*Journal of*
*Computational and Theoretical Nanoscience*
Vol. 11, 1–9, 2014

# Validation of Computational Methods Applied in Molecular Modeling of Artemisinin with Antimalarial Activity

Cleydson Breno Rodrigues dos Santos[1, 2, 3, *], Josinete Braga Vieira[3], Adriane da Silva Formigosa[3], Elizabeth Viana Moraes da Costa[1], Mayara Tânia Pinheiro[1, 2], Jocivânia Oliveira da Silva[1, 3], Williams Jorge da Cruz Macêdo[4], and José Carlos Tavares Carvalho[1, 2]

[1] *Laboratory of Drug Research, School of Pharmaceutical Sciences, Federal University of Amapá, Rod JK km 2, Macapá-Amapá, 68902-280, Brazil*
[2] *Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Federal University of Amapá, Rod JK km 2, Macapá-Amapá, 68902-280, Brazil*
[3] *Laboratory of Modelling and Computational Chemistry, Federal University of Amapá, Rod JK km 2, Macapa, Amapá 68902-280, Brazil*
[4] *Secretary of State for Public Health, Central Laboratory of the State of Pará-LACEN/PA, Rod Augusto Montenegro, km 10, Icoaraci, 66823-010, Belém-Pará-Amazon-Brazil*

Artemisinin (qinghaosu) has a unique structure bearing a stable endoperoxide lactone (1,2,13-trioxane) totally different from previous antimalarial in its structure and mode of action, are isolated from Artemisia annua is a composed of a remarkable life and antimalarial effective against Plasmodium falciparum and cerebral malaria. We propose a combination of chemical quantum methods and multivariate analysis to study the geometric parameters of artemisinin in the region endoperoxide of the ring (1,2,13-trioxane), in order to be effective in selecting the method and level of theory when compared with data crystallographic, aiming to classify and correlate. The most important geometrical parameters selected by principal component analysis (PCA) were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The results of PCA showed that the model was built with three main components (3PCs), explains 97.0861% of the total variance. The level of theory HF/6-31G** show high similarity with the experimental data assuming that the combination of *ab initio* method can be used for modeling the molecular structure of artemisinin and its derivatives antimalarial drugs with mechanism of action in the ring region endoperoxide.

**Keywords:** Artemisinin, HF/6-31G**, Molecular Modeling, Quantum Chemical Methods, Multivariate Analysis.

RESEARCH ARTICLE

## 1. INTRODUCTION

The evolution of computational chemistry is one of the most important advances in the design and discovery of new drugs has been the use of molecular modeling (MM). Currently, MM is an indispensable tool not only in the process of drug discovery, but also the optimization of existing prototypes and the rational design of drug candidates.[1–4] According to IUPAC, the MM is the investigation of molecular structures and properties by the use of computational chemistry and graphical visualization techniques in order to provide a three-dimensional representation under a given set of circumstances.[2] The nature of the molecular properties used and the extent to which they describe the structural features of molecules can be related to biological activity, which is an important part of any QSAR studies.

The molecular properties represent a means of chemical information, contained in the molecular structure of the compound. This information is transformed and codified for lot of problems chemical, pharmacological and toxicological of the relationship between quantitative structure-activity and structure-property studies (QSAR, QSPR). The molecular properties take into account different aspects

*Author to whom correspondence should be addressed.

of chemical information, this information may be through experiments, theoretical calculations or simple counting, considering the entire molecule, fragments or functional groups, knowledge of 3-D structure of the molecule or its molecular graphics or simply its formula, information defined by scalar values, vectors or scalar fields.[5]

The great development of MM in recent years was due largely to the advancement of computational resources in terms of hardware (speed of computation) and software (computer programs), in addition to advances in computational chemistry, nuclear magnetic resonance, ray crystallography-X, biochemistry and molecular biology. This allowed major contribution to the discovery of drug candidates, leading to rapid progress in research and attracting the interest of both academia and pharmaceutical industries.[1, 4]

Silva et al.[6] used density functional theory (DFT) calculations (B3PW91/DGDZVP) to determine $^{13}C$ and $^{1}H$ nuclear magnetic resonance (NMR) chemical shifts for the two dihydrochalcones: 3,4,5-trimethoxydihydrochalcone and 2,3,4,4-tetramethoxydihydrochalcone. The experimental and theoretical NMR data were analyzed by simple linear regression and the more relevant parameters were selected. In additional, other statistical parameters (correlation coefficients, significance and predictability) were available to judge the quality of the calculations. Finally, the statistical analysis show good correlation experimental and theoretical NMR data with high predictive power.

Ibrahim et al.[7] used Semiemperical molecular modeling technique is applied to assess the interaction of amino acids (alanine, asparagines, aspartic, arginine, cysteine, glutamine, glycine and tryptophan) with chitosan. Results indicate the selectivity of chitosan furthermore we introduce the site whereas amino acid could interact with chitosan. Chitosan is interacting with amino acid through $NH_2$ group. It is concluded that chitosan is acting with amino acids like protein interaction which dedicate chitosan for many applications in the biological system.

The Malaria is one of the most common diseases in tropical countries. More than 300 million infections and millions of malaria deaths occur annually worldwide. The rapid spread of resistance to current quinoline antimalarial has made malaria a major global problem, so it is essential to seek new drugs against malaria and understand its mechanism of action for treating patient.[8]

The treatment of malaria has been used in traditional Chinese medicine for more than two million years. The medicine used is the artemisinin (qinghaosu) and is extracted from the plant Artemisia annua L, used to combat diseases of 52 species of the People's Republic of China. Artemisinin (qinghaosu) has a unique structure (Fig. 1) bearing a stable endoperoxide lactone (1,2,13-trioxane) totally different from previous antimalarial in its structure and mode of action, are isolated from Artemisia annua is a composed of a remarkable life and antimalarial effective against Plasmodium falciparum and cerebral malaria.[8, 9]

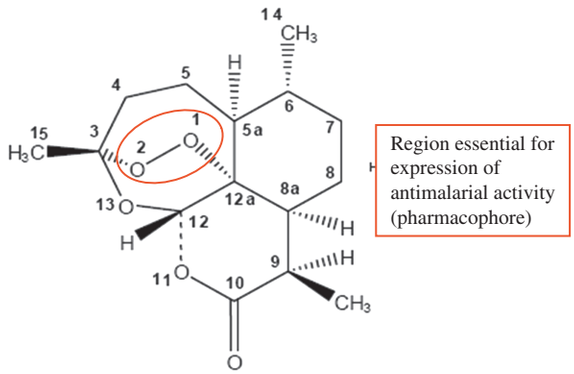


**Fig. 1.** Artemisinin (structure).

Artemisinin and its derivatives induce a rapid reduction of the number of parasites when compared with other known drugs. Consequently, they are of particular interest to severe malaria. The first decline in the number of parasites is also beneficial for combination therapies. This led to an enormous interest in the mechanism of action, chemistry and drug development of this new class of antimalarials. The group endoperoxide is essential for the antimalarial activity and is mediated by activated oxygen (superoxide, $H_2O_2$ and/or hydroxyl radicals) or carbon free-radicals.[9–12]

Carvalho et al.[13] studied with B3LYP/6-31G** level of theory the artemisinina and 31 analogues with anti leishmanicidal activity against *Leishmania donovani,* and proposed a set of 13 artemisinins, 7 less active and 6 that have not been tested, and of these six, one was expected to be more active against *L. donovani.* In this study, maps Electrostatic Potential (MEP) were used in an attempt to identify key structural features of artemisinin and analogs, and mode of interaction with its receptor (heme). The chemometric methods: PCA, HCA, SDA, KNN and SIMCA were used to reduce dimensionality and investigate which subset of descriptors are responsible for the classification of the activity anti leishmanicidal as major and minor.

Figueiredo et al.[14] studied by computational chemistry dispiro-1,2,4-trioxolanes with antimalarial activities against *K*-1 strains of Plasmodium falciparum. Molecules were optimized with B3LYP/631G* method. A predictive model was generated by PLS method, with three latent variables explaining 99.8% of the total variance, $Q^2 =$ 087, $R^2 = 0.85$, obtained for 16/4 molecules in the training/external validation set. The descriptors selected for the model were the binding free energy, logarithm of octanol-water partition coefficient (log*P*) and molecular volume.

In this work, molecular modeling of the artemisinin with activities antimalarial, tested *in vitro* against human malaria Plasmodium falciparum, was modeled and constructed by following the strategy primarily based on the knowledge that the group endoperoxide the artemisinin present is responsible for its antimalarial activity. The artemisinin molecule was modeled in three levels of theory

RESEARCH ARTICLE

and methods, were performed principal component analysis (PCA) and hierarchical cluster analysis (HCA), in order to evaluate which geometric parameters the artemisinin in the region of ring endoperoxide (1,2,13-trioxane), may be more effective in the choice of a method and theory levels studied, as compared to the crystallographic data.

## 2. METHODOLOGY

### 2.1. Molecular Modeling

The molecular modeling of artemisinin was built following the strategy described: initially the structure of artemisinin was performed with the program GaussView 3.0,[15] and optimized with methods and levels of theory different-Semi-empirical (AM1, PM3 and ZINDO), *ab initio*/Hartree-Fock (HF/6-31G, HF/6-31G* and HF/6-31G**) and DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**), implemented with the program Gaussian 03.[16] These calculations were performed to find the method and level of theory with the best fit between the computational time and accuracy of the information on the experimental data.[17] The experimental structure of artemisinin was taken from the Cambridge Structural Database CSD, with REFCODES: QNGHSU10, crystallographic *R* factor 3.6.[18] The numbering of the atoms used in this study is shown in Figure 1 (artemisinin).

## 3. CHEMOMETRICS

### 3.1. Principal Component Analysis (PCA) e Hierarchical Cluster Analysis (HCA)

The multivariate analysis is when measurements are made on a number of objects, the results are usually arranged in a matrix, which is called the data matrix. The measures (in this study the geometrical parameters, concerning the methods and different levels of theory) are placed in columns, and objects (in this study the quantum-chemical methods applied) are associated with the lines. The step multivariate analysis was accomplished with the program Piroutte 3.10.[19]

Principal component analysis (PCA) is a method of data compression based on the correlation between the variables, the compression of data generates a small set of variables which are known as principal components, and they are mutually orthogonal, so that the correlation between the variables does not limit its application, unlike the multiple linear regression that is sensitive to the presence of highly correlated variables because the regression coefficients become unstable and meaningless.[20, 21] For a multidimensional data matrix, multivariate statistical methods are needed for standardization autoscaled employ in order to treat all variables with the same degree of importance for understanding such data in its entirety. The main objective of the analysis is to show the data in a multidimensional space of low dimensionality with minimal loss of information overall.[22]

The PCA was performed with autoscaled processing, with a maximum of three factors (3 PCS) using the procedure "leave-one-out method validation and cross-validation." The data matrix was constructed with dimensions $10 \times 18$, where each column was associated with three (03) methods and nine (09) level of theory, one (01) of columns relating to the geometric parameters experimental,[17] and each line 18 represents the geometric parameters of 1,2,13-trioxane ring (bond lengths, bond angles, and torsion angles). The final result of PCA is to select a small number of geometric parameters which artemisinin can best be related to dependent variable, in this case the standard deviation of the various methods and levels of theory.

As in PCA results of hierarchical cluster analysis (HCA) are qualitative, being arranged in the form of a dendrogram thus view the methods studied or the variable (geometric parameters of artemisinin) in a two-dimensional space, which illustrates the merger or divisions made in each successive stage analysis. Samples (methods and levels of theory) are represented by the bottom branch of the dendrogram. The similarity between the agglomerates is given by the length of its branches, so that the methods and theory has low levels of similarity have long branches, while such methods and theory levels of high similarity have short branches.[20] In HCA the distance between these variables is calculated and transformed into a similarity matrix *S*. A hierarchical cluster analysis aims to show the data in such a way to accentuate their natural groupings and patterns. The statistical analysis required in this study to group the methods and theory levels similarly in their respective categories. HCA is a statistical method developed for this purpose. HCA was performed with autoscaled processing, with euclidean distance and the incremental method.

## 4. RESULTS AND DISCUSSION

### 4.1. Method and Basis Set for the Description of the Geometries of Artemisinin

We determined the theoretical geometric parameters of artemisinin in the region endoperoxide of the ring 1,2,13-trioxane (bond length, bond angle and angle of twist of forming this ring atoms), with objective of assess the quality of the wave function of the molecular theoretical and experimental geometrical parameters of the ring 1,2,13-trioxane artemisinin shown in Table I.

According to Table I shows that the method *ab initio*/Hartree-Fock all three sets of base (HF/6-31G**, HF/6-31G*, HF/6-31G) describe well all the structural parameters in magnitude and sign when compared with the experimental values, contrary the semi-empirical methods

RESEARCH ARTICLE

**Table I.** Theoretical and experimental parameters of the 1,2,13-trioxane ring in artemisinin.

| | Semi-emperical | | | Hartree-fock/HF | | | DFT/B3LYP | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameters[a] | AM1[23] | PM3[23] | ZINDO[23] | 6-31G[23] | 6-31G*[24] | 6-31G**[b] | 3-21G[25] | 3-21G*[25] | 3-21G**[25] | Experimental[17] |
| Bond length (Å) | | | | | | | | | | |
| O1O2 | 1.288 | 1.544 | 1.237 | 1.447 | 1.391 | 1.390 | 1.524 | 1.524 | 1.524 | 1.469 |
| O2C3 | 1.447 | 1.403 | 1.400 | 1.435 | 1.393 | 1.396 | 1.455 | 1.455 | 1.454 | 1.416 |
| C3O13 | 1.427 | 1.428 | 1.396 | 1.435 | 1.388 | 1.408 | 1.473 | 1.473 | 1.472 | 1.445 |
| O13C12 | 1.416 | 1.403 | 1.392 | 1.403 | 1.400 | 1.376 | 1.430 | 1.430 | 1.430 | 1.379 |
| C12C12a | 1.537 | 1.555 | 1.513 | 1.533 | 1.533 | 1.532 | 1.535 | 1.535 | 1.535 | 1.523 |
| C12aO1 | 1.468 | 1.426 | 1.416 | 1.469 | 1.429 | 1.429 | 1.504 | 1.504 | 1.504 | 1.461 |
| Bond angle (°) | | | | | | | | | | |
| O1O2C3 | 112.530 | 110.340 | 114.310 | 108.800 | 106.100 | 109.460 | 105.590 | 105.590 | 105.480 | 108.100 |
| O2C3O13 | 103.600 | 104.810 | 105.370 | 106.760 | 110.800 | 107.800 | 108.220 | 108.220 | 108.250 | 106.600 |
| C3O13C12 | 115.480 | 116.010 | 115.843 | 117.300 | 112.800 | 115.300 | 113.200 | 113.200 | 113.200 | 114.200 |
| O13C12C12a | 113.510 | 115.200 | 113.270 | 112.280 | 108.700 | 112.300 | 113.300 | 113.300 | 113.230 | 114.500 |
| C12C12aO1 | 111.070 | 113.180 | 107.290 | 110.910 | 110.500 | 110.545 | 112.410 | 112.410 | 112.470 | 110.700 |
| C12aO1O2 | 113.740 | 112.290 | 118.380 | 113.240 | 112.700 | 112.700 | 109.620 | 109.620 | 109.590 | 111.200 |
| Torsion angle (°) | | | | | | | | | | |
| O1O2C3O13 | −77.800 | −73.310 | −70.403 | −71.840 | −73.369 | −73.400 | −76.610 | −76.610 | −76.740 | −75.500 |
| O2C3O13C12 | 42.070 | 52.700 | 36.370 | 33.390 | 31.034 | 31.100 | 33.750 | 33.750 | 33.720 | 36.000 |
| C3O13C12C12a | 11.400 | 2.811 | 17.420 | 25.320 | 27.432 | 27.400 | 29.059 | 29.060 | 29.080 | 25.300 |
| O13C12C12aO1 | −41.770 | −40.510 | −46.610 | −49.410 | −50.100 | −50.143 | −52.190 | −52.190 | −52.030 | −51.300 |
| C12C12aO1O2 | 12.050 | 19.940 | 18.110 | 12.510 | 10.900 | 10.924 | 9.060 | 9.600 | 9.340 | 12.700 |
| C12aO1O2C3 | 47.050 | 35.630 | 40.130 | 46.700 | 48.700 | 48.674 | 51.060 | 51.060 | 51.320 | 47.800 |
| Standard deviation | 4.776 | 8.388 | 4.372 | 1.663 | 2.484 | 1.762 | 1.915 | 1.855 | 1.987 | – |

*Notes*: [a]The atoms are numbered according to Figure 1; [b]Valence basis set separately validated for obtaining the calculations of molecular properties; [17]J. N. Lisgarten, et al., *Journal of Chemical Crystallography* 28, 539 **(1998)**; [23]J. C. Pinheiro, et al., *Journal Molecular Structure* (*THEOCHEM*) 572, 35 **(2001)**; [24]C. Thomsom, et al., *International Journal of Quantum Chemistry: Quantum Biology Symposium* 18, 231 **(1991)**; [25]A. D. Becke, *J. Chem. Phys.* 98, 5648 **(1993)**.

RESEARCH ARTICLE

(AM1, PM3 and ZINDO) and DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**), which show the standard deviation 4.776, 8.388 and 4.372 for the semi-empirical and 1.915, 1.855 and 1.987 for the DFT, respectively. When comparing the method *ab initio*/ Hartree-Fock, we can see that both levels of theory (6-31G and 6-31G**) have the lowest standard deviations in relative to semi-empirical methods and DFT, with values of 1.663 and 1.762, respectively, having a variation of about 0.099 between them.

### 4.2. PCA Method

The theoretical and experimental parameters of the 1,2,13-trioxane ring of artemisinin were used with objective to identify, through PCA and HCA, which optimize the geometry of artemisinin in different methods and levels of theory defines the results closer to experimental data. The advantage of using PCA and HCA methods in the present study was that all structural parameters are considered simultaneously, and that takes into account the correlations between them.

The PCA results showed that descriptors four (04) most important related to ring trioxane were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The values of geometric parameters selected by principal component analysis, standard deviation and Pearson correlation matrix are shown in Table II. These geometric parameters are responsible for the separation of methods and levels in three classes: semi empirical, *ab inito*/Hartree-Fock and DFT. The four (04) geometrical parameters related to trioxane ring are identified by atoms with the measures O13C12, which is the interplanar distance between the two atoms (bond length), O1O2C3 is the angle of bond between these three atoms, C3O13C12C12a and C12C12aO1O2 are related or dihedral angle of torsion.

Table II shows the correlation matrix between the geometric parameters and the standard deviation, it is noted that the correlation between the geometric parameters is less or equal to 0.827, while the correlation between the geometric parameters and the standard deviation is less or equal to 0.914. The geometric parameters selected with the PCA technical represent characteristics necessary to validate the best method and level of theory.

The results of model selection are shown in Table III, and show that the model was constructed with three main components (3PCs), where the first principal component (PC1) describes 24.3990% of the total information, the second main component (PC2) describes 8.0043%, and third (PC3) 2.5477%. Further in this table, we observe that PC1 contains 67.7751% of the original data, the first two (PC1 + PC2) 90.0093% and first three (PC1 + PC2 + PC3) can explain 97.0861% of the total information, losing only 2.9139% of original information. In the same table, it is verified that the geometrical parameters O1O2C3 (0.5447) and C12C12aO1O2 (0.5657) are the main contributors

**Table II.** Geometric parameters selected by principal component analysis, standard deviation and Pearson correlation matrix.

| Methods | O13C12 | O1O2C3 | C3O13C12C12a | C12C12aO1O2 | Standard deviation |
|---|---|---|---|---|---|
| AM1 | 1.416 | 112.530 | 11.400 | 12.050 | 4.776 |
| PM3 | 1.403 | 110.340 | 2.811 | 19.940 | 8.388 |
| ZINDO | 1.392 | 114.310 | 17.420 | 18.110 | 4.372 |
| 6-31G | 1.403 | 108.800 | 25.320 | 12.510 | 1.663 |
| 6-31G* | 1.400 | 106.100 | 27.432 | 10.900 | 2.484 |
| 6-31G** | 1.376 | 109.460 | 27.400 | 10.924 | 1.762 |
| 3-21G | 1.430 | 105.589 | 29.059 | 9.060 | 1.915 |
| 3-21G* | 1.430 | 105.590 | 29.060 | 9.600 | 1.855 |
| 3-21G** | 1.430 | 105.480 | 29.080 | 9.340 | 1.987 |
| Experimental | 1.379 | 108.100 | 25.300 | 12.700 | 0.000 |
| O13C12 | | −0.465 | 0.140 | −0.418 | 0.068 |
| O13C3 | | | −0.700 | 0.727 | 0.540 |
| C3O13C12C12a | | | | −0.827 | −0.914 |
| C12C12aO1O2 | | | | | 0.758 |

to PC1, while that the geometrical parameters O13C12 (0.8771) and C12C12aO1O2 (0.1213) are the main contributors to PC2.

As the principal components can be written as a linear combination of selected geometrical parameters, mathematical expressions for PC1 and PC2 are shown below:

$$PC1 = -0.3296(\text{O13C12}) + 0.5447(\text{O1O2C3}) - 0.5239(\text{C3O13C12C12a}) + 0.5657(\text{C12C12aO1O2}) \quad (1)$$

$$PC2 = 0.8771(\text{O13C12}) - 0.0404(\text{O1O2C3}) - 0.4628(\text{C3O13C12C12a}) + 0.1213(\text{C12C12aO1O2}) \quad (2)$$

Figure 2 shows the scores for 03 (three) methods and 09 (nine) levels of theory, correlating with the experimental geometry parameter of 1,2,13-trioxane ring. According to the score shows that the methods are broken into three groups according PC1. The semi-empirical method (AM1 and PM3 ZINDO) are located on the right side, while the method ab inito/Hartree-Fock and experimental are located in the central part of Figure 2. Already the DFT/B3LYP method are located in the upper left. Furthermore, we can see that the method *ab initio* HF/6-31G, HF/6-31G* and HF/6-31G** are the closest experimental geometry parameter indicating that any one of them can be used in the development of calculations of molecular properties.

Figure 3 shows the loading for four (4) major classification descriptors of the methods and theory levels

**Table III.** Principal component analysis of the selection model of levels of theory.

| | Principal component | | |
|---|---|---|---|
| | PC1 | PC2 | PC3 |
| Variance (%) | 24.3990 | 8.0043 | 2.5477 |
| Variance cumulative (%) | 67.7751 | 90.0093 | 97.0861 |
| | Contribution | | |
| Molecular descriptors | PC1 | | PC2 |
| O13C12 | −0.3296 | | 0.8771 |
| O1O2C3 | 0.5447 | | −0.0404 |
| C3O13C12C12a | −0.5239 | | −0.4628 |
| C12C12aO1O2 | 0.5657 | | 0.1213 |

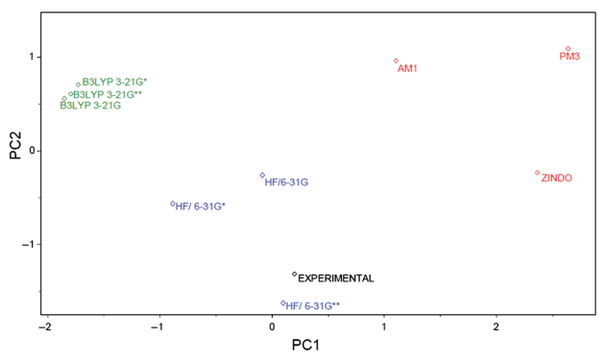


**Fig. 2.** Plot of the scores PC1–PC2 for 03 (three) methods and 09 (nine) levels of theory, correlating with the experimental geometry parameter of 1,2,13-trioxane ring.

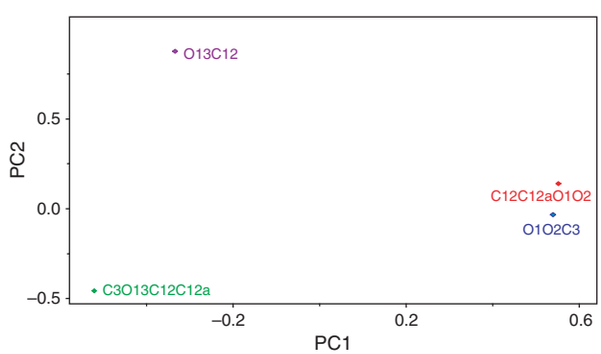


**Fig. 3.** Plot of PC1–PC2 loadings using four (04) with geometric parameters selected principal components analysis to three (3) methods and nine (09) levels of theory, correlating with the experimental geometry parameter ring 1,2,13-trioxane.

RESEARCH ARTICLE

are observed in the score that the geometrical parameters regarding the methods and levels of theory DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) and semi-empirical (AM1 and PM3) have main contribution of the geometric parameters O13C12 and C12C12aO1O2, and these are responsible for moving the methods and theory levels for the upper score. While the method *ab initio* (HF/6-31G, HF/6-31G* and HF/6-31G**) and semi-empirical (ZINDO) have a high contribution of the geometric parameters O1O2C3 and C3O13C12C12a, which are responsible in displacing methods and theory levels to the bottom of the score. Also in Figure 3, note that the larger the contribution of the geometric parameters O13C12 and C12C12aO1O2 the second main component the greater the score, and thus the method and levels of theory are less efficient, because if the distance experiment. The geometric parameters O1O2C3 and C3O13C12C12a contribute in lesser degree, by weight have a negative PC2, demonstrating that the methods and theory levels in general have higher values of the geometric parameters.

The geometric parameters O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2 are of great importance in our study, since according to the proposal made by Jefford and colleagues of the heme iron attacks the artemisinin O1 position and generates a free radical in position O2 (Fig. 4), after bond C3–C4 be broken into a radicular carbon at C4.[26] This free radical C4 has been suggested as an important substance in the antimalarial activity.[27] Study of molecular docking of artemisinin and its receptor, heme, made by Tonmunphean, Parasuk and Kokpol also indicated that the heme iron interacts with O1 more preferably to O2,[28] which sets the importance of geometric parameter O1O2C3, selected this model to be associated with mechanism of action suggested.

Elhaes et al.[29] reported that the possible interactions of nanomaterials with living cells must be of concern, and studies have shown that the interaction between Epoxides C60 (C60–O) and hemoglobin is more likely to happen to adsorb and complex state. Djemil et al.[30] studied the structural aspects for the complexation of dopamine (DA) and epinephrine (EP) to $\beta$-CD were Explored by using PM6, HF and ONIOM methods. The structures show the presence of several intermolecular hydrogen bond interactions that were studied on the basis of NBO analysis employed to quantify the donor–acceptor interactions between the guest molecules and $\beta$-CD.

Table IV shows the geometrical parameters selected by principal component analysis, method and theory levels, variation of geometrical parameters in relation to the experimental data ($\Delta$ and $\Delta\%$). In the semi-empirical and DFT methods there isn't good agreement between theoretical and experimental values for the torsion angles, especially the angles formed by atoms C3O13C12C12a and C12C12aO1O2.

In semi-empirical method AM1, PM3 and ZINDO shows deviations $\Delta = -13.900°$ ($\Delta\% = -54.940$), $\Delta = -22.489°$ ($\Delta\% = -88.889$) and $\Delta = -7.879°$ ($\Delta\% = -31.146$) in relative to torsion angles C3O13C12C12a, respectively. For the torsion angles C12C12aO1O2 shows deviations $\Delta = -0.650°$ ($\Delta\% = -5.118$), $\Delta = -7.240°$ ($\Delta\% = 57.007$) and $\Delta = 5.410°$ ($\Delta\% = 42.598$). In relation

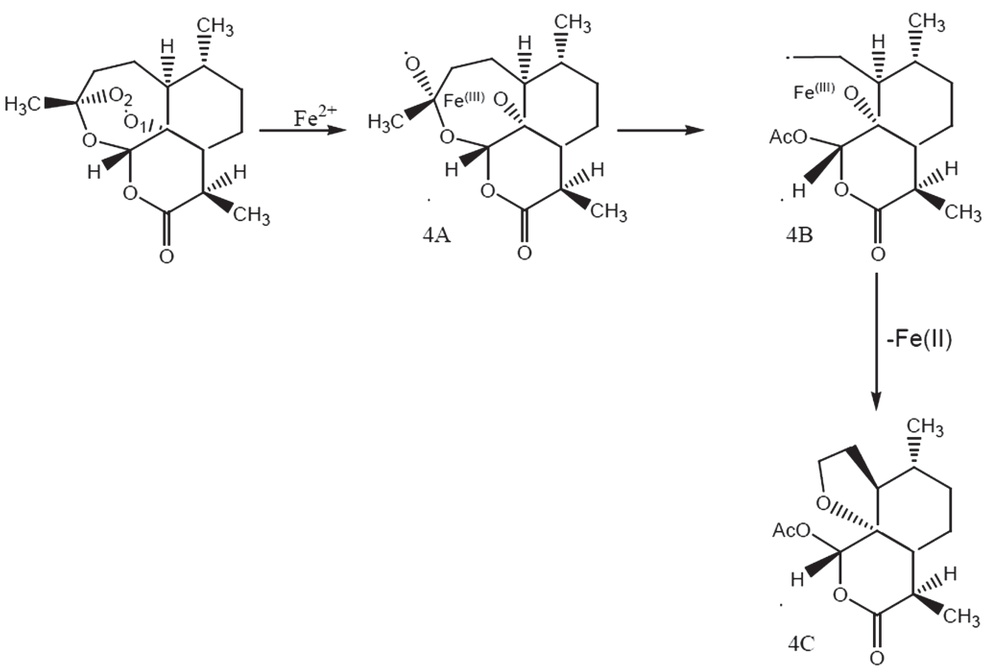


**Fig. 4.** Mechanism of action of artemisinin.

**Table IV.** Geometric Parameters selected by principal component analysis, method and theory levels, the variation of geometrical parameters with respect experimental data (Δ and Δ%).

| | Semi-emperical | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameters[a] | AM1 | ΔAM1 | Δ (%) | PM3 | $\Delta_{PM3}$ | Δ (%) | ZINDO | $\Delta_{\text{ZINDO}}$ | Δ (%) | Experimental |
| Bond length (Å) | | | | | | | | | | |
| O13C12 | 1.416 | 0.037 | 2.683 | 1.403 | 0.024 | 1.740 | 1.392 | 0.013 | 0.942 | 1.379 |
| Bond angle (°) | | | | | | | | | | |
| O1O2C3 | 112.530 | 4.430 | 4.098 | 110.340 | 2.240 | 2.072 | 114.310 | 6.210 | 5.744 | 108.100 |
| Torsion angle (°) | | | | | | | | | | |
| C3O13C12C12a | 11.400 | −13.900 | −54.940 | 2.811 | −22.489 | −88.889 | 17.420 | −7.880 | −31.146 | 25.300 |
| C12C12aO1O2 | 12.050 | −0.650 | −5.118 | 19.940 | 7.240 | 57.007 | 18.110 | 5.410 | 42.598 | 12.700 |
| | *Ab initio*/Hartree-fock/HF | | | | | | | | | |
| Parameters[a] | 6-31G | $\Delta_{6\text{-}31G}$ | Δ% | 6-31G* | $\Delta_{6\text{-}31G*}$ | Δ (%) | 6-31G** | $\Delta_{6.31G**}$ | Δ (%) | Experimental |
| Bond length (Å) | | | | | | | | | | |
| O13C12 | 1.403 | 0.024 | 1.740 | 1.400 | 0.021 | 1.522 | 1.376 | −0.003 | −0.217 | 1.379 |
| Bond angle (°) | | | | | | | | | | |
| O1O2C3 | 108.800 | 0.700 | 0.647 | 106.100 | −2.000 | −1.850 | 109.460 | 1.360 | 1.258 | 108.100 |
| Torsion angle (°) | | | | | | | | | | |
| C3O13C12C12a | 25.320 | 0.020 | 0.079 | 27.432 | 2.132 | 8.426 | 27.400 | 2.100 | 8.300 | 25.300 |
| C12C12aO1O2 | 12.510 | −0.190 | −1.496 | 10.900 | −1.800 | −14.173 | 10.924 | −1.776 | −13.984 | 12.700 |
| | DFT/B3LYP | | | | | | | | | |
| Parameters[a] | 3-21G | $\Delta_{3\text{-}21G}$ | Δ (%) | 3-21G* | $\Delta_{3\text{-}21G*}$ | Δ (%) | 3-21G** | $\Delta_{3.21G**}$ | Δ(%) | Experimental |
| Bond length (Å) | | | | | | | | | | |
| O13C12 | 1.430 | 0.051 | 3.698 | 1.430 | 0.051 | 3.698 | 1.430 | 0.051 | 3.698 | 1.379 |
| Bond angle (°) | | | | | | | | | | |
| 1O2C3 | 105.589 | −2.511 | −2.322 | 105.590 | −2.510 | −2.321 | 105.480 | −2.620 | −2.423 | 108.100 |
| Torsion angle (°) | | | | | | | | | | |
| C3O13C12C12a | 29.059 | 3.759 | 14.857 | 29.060 | 3.760 | 14.861 | 29.080 | 3.780 | 14.940 | 25.300 |
| C12C12aO1O2 | 9.060 | −3.640 | −28.66 | 9.600 | −3.100 | −24.409 | 9.340 | −3.360 | −26.456 | 12.700 |

*Notes*: Δ = Theoretical-experimental; Δ (%) = $\delta \times 100$/experimental.

to the DFT (3-21G, 3-21G* and 3-21G**) the torsion angles formed by atoms C3O13C12C12a, have deviations Δ = 3.759° (Δ% = 14.857), Δ = 3.760° (Δ% = 14.861) and Δ = 3.780° (Δ% = 14.940), respectively. For torsion angles C12C12aO1O2 present deviations Δ = −3.640° (Δ% = −28.660), Δ = −3.100° (Δ% = −24.409) and Δ = −5.410° (Δ% = 42, 598). Already in the method *ab initio*/Hartree-Fock is seen that the three levels (HF/6-31G, HF/6-31G* and HF/6-31G**) show excellent results for the bond length O13C12 with deviations Δ = 0.024 Å (Δ% = 1.740) to HF/6-31G, Δ = 0.021 Å (Δ% = 1.522) to HF/6-31G* and Δ = −0.003 Å (Δ% = −0.217) to HF/6-31G**, as shown in Table IV.

In Table IV, we highlight the comparison *ab initio* method with levels of theory HF/6-31G and HF/6-31G**, where the valence basis set HF/6-31G obtained satisfactory results, but the base HF/6-31G** showed excellent results in sign and magnitude with respect to the bond length and bond angle O13C12 O1O2C3. As the angles of twists or dihedral angle showed a good agreement with the experimental values reported in the literature, showing that the HF/6-31G** basis in this parameter are close to the crystallographic data in the region of artemisinin endoperoxide ring.

### 4.3. HCA Method

A hierarchical cluster analysis (HCA) was used to validate the method and level of theory more appropriate for future calculations of molecular properties with greater accuracy in your results. The analysis was obtained in the form of a dendrogram as shown in Figure 5. It is observed in this figure three methods: Semi-empirical (AM1, PM3 and ZINDO), DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) and *ab initio*/Hartree-Fock (HF/6-31G, HF/6-31G* and HF/6-31G**). We note that the semi-empirical method (AM1, PM3 and ZINDO), has long branches featuring low similarity with the experimental. However, the levels of theory of DFT method (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) have high similarity among them, because they have short branches. Since the levels of theory of *ab initio* method (HF/6-31G, HF/6-31G* and HF/6-31G**) shows high similarity among them and independent of other methods and levels of theory the HF/6-31G** has

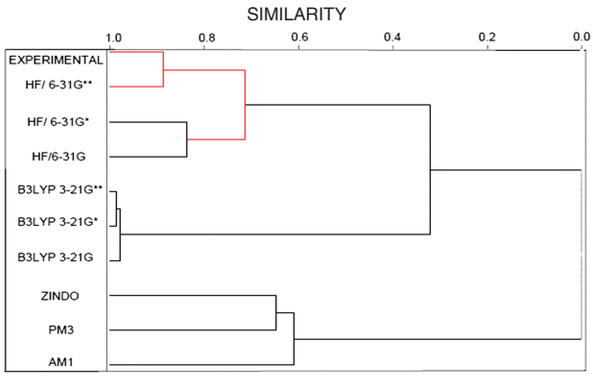


**Fig. 5.** HCA dendogram for 03 (three) methods and 09 (nine) levels of theory, correlating with the experimental geometry parameter of 1,2,13-trioxane ring.

high similarity with the experimental, highlighted by red color branch. Therefore, as the results of combination of *ab initio* method with the level of theory 6-31G** can be used for molecular modeling of the structure of artemisinin and derivatives antimalarials, with mechanical action in the region of the endoperoxide ring.

The results of theoretical methods and experimental were distributed similar to those obtained with PCA. HCA confirms the results of PCA.

## 5. CONCLUSIONS

The multivariate analysis technique PCA and HCA were of great importance, because enable the classification of the methods and levels of theory in three groups: semi-empirical, *ab initio*/Hartree-Fock and DFT methods. Where the total geometrical parameters calculated was 18 (eighteen), among the parameters analyzed and selected the most important in the classification of methods and levels of theory, after analysis of the correlation matrix between the geometric parameters and standard deviation were four (04) parameters geometric related to trioxane ring, and identified by atoms: O13C12 which is the interplanar distance between the two atoms (bond length); O1O2C3 which is the angles length between these three atoms; C3O13C12C12a and C12C12aO1O2 are related dihedral angle or torsion. The results of HCA were similar to those obtained with PCA.

The utilization of semi-empirical calculations, *ab initio*/ Hartree-Fock and DFT was possible because we have a molecule size is not very large, and thus the calculations involving quantum chemistry were developed at different levels of theory and methods. After the analyzes of PCA and HCA found that the Hartree-Fock method with the basis set separate valence 6-31G** is of good quality, and is suitable for molecular modeling studies of the structure of artemisinin, and to describe the conformation of derivatives antimalarial with mechanism of action in the ring region endoperoxide. The *ab initio* HF/6-31G** method can be used for futures calculations of molecular properties which represent a means of chemical information, contained in the molecular structure of the compound on the development of chemical problems studies, pharmacological and toxicological studies in quantitative relationship between structure-activity and structure-property (QSAR, QSPR).

**Acknowledgments:** We gratefully acknowledge the final support of the Brazilian Agency Conselho Nacional de Desenvolvimento Científico e Tecnológico. We also thank the Instituto de Ciências da Saúde of Federal University of Pará for the use of the GaussView and Gaussian softwares. The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity—Network BIONORTE and to the Laboratório de Modelagem e Química Computacional, Federal University of Amapá for the computational support.

## References

1. N. C. Cohen, Guidebook on Molecular Modeling in Drug Design, Academic Press, San Diego **(1996)**.
2. C. M. R. Sanťanna, G. de termos usados no planejamento de fármacos, Recomendações da IUPAC para 1997, Quim. Nova **(2002)**, Vol. 25, pp. 505–12.
3. I. Carvalho, A. D. L. Borges, and L. S. C. Bernardes, *J. Chem. Educ.* 82, 588 **(2005)**.
4. C. G. Wermuth, The Practice of Medicinal Chemistry, 3rd edn., Academic Press, London **(2009)**.
5. R. Todeschini and P. Gramatica, *Perspectives in Drug Discovery and Design* **(1998)**.
6. O. S. de Silva, J. C. M. Corrêa, R. H. Bitencourt, R. W. Monteiro, J. Lameira, S. L. Santos, S. P. G. M. Guilhon, and S. B. D. Brasil, *J. Comput. Theor. Nanosci.* 9, 953 **(2012)**.
7. A. M. Ibrahim and A. El-Din Alaa Gawad, *J. Comput. Theor. Nanosci.* 9, 1120 **(2012)**.
8. M. Srivastava, H. Singh, and P. K. Naik, *Current Research Journal of Biological Sciences* 2, 83 **(2010)**.
9. D. L. Klayman, *Science* 228, 1049 **(1985)**.
10. S. C. Vonwiller and R. K. Haynes, *Tetrahedron. Lett.* 37, 253 **(1996)**.
11. M. Jung, *Curr. Med. Chem.* 1, 35 **(1994)**.
12. J. N. Cumming, P. Ploypradith, and G. H. Posner, *Adv. Pharmacol.* 37, 253 **(1997)**.
13. C. de José Ribamar Carvalho, V. E. J. Ferreira, P. J. Barbosa, S. Da Maycon Lobato, F. C. C. Meneses, M. M. Soeiro, S. De Márcio Farias, O. R. C. De Almeida, C. K. Ventura, C. J. Pinheiro, and R. De Lourdes Andréia Pinheiro, *J. Comput. Theor. Nanosci.* 8, 2193 **(2011)**.
14. F. A. Figueiredo, V. E. J. Ferreira, P. J. Barbosa, C. Da Jorge Williams Macedo, G. G. Da Maria Cristino, S. Da Maycon Lobato, C. J. Pinheiro, and A. T. R. Serra, *Journal of Computational and Theoretical Nanoscience* 8, 1847 **(2011)**.
15. Gaussview, 1.0, Gaussian, Inc., Pittsburgh, PA **(1997)**.
16. M. J. Frisch, Gaussian 98-Revision A.11, Gaussian, Inc., Pittsburgh, PA **(2001)**.
17. J. N. Lisgarten, *Journal of Chemical Crystallography* 28, 539 **(1998)**.
18. F. H. Allen, *Acta Cryst. B* 58, 380 **(2002)**.
19. Infometrix, Inc., Pirouette 3.01, Woodinville, WA **(2001)**.
20. M. M. C. Ferreira, *Journal of the Brazilian Chemical Society* 13, 742 **(2002)**.

RESEARCH ARTICLE

**21.** M. M. C. Ferreira, C. A. Montanari, and A. C. Gaudio, *Química Nova.* 25, 439 **(2002)**.
**22.** R. Franke and A. Gruska, Multivariate data analysis of chemical and biological data, Chemometrics Methods in Molecular Design, edited by H. van de Waterbeemd, VCH, New York **(1996)**, p. 115.
**23.** J. C. Pinheiro, M. M. C. Ferreira, and O. A. S. Romero, *Journal Molecular Structure* (*THEOCHEM*) 572, 35 **(2001)**.
**24.** C. Thomson, M. Cory, and M. Zerner, *International Journal of Quantum Chemistry: Quantum Biology Symposium* 18, 231 **(1991)**.
**25.** A. D. Becke, *J. Chem. Phys.* 98, 5648 **(1993)**.
**26.** C. W. Jefford, *Helvetica Chimica Acta* 79, 1475 **(1996)**.
**27.** G. H. Posner, J. N. Cummings, P. Ploypradith, and C. O. Oh, *J. Am. Chem. Soc.* 117, 5885 **(1995)**.
**28.** S. Tonmunphean, V. Parasuk, and S. Kokpol, *Journal of Molecular Modeling* 7, 26 **(2001)**.
**29.** H. Elhaes, O. Osman, and M. Ibrahim, *J. Comput. Theor. Nanosci.* 9, 901 **(2012)**.
**30.** R. Djemil and D. Khatmi, *J. Comput. Theor. Nanosci.* 9, 1571 **(2012)**.

Received: 12 December 2012. Accepted: 7 January 2013.