Journal of Biophysical Chemistry, 2023, 14, 1-29
https://www.scirp.org/journal/jbpc
ISSN Online: 2153-0378
ISSN Print: 2153-036X

Design of N-11-Azaartemisinins Potentially Active against *Plasmodium falciparum* by Combined Molecular Electrostatic Potential, Ligand-Receptor Interaction and Models Built with Supervised Machine Learning Methods

Jeferson Stiver Oliveira de Castro[1], José Ciríaco Pinheiro[2,3*], Sílvia Simone dos Santos de Morais[4], Heriberto Rodrigues Bitencourt[5], Antonio Florêncio de Figueiredo[1], Marcos Antonio Barros dos Santos[6], Fábio dos Santos Gil[2], Ana Cecília Barbosa Pinheiro[7]

[1]Instituto de Educação, Ciência e Tecnologia do Pará, Castanhal, PA, Brasil
[2]Laboratório de Química Teórica e Computacional, Universidade Federal do Pará, Belém, PA, Brasil
[3]Instituto Amazônia dos Saberes, São Luís, MA, Brasil
[4]Universidade do Estado do Amapá, Macapá, AP, Brasil
[5]Laboratório de Síntese, Universidade Federal do Pará, Belém, PA, Brasil
[6]Universidade do Estado do Pará, Belém, PA, Brasil
[7]Fundação Santa Casa de Misericórdia do Pará, Belém, PA, Brasil
Email: *ciriaco@ufpa.br

How to cite this paper: de Castro, J.S.O., Pinheiro, J.C., dos Santos de Morais, S.S., Bitencourt, H.R., de Figueiredo, A.F., dos Santos, M.A.B., dos Santos Gil, F. and Pinheiro, A.C.B. (2023) Design of N-11-Azaartemisinins Potentially Active against *Plasmodium falciparum* by Combined Molecular Electrostatic Potential, Ligand-Receptor Interaction and Models Built with Supervised Machine Learning Methods. *Journal of Biophysical Chemistry*, **14**, 1-29.
https://doi.org/10.4236/jbpc.2023.141001

Received: January 7, 2023
Accepted: February 25, 2023
Published: February 28, 2023

Abstract

N-11-azaartemisinins potentially active against *Plasmodium falciparum* are designed by combining molecular electrostatic potential (MEP), ligand-receptor interaction, and models built with supervised machine learning methods (PCA, HCA, KNN, SIMCA, and SDA). The optimization of molecular structures was performed using the B3LYP/6-31G* approach. MEP maps and ligand-receptor interactions were used to investigate key structural features required for biological activities and likely interactions between N-11-azaartemisinins and heme, respectively. The supervised machine learning methods allowed the separation of the investigated compounds into two classes: *cha* and *cla*, with the properties ε_{LUMO+1} (one level above lowest unoccupied molecular orbital energy), $d(C_6\text{-}C_5)$ (distance between C_6 and C_5 atoms in ligands), and TSA (total surface area) responsible for the classification. The insights extracted from the investigation developed and the chemical intuition enabled the design of sixteen new N-11-azaartemisinins (prediction set), moreover, models built with supervised machine learning methods were applied to this prediction set. The result of this application showed twelve new promising N-11-azaartemisinins

for synthesis and biological evaluation.

Keywords

Antimalarial Design, MEP, Ligand-Receptor Interaction, Supervised Machine Learning Methods, Models Built with Supervised Machine Learning Methods

1. Introduction

Malaria is a potentially fatal disease caused by parasites of the genus Plasmodium, transmitted by the female Anopheles mosquito. According to the literature [1], in 2020, ~241 million cases of malaria occurred worldwide, compared to ~227 million cases in 2019 and ~231 million cases in 2017; with the majority of these cases being concentrated in the African continent (~228 million or 95% of the incidence), followed by Southeast Asia (~2% of the total cases).

There are four species of human malaria (*P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*). *P. falciparum* is the most prevalent parasite in the African region with ~82.2% of cases, in Southeast Asia (~10.0%), in the Eastern Mediterranean (~4.9%), and in the Western Pacific (~1.8%). *P. vivax* predominates in the Americas region, representing ~1% of malaria cases. *P. falciparum* is the most dangerous, as it multiplies very quickly in the bloodstream and causes severe anemia [1]. In addition, it has cytoadherence properties that favor the sequestration of the parasite in the brain microcirculation, which can lead to death in some patients. In several regions of the world, it has shown resistance to antimalarial compounds of different chemical classes [2].

In Brazil, the Legal Amazon, which comprises the states of Amazonas, Acre, Maranhão, Pará, Rondônia, and Roraima, accounts for ~99.7% of all malaria cases. In order to understand the reasons for this high rate of the disease in the Amazon region, it is necessary to understand how this area was occupied in the past [3].

Although Malaria is a very old infectious disease, it still generates serious public health problems, threatening its control. One of the main reasons for this fact refers to the ability of the *Plasmodium protozoan* to resist the discovered drugs [2].

Currently, efforts continue in the search for medicines [4]-[17] that can help to overcome this disease that still afflicts a large part of humanity. The World Health Organization (WHO) recommends 14 drugs for curative treatment of malaria and 4 drugs for prophylactic treatment, with these treatments being formulated as a single drug or as combinations, and artemisinin and its derivatives appear as essential in these formulations [7].

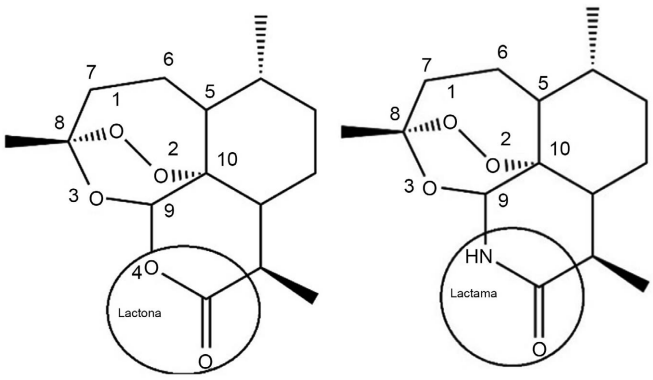
Among the synthesized artemisinin derivatives that show efficiency in combating *P. falciparum*, 11-azaartemisinin and its N-substituted derivatives have attracted the attention of researchers [18] [19] [20] [21], as they have great advantages over artemisinin and other derivatives already used in the treatment of malaria (dihy-

droartemisinin, artesunate and arthemeter). These derivatives are easily prepared from artemisinin (**Figure 1(a)**) and some of them show remarkable thermal stability [18]. Its starting compound (11-azaartemisinin) contains a six-membered lactam unit other than the lactone unit of artemisinin (**Figure 1(b)**).

Chemically, lactam is much more stable in acidic or basic conditions than lactone, due to lower ring deformation and reduced electrophilicity in the carbonyl carbon atom, because of the presence of the adjacent nitrogen atom, which is an electron donor. They are more stable under acidic conditions, such as in the stomach and in the blood stream at pH 7.4, showing superior bioavailability compared to artemisinin [22] [23].

In this article, N-11-azaartemisinin derivatives were investigated with the following approaches: molecular electrostatic potential (MEP) [24] [25] [26] [27], ligand-receptor interaction [28] [29], and supervised machine learning methods [30] [31] [32] [33]. In a first moment, MEP maps were constructed, evaluated and used in the assumption to identify the key features of N-11-azaartemisisnins that are necessary for their activities and to investigate their probable interactions with a receptor through recognition in a biological process. Next, the interactions between N-11-azaartemisinins and the heme receptor were investigated and the interaction energies were correlated with the biological activities of the molecules. Subsequently, supervised machine learning methods were used to investigate the molecular properties that best classify N-11-azaartemisinins into two classes: compounds with high activity (*cha*) and compounds with low activity (*cla*), respectively, giving rise to the pattern recognition models: principal component analysis (PCA); hierarchical cluster analysis (HCA); K-nearest neighbor (KNN); soft independent modeling of class analogy (SIMCA); and stepwise discriminant analysis (SDA).

The information extracted from each step of the investigation, along with the chemical intuition, led to the design of new N-11-azaartemisinins *cha* that were evaluated by the models built with supervised machine learning methods previously designed to establish the most promising compounds for syntheses and biological evaluation.



(a) (b)

Figure 1. 2D structure for artemisinin (a) and 11-azaartemisinin (b).

2. Computational Procedure

The 3D structure of artemisinin encoded in CCDC-691593 [34] was optimized with the method DFT/B3LYP [35] [36]/6-31G [37] and 6-31G* [38] available in the GAUSSIAN 09 software [39], and the computed results were compared to experimental data from the literature [40]. In this procedure, the B3LYP/6-31G method was selected for the subsequent electronic structure calculations step of the research.

With the optimized 3D B3LYP/6-31G structure of artemisinin, the structures of nineteen N-11-azaaretemisinins derivatives from the literature [41] were constructed and optimized with the same theoretical approach. For the optimized derivatives, calculations of MEPs [24] [25] [26], with subsequent obtaining of MEP maps [42], and of the interaction with molecular receptor (ligand-receptor), through the AUTODOCK software [43], were carried-out.

The optimized geometries of N-11-azaartemisinns also allowed calculations of molecular properties used in the chemometric step of the research to identify descriptors capable of separating these compounds into two classes: compounds with high activity (*cha*) and compounds with low activity (*cla*). Chemometric approaches were carried out with artificial intelligence methods (pattern recognition methods) [44] [45] [46] [47] [48], available in the PIROUETTE [49] and MINITAB [50] softwares, respectively.

2.1. About the Investigated N-11-Azaartemisinis

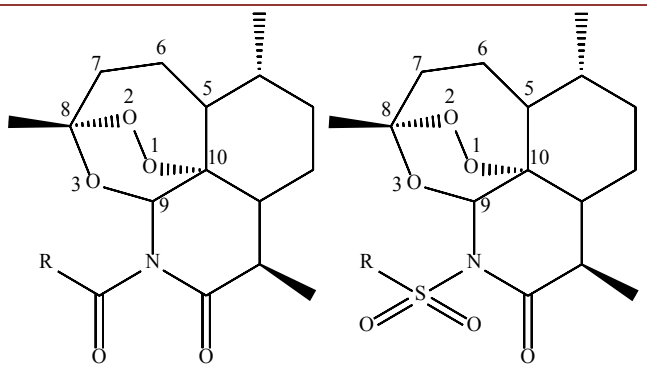
The nineteen N-11-azaartemisinins investigated have biological activity against the K-1 strain of *P. falciparum* resistant to chloroquine, pyrimethamine and cycloguanil. The relative activities were obtained with the expression: relative IC_{50} = IC_{50} (artemisinin)/IC_{50} (N-11-azaartemisinin) [51], where IC_{50} corresponds to 50% of the inhibitory concentration of the compounds. The following hypothesis was considered in the research: N-11-azaartemisinins with relative IC50 ≥ 0.24 correspond to *cha* (**1-8**) and N-11-azaartemisinins with relative IC50 < 0.24 correspond to *cla* (**9-19**).

Table 1 shows the N-11-azaartemisinins (R-substituted N-carbonyl and N-sulfonyl-11-azaartemisinins derivatives) and their respective IC50's. By inspecting the structure-activity relationship it is possible to obtain some indication of how they are correlated. The N-carbonyl derivatives **4** and **5**, which have a Nitro substituent (-NO_2) attached at different positions on the aromatic ring, show higher antimalarial activity than artemisinin. Derivative **6**, an acylurea, is the compound with the highest activity against *P. falciparum* among the N-11-azaartemisinins, about 1.5 times more active than artemisinin. In the N-sulfonyl derivatives, which present monosubstituted aromatic rings (**9**-**14**), it is observed that the activity increases with the increase of the electronegativity of the substituent.

Also in Table 1, in the N-sulfonyl compounds (**15** and **16**), whose ring is disubstituted, derivative **15**, which has two electronegative substituents (Cl and

NO_2), shows activity around 11 times greater than compound **16**. Derivatives **12** and **13**, whose only structural difference is in the position of the Nitro group attached to the ring, the change in the position of this group did not significantly affect the antimalarial activity.

Table 1. N-Carbonyl and N-sulfonyl-11-azaartemisinin derivatives of the training set with their substituents and respective IC_{50} (ng/mL) and relative IC_{50}[a].



(1 to 6) (7 to 19)

Compounds	R	IC_{50}	Relative IC_{50}	Antimalarial activity
1	CH_3CH_2—	1.0	0.90	*cha*
2	$CH_3(CH_2)_2$—	2.0	0.45	*cha*
3	$CH_3(CH_2)_4$—	2.0	0.45	*cha*
4	4'-$O_2NC_6H_4$—	0.6	1.5	*cha*
5	3'-$O_2NC_6H_4$—	0.6	1.5	*cha*
6	4'-$O_2NC_6H_4NH$—	0.4	2.25	*cha*
7	CH_3—	3.2	0.28	*cha*
8	CH_3CH_2—	3.7	0.24	*cha*
9	— 4'-FC_6H_4—	8.0	0.11	*cla*
10	4'-ClC_6H_4—	17	0.05	*cla*
11	4'-$CH_3C_6H_4$—	37	0.02	*cla*
12	4'-$O_2NC_6H_4$—	9.1	0.10	*cla*
13	2'-$O_2NC_6H_4$—	10	0.09	*cla*
14	3'-N÷CC_6H_4—	50	0.02	*cla*
15	4'-Cl-3'-$O_2NC_6H_3$—	3.8	0.23	*cla*
16	3',4'-$(CH_3O)_2C_6H_3$—	40	0.02	*cla*
17	4'-$CH_3SO_2C_6H_4$—	13	0.07	*cla*
18	4'-C_6H_5-C_6H_4—	45	0.02	*cla*
19	5'-Cl-2'-tienil	40	0.02	*cla*

[a]Relative IC_{50} = IC_{50} of artemisinin/IC_{50} of derivative, where IC_{50} corresponds to 50% of the inhibitory concentration. The following hypothesis was considered: where relative $IC_{50} \geq 0.24$ corresponds to compounds with high activity (*cha*) and relative $IC_{50} < 0.24$ corresponds to compounds with low activity (*cla*).

Another interesting aspect to notice in the analysis of the structure-activity relationship of the N-11-azaartemisinins in Table 1 is the comparison between derivatives **4** and **12**. Both have the Nitro group as a substituent on the aromatic ring in the same position. However, a biological activity is observed for compound **4** (N-carbonyl) about 23 times greater than that of compound **12** (N-sulfonyl). Derivatives **1** and **8** have the same number of carbon atoms in their chains, however, it is noted that compound **1**, an N-carbonyl compound, is 4 times more active than derivative **8**, an N-sulfonyl compound.

By comparing the structure-activity relationship of the N-11-azaartemisinins in Table 1, it is noted that the N-carbonyl-11-azaartemisinin derivatives are much more biologically active than the N-sulfonyl derivatives, indicating a possible contribution of carbonyl in the N-carbonyl derivatives for antimalarial activity.

2.2. The Biological Recognition Process through Molecular Electrostatic Potential

In the biological recognition process, the receptor molecule "recognizes" the approach of another molecule with key features to promote mutual interaction. It is understood that such a recognition should typically occur when the drug (substrate) and receptor (enzyme) are in relatively considerable separation, and precede the formation of any covalent bond [26].

Molecular electrostatic potential has the physical meaning of how the molecule is perceived by its surroundings. Searching through this approach, the main key features that determine the occurrence or not of a recognition is a natural behavior in research associated with this tool. Furthermore, the aforementioned tool proved to be an effective means of analyzing and elucidating recognition processes [26].

MEP calculations were carried out using the approach reported by Politzer & Murray (Equation (1)) [52] through electron density and the MEP maps obtained with the MOLEKEL software [42].

$$V(r) = \sum_{A} \frac{Z_A}{|R_A - r|} - \int \frac{(r')dr'}{|r' - r|} \tag{1}$$

where Z_A is the charge on nucleus A, located at R_A, and $\rho(r)$ is the electronic density of the atom or molecule.

2.3. Ligand-Receptor Interaction

The study of ligand-receptor interaction (molecular docking) investigates in detail how a small molecule (ligand) interacts in the binding region with a macromolecule (receptor), generating important information for understanding the biological activity of a given drug. In addition, it makes possible to visualize the way in which the two molecules interact and to quantify this interaction. The process begins when the structures are not yet connected, the ligand seeks its best position in the receptor's active site, generating a certain number of con-

formational possibilities in the search for the best ligand-receptor anchoring [29] [53] [54].

The ligand-receptor molecular recognition involves multiple steps of conformational accommodation, resulting in the most favorable mode of interaction enthalpic and entropically. Equation (2) can estimate these processes, through the Gibbs binding free energy (ΔG); which is related to the inhibition constant (K_i), determined experimentally. In the Equation (2), ΔH is the enthalpy change, T is the absolute temperature, ΔS is the entropy change, and R is the universal gas constant.

$$\Delta G = \Delta H - T\Delta S = RTlnK_i \quad (2)$$

In the molecular docking study, the heme [55] [56] was used as a receptor for artemisinin, 11-azaartemisinin, and N-11 azaartemisinins, and its 3D structure was taken from the Protein Data Bank (PDB) RCSB, 1A6M [56].

2.4. Supervised Machine Learning Methods

To identify the molecular descriptors capable of separating N-11-azaartemisinins into the *cha* and *cla* classes, respectively, supervised machine learning methods were used: Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), K-Nearest Neighbor method, (KNN), Soft Independent Modeling of Class Analogy (SIMCA) method [44] [45] [46], and Stepwise Discriminant Analysis (SDA) [45] [46] [47] [48]. For a description of these methods, consult the cited references [44] [45] [46] [47] [48].

2.5. Calculated Molecular Properties

The molecular properties calculated and used as descriptors were steric properties: bond lengths, bond angles, and torsion angles; electronic properties: total energy (TE), ε_{HOMO} (highest occupied molecular orbital energy), ε_{HOMO-1} (one level below to highest occupied molecular orbital energy), ε_{HOMO-2} (two level below to highest occupied molecular orbital energy), ε_{HOMO-3} (three levels below to highest occupied molecular orbital energy), absolute value of ε_{LUMO} (lowest unoccupied molecular orbital energy), absolute value of ε_{LUMO+1} (one level above lowest unoccupied molecular orbital), absolute value of ε_{LUMO+2} (two levels above lowest unoccupied one), absolute value of ε_{LUMO+3} (three levels above lowest unoccupied molecular orbital); Mulliken's electronegativity ($\chi = 1/2(\varepsilon_{HOMO} - \varepsilon_{LUMO})$), molecular hardness ($\eta$ = I – AE/2), where I is ionization potential and AE electron affinity; molecular softness ($1/\eta$), defined as the inverse of molecular hardness; dipole moment (μ), energy GAP (GAP = $\varepsilon_{HOMO} - \varepsilon_{LUMO}$), and atomic charges on an Nth atom (Q_N); physicochemical properties: molecular polarizability (POL), Molecular refractivity (MR), hydration energy (HE), partition coefficient octanol-water (log P), molecular mass, total surface area (TSA), and molecular volume (VOL).

Additionally, to represent different sources of chemical information about size, symmetry, and distribution of atoms in the molecule, holistic properties

were also calculated and considered as descriptors.

All properties were considered in the most stable conformation of each molecule and the descriptors were computed to given information about the influence of steric, electronic, hydrophobic, and hydrophilic features on the antimalarial activity of the studied N-11-azaartemisinins.

The steric and electronic, physicochemical, and holistic properties were performed with the GAUSSIAN 09 [39], CHEMPLUS [57], and DRAGON [58] softwares, respectively.

3. Results and Discussion

3.1. Theoretical Geometry

Table 2 shows the geometry obtained with the B3LYP/6-31G, B3LYP/6-31G*, and experimental methods, and the discrepancies between the theoretical and experimental values. It can be seen that the bond lengths when compared to the literature data are very well described with both methods, with the B3LYP/6 31G* method (this work) showing a slightly better performance in the description of this geometric parameter, especially considering the importance of the distance between O_1 and O_2 in the antimalarial activity of artemisinin [59] [60] [61] [62].

3.2. Molecular Electrostatic Potential Maps for Artemisinin, 11-Azaartemisinin and N-11-Azaartemisinins

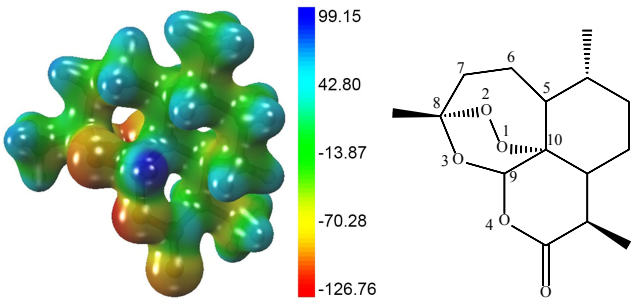
The literature reports the pharmacophore 1,2,4-trioxane as crucial to artemisinin activity and its derivatives [2] [59] [60] [61] [63]. **Figure 2** shows the 2D structures and the molecular MEP maps for artemisinin (**Figure 2(a)**) and 11-azaartemisinin (**Figure 2(b)**), respectively. In this figure, 2D structure, it can be noted that the replacement of O_4 (artemisinin, IC_{50} = 0.903 ng/mL [49] or 0.903/0.904 = 1) by N (11-azaartemisinin, IC = 2.656 ng/mL [19] or 2.656/0.903 = 2.94) maintains the endoperoxide link necessary for antimalarial activity. Also in this figure, the MEP maps of the compounds show considerable similarities, with decrease in electron density in the endoperoxide region, evidenced by the decrease in the biological activity of 11-azaartemisinin; with artemisinin exhibiting minimum and maximum electrostatic potential values of −126.76 and +99.15 kcal/mol, respectively, while in 11-azaartemisinin these values are around −130.52 and 120.48 kcal/mol.

All N-11-azaartemisinins shown in **Table 1** (see Section 2.1. About the investigated N-11-azaartemisinis) also exhibit the endoperoxide linkage necessary for antimalarial activity. **Figure 3** shows the MEP maps for the N-11-azaartemisinins of the training set obtained by the inclusion of substituents in the N atom of the lactam function (1-19). As one can see in this figure, the MEP maps are similar to artemisinin and 11-azaartemisinin in the 1, 2, 4 trioxane ring region, with the electron density of some molecules more concentrated in this region, indicating greater biological activity. N-11-azaartemisinins are susceptible to electrophilic attack in the most negative MEP region, –130.52 to –114.21 kcal/mol, respectively.

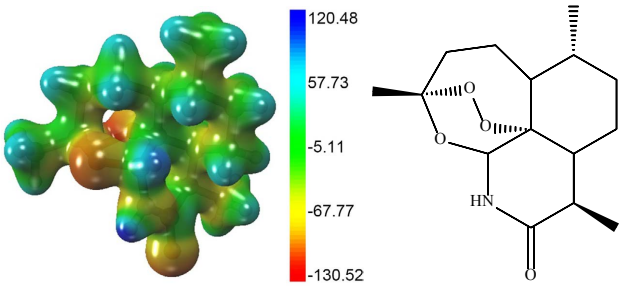
Table 2. Theoretical and experimental geometry of the 1,2,4-trioxane ring for artemisinin.

Geometry[a]	Methods Theoretical			Discrepancy (Δ)[c]	
	B3LYP/6-31G	B3LYP/6-31G* (this work)	Exp[b]	Δ_1	Δ_2
Bond length (A)					
C_{10}-O_1	1.499	1.455	1.461	3.8×10^{-2}	-6×10^{-3}
O_1-O_2	1.524	1.460	1.469	5.5×10^{-2}	-9×10^{-3}
O_2-C_8	1.452	1.414	1.416	3.6×10^{-2}	-4×10^{-3}
C_8-O_3	1.473	1.441	1.445	2.8×10^{-2}	-4×10^{-3}
O_3-C_9	1.425	1.396	1.379	4.6×10^{-2}	1.7×10^{-2}
C_9-C_{10}	1.538	1.539	1.523	1.5×10^{-2}	1.6×10^{-2}
C_{10}-C_5	1.533	1.555	1.534	1.9×10^{-2}	2.1×10^{-2}
C_5-C_6	1.552	1.548	1.527	2.5×10^{-2}	2.1×10^{-2}
C_6-C_7	1.544	1.540	1.520	2.4×10^{-2}	2.0×10^{-2}
C_7-C_8	1.544	1.547	1.510	3.4×10^{-2}	3.7×10^{-2}
Bon angle (˚)					
C_{10}-O_1-O_2	111.407	111.609	111.200	2.07×10^{-1}	4.09×10^{-1}
O_1-O_2-C_8	107.304	108.261	108.100	-7.96×10^{-1}	1.61×10^{-1}
O_2-C_8-O_3	107.738	108.487	106.600	1.14×10^{0}	1.89×10^{0}
C_8-O_3-C_9	114.996	114.069	114.200	7.96×10^{-1}	-1.31×10^{-1}
O_3-C_9-C_{10}	113.641	113.264	114.500	-8.59×10^{-1}	-1.24×10^{0}
C_9-C_{10}-O_1	111.751	111.323	111.700	5.1×10^{-2}	-3.77×10^{-1}
Dihedral angle (˚)					
C_{10}-O_1-O_2-C_8	46.883	47.915	47.800	-9.17×10^{-1}	9.3×10^{-2}
O_1-O_2-C_8-O_3	−73.464	−73.450	−75.500	2.04×10^{0}	1.55×10^{0}
O_2-C_8-O_3-C_9	34.986	32.892	36.00	-1.01×10^{0}	-3.11×10^{0}
C_8-O_3-C_9-C_{10}	26.252	27.351	25.300	9.52×10^{-1}	2.04×10^{0}
O_3-C_9-C_{10}-O_2	−51.202	−51.180	−51.300	9.8×10^{-2}	1.20×10^{-1}
C_9-C_{10}-O_1-O_2	12.765	11.671	12.700	6.5×10^{-2}	-1.03×10^{0}

[a]The atoms were labeled according to **Figure 2**. [b]Ref [40]. [c]Δ = Theoretical (6-31G and 6-31G* basis sets, respectively) – Experimental data.



(a)



(b)

Figure 2. 2D structure and MEP (kcal/mol) maps for artemisinin (a) and 11-azaartemisinin (b), respectively.

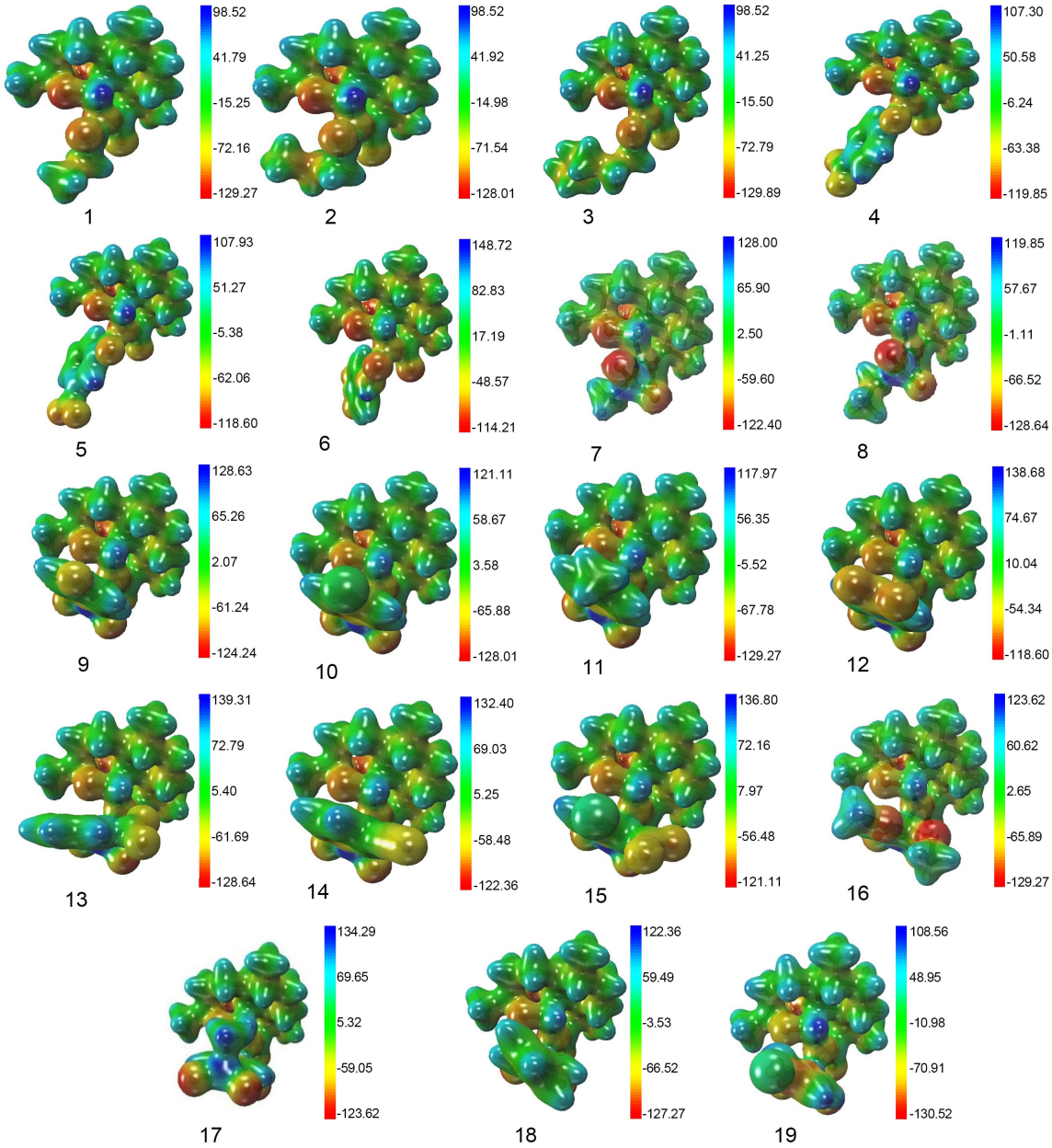
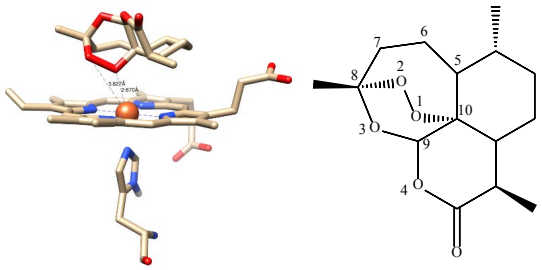


Figure 3. MEP (kcal/mol) maps for N-11-azaartemisinins (training set).

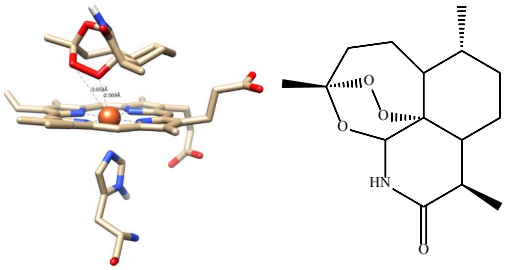
3.3. Interaction between Artemisinin, 11-Azaartemisinin, N-11-Azaartemisinins and the Heme Receptor

Since all N-11-azaartemisinins were derived from 11-azaartemisinin, a previous study of the interaction between artemisinin and 11-azaartemisinin with the biological receptor was performed. Figure 4 shows the 2D structures and the respective interactions of artemisinin and 11-azaartemisinin with heme. According to this figure, the ligand-heme interaction energies for these molecules correspond to −5.24 and −6.37 kcal/mol, respectively. The d(Fe-O_1) and d(Fe-O_2) distances for artemisinin are equal to 2.670 and 3.827 Å, respectively, with d(Fe-O_1) very close to the value 2.7 reported by other studies [64] [65], reinforcing the perspective of O_1 preferential binding of the trioxane ring to Fe^{2+} heme. For 11-azaartemisinin, these distances are 2.569 and 3.659 Å, respectively; with lower d(Fe-O_1) value for 11-azaartemisinin, reflecting its higher interaction energy and lower biological activity.

In Figure 5, the interactions between N-11-azaartemisinins (1-19) and heme are shown. As can be seen, also in this figure, O_1 is preferentially oriented to Fe^{2+}-heme in the totality of N-11-azaartemisinins. Table 3 shows the values of the d(Fe-O_1) and d(Fe-O_2) distances and the ligand-heme interaction energies. From this table, it can be seen that d(Fe-O_1) is always smaller than d(Fe-O_2) in all interactions and that, in general, d(Fe-O_1) has higher values for N-11-azaartemesinins *cha*, when compared with the corresponding *cla.* Also, according to Table 3, it is possible to associate, in general, higher values of d(Fe-O_1) and lower values for the interaction energy of N-11-azaartemisinins *cha* to biological processes occurring with increased biological activity of these compounds.



(a)



(b)

Figure 4. 2D structures and interactions (kcal/mol) of artemisinin (a) and 11-azaartemisinin (b) with heme receptor, respectively. Colors associated with atoms: golden (C), blue (N), red (O), and orange (Fe).

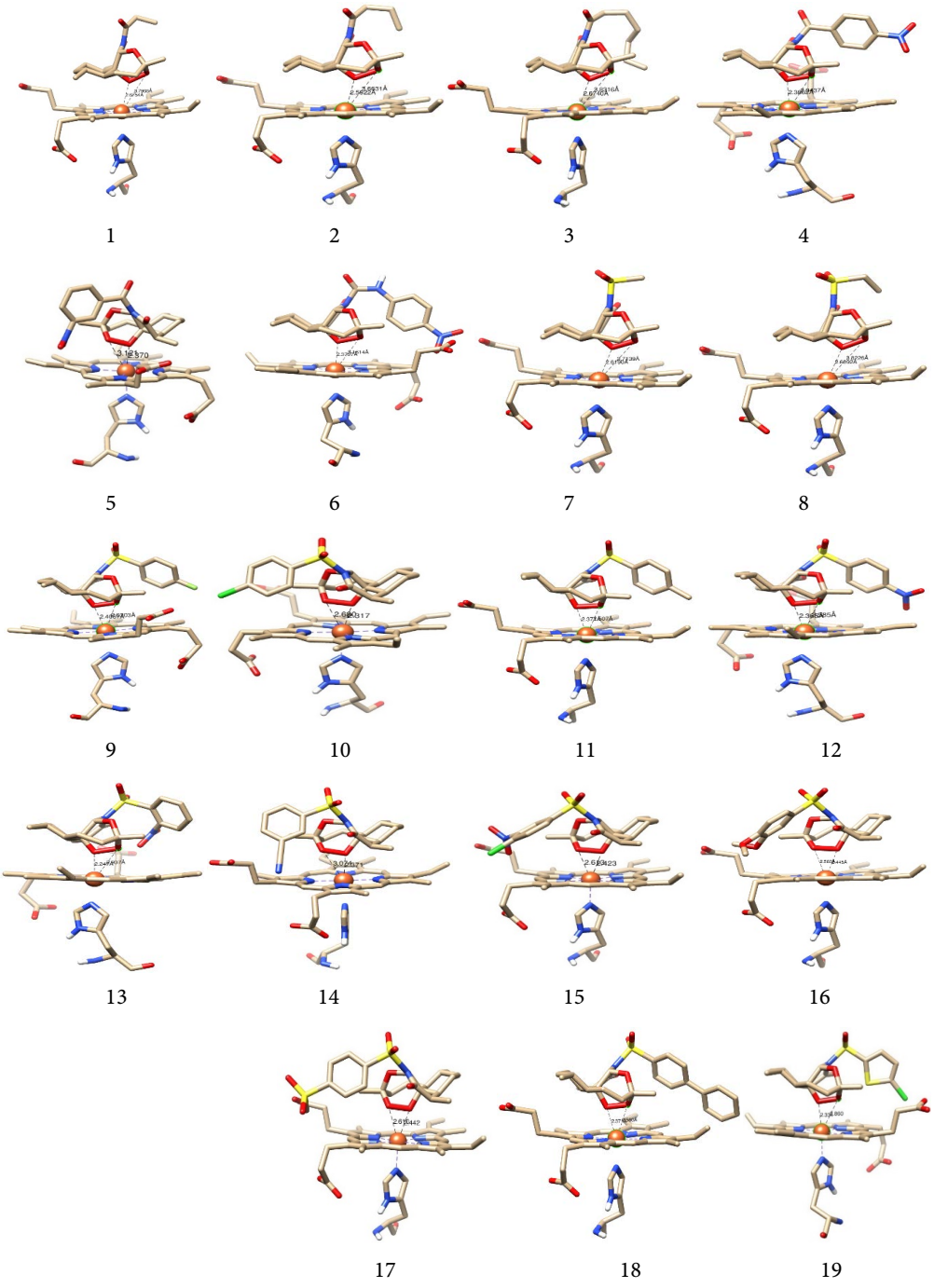


Figure 5. Interactions (kcal/mol) between N-11-azaartemisinins and the heme receptor (training set). Colors associated with atoms: golden (C), blue (N), red (O), light green (F), yellow (S), intense green (Cl), and orange (Fe).

3.4. Supervised Machine Learning Methods

With the molecular properties calculated, a matrix of descriptors of dimension 19 × 120 (number of N-11-azaartemisinins (lines) *versus* molecular properties (columns)) was constructed and the exploratory analyzes of the autoscaled data were performed with the supervised learning methods PCA and HCA. Data

Table 3. Values of d(Fe-O_1), d(Fe-O_2), absolute value of the difference between d(Fe-O_1) and d(Fe-O_2) (Δ) and interaction energy of the ligand-heme complex for the training set.

Compouds	d(Fe-O_1)[a]	d(Fe-O_2)[a]	Δ[b]	Energy interaction[c]
1	2.675	3.797	1.05	−4.90
2	2.592	3.663	1.07	−4.79
3	2.674	3.832	1.16	−4.71
4	2.392	3.007	0.62	−6.17
5	2.370	3.121	0.75	−5.52
6	2.380	3.161	0.78	−5.80
7	2.619	3.714	1.10	−5.09
8	2.609	3.823	1.21	−5.96
9	2.407	2.630	0.22	−6.51
10	2.317	2.680	0.36	−6.99
11	2.371	2.507	0.14	−6.85
12	2.365	2.585	0.22	−6.63
13	2.247	2.937	0.69	−6.20
14	2.371	3.074	0.70	−5.41
15	2.423	2.613	0.19	−5.04
16	2.445	2.566	0.12	−6.49
17	2.442	2.610	0.17	−5.04
18	2.371	2.560	0.19	−5.83
19	2.331	2.860	0.53	−6.85

[a]Values in Å. [b] $\Delta = \left| d\left(\text{Fe-O}_1\right) - d\left(\text{Fe-O}_2\right) \right|$. [c]Values in kcal/mol.

compression was performed considering the Pearson r correlation coefficient > 0.82 and, consequently, low correlation between two descriptors, with one of them being randomly excluded from the matrix for theoretically describing the same property. With this procedure, a considerable reduction in the size of the initial matrix was verified and, then, the PCA exploratory method was applied.

3.4.1. PCA Model

After several attempts, the 11-N-azaartemisinins were separated, with the help of three variables, into two classes: *cha* class and *cla* class. **Table 4** shows the descriptor matrix selected in the PCA for the training set and respective values of the biological activities and the correlation matrix between the descriptors. As can be seen, the correlation between the descriptors is less than 0.642.

The molecular properties responsible for the separation of the *cha* and *cla* classes were: $\varepsilon_{\text{LUMO+1}}$, TSA, and the distance between carbon 6 and carbon 5 atoms, d(C_6-C_5) of the molecules, respectively. The separation is done at PC1 and explains 89% of the total information, with PC1 and PC2 retaining 70% and 19% of this information, respectively. According to **Table 4**, in general, N-11-azaarrtemisinins of the *cha* class exhibit lower values for $\varepsilon_{\text{LUMO+1}}$ combined with lower values for

$d(C_6\text{-}C_5)$ and TSA compared to those of the *cla* class.

Figure 6 shows the plots of scores (a) and loadings (b) and the dendrogram (c) obtained with PCA and HCA, respectively, for N-11-azaartemisinins (training set). In Figure 6, the N-11-azaartemisinins of the *cha* class (**1-8**) are located to the left, Figure 6(a), due to the displacement in this direction produced by the properties ε_{LUMO+1} and TSA, while the N-11 azaartemisinins of the *cla* class (**9-19**) are shifted to the right of the same figure due to the action of the property $d(C_5\text{-}C_6)$, Figure 6(b).

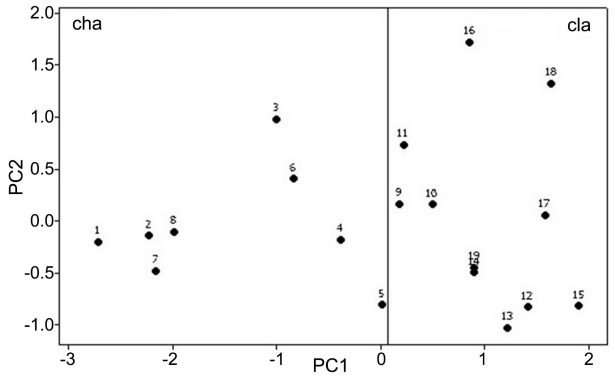
Table 5 shows the loadings matrix for the selected descriptors. According to this table, it is possible to classify new N-11-azaartemisinins with *cha* using Equation (3), reaffirming the information previously extracted from Table 4, *i.e.*, new N-11-azaartemisinins can be classified as *cha* combining lower values for ε_{LUMO+1}, $d(C_6\text{-}C_5)$, and TSA, respectively.

$$PC1 = 0.543\varepsilon_{LUMO+1} + 0.614d\left(C_6\text{-}C_5\right) + 0.573TSA \quad (3)$$

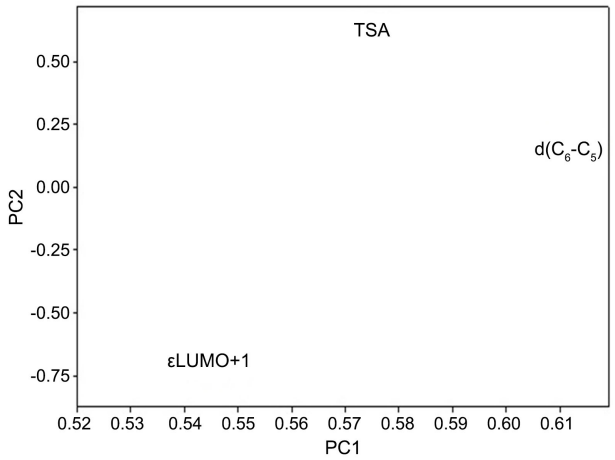
Table 4. Matrix of descriptors selected in the PCA of the training set with their respective values.

Compounds	ε_{LUMO+1}[a]	$d(C_6\text{-}C_5)$[b]	TSA[c]	Relative IC_{50}	Antimalarial activity
1	19.15	1.550	526.1	0.90	*cha*
2	20.44	1.551	529.2	0.45	*cha*
3	20.44	1.551	610.1	0.45	*cha*
4	39.01	1.551	604.4	1.50	*cha*
5	49.52	1.551	602.9	1.50	*cha*
6	30.90	1.5505	618.4	2.25	*cha*
7	23.18	1.551	513.3	0.28	*cha*
8	22.57	1.551	541.6	0.24	*cha*
9	32.45	1.553	585.3	0.11	*cla*
10	36.75	1.553	605.2	0.05	*cla*
11	27.55	1.553	608.2	0.02	*cla*
12	56.10	1.553	613.5	0.10	*cla*
13	55.13	1.553	592.9	0.09	*cla*
14	48.95	1.553	611.0	0.02	*cla*
15	60.22	1.553	630.9	0.24	*cla*
16	23.55	1.553	663.4	0.02	*cla*
17	48.86	1.553	649.1	0.07	*cla*
18	35.99	1.553	687.0	0.02	*cla*
19	46.26	1.553	599.3	0.02	*cla*
Descriptor					
ε_{LUMO+1}		0.565	0.436		
$d(C_6\text{-}C_5)$			0.641		

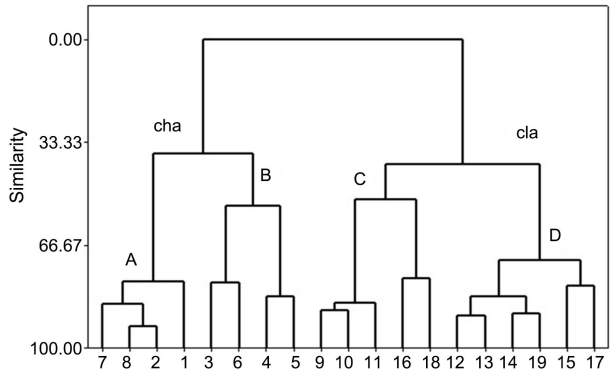
[a]Values in kcal/mol. [b]Values in Å. [c]Values in Å^2.



(a)



(b)



(c)

Figure 6. Scores (a) and loading (b) plots and dendrogram (c) for N-11-azaartemisinins (training set).

Table 5. Variable matrix for PC1, PC2 and PC3.

Variable	PC1	PC2	PC3
ε_{LUMO+1}	0.543	−0.785	−0.298
$d(C_6\text{-}C_5)$	0.614	0.129	0.779
TSA	0.573	0.606	−0.552
Variance	70%	19%	11%

3.4.2. HCA Model

Figure 6(c) shows the dendrogram obtained with the HCA, incremental method, for the training set N-11-azaartemisinins. According to this figure, N-11-azaartemisinins are distributed in two main clusters. The cluster on the left, corresponding to N-11-azaartemisinins with *cha*, which is formed by two clusters, namely: A and B. Cluster A consists of N-carbonyl and N-sulfonyl derivatives with alkyl substituents of one to three carbons in the R position (Table 1). For these N-11-azaartemisinins, the increasing of the number of carbons in the chain produced a decrease in the biological activity. In *cha* class, these N-11-azaartemisinins are not among the most efficient compounds against malaria and have lower TSA and $d(C_6\text{-}C_5)$ and lower values for ε_{LUMO+1}. Cluster B is formed by N-11 azaartemisinins among the most active compounds with *cha*. Structurally, all azaartemisinins have an aromatic ring (4, 5, and 6) as substituents in R, except for derivative 3, which has an alkyl chain with 5 carbons, being the largest aliphatic chain among all the N-11 azaartemisinins investigated. The N-11-azaartemisinins 4 and 5 are N-carbonyl whose the only difference between them is in the position of the nitro group attached to the aromatic ring in R. In the *cha* class, they show the highest TSA values and lower than those with *cla*.

The cluster on the right in Figure 6(c) is also made up of two smaller clusters, C and D. All the N-11-azaartemisinins in cluster C (9, 10 and 11) are N-sulfonyls that have a monosubstituted aromatic ring in the R position 4, differing only in the substituents used, namely: Fluorine (F), Chlorine (Cl) and methyl (Me), respectively. In comparison with one of the N-11-azaartemisinin with *cha* (4), an N-carbonyl with the same substitution in R by a monosubstituted aromatic ring in position 4, but with a different substituent (nitro), shows a great increase in the biological activity, confirming that N-carbonyls present better results as antimalarial and suggesting that the presence of the nitro group also contributes to this better performance. In cluster C, all N-11-azaartemisinins have the highest TSA in the training set.

In cluster D, which contains N-11-azaartemisinins with lower biological potency, it can be noted that all compounds are N-sulfonyl derivatives with monosubstitution (12, 13, and 14) and disubstitution (15 and 16) in the aromatic ring, two rings isolates (18) and 5-Chloro-thienyl (19) in R. In this cluster, N-11-azaartemisinins show the highest values for ε_{LUMO+1} and $d(C_6\text{-}C_5)$ of the training set.

Finally, it can be reported that the HCA results confirm the PCA.

3.4.3. KNN Model

The KNN study was developed with the variables ε_{LUMO+1}, $d(C_6\text{-}C_5)$ and TSA selected by the PCA and HCA for the N-11 azaartemisins of the training set. Table 6 shows the results of classification using this method. The correct information was 100% for 1 KNN, 2 KNN, 3 KNN and 4 KNN, that is, all derivatives were correctly classified within the predicted classes, indicating that the classifications (1 KNN, 2 KNN, 3 KNN and 4 KNN) obtained with the selected descriptors provide good predictive capacity. The KNN model was built with the four nearest neighbors [66].

Table 6. Classification of the training set with the KNN method.

		Incorrectly classified compounds			
Class	Number of Compounds	1 KNN	2 KNN	3 KNN	4 KNN
cha	8	0	0	0	0
cla	11	0	0	0	0
Total	19	0	0	0	0
Percentage of correct information		100	100	100	100

3.4.4. SIMCA Model

The SIMCA model was built with the selected descriptors in PCA, HCA and KNN and three PCs. **Table 7** shows the classification matrix obtained by the SIMCA method. The correct information in the classification was 100%. Furthermore, the SIMCA model shows good discriminating power, evidenced by high distances and high residuals between classes, **Figure 7**.

3.4.5. SDA Model

The SDA approach provided the achievement of discriminant functions with the variables $\varepsilon_{\text{LUMO+1}}$, $d(\text{C}_6\text{-C}_5)$ and TSA showing significant contributions in the classification methodology of N-11-azaartemisinins. Equations (4a) and (4b) show these functions for the classes with *cha* and *cla*, respectively:

$$\text{cha class}: -0.063\varepsilon_{\text{LUMO+1}} - 19.401d\left(\text{C}_6\text{-C}_5\right) + 0.749\text{TSA} - 10.519 \quad (4a)$$

$$\text{cla class}: +0.046\varepsilon_{\text{LUMO+1}} + 14.110d\left(C_6\text{-C}_5\right) - 0.545\text{TSA} - 5.564 \quad (4b)$$

With the discriminant functions and the values of the selected variables for the N-11-azaartemisinins, a classification matrix was obtained for the training set. **Table 8** shows the SDA classification matrix with the respective percentage of correct answers.

The reliability of the SDA model was evaluated through the cross-validation test, the leave-one-out technique, which consisted of omitting an N-11-azaaratemisinin from the data set and building the discriminant functions with the other samples. Subsequently, this omitted N-11-azaartemisinin was classified with the generated discriminant functions, and the procedure was repeated until the last N-11 azaartemisinin in the data set was omitted. **Table 9** shows the SDA classification matrix using the cross-validation procedure.

With the SDA model, an allocation rule was established when new N-11-azaartemisinins with *cha* or *cla* investigated: 1) initially calculate, for the new molecule, the value of the properties used in the construction of the model, 2) consider the autoscaled values of these molecular properties (descriptors) in the discriminant functions, Equations 4(a) and 4(b), and 3) verify which discriminant function has the highest value. The new N-11-azaartemisinin is ch*a* if it is related to the discriminant functions of the *cha* class and vice versa.

Once the properties selected in the study with supervised machine learning methods (PCA, HCA, KNN, SIMCA and SDA) proved to be more important in the description of the antimalarial activity of N-11-azaartemisinins, some con-

siderations that lead to the understanding of the behavior of the class *cha* become relevant. In this context, in Table 4 it can be seen that, in general, the N-11-azaartemisinins of the *cha* class have ε_{LUMO+1} values lower than those of the N-11-azaartemisinins of the *cla* class. Conditioning to the LUMO+1 of the N-11-azaartemisinins the possibility of interaction with the heme receptor orbital that presents the greatest contribution of electron density [67], an electron transfer reaction will be important in the mechanism of action of ligands (*cha* class) in the biological process.

The bond length is a geometric parameter related to the steric availability of the atoms involved in the bond [68]. According to Table 4, the N-11-azaartemisinins of the *cha* class have $d(C_6\text{-}C_5)$ values lower than those of the *cla* class. This may be indicative of the importance of the steric distribution of atoms of the N-11-azaartemisinins of the *cha* class in their possible mechanism of action in the biological process.

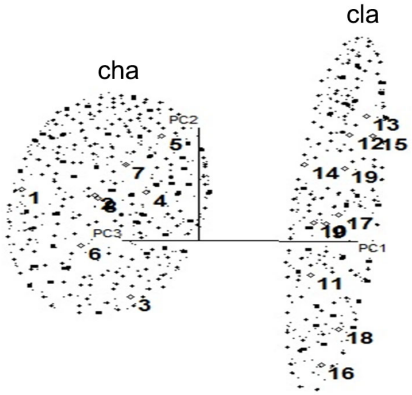


Figure 7. 3D score plot showing boundaries classes for N-11 azaartemisinins (training set)

Table 7. Classification matrix obtained with the SIMCA method.

Class	Number of Compounds	SIMCA Classification
cha	08	08
cla	11	11
Total	19	19
Percentage of correct information		100

Table 8. Classification matrix of the training set compounds with SDA.

Class	Number of compounds	Correct classification	
		cha	*cla*
cha	8	8	0
cla	11	0	11
Total	19	8	11
Percentage of correct information		100	100

Table 9. Classification matrix using SDA with Cross-Validation.

Class	Number of compounds	Correct classification	
		cha	*cla*
cha	8	8	0
cla	11	0	11
Total	19	8	11
Percentage of correct information		100	100

The TSA property determines the solvent accessible area of a molecule. **Table 4** shows that N-11-azaartemisinins belonging to the *cha* class exhibit lower values for TSA when compared to those of the *cla* class. This is an indication that in biological processes involving N-11-azaartemisinins and heme, hydrophilic interactions may be important in the mechanism of action of these molecules.

3.4.6. Combining MEP, Ligand-Receptor Interaction and Supervised Machine Learning Methods in the Design of New N-11-Azaartemisinins

The insights gained from studies of MEP, ligand-receptor interaction and supervised machine learning methods, together with chemical intuition, enabled the design of sixteen new N-11-azaartemisinins (prediction set) potentially active against *P. falciparum*. **Table 10** shows the N-11-azaartemisinins of the prediction set.

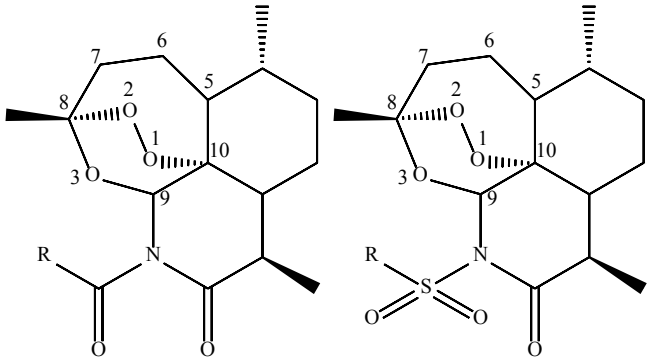
The models obtained with supervised machine learning methods were applied to the prediction set and **Table 11** shows the results of this application. In this table, N-11-azaartemisinins **20-26** and **31-35** were considered as *cha* by all models, while N-11-azaartemisinins 27 and 30 were reported as *cla* by these models. N-11-azaartemisinins **28** and **29** were classified as *cha* by the KNN model and *cla* by the PCA, HCA and SDA models, with no classification by the SIMCA model.

The results of the application of models built with supervised machine learning methods to the prediction set show twelve N-11-azaartemisinins: **20, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34,** and **35** potentially in the *cha* class as the most promising for future syntheses and biological assays, which will be able to validate the methodology proposed in this research for the design of these new N-11-azaartemisinins with activity against *P. falciparum.*

In **Table 12**, the molecular properties of the N-11-azaartemisinins of the prediction set are reported. In this table, similar to **Table 4**, it can be scrutinized that ε_{LUMO+1}, $d(C_6\text{-}C_5)$ and TSA, in general, present lower values for *cha*, when compared to *cla*.

Table 13 shows the values of $d(Fe\text{-}O_1)$, $d(Fe\text{-}O_2)$ distances, absolute values of the differences between $d(Fe\text{-}O_1)$ and $d(Fe\text{-}O_2)$, and interaction energies of the ligand-heme complex for the N-11-azaartemisinins in the prediction set.

Table 10. Classification matrix using SDA with Cross-Validation.



(20 to 26) (27 to 35)

Compounds	R
20	4'-$HOCC_6H_4NH$—
21	4'-NCC_6H_4NH—
22	4'-$F_3CC_6H_4NH$—
23	4'-$HOOCC_6H_4NH$—
24	4'-$HOCC_6H_4$—
25	4'-$H_2NOCC_6H_4$—
26	4'-$HOOCC_6H_4$—
27	3'-$HOCC_6H_4$—
28	*α*-$O_2NC_{10}H_6$—
29	*α*-$HOCC_{10}H_6$—
30	*α*-$HOCC_{10}H_6$—
31	$HOCCHCHCH_2$—
32	HOCCHCH—
33	HOCNHCHCH—
34	$CH_3CHCOOCHCH$—
35	HOCCHCHNH—

3.4.7. MEP Maps, Ligand-Heme Interaction and Position of the LUMO+1 Orbital for Some N-11-Azaartemisinins from the Training and Prediction Sets

Figure 8 reports 2D structure (a), MEP maps (b), ligand-heme interaction (c), and position of the LUMO+1 orbital (d) in N-11-azaartemisinins *cha* (6 and 11) of the set of training and *cla* (20 and 27) of the prediction set. Scrutinizing this figure, it can be noticed that the MEP (b) maps of the N-11-azaartemisinins *cha* in the two sets of molecules studied (11 and 27, respectively) are similar to artemisinin (Figure 2(a)), 11 azaartemisinin (Figure 2(b)) and N-11-azaartemisinins (Figure 3) in the region of the endoperoxide group. In Figure 8(a) and Figure 8(b), the N-11-azaartemisinins can be attacked by nucleophiles in this region. Furthermore, as can be seen in Figure 8(b), the MEPs of N-11-azaartemisinins *cha* (11 and 27) present more negative values (−129.27 and −124.25 kcal/mol) when compared to the values (−114.21 and −121.74 kcal/mol) of N-11-azaatemisinins *cla* (6 and 20).

Table 11. Results of applying PR models to the prediction set.

Compounds	PCA model	HCA model	KNN model	SIMCAmodel	SDA model
20	*cha*	*cha*	*cha*	*cha*	*cha*
21	*cha*	*cha*	*cha*	*cha*	*cha*
22	*cha*	*cha*	*cha*	*cha*	*cha*
23	*cha*	*cha*	*cha*	*cha*	*cha*
24	*cha*	*cha*	*cha*	*cha*	*cha*
25	*cha*	*cha*	*cha*	*cha*	*cha*
26	*cha*	*cha*	*cha*	*cha*	*cha*
27	*cla*	*cla*	*cla*	*cla*	*cla*
28	*cla*	*cla*	*cla*	*0*	*cla*
29	*cla*	*cla*	*cha*	*0*	*cla*
30	*cla*	*cla*	*cla*	*cla*	*cla*
31	*cha*	*cha*	*cha*	*cha*	*cha*
32	*cha*	*cha*	*cha*	*cha*	*cha*
33	*cha*	*cha*	*cha*	*cha*	*cha*
34	*cha*	*cha*	*cha*	*cha*	*cha*
35	*cha*	*cha*	*cha*	*cha*	*cha*

Table 12. Molecular properties of the N-11-azartemisinins in the prediction set.

Compounds	ε_{LUMO+1}	$d(C_6\text{-}C_5)^b$	TSA[c]
20	27.22	1.551	614.5
21	28.10	1.550	615.6
22	28.16	1.550	626.5
23	26.51	1.551	627.4
24	26.98	1.551	601.4
25	25.44	1.551	618.3
26	25.56	1.551	614.9
27	53.00	1.553	646.3
28	55.73	1.553	642.7
29	51.36	1.553	636.7
30	51.26	1.553	605.5
31	42.22	1.551	605.5
32	51.14	1.551	567.8
33	35.36	1.551	595.3
34	50.50	1.551	624.8
35	26.76	1.551	572.9

[a]Values in kcal/mol. [b]Values in Å. [c]Values in $Å^2$

Table 13. Values of d(Fe-O_1), d(Fe-O_2), absolute value of the difference between d(Fe-O_1) and d(Fe-O_2), and interaction energy of the ligand-heme complex for the N-11-azaartemisinins in the prediction set.

Compounds	d(Fe-O_1)[a]	d(Fe-O_2)[a]	Δ[b]	Energy interaction[c]
20	2.336	3.105	0.77	−7.14
21	2.319	2.910	0.60	−7.45
22	2.359	3.164	0.80	−7.08
23	2.397	3.113	0.71	−6.57
24	2.351	2.984	0.63	−5.45
25	2.544	2.737	1.07	−7.21
26	2.387	2.890	0.50	−6.14
27	2.350	3.160	1.00	−5.12
28	2.395	2.583	0.19	−6.68
29	2.480	2.660	0.18	−5.33
30	2.510	2.620	0.11	−7.31
31	2.389	3.139	0.75	−5.92
32	2.511	3.600	1.09	−5.77
33	2.524	3.635	1.11	−5.39
34	2.458	3.483	1.03	−5.83
35	2.540	3.670	1.13	−6.42

[a]Values in Å. [b] $\Delta = \left| d\left(\text{Fe-O}_1\right) - d\left(\text{Fe-O}_2\right) \right|$. [c]Values in kcal/mol.

Still in **Figure 8(c)**, as one can see, the N-11-azaartemisinins-heme interaction of the *cha* class, in the training (6) and prediction (20) sets, are also similar to those verified in artemisinin (**Figure 4(a)**), 11-azaartemisinin (**Figure 4(b)**) and N-11-azaartemisinins (**Figure 3**). The distances d(Fe-O_1) and d(Fe-O_2) in the N-11-azaartemisinins in **Figure 8(c)**, class *cha*, lie between 2.370 and 2.675 Å, 3.007 and 3.832 Å, respectively. For the N-11-azaartemisinins, *cla* class, these distances are, respectively, 2.247 and 2.445 Å, 2.507 and 3.074 Å.

The distinction between the *cha* and *cla* classes can be evidenced when the LUMO+1 orbitals are compared, as shown in **Figure 8(d)**, where one can see representative N-11-azaartemisinins, *cha* (6) and *cla* (11), from the set of training, and N-11 azaartemisinins, *cha* (20) and *cla* (27), from the prediction set. The LUMO+1 orbital lobes in N-11 azaartemisinins *cha* are positioned primarily on the atoms of the 1,2,4-trioxane ring. The lobes in all these compounds are directed to the possible position of heme, indicating the importance of the interaction with the delocalized heme system. N-11-azaartemisinins *cla* exhibit LUMO+1 orbital lobes concentrated in some atoms of the substituents.

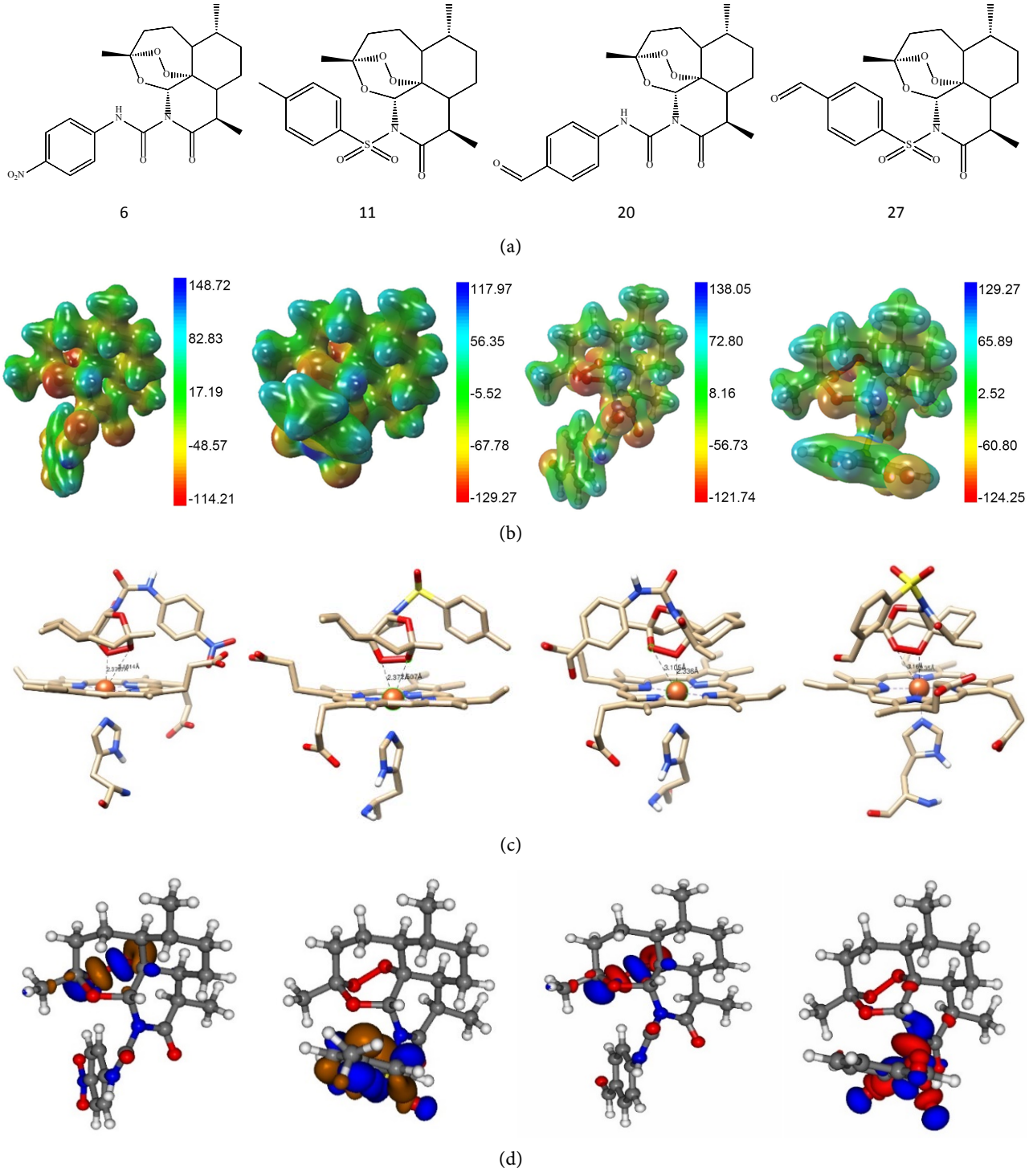


Figure 8. 2D structures (a), MEP (kcal/mol) maps (b), interactions (kcal/mol) between N-11-azaartemisinins and the heme receptor (c), and LUMO+1(d) for 6, 11, 20, and 27 N-11-azaartemisinins, respectively.

4. Conclusions

For the investigation carried out for N-11-azaartemisinins, the approximation base set B3LYP/6-31* was indicated as the most suitable in the preliminary study with geometric parameters of artemisinin.

For N-11-azaartemisinins, the MEP maps are similar to artemisinin and 11-azaartemisinin in the 1,2,4-trioxane ring region, with the electron density of some molecules more concentrated in this region, indicating greater biological activity. N-11-azaartemisinins are susceptible to electrophilic attacks in the most negative MEP region, –130.52 to –114.21 kcal/mol, respectively.

The interactions between N-11-azaartemisinins (1-19) and heme in their entirety occur preferentially orienting Fe^{2+} to heme. The distance d(Fe-O_1) is smaller than d(Fe-O_2) in all ligand-heme interactions and, in general, d(Fe-O1) has higher values for N-11-azaartemesinins *cha*, when compared to N-11-azaartemisinins *cla*. It is also verified, in general, higher values of d(Fe-O_1) and lower values for the interaction energy in N-11-azaartemisinins *cha* for biological processes that occur with increased biological activity of these compounds.

The application of supervised machine learning methods (PCA, HCA, KNN, SIMCA, and SDA) led to the separation of N-11-azaartemisinins into classes *cha* and *cla*, with properties ε_{LUMO+1}, d(C6-C5), and TSA being responsible for classification. Each of these properties can be associated with relevant aspects in the mechanism of action of the ligands of the *cha* class, that is:

1) the LUMO+1 orbital of the N-11-azaartemisinins can interact with the heme orbital that presents the greatest contribution to the electron density, characterizing the importance of the charge transfer reaction in the action of the ligands of the *cha* class in the biological process;

2) the d(C_6-C_5) property may indicate the importance of the steric distribution of atoms of the N-11-azaartemisinins of the *cha* class in the biological process;

3) the TSA property may indicate that in the biological process involving N-11-azaartemisinins and heme, hydrophobic interactions may be relevant in the mechanism of action of molecules of the *cha* class.

The insights resulting from the investigation of N-11-azaartemisinins with MEP, ligand-receptor interaction, supervised machine learning methods, and chemical intuition, led to the design of 16 new molecules (20-35). The application of the models obtained with the supervised learning methods (PCA, HCA, KNN, SIMCA and SDA) to this prediction set showed 12 new N-11-azaartemisinins (20-26, and 31-35) promising for syntheses (in progress in the Synthesis Laboratory of the Federal University of Pará) and biological evaluation, which may contribute to the validation of the results of this investigation in the future.

Acknowledgements

The authors would like to thank the Swiss Center for Scientific Computing for use of MOLEKEL software. We also employed computing facilities at the Laboratório de QuímicaTeórica e Computacional (LQTC) at the Universidade Federal do Pará (UFPA).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] World Health Organization (2021) World Malaria Report 2021. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021

[2] World Health Organization (2018) Artemisinin Resistance and Artemisinin-Based Combination Therapy Efficacy: Status Report. https://apps.who.int/iris/handle/10665/274362

[3] Ferreira, M.U. and Castro, M.C. (2016) Challenges for Malaria Elimination in Brazil. *Malaria Journal*, **15**, 284. https://doi.org/10.1186/s12936-016-1335-1

[4] Fröhlich, T., Karagöz, A.Ç., Reiter, C. and Tsogoeva, S.B. (2016) Artemisinin-Derived Dimers: Potent Antimalarial and Anti-Cancer Agents. *Journal of Medicinal Chemistry*, **59**, 7360-7388. https://doi.org/10.1021/acs.jmedchem.5b01380

[5] Abbasitabar, F. and Zare-Shahabadi, V. (2017) QSAR Study of Artemisinin Analogues as Antimalarial Drugs by Neural Network and Replacement Method. *Drug Research* (*Stuttg*), **67**, 476-484. https://doi.org/10.1055/s-0043-108553

[6] Yamansarova, E.Y., Kazakova, D.V., Medvedeva, N.I., Khusnutdinovaa, E.F., Kazakovaa, O.B., Legostaevaa, Y.V., Ishmuratova, G.Y., Huongc, L.M.H., Hac, T.T.H., Huongc, D.T. and Suponitskyd, K.Y. (2018) Synthesis and Antimalarial Activity of 3'-Trifluoromethylated 1,2,4-Trioxolanes and 1,2,4,5-Tetraoxane Based on Deoxycholic Acid. *Steroids*, **129**, 17-23. https://doi.org/10.1016/j.steroids.2017.11.008

[7] Tse, E.G., Korsik, M. and Todd, M.H. (2019) The Past, Present and Future of Anti-Malarial Medicines. *Malaria Journal*, **18**, 93. https://doi.org/10.1186/s12936-019-2724-z

[8] Yousefnejad, S., Mahboubifar, M. and Eskandari, R. (2019) Quantitative Structure-Activity Relationship to Predict the Anti-Malarial Activity in a Set of New Imidazolopiperazines Based on Artificial Neural Networks. *Malaria Journal*, **18**, 310. https://doi.org/10.1186/s12936-019-2941-5

[9] Palla, D., Antoniou, A.I., Baltas, M., Menendez, C., Grellier, P., Mouray, E. and Athanassopoulos, C.M. (2020) Synthesis and Antiplasmodial Activity of Novel Fosmidomycin Derivatives and Conjugates with Artemisinin and Aminochloroquinoline. *Molecules*, **25**, 4858. https://doi.org/10.3390/molecules25204858

[10] Chang, C.M. (2020) A Quantitative Structure-Activity Relationship Study on the Antimalarial Activities of 4-Aminoquinoline, Febrifugine and Artemisinin Compounds. *International Journal Quantitative Structure-Property Relationships*, **5**, 63-79. https://doi.org/10.4018/IJQSPR.2020010104

[11] Nguyen, P.T.V., Dat, T.V., Mizukami, S.M., Nguyen, D.L.H., Mosaddeque, F., Kim, S. N., Nguyen, D.H.B., Đinh, O.T., Vo, T.L., Nguyen, G.L.T., Duong, C.Q., Mizuta, S., Dao Ngoc Hien Tam, D.N.H., Truong, M.P.T., Huy, N.T. and Hirayama, K. (2021) 2D-Quantitative Structure-Activity Relationships Model Using PLS Method for Anti-Malarial Activities of Anti-Haemozoin Compounds. *Malaria Journal*, **20**, 264. https://doi.org/10.1186/s12936-021-03775-2

[12] Patel, O.P.S., Beteck, R.M. and Legoabe, L.J. (2021) Exploration of Artemisinin Derivatives and Synthetic Peroxides in Antimalarial Drug Discovery Research. *European Journal of Medicinal Chemistry*, **213**, Article ID: 113193. https://doi.org/10.1016/j.ejmech.2021.113193

[13] Jahan, M., Leon, F., Fronczek, F.R., Elokely, K.M, Rimoldi, J., Khan, S.I. and Avery, M.A. (2021) Structure-Activity Relationships of the Antimalarial Agent Artemisinin 10. Synthesis and Antimalarial Activity of Enantiomers of rac-5β-Hydroxy-D-Se-

coartemisinin and Analogs: Implications Regarding the Mechanism of Action. *Molecules*, **26**, 4163. https://www.mdpi.com/1420-3049/26/14/4163 https://doi.org/10.3390/molecules26144163

[14] Tam, D.N.H., Tawfk G.M., El Qushayri, A.E., Mehyar, G.M., Istanbuly, S., Karimzadeh, S., Tu, V.L., Tiwari, R., Dat, T.V., Nguyen, P.T.V., Hirayama, K. and Tien Huy, N.T. (2020) Correlation between Anti-Malarial and Anti-Haemozoin Activities of Anti-Malarial Compounds. *Malaria Journal*, **19**, 298. https://doi.org/10.1186/s12936-020-03370-x

[15] Bansal, M., Uapadhyay, C., Poonam, Kumar, S. and Rathi, B. (2021) Phthalimide Analogs for Antimalarial Drug Discovery. *RSC Medicinal Chemistry*, **12**, 1854-1867. https://doi.org/10.1039/D1MD00244A

[16] Pal, K., Raza, M.K., Legac, J., Rahman, M.A., Manzoor, S., Rosenthal, P.J. and Hoda, N. (2021) Design, Synthesis Crystal Structure and Anti-Plasmodial Evaluation of Tetrahydrobenzo[4,5]thieno[2,3-d]pyrimidine Derivatives. *RSC Medicinal Chemistry* **12**, 970-981. https://doi.org/10.1039/D1MD00038A

[17] Brito, D., Marquez, E., Rosas, F. and Rosas, E. (2022) Predicting New Potential Antimalarial Compounds by Using Zagreb Topological Indices. *AIP Advances*, **12**, Article No. 04507. https://doi.org/10.1063/5.0089325

[18] Harmse, R., Wong, H.N., Smit, F., Haynes, R.K. and N'Da, D.D. (2015) The Case for Development of 11-Aza-Artemisinins for Malaria. *Current Medicinal Chemistry*, **22**, 3607-3630. https://doi.org/10.2174/0929867322666150729115752

[19] Harmse, R., Coertzen, D., Wong, H.N., Smit, F.J., van der Watt, M.E., Reader, J., Nondaba, S.H., Birkholtz, L.-M., Haynes, R.K. and N'D, D.D. (2017) Activities of 11-Azaartemisinin and N-Sulfonyl Derivatives against Asexual and Transmissible Malaria Parasites. *Chemistry Medicinal Chemistry*, **12**, 2086-2093. https://doi.org/10.1002/cmdc.201700599

[20] Nisar, M., Haynes, R.K., Sung, H.H. and Williams, I.D. (2017) Mechanochemical Conversion of 11-Azaartemisinin into Pharmaceutical Cocrystals with Improved Solubility. Mechanochemical Conversion of 11-Azaartemisinin into Pharmaceutical Cocrystals with Improved Solubility. *Acta Crystallographica*, **A73**, a268. https://doi.org/10.1107/S0108767317097367

[21] Kumaria, A., Karnataka, M., Singhb, D., Shankarb, R., Jatc, J.L., Sharmad, S., Yadavd, D., Shrivastavae, R. and Vermaa, V.P. (2019) Current Scenario of Artemisinin and Its Analogues for Antimalarial Activity. *European Journal of Medicinal Chemistry*, **163**, 804-829. https://doi.org/10.1016/j.ejmech.2018.12.007

[22] Torok, D., Ziffer, H., Meshnick, S.R., Pan, X.-Q. and Ager, A. (1995) Syntheses and Antimalarial Activities of N-Substituted ll-Azaartemisinins. *Journal of Medicinal Chemistry*, **38**, 5045-5050. https://doi.org/10.1021/jm00026a012

[23] Bach, R.D. and Dmitrenko, O. (2005) The Effect of Carbonyl Substitution on the Strain Energy of Small Ring Compounds and Their Six-Member Ring Reference Compounds. *The Journal of American Chemical Society*, **128**, 4598-4611. https://doi.org/10.1021/ja055086g

[24] Bonaccorsi, R., Scrocco, E. and Tomasi, J. (1970) Molecular SCF Calculations for the Ground State of Some Three-Membered Ring Molecules: $(CH_2)_3$, $(CH_2)_2NH_2$, $(CH_2)_2NH_2^+$, $(CH_2)_2O$, $(CH_2)_2S$, $(CH)_2CH_2$, and N_2CH_2. *The Journal of Chemical Physics*, **52**, 5270. https://doi.org/10.1063/1.1672775

[25] Scrocco, E. and Tomasi, J. (2005) The Electrostatic Molecular Potential as a Tool for the Interpretation of Molecular Properties. In: Davison, A. and Dewar, M.J.S., Eds., *New Concepts II*, Topics in Current Chemistry, Vol. 42. Springer, Berlin, 95-170.

[26] Politzer, P. and Truhlar, G. (1981) Chemical Applications of Atomic and Molecular Electrostatic Potentials. Plenum Press, New York. https://doi.org/10.1007/978-1-4757-9634-6

[27] Politzer, P., Laurence, P.R. and Jayasuriya, K. (1985) Molecular Electrostatic Potentials: An Effective Tool for the Elucidation of Biochemical Phenomena. *Environmental Health Perspectives*, **61**, 191-202. https://doi.org/10.1289/ehp.8561191

[28] Pinzi, L. and Rastelli, G. (2019) Molecular Docking: Shiffting Paradigms in Drug Discovery. *International Journal of Molecular Sciences*, **20**, 4331. https://doi.org/10.3390/ijms20184331

[29] Stanzione, F., Giangreco, I. and Cole, J.C. (2021) Use of Molecular Docking Computational Tools in Drug Discovery. In: Witty, D.R. and Cox, B., Eds., *Progress in Medicinal Chemistry*, Elsevier, New York, 273-343. https://doi.org/10.1016/bs.pmch.2021.01.004

[30] Jurs, P.C., Kowalski, B.R. and Isenhour, T.L. (1969) Investigation of Combined Patterns from Diverse Analytical Data Using Computerized Learning Machines. *Analytical Chemistry*, **41**, 1949-1953. https://doi.org/10.1021/ac50159a027

[31] Zhao, S. (2021) Prediction of Protein Expression and Growth Rates by Supervised Machine Learning. *Natural Science*, **13**, 301-330. https://doi.org/10.4236/ns.2021.138025

[32] Livingstone, D.J. (1991) Pattern Recognition Methods in Rational Drug Design. *Methods in Enzymology*, **203**, 613-638. https://doi.org/10.1016/0076-6879(91)03032-C

[33] Varmuza, K. (2018) Methods for Multivariate Data Analysis. In: Engel, T. and Gasteiger, J., Eds., *Chemoinformatics*: *Basic Concepts and Methods*, Wiley-VCH, Weinheim, 339-437.

[34] Allen, F.H. (2002) The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallographica B*, **58**, 380-388. https://doi.org/10.1107/S0108768102003890

[35] Becke, A.D. (1993) Density-Functional Thermochemistry. III. The Role of Exact Exchange. *The Journal of Chemical Physics*, **98**, 5648-5652. https://doi.org/10.1063/1.464913

[36] Lee, C., Yang, W. and Parr, R.G. (1988) Development of the Colic-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B*, **37**, 785-789. https://doi.org/10.1103/PhysRevB.37.785

[37] Binkley, J.S., Pople, J.A. and Hehre, W.J. (1980) Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. *Journal of the American Chemical Society*, **102**, 939-947. https://doi.org/10.1021/ja00523a008

[38] Hehre, W.J. (1976) *Ab Initio* Molecular Orbital Theory. *Accounts of Chemical Research*, **9**, 399-406. https://doi.org/10.1021/ar50107a003

[39] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J. R., Montgomery, J. A., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, O., Baboul, A.G., Clifford,

S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C. and Pople, J.A. (1998) Gaussian 98, Revision A.6. Gaussian, Inc., Pittsburgh.

[40] Lisgarten, J.N., Potter, B.S., Bantuzeko, C. and Palmer, R.A. (1998) Structure, Absolute Configuration, and Conformation of the Antimalarial Compound, Artemisinin. *Journal of Chemical Crystallography*, **28**, 539-543. https://doi.org/10.1023/A:1023244122450

[41] Haynes, R.K., Wong, H.-N., Lee, K.-W., Lung, C.-M., Shek, L.Y., Williams, I.D., Croft, S.L., Vivas, L., Rattray, L., Stewart, L., Wong, V.K.W. and Ko, B.C.B. (2007) Preparation of N-Sulfonyl- and N-Carbonyl-11-Azaartemisinins with Greatly Enhanced Thermal Stabilities: *In Vitro* Antimalarial Activities. *Chemistry Medicinal Chemistry*, **2**, 1464-1472. https://doi.org/10.1002/cmdc.200700065

[42] Flukiger, P., Luthi, H.P., Portmann, S. and Weber, J. (2000-2001) Molekel. Swiss Center for Scientific Computing, Mano.

[43] Goodsell, D.S., Morris, G.M. and Olson, A.J. (1996) Automated Docking of Flexible Ligands: Applications of AutoDock. *Journal of Molecular Recognition*, **9**, 1-5. https://doi.org/10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6

[44] Beebe, K.R., Pell, R.J. and Seasholtz, M.B. (1998) Chemometrics: A Practical Guide. Wiley & Sons, New York.

[45] Oliveira, L.F.S., Cordeiro, H.C., Brito, H.G., Pinheiro, A.C.B., Santos, M.A.B., Bitencourt, H.R., Figueiredo, A.F., Araújo, J.J.O., Gil, F.S., Farias, M.S., Barbosa, J.P. and Pinheiro, J.C. (2021) Molecular Electrostatic Potential and Pattern Recognition Models to Design Potentially Active Pentamidine Derivatives against *Trypanosoma brucei* Rhodesiense. *Research, Society and Development*, **10**, e261101220207. https://doi.org/10.33448/rsd-v10i12.20207

[46] Araújo, J.J.O., de Miranda, R.M., Castro, J.S.O., Figueiredo, A.F., Pinheiro, A.C.B., Santos Morais, S.S., Santos, M.A.B., Pinheiro, A.L.R., Gil, F.S., Bitencourt, H.R., Alves, G.N.R. and Pinheiro, J.C. (2023) Designing Artemisinins with Antimalarial Potential, Combining Molecular Electrostatic Potential, Ligand-Heme Interaction and Multivariate Models. *Computational Chemistry*, **11**, 1-23. https://doi.org/10.4236/cc.2023.111001

[47] Johnson, R.A. and Wichem, D.W. (1992) Applied Multivariate Statistical Analysis. Prentice-Hall, Hoboken.

[48] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) Multivariate Analysis. Academic Press, New York.

[49] Pirouette 3.01 (2001) Infometrix, Inc., Woodinville.

[50] Minitab, Inc. (2009) Minitab Statistical Software, Release 16 for Windows. State College.

[51] Njogu, P.M., Gut, J., Rosenthal, P.J. and Chibale, K. (2013) Design, Synthesis, and Antiplasmodial Activity of Hybrid Compounds Based on (2R,3S)-N-Benzoyl-3-Phenylisoserine. *ACS Medicinal Chemistry Letters*, **4**, 637-641. https://doi.org/10.1021/ml400164t

[52] Politzer, P. and Murray, J. (2021) The Neglected Nuclei. Molecules, **26**, 2982. https://doi.org/10.3390/molecules26102982

[53] Thomsen, R. (2003) Flexible Ligand Docking Using Evolutionary Algorithms: Investigating the Effects of Variation Operators and Local Search Hybrids. *Biosystems*, **72**, 57-73. https://doi.org/10.1016/S0303-2647(03)00135-7

[54] Rawe, S.L. (2020) Artemisin and Artemisinin-Related Agents. In: Patrick, G.L, Ed., *Antimalarial Agents*: *Design and Mechanism of Action*, Elsevier, London, 99-132. https://doi.org/10.1016/B978-0-08-101210-9.00004-4

[55] Chovsky´, J.V., Chu, K.C., Berendzen, J., Sweet, R.M. and Schlichting, I. (1999) Crystal Structures of Myoglobin-Ligand Complexes at Near-Atomic Resolution. *Biophysical Journal*, **77**, 2153-2174. https://doi.org/10.1016/S0006-3495(99)77056-6

[56] Protein Data Bank. The Rutgers State University. Piscataway, New Jersey. http://www.resb.org/pdb

[57] ChemPlus: Modular Extensions to HyperChem Release 8.06. (2008) Molecular Modeling for Windows. Hyperchem, Inc., Gainesville.

[58] Todeschini, R. and Consonni, V. (2009) Molecular Descriptors for Chemoinformatics. Wiley-VCH, New York. https://doi.org/10.1002/9783527628766

[59] Bernardinelli, G., Jefford, C.W., Marie, D., Thomson, C. and Weber, J. (1994) Computational Studies of the Structures and Properties of Potential Antimalarial Compounds Based on the 1,2,4-Trioxane Ring Structure. I. Artemisinin-Like Molecules. *International Journal of Quantum Chemistry*: *Quantum Biology Symposium*, **21**, 117-131. https://doi.org/10.1002/qua.560520710

[60] Posner, G.H, Cumming, J.N., Ploypradith, P. and Oh, C.H. (1995) Evidence for Fe(IV) = O in the Molecular Mechanism of Action of the Trioxane Antimalarial Artemisinin. *Journal of the American Chemical Society*, **117**, 5885-5886. https://doi.org/10.1021/ja00126a042

[61] Jefford, C.W., Grigorov, M., weber, J., Lüthi, H.P. and Troncher, J.M.J. (2000) Correlating the Molecular Electrostatic Potentials of Some Organic Peroxides with Their Antimalarial Activities. *Journal of Chemical Information and Computer Sciences*, **40**, 354-357. https://doi.org/10.1021/ci990276u

[62] Cardoso, F.J.B., Figueiredo, A.F., Lobato, M.S., Miranda, R.M., Almeida, R.C.O. and Pinheiro, J.C. (2008) A Study on Antimalarial Artemisinin Derivatives Using MEP Maps and Multivariate QSAR. *Journal of Molecular Modeling*, **14**, 39-48. https://doi.org/10.1007/s00894-007-0249-9

[63] Meunier, B. and Robert, A. (2010) Heme as Trigger and Target for Trioxane-Containing Antimalarial Drugs. *Accounts of Chemical Research*, **43**, 1444-1451. https://doi.org/10.1021/ar100070k

[64] Cheng, F., Shen, J., Luo, X., Zhu, W., Gu, J., Ji, R., Jiang, H. and Chen, K. (2002) Molecular Docking and 3-D-QSAR Studies on the Possible Antimalarial Mechanism of Artemisinin Analogues. *Bioorganic Medicinal & Chemistry*, **10**, 2883-2891. https://doi.org/10.1016/S0968-0896(02)00161-X

[65] Tonmunphean, S., Parasuk, V. and Kokpol, S. (2001) Automated Calculation of Docking of Artemisinin to Heme. *Journal of Molecular Modeling*, **7**, 26-33. https://doi.org/10.1007/s008940100013

[66] Ferreira, M.M.C. (2015) Químiometria: Conceitos, Métodos e Aplicações. UNICAMP, Campinas. https://doi.org/10.7476/9788526814714

[67] Bulat, F.A., Murray, J.S. and Politzer, P. (2021) Identifying the Most Energetic Electrons in a Molecule: The Highest Occupied Molecular Orbital and the Average Local Ionization Energy. *Computational and Theoretical Chemistry*, **1199**, Article ID: 113192. https://doi.org/10.1016/j.comptc.2021.113192

[68] Todeschini, R. and Consonni, V. (2000) Handbook of Molecular Descriptors. Wiley, New York. https://doi.org/10.1002/9783527613106