarial activities present similar patterns in their *MEP* maps. Such a method enables a qualitative analysis to be used to localize reactive sites in a molecule and the role played by both the electronic and steric (size/shape) effects on its potency. It is worthwhile to point out that visualization of *MEP* maps gives qualitative information on molecules such as the behavior in the interaction between ligand and receptor. *MEP* maps for artemisinin, derivatives and heme were computed from the atomic charge at the HF/3-21G** level using the Gaussian 98 program and the results are displayed with Molekel software.

Molecular docking

The interaction between the ligands and the receptor was studied with aid of molecular docking in order to find the best geometry for the complex formed between these two molecules. The geometry of artemisinin and the derivatives (ligands) was designed with the HF/3--21G** level of theory, whereas the geometry of heme (receptor) was obtained from the protein data bank (PDB) RCSB, identified by the code 1A6M, according to the work of Vojtechovsky *et al*.[35] The arrangement in the docking calculation took into account the presence of the proximal histidine residue under the plane of the porphyring ring. This histidine unity is, as usual, coordinated perpendicularly to Fe^{2+} through its sp^2 nitrogen atom of its imidazole ring. Such an arrangement will allow the Fe^{2+} to attain a nearly octahedral hexacoordinated arrangement after binding to the artemisinin molecule.[22] The orientation of the ligand was set just above the plane of heme. Then, for each ligand/receptor interaction, 20 (twenty) conformations were calculated and the most probable one, based on the lowest energies of interaction, was selected to evaluate the distance between the O1 atom from the ligand and the iron atom from heme, *d*(Fe–O1), and the distance between the O2 atom from ligand and the iron atom from heme, *d*(Fe–O2). Automated docking calculations were performed to develop possible conformations for the complex employing the Lamarckian Genetic Algorithm implemented in the package Autodock 4.0. This program starts the docking displaying the ligand in an arbitrary conformation and position and look for the favorable dockings with the receptor using both simulating annealing and genetic algorithms. AutoDock uses a random number generator to create new poses for the ligand during its search and estimates the free energy of binding of a ligand to its target. The resulting conformations were ranked in order of increasing binding energy of the lowest binding energy conformation in each cluster.

Multivariate analysis

A multivariate analysis was performed in order to extract meaningful information efficiently from the data computed. However, prior to performing the exploratory data analysis, all variables were auto-scaled as a preprocessing so that they could be standardized and this way could have the same importance regarding the scale. Furthermore, given the large quantity of multivariate data available, it was necessary to reduce the number of variables. Thus, if two any descriptors had a high Pearson correlation coefficient ($r > 0.8$), one of the two was randomly excluded from the matrix, since theoretically they describe the same property,[36] that is, they also have a high correlation with antimalarial activity and it is sufficient to use only one of them as an independent variable in a predictive model. Moreover those properties that showed either the same values for most of the samples or a small correlation with activity ($r < < 0.2$) were also eliminated.

After this data compression, two complementary methods for exploratory data analysis, HCA and PCA, were employed to study intersample and intervariable relationships and to select the properties that better contribute to the classification of the compounds into two groups: one for more potent analogs and other for less potent analogs. The PCA was employed to reduce the dimensionality of the data, find descriptors that could be useful in characterizing the behavior of the compounds acting against malaria and look for natural clustering in the data and outlier samples. While processing PCA, several attempts to obtain a good classification of the compounds were made. At each attempt, the score and loading plots were analyzed based on the variables employed in the analysis. The score plot gives information about the compounds (similarity and differences). The loading plot gives information about the variables (how they are connected to each other and which are the best to describe the variance in the original data). Then the descriptors selected by PCA were used to perform

HCA and PLS. The objective of HCA was also to present the compounds distributed in natural groups and the results confirm the PCA results. Thus, several approaches to establish links between samples/cluster were attempted. All of them were of an agglomerative type, since each sample was first defined as its own cluster, then others were grouped together to form new clusters until all samples were part of a single cluster.

The final purpose of the multivariate analysis was the construction of a mathematical model to predict antimalarial activity. The samples selected to compose the external validation set were **5**, **8**, **15** and **19**. In order to evaluate the statistical significance of the model, some validation parameters were calculated as recommended,[24,36,37] which included the total variance explained, R^2 (correlation between the estimated values predicted by the model built with the full data set and actual values of y), Q^2 (the cross-validated correlation coefficient), *PRESS* (prediction residual error sum of squares), *SEP* (standard error of the prediction), s (standard deviation) and F (Fisher test). The statistical analysis, including both calculations and plots was realized with Pirouette software.[38]

RESULTS AND DISCUSSION

Molecular electrostatic potential maps

The molecules in the training set have similar *MEP* maps (Fig. 3). They display contour surfaces close to that of the 1,2,4-trioxane ring, which is characterrized by negative electrostatic potentials (red color), on which the lowest values for charge were about –0.30 au (atomic unit). Such a characteristic indicates a concentration of the electron density due to the lone electron pairs on the oxygen atoms (O_1, O_2 and O_4). These molecules also have contour surfaces characterrized by positive electrostatic potentials (blue color), the highest values of which are about 0.30 au. The distribution of electron density on the molecules around the trioxane ring induces their activity against malaria, a belief supported by the fact that the complexation of artemisinin with heme involves particularly the interaction between the peroxide bond, the most negatively charged zone on the ligand, and Fe^{2+}, the most positively charged zone on heme (the receptor molecule).[14,23] The *MEP* map for heme (Fig. 4) displayed a contour surface around the porphyrin ring associated to negative electrostatic potentials (red color) and the zone above Fe^{2+} is presented as having a positive electrostatic potential (blue color). Hence, the presence of a surface in red close to the trioxane ring suggests these compounds have a reactive site for electrophilic attack and must possess antimalarial potency; consequently they are interesting to be investigated. Thus, in the case of an electrophilic attack of the iron of heme against an electronegative zone, this attack has a great preference to occur through the involvement of the endoperoxide linkage. Such a pattern of *MEP* maps is an indication that the compounds in the test set are all active against the malarial strain *P. falciparum* (W-2 clone). Thus by analyzing *MEP* maps, the selection of inactive compounds is avoided.

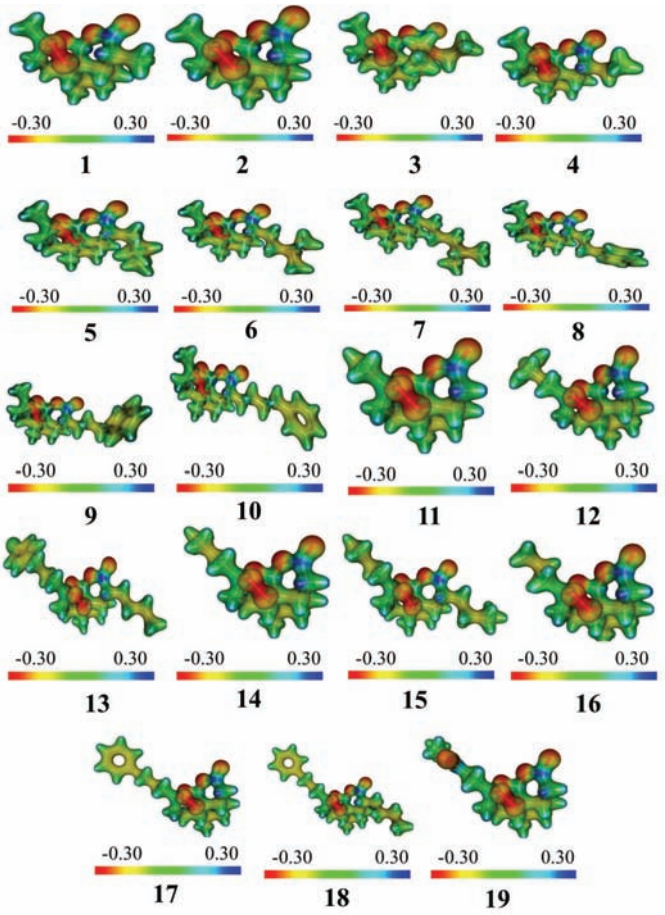

Fig. 3. *MEP* maps (values in atomic unit) for artemisinin and derivatives with antimalarial activity against W-2 strains of *P. falciparum*.

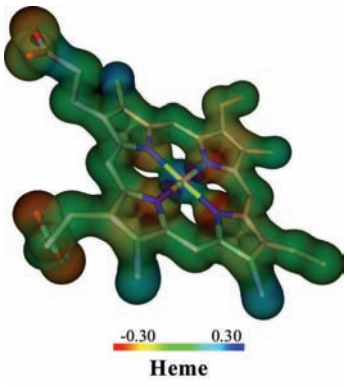

Fig. 4. *MEP* map (values in atomic unit) for heme.

Molecular docking

Docking calculations showed that the entire molecules of the ligands are placed parallel to the plane of the porphyrin ring of heme and the polar part of the ligands containing peroxide bond is directed towards the polar part of the heme system containing Fe^{2+}. This interaction for compounds **2** and **14**, the most active and the least active in the training set, respectively, is visualized in Fig. 5. Such orientations were assumed as the most favorable and so represent the real system under investigation, considering they were chosen based on the lowest free-energy of binding (interaction energy). For compounds in the training set, the values of d(Fe–O_1) ranged from 2.48 to 2.89 Å; however this interval for the d(Fe–O_2) distances ranged from 2.79 to 3.93 Å. For artemisinin (**1**), the d(Fe–O_1) calculated distance was found to be 2.68 Å, which is very close to the value reported (2.7 Å) in other theoretical studies.[40,41] A clear trend occurs involving the interatomic separation between Fe^{2+} and the oxygen atom in the trioxane ring because the distances are shorter for the O_1 atom than for the O_2 atom. This result reinforces the conception that the O_1 atom from artemisinin binds to Fe^{2+} from heme more preferably than the O_2 atom. Compounds **9**, **10**, **17** and **19** have higher activity than artemisinin and also higher values of d(Fe–O_1). They have a large substituent which certainly causes repulsion due to the steric effect which prevents them from approaching closer to the heme. Now considering the interaction energy for the ligand/receptor complex, it showed a poor linear correlation with activity (r = –0.04) and ranged from –6.80 to –5.53 kcal mol^{-1}. In fact, it was verified that even though some orientations were associated with the lowest interaction energy, they seemed strongly favorable with respect to being active against malaria for they presented the endoperoxide bond apart from Fe^{2+}.

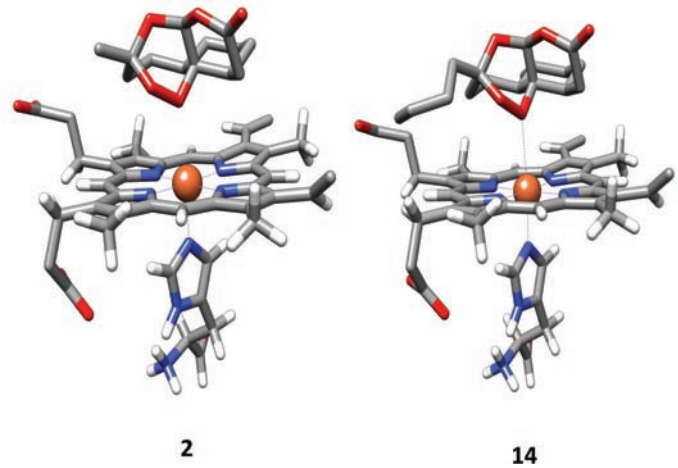

Fig. 5. Molecular docking between heme and artemisinin derivatives. Molecules **2** and **14** are the most and the least potent in the training set, respectively.

Nowadays, the most accepted mechanism of antimalarial action involves the formation of a complex between heme and artemisinin derivatives in which the iron of heme interacts with O_1 of the endoperoxide.[40] Moreover substituent and conformation effects may affect the charge distribution at the oxygen and even the Fe–O_1 bond.[9] An increase in the polar area of artemisinin increases the space with polar interactions between heme, ligand and globin.

PCA method

The first three principal components, PC1, PC2 and PC3 explained 57.89, 24.23 and 17.88 % of the total variance, respectively. Fig. 6 shows the PC1–PC2 scores for the samples **1**–**19**, which are distributed into two distinct separated zones in PC1. The left side has samples with the lowest PC1 values, corresponding to the more potent analogs (associated with a plus sign), whereas the right side has samples with the highest PC1 values, corresponding to less potent analogs (associated with a minus sign). The molecular parameters of the training set responsible for such a distinct classification were the radial distribution function 6.0 weighted by the atomic Sanderson electronegativities, *RDF060e*,[39] hydration energy, *HE*, and *d*(Fe–O_1). They were selected among the descriptors initially computed and are believed to be closely related to the investigated biological response. According to Table I, the Pearson correlation coefficient between log *RA* and these variables are moderate: *RDF060e* (–0.55), *HE* (–0.71) and *d*(Fe–O_1) (0.68); however, between the variables it is, in absolute values, less than 0.45. The other variables were not selected because they either had a poor linear correlation with activity or they did not give a distinct separation of the two classes. The PC1–PC2 loadings (Fig. 7) revealed that the compounds with a higher activity have the main contribution from the *d*(Fe–O_1) descriptor, while the compounds with lower activities have the major influence from the *RDF060e* and *HE* descriptors.

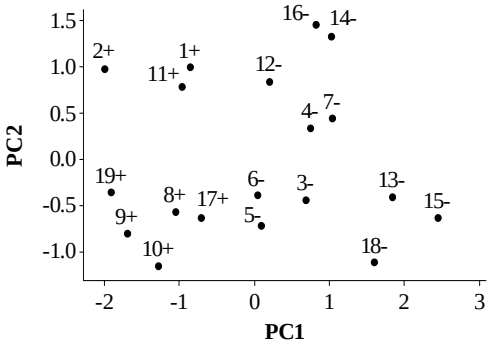

Fig. 6. Plot of PC1–PC2 scores for artemisinin and derivates with antimalarial activity against W-2 strains of *P. falciparum*. Plus sign for more potent analogs and minus sign for less potent analogs.

The radial distribution function descriptors are based on the distance distribution of the molecules and are related to the three-dimensional arrangement of

the atoms. Thus, the *RDF* of an ensemble of *n* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius *R*,[39] The spherical volume comprises the inner part of the molecule. For *RDF060e* this radius was fixed at 6.0 Å from the geometrical center of each molecule under investigation. High values for *RDF060e* are attributed to the presence of electronegative atoms in the inner part of the molecule limited by this spherical volume.

TABLE I. Values of the three properties (descriptors) that classify artemisinin and its 18 derivatives (training set) and their respective experimental log *RA* (W-2) values. The table also shows the correlation matrix (+: more active analogs, –: less active analogs)

Compound	*RDF060e*	*HE*[a] / kcal mol^{-1}	*d*(Fe–O_1) / Å	log *RA* (ref. 6)
1+	21.89	–3.00	2.68	0.00
2+	17.35	–3.52	2.79	0.81
3–	32.87	–1.38	2.73	–0.42
4–	31.00	–2.06	2.59	–0.080
5–	32.28	–1.89	2.82	–0.61
6–	29.53	–1.44	2.81	–0.010
7–	29.98	–1.09	2.62	–0.14
8+	31.28	–4.48	2.74	0.00
9+	28.22	–4.02	2.89	0.64
10+	31.27	–3.72	2.89	0.48
11+	22.00	–2.83	2.73	0.050
12–	29.78	–3.92	2.48	–0.060
13–	40.06	–2.14	2.51	–0.49
14–	27.30	–1.82	2.48	–0.78
15–	40.03	–0.17	2.60	–0.59
16–	25.58	–1.77	2.50	–0.39
17+	31.95	–3.86	2.75	0.31
18–	42.27	–2.19	2.60	–0.51
19+	25.76	–4.37	2.85	0.27
Descriptor			Correlation coefficient	
RDF060e	–	0.38	–0.28	–0.55
HE	–	–	–0.44	–0.71
d(Fe–O_1)	–	–	–	0.68

[a]1 kcal = 4.184 kJ

In a similar study, Pinheiro *et al.*[9] performed an investigation on artemisinin derivatives with antimalarial activities against *P. falciparum* with the aid of computation and multivariate analysis and selected an *RDF* descriptor. Their work indicated the *RDF030m* (the radial distribution function centered at 3.0 Å interatomic distance and weighted by atomic mass) as a property closely related to antimalarial potency of molecules. They associated the compounds with high activity to low values of *RDF030m*. The more potent analogs exhibited, in general, lower values of *RDF060e* and the most potent **2** presented the lowest *RDF060e*. In addition, compounds **1**, **2** and **11** with small substituents (hydrogen atom, methyl or

ethyl) presented lower values for *RDF060e*. The presence of large substituents at the R and R1 position (**13**, **15** and **18**) caused an increase in *RDF060e* due to the presence of more electronegative atoms.

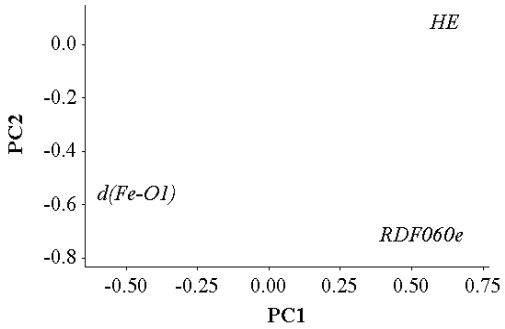

Fig. 7. Plot of the PC1–PC2 loadings with the three descriptors selected to build the PLS model of artemisinin and derivatives with biological activity against W-2 strains of *P. falciparum*.

Regarding the *HE*, more potent analogs showed trends in having more negative energies involved in the hydration process. The *HE* descriptor is a physicochemical property that is a measure of the energy released when water molecules surrounds certain molecules. Its presence means that the mechanism of molecules against malaria is dependent on a hydration process which is related to solubility. Actually, the study of Costa *et al.*[40] led to the conclusion that the presence of water changed the dihedral angle involved in the complex heme–artemisinin (C–Fe––O_1–O_2). Thus, this effect is believed to influence the process of molecular recognition between artemisinin and derivatives and heme in aqueous biological systems. Finally, the selection of the *d*(Fe–O_1) descriptor suggests that the action of drugs against malaria depends on electrophilic attack on the endoperoxide bond, particularly on the O_1 atom. This result was confirmed by both an analysis of the *MEP* maps and by molecular docking as discussed before.

HCA method

The best approach chosen in HCA to group samples into two main classes (one for more potent analogs and the other for less potent analogs) was based on the Euclidean distance and the average group method.[38] This method established links between samples/cluster. The distance between two clusters was computed as the distance between the average values (the mean vector or centroids) of the two clusters. The descriptors employed in HCA were the same in PCA, that is, *RDF060e*, *HE* and *d*(Fe–O_1). The representation of clustering results is shown by the dendrogram in Fig. 8, which depicts the similarity of samples. The branches on the bottom of the dendrogram represent single samples. The length of the branches linking two clusters is related to their similarity. Long branches are related to low similarity while short branches mean high similarity. On the scale of similarity, a value of 100 is assigned to identical samples and a value of 0 to the most dissimilar samples.

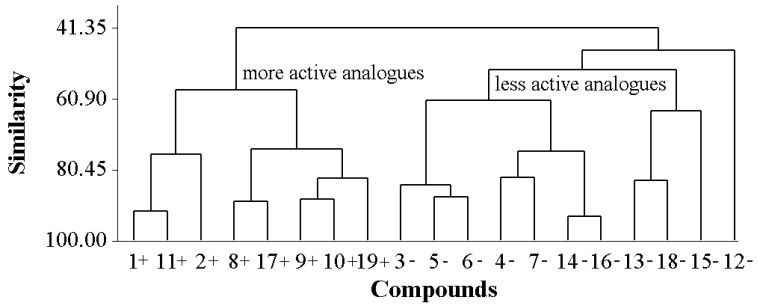

Fig. 8. HCA dendrogram for artemisinin and derivatives with biological activity against W-2 strains of *P. falciparum.* Plus sign for more potent analogs and minus sign for less active compounds.

The dendrogram shows, like the PCA plot, the compounds separated into two different classes according to their activities with no sample incorrectly classified. More active compounds are on the left side and are divided into two groups. In one group consisting of **1**, **2** and **11**, the most active compound 2 is present. The substituents are small (hydrogen atom, methyl or ethyl) and the values for *RDF060e* are the lowest. They present the largest range for activity since it varies from the highest value for log *RA*, 0.81 to 0.00, the activity of artemisinin. The other group consisting of **8**–**10**, **17** and **19** has structures presenting a hydrogen atom or methyl as one of the substituents whereas the other substituent has either a phenyl attached at the end of its alkyl chain with two to four carbons or an ester functional group. They present the most negative values of *HE*. Now considering the less active compounds, there are three main groups. One of them consisting of **13**, **15** and **18** has a phenyl attached at the end of an alkyl chain of three or four carbons in length at R and *n*-butyl at R1; the exception is **15** which has an *n*-butyl at both positions. These analogs present the highest *RDF060e*. A second group consisting of **3**, **5** and **6** has a methyl at R and an alkyl chain of two to five carbons at R1. It is worthwhile mentioning that analog **3** is the only one that has a substitution at R2. A third group consisting of **4**, **7**, **12**, **14** and **16** has compounds with hydrogen atom or methyl as one of the substituents while the other substituent has an alkyl chain length of three to six carbons. Compound **12** is classified as outlier. It presents the most negative hydration energy among the compounds classified as less active. This compound has only a substitution at R, which is a phenyl group with a chlorine atom in the *p* position attached at the end of its alkyl chain. Others compounds classified as less active, the substitution of which occurs only at R, have completely different substituents, they are **14** (*n*-butyl) and **16** (*i*-butyl). Another conclusion from cluster analysis is that the presence of either an *n*-butyl or *i*-butyl at R or R1 causes a great decrease in antimalarial activity, as is verified by analogs **5**, **13**–**16** and **18**.

PLS method

The model built with the aid of the PLS method was based on three latent variables and 15 compounds selected from the training set. The equation which relates the descriptors and biological activity is:

$$\log RA = -0.30RDF060e - 0.45HE + 0.49d(\mathrm{Fe–O_1}) \quad (2)$$

The traditional validation parameters expressed in QSAR terms are: total variance explained = 100 %, R^2 = 0.86, Q^2 = 0.85, *PRESS* = 0.42, *SEP* = 0.17, *s* = = 0.17 and $F_{3,11}$ = 22.52. They support the fact that the model is efficient and hence satisfactory if the complexity of the antimalarial mechanisms and the small number of descriptors (three) selected to build the QSAR model are considered.

According to the PLS equation, the three variables present almost the same magnitude of the coefficients of regression (in absolute value). The model reveals that to achieve high biological potency against *P. falciparum*, a combination of higher values of $d(\mathrm{Fe–O_1})$, lower values of *RDF060e* and more negative values for *HE* are necessary. The predicted and experimental activities reveal a good numerical comparison (Table II). The plot of the correlation between the measured and the predicted log *RA* is shown in Fig. 9. The PLS plot identifies compounds with higher activity (located up) and compounds with lower activity (located down) separately, including compounds from the external validation set. Such classifi-

TABLE II. Experimental and estimated antimalarial activity (log *RA*) by PLS and residuals for the compounds from the training set

Compound	Experimental (ref. 6)	Predicted[a]	Experimental – Predicted
1	0.00	0.22	–0.22
2	0.81	0.60	0.21
3	–0.42	–0.26	–0.16
4	–0.080	–0.29	0.21
5[b]	–0.61	–0.040	–0.57
6	–0.010	–0.050	0.040
7	–0.14	–0.43	0.29
8[b]	0.00	0.41	–0.41
9	0.64	0.61	0.030
10	0.48	0.48	0.00
11	0.050	0.26	–0.21
12	–0.060	–0.080	0.020
13	–0.49	–0.60	0.11
14	–0.78	–0.44	–0.34
15[b]	–0.59	–0.85	0.26
16	–0.39	–0.38	–0.010
17	0.31	0.28	0.030
18	–0.51	–0.49	–0.020
19[b]	0.27	0.67	–0.40

[a]PLS model using three latent principal components and leave-one-out cross-validation; [b]compounds from the external validation set

cation is the same as that achieved by the PCA and HCA methods: no sample appearing in a different class.

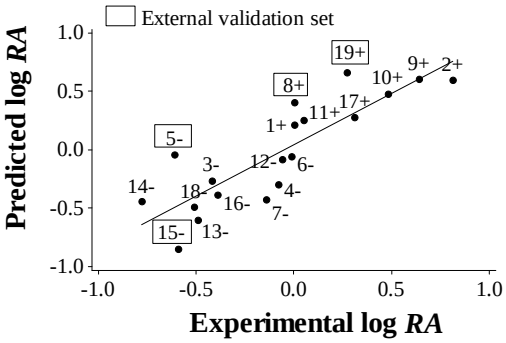

Fig. 9. Experimental *vs.* predicted log *RA* (W-2) for artemisinin and derivatives plotted by PLS model using three latent variables. External validation compounds included.

CONCLUSIONS

The HF method and the 3-21G** basis set revealed themselves to be adequate to optimize the structures of artemisinin and derivatives and consequently their study. *MEP* maps characterized the region of the 1,2,4-trioxane ring for artemisinin and derivatives as a region of negative electrostatic potential. Molecular docking calculations revealed that the Fe^{2+} ion from heme attacks the O_1 from artemisinin more preferably than O_2. The PCA and HCA methods enabled the compounds in the training set to be classified into two groups according to their degree of antimalarial activity against *P. falciparum*. The descriptors *RDF060e*, *HE* and *d*(Fe–O_1) were the properties responsible for distinguishing compounds with higher or lower antimalarial activity. The combination of these structural attributes is believed to govern the antimalarial response of the compounds studied in this work. The PLS model revealed that the more active compounds have higher values of *d*(Fe–O_1), lower values of *RDF060e* and more negative values of *HE*. By this strategy, useful information was obtained that could be of use in experimental processes of syntheses and biological evaluation in order to find new efficient drugs against malaria.

Acknowledgements. We acknowledge the financial support of the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico. We also thank the Instituto de Química-Araraquara for the use of the GaussView software and the Swiss Center for Scientific Computing for the use of the Molekel software. We employed the computing facilities at the Centro Nacional de Processamento de Alto Desempenho, Universidade Estadual de Campinas, and at the Laboratório de Química Teórica e Computacional, Universidade Federal do Pará.

ИЗВОД

СТУДИЈА НОВИХ АНТИМАЛАРИЈСКИХ АРТЕМИЗИНИНА ПОМОЋУ МОЛЕКУЛСКОГ МОДЕЛИРАЊА И МУЛТИВАРИЈАНТНЕ АНАЛИЗЕ

JOÃO E. V. FERREIRA[1], ANTONIO F. FIGUEIREDO[1], JARDEL P. BARBOSA[1], MARIA G. G. CRISTINO[1], WILLIAMS J. C. MACEDO[1], OSMARINA P. P. SILVA[1], BRUNO V. MALHEIROS[1], RAYMONY T. A. SERRA[2] и JOSE CIRIACO-PINHEIRO[1]

[1]*Laboratório de Química Teórica e Computacional, Faculdade de Química, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Avenida Augusto Correa, 01, CP 101101, CEP 66075-110, Belém, PA, Amazônia и* [2]*Centro de Ciências Biológicas e da Saúde, Universidade Federal do Maranhão, CEP 65085-580, São Luis, MA, Brasil*

Артемизинин и његових 18 деривата који имају антималарично дејство на W-2 сој *Plasmodium falciparum*-а проучаване су методама квантне хемије и мултиваријантне анализе. Оптимизација геометрије је изведена помоћу Хартри-Фокове (ХФ) теорије са базним сетом 3-21G** . Мапе молекулског електростатичног потенцијала (*MEP*) и молекулско доковање су употребљени за истраживање интеракција између лиганда и рецептора (хема). Анализа главне компоненте (*Principal Component Analysis*, PCA) и хијерархијска кластерска анализа (*Hierarchical Cluster Analysis*, HCA) су примењени да би се изабрали најважнији дескриптори у односу на активност. Помоћу методе парцијалних најмањих квадрата (*Partial Least Square*, PLS) генерисан је одговарајући предиктивни модел. Конструкција тог модела базирана је на 15 молекула и још 4 коришћена као сет за спољашњу валидацију; параметри валидације су Q^2 = 0,85 и R^2 = 0,86. Модел укључује следеће молекулске параметре: радијалну функцију расподеле, *RDF060e*, енергију хидратације, *HE*, и растојање између атома O_1 у лиганду и атома гвожђа у хему, *d*(Fe–O_1). На тај начин могу се синтетисати нови деривати на основу мапа *MEP* и PLS анализе.

(Примљено 26. јануара, ревидирано 12. јуна 2010)

REFERENCES

1. D. M. Opsenica, B. A. Šolaja, *J. Serb. Chem. Soc.* **74** (2009) 1155
2. R. Arav-Boger, T. A. Shapiro, *Annu. Rev. Pharmacol. Toxicol.* **45** (2005) 565
3. N. J. White, *J. Clin. Invest.* **113** (2004) 1084
4. R. G. Ridley, *Nature* **424** (2003) 887
5. P. M. O'Neill, N. L. Searle, K. W. Kan, R. C. Storr, J. L. Maggs, S. A. Ward, K. Raynes, B. K. Park, *J. Med. Chem.* **42** (1999) 5487
6. J. R. Woolfrey, M. A. Avery, A. M. Doweyko, *J. Comput.-Aided Mol. Des.* **12** (1998) 165
7. M. A. Rafiee, N. L. Hadipour, H. Naderi-Manesh, *J. Chem. Inf. Model.* **45** (2005) 366
8. G. Bernardinelli, C. W. Jefford, D. Marić, C. Thomson, J. Weber, *Int. J. Quantum Chem.: Quantum Biol. Symp.* **21** (1994) 117
9. J. C. Pinheiro, R. Kiralj, M. M. C. Ferreira, O. A. S. Romero, *QSAR Comb. Sci.* **22** (2003) 830
10. J. C. Pinheiro, M. M. C. Ferreira, O. A. S. Romero, *J. Mol. Struct. Theochem* **572** (2001) 35
11. F. J. B. Cardoso, A. F. Figueiredo, M. S. Lobato, R. M. Miranda, R. C. O. Almeida, J. C. Pinheiro, *J. Mol. Modell.* **14** (2008) 39
12. F. J. B. Cardoso, R. B. Costa, A. F. Figueiredo, J. P. Barbosa, I. Nava-Junior, J. C. Pinheiro, O. A. S. Romero, *Internet Electon. J. Mol. Des.* **6** (2007) 122

13. R. Guha, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **44** (2004) 1440
14. F. Cheng, J. Shen, X. Luo, W. Zhu, J. Gu, R. Ji, H. Jiang, K. Chen, *Bioorg. Med. Chem.* **10** (2002) 2883
15. A. K. Bhattacharjee, K. A. Carvalho, D. Opsenica, B. A. Šolaja, *J. Serb. Chem. Soc.* **70** (2005) 329
16. A. Doweyko, *J. Comp.-Aided Mol. Des.* **22** (2008) 81
17. D. J. Creek, D. K. Chalmers, W. N. Charman, B. J. Duke, *J. Mol. Graphics Modell.* **27** (2008) 394
18. P. M. O'Neill, *Expert Opin. Invest. Drugs* **14** (2005) 1117
19. G. A. Biagini, P. M. O'Neill, P. G. Bray, S. A. Ward, *Curr. Opin. Pharmacol.* **5** (2005) 473
20. R. K. Haynes, S. C. Vonwiller, *Tetrahedron Lett.* **37** (1996) 253
21. G. H. Posner, D. Wang, J. N. Cumming, C. H. Oh, A. N. French, A. L. Bodley, T. A. Shapiro, *J. Med. Chem.* **38** (1995) 2273
22. J. Q. Araújo, J. W. M. Carneiro, M. T. Araújo, F. H. A. Leite, A. G. Taranto, *Bioorg. Med. Chem.* **16** (2008) 5021
23. C. W. Jefford, *Curr. Med. Chem.* **8** (2001) 1803
24. P. Gramatica, *QSAR Comb. Sci.* **26** (2007) 694
25. *GaussView 1.0*, Gaussian, Inc., Pittsburgh, PA, 1997
26. C. C. J. Roothaan, *Rev. Mod. Phys.* **23** (1951) 23, 69
27. J. S Binkley, J. A. Pople, W. J. Hehre, *J. Am. Chem. Soc.* **102** (1980) 939
28. *Gaussian 98, revision A.7*, Gaussian, Inc., Pittsburgh, PA, 1998
29. E. Estrada, E. Molina, *J. Mol. Graphics Modell.* **20** (2001) 54
30. *Virtual Computational Laboratory*, VCCLAB 2005, http://www.vcclab.org (accessed January 2010)
31. I. V. Tetko, J. Gasteiber, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Radchenko, N. S. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko, *J. Comput.-Aided Mol. Des.* **19** (2005) 453
32. *Auto-Dock 4.0*, The Scripps Research Institute, Department of Molecular Biology, MB-5, La Jolla, CA, 2007
33. *Molekel 4.3*, Swiss Center for Scientific Computing, Manno, Switzerland, 2000
34. *ChemPlus, Modular Extensions to HyperChem, Release 6.02*, Molecular Modeling for Windows, Hyper, Inc., Gainesville, 2000
35. J. Vojtechovsky, K. Chu, J. Berendzen, R. M. Sweet, I. Schlichting, *Biophys. J.* **27** (1999) 2153
36. M. M. C. Ferreira, C. A. Montanari, A. C. Gaudio, *Quim. Nova* **3** (2002) 439
37. A. Golbraikh, A. Tropsha, *J. Mol. Graphics Modell.* **20** (2002) 269
38. *Pirouette 3.01*, Infometrix, Inc., Woodinville, WA, 2001
39. M. C. Hemmer, V. Steinhauer, J. Geisteiger, *Vib. Spectrosc*. **19** (1999) 151
40. M. S. Costa, R. Kiralj, M. M. C. Ferreira, *Quim. Nova* **30** (2007) 25
41. S. Tonmunphean, V. Parasuk, S. Kokpol, *J. Mol. Modell.* **7** (2001) 7 26.