Article

The Correlation Between Electronic Structure and Antimalarial Activity of Tetrahydropyridines

Oscar A. Naranjo-Montoya,[a] Lucas M. Martins,[b] Luiz C. da Silva-Filho,[c]
Augusto Batagin-Neto,[d] and Francisco C. Lavarda[e]*

[a]*Programa de Física, Facultad de Ciencias Basicas y Tecnologias, Universidad del Quindío,
Carrera 15 Calle 12 Norte Armenia, Quindío, Colombia*

[b]*Programa de Pós-Graduação em Ciência e Tecnologia de Materiais and* [c]*Departamento de
Química, Universidade Estadual Paulista (UNESP), 17033-360 Bauru-SP, Brazil*

[d]*Campus Experimental de Itapeva, Universidade Estadual Paulista (UNESP),
Rua Geraldo Alckmin 519, 18409-010 Itapeva-SP, Brazil*

[e]*Faculdade de Ciências, Departamento de Física, Universidade Estadual Paulista (UNESP),
17033-360 Bauru-SP, Brazil*

In this study, correlations between electronic structure and the antimalarial activities reported for a group of 21 tetrahydropyridines were evaluated by multivariate methods. Simple and multiple linear regressions, principal component analyses, and linear discriminant analyses were employed to evaluate possible correlations between quantum-mechanical descriptors and pharmacological activity. The results obtained provide useful information on the position of active sites of tetrahydropyridines and identify some specific features associated with active derivatives. Based on these findings, simple substitution rules are proposed for designing more efficient tetrahydropyridine derivatives.

Keywords: antimalarials, tetrahydropyridines, electronic structure, molecular modeling, quantitative structure–activity relationship

Introduction

According to the World Health Organization (WHO), malaria is one of the most deadly diseases caused by parasites. This acute febrile illness is typical of tropical and subtropical regions, and is currently endemic in 99 countries. In 2012 alone, more than 207 million cases of malaria were reported around the world, of which 627,000 culminated in death.[1]

Malaria is caused by four different species of protozoan parasites, of which *Plasmodium falciparum* is the most dangerous, being responsible for the lethal form of the disease. Unfortunately, *P. falciparum* is also the most resistant specie to currently employed treatments.[2]

Malaria thus continues to present a worrying threat. Despite several studies regarding possible vaccine candidates, there are still no licensed vaccines, and continuous preventive treatments are the only way to reduce the frequency of infection.[1,3] Furthermore, a gradual increase in resistance of some parasite strains to the currently employed drugs has been observed, which makes the search for new antimalarial compounds all the more relevant.

In this context, tetrahydropyridines (THP) derivatives have shown to be promising compounds. The latter are commonly prepared by highly efficient and economical multicomponent reactions (MCR).[4] A wide range of derivatives with varied characteristics can be easily synthesized simply by adjusting the initial substrates. In particular, metal-free compounds can be obtained when organic catalysts are employed, and these are of great interest for medical applications.[4]

In general, THP-based compounds present several distinct biological properties, such as antiparasitic, antimicrobial, anticancer, antiviral, and so on.[4,5] Recent studies performed by Misra *et al.* have shown that these compounds display promising antimalarial activity.[4] In their

*e-mail: abatagin@itapeva.unesp.br

work, organocatalysed MCR methods were employed to synthesize distinct THP derivatives, based on a varied set of aromatic aldehydes, anilines, and b-keto ester precursors. Antimalarial activities were assayed against a resistant 3D7 strain of *P. falciparum* and high activity was observed, even at low drug concentration.[4]

Despite these promising results, the mechanism of the antimalarial activity of THP derivatives remains incompletely understood. Indeed, to the best of our knowledge, no previous quantitative structure-activity relationship (QSAR) studies have been performed for this set of compounds. Only basic aspects regarding the nature of the substituents have been experimentally studied, and it is not clear how the findings could lead to more active compounds.

A quantitative structure-activity correlation could help in the molecular modeling of new compounds with improved properties, or at least restrict the number of derivatives to evaluate. Varied multidisciplinary efforts have been made to establish such relationships. In particular, research into correlating electronic properties with the biological activity of the compounds has shown interesting results for both predicting new active derivatives and outlining possible mechanisms associated with them.[6–10]

In this study, we have evaluated possible relationships between electronic structure data from quantum-mechanics calculations and the antimalarial activity of the 21 derivatives reported by Misra *et al.*[4] Distinct multivariate methods were employed for statistical analyses: simple and multiple linear regression (SLR and MLR), principal component analysis (PCA), and linear discriminant analysis (LDA). The results obtained allow us to outline some specific features expected in very active derivatives and to suggest simple substitutions that could lead to compounds with improved biological activity. The analyses also provide useful information on the position of active sites on the basic structure of THP, which could improve understanding of the inhibitory mechanism displayed by this molecule.

Materials and Methods

Experimental data

In this report, a set of 21 THP derivatives, whose antimalarial activities were reported by Misra *et al.*,[4] was evaluated to investigate possible quantitative structure-activity relationships (QSAR). Figure 1 shows the basic structure common to all the compounds studied here (properly numbered for QSAR studies). The nature of the substituents R^1, R^2, and R^3, as well as the biological activity (percentage of inhibition of Schizonts of *P. falciparum*, *%ISPf*; and drug concentration necessary to promote 50%
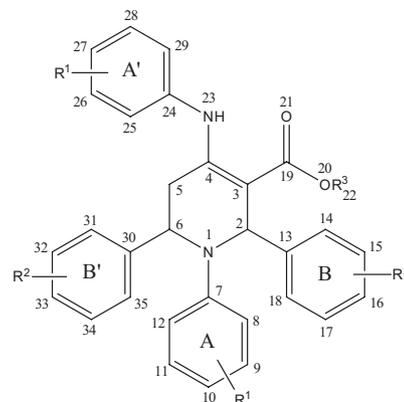

Figure 1. Basic structure of THP derivatives and numeration employed in QSAR studies (R^1, R^2, and R^3 are defined in Table 1).[4]

inhibition of Schizonts of *P. falciparum*, IC_{50}), are outlined in Table 1. For simplicity, the same molecule numbers of the experimental work were adopted.

The first biological descriptor presented in Table 1 is the percentage of inhibition of Schizonts of *P. falciparum* (*%ISPf*).[4] In the experimental work, this descriptor was evaluated at two stages. Initially, the activities of all the compounds (from **1** to **21**) were evaluated at four distinct concentrations: 10, 5, 2.5, and 1.25 μg mL^{-1}. Subsequently, additional experiments employing reduced concentrations (0.78, 0.39, 0.19, 0.09, and 0.05 μg mL^{-1}) were conducted for compounds that had presented 100% efficiency in the first stage (higher concentrations).

In this study, only the *%ISPf* values evaluated at 1.25 μg mL^{-1}, *%ISPf$_{1.25}$* (presented in Table 1), were considered for classificatory analyses (PCA and LDA). This choice was based on the fact that, in addition to characterizing the activity of most of the compounds, the *%ISPf* values at this concentration better discriminate between active and potentially inactive compounds. For information about *%ISPf* at distinct concentrations see reference 4.

However, as can be observed in Table 1, the biological index, *%ISPf*, reported in the experimental work is not an absolute descriptor of the compound's activity. For instance, it is impossible to distinguish the antimalarial potency of the compounds **1**, **3**, **4**, **8**, **12**, **16**, **18**, and **20** at 1.25 μg mL^{-1}, since they all produce 100% inhibition. This finding is evidenced at lower concentrations (0.05 μg mL^{-1}), where the dissimilarity of the activity of these compounds is revealed.[4] Consequently, preliminary fittings were performed in order to get more appropriate descriptors for predictive studies (SLR and MLR). For this purpose, we consider that the biological activity (*%ISPf*) dependence with drug concentration (*C*) may be described by a logistic equation:

$$\%ISPf = \frac{100}{1 + \left(C / IC_{50} \right)^{-p}} \qquad (1)$$

Table 1. Description of the THP derivatives and biological activity descriptors: percentage of inhibition of Schizonts of *P. falciparum* at 1.25 µg mL^{-1} (*%ISPf$_{1.25}$*),[4] and *IC$_{50}$* values obtained by logistic fits (see Supplementary Information)

Compound	R^1	R^2	R^3	Biological activity	
				%ISPf$_{1.25}$	*IC$_{50}$*
1	4-Chloro	4-Bromo	Methyl	100	0.147
2	H	4-Fluoro	Methyl	0	10.315
3	4-Methoxy	4-Fluoro	Methyl	100	0.116
4	H	4-Bromo	Methyl	100	0.046
5	4-Methoxy	3-Chloro	Methyl	46	1.326
6	4-Methoxy	4-Methoxy	Methyl	33	1.287
7	4-Chloro	4-Methoxy	Methyl	86	0.319
8	4-Bromo	4-Methoxy	Methyl	100	0.184
9	4-Chloro	4-Fluoro	Methyl	25	1.562
10	H	4-Benzyloxy	Methyl	93	1.098
11	H	H	Methyl	0	13.910
12	H	3-Chloro	Methyl	100	0.227
13	4-Methoxy	3-Chloro	Methyl	46	1.303
14	4-Chloro	Thiophene-2-carboxaldehyde	Ethyl	16	2.412
15	4-Methoxy	Thiophene-2-carboxaldehyde	Ethyl	86	0.398
16	4-Chloro	3-Chloro	Ethyl	100	0.126
17	4-Methoxy	H	Ethyl	0	4.465
18	4-Chloro	H	Ethyl	100	0.544
19	H	Pyridine-3-carboxaldehyde	Ethyl	25	1.550
20	4-Methoxy	4-Bromo	Ethyl	100	0.076
21	Aldehyde	Aniline	Methyl	33	1.430

where p represents the slope of the curve at its midpoint, and IC_{50} value is the targeted activity index that represents the drug concentration necessary to promote 50% inhibition of Schizonts of *P. falciparum*. Since this descriptor is obtained from the drug response at distinct concentrations, it better describes the real activity of the compounds. The IC_{50} values derived from the fittings are presented in the last column of Table 1 (for details about the fittings see Supplementary Information).

By analyzing the values of IC_{50}, one can see that, although the *%ISPf$_{1.25}$* indexes are not good quantitative descriptors of the THP's biological activity, they are indeed good indexes for classifying the compounds into active and non-active subsets, justifying their use in the classificatory studies.

Electronic structure calculations

Geometry optimization

The geometry optimizations were performed by considering ten distinct initial structures for each molecule. Molecular dynamics (MD) calculations were performed at relatively high temperature with a view to obtaining weakly correlated structures. The MD simulations were performed with the aid of the Gabedit computational package[11] by considering the molecules in contact with a reservoir at 1000 K, during 1 ps (steps with 0.01 ps).[12]

In order to avoid possible convergence problems, all the resulting structures were subsequently pre-optimized using a PM6 (Parametric Method 6) semi-empirical quantum-mechanical method in a Restricted Hartree-Fock (RHF) approach.[13] These pre-optimizations were performed until relatively low gradients were obtained (0.01), such that the resulting structures were as close as possible to the equilibrium conformation of the compounds. The MOPAC2012 computational package was employed.[14,15]

After pre-optimizations, a comparative analysis of the total energy values, E_T, (derived from PM6 calculations) associated with each compound's conformer was conducted to identify the most stable structure ($E_T = E_{T(min)}$) and other structures with sufficiently low E_T values (specifically $E_T < E_{T(min)} + k_B T_{300}$). All the selected structures were then compared by evaluating the root mean square deviation of atomic positions (RMSD-AP) of each conformer

in relation to the most stable structure. This parameter defines the degree of similarity between the geometries of the conformers considered. If they present quite similar geometries, RMSD-AP is a low value, and then it is enough to consider only one of the structures in the study of electronic properties. If, however, they are dissimilar, both the structures should be considered.

In the present work, RMSD-AP = 2 Å was a cutoff parameter for defining molecular similarity. This criterion is commonly adopted in studies of docking with small ligand molecules, where simulated and experimental conformations are compared.[16] In our case, conformers with RMSD-AP < 2 Å were considered as equivalent structures, and the lowest energy conformation alone was evaluated in the next steps. The Qmol computational package was employed in all RMSD-AP analyses,[17] and the positions of heavy atoms only (not hydrogen) were considered.

As the last step of the geometry study, all the (pre-optimized) relevant structures, obtained from RMSD-AP analyses, were fully optimized in a DFT approach. The optimization was performed *in vacuo*, employing a Becke's LYP (B3LYP) exchange-correlation (XC) functional,[18,19] and 6-31G basis set. All the calculations in this step were carried out with the GAMESS computational package.[20]

Single point calculations

Single point (SP) calculations were performed for each of the 21 selected and optimized structures (as well as possible additional low energy conformers), with the aim of extracting electronic information on the compounds.

A collection of 195 indexes, mainly related to the energy, bond orders, electric charge, electric dipole, and frontier molecular orbitals, was obtained and organized in a single data file for subsequent statistical analyses. Table 2 presents a brief description of each electronic index.

For the cases where two or more structures were obtained from the RMSD-AP study, each electronic descriptor, A, was estimated by the expected value, $\langle A \rangle$, given by:

$$\langle A \rangle = \frac{\sum A_i e^{\frac{\Delta E_T^i}{k_B T}}}{\sum e^{\frac{\Delta E_T^i}{k_B T}}} \qquad (2)$$

where n represents the total number of dissimilar structures of a given compound; ΔE_T^i represents the total energy difference between the i-th structure and the most stable structure ($E_T = E_{T(min)}$); k_B represents the Boltzmann constant; and T is the temperature (considered equal to 300 K).

Table 2. Electronic descriptors employed in this work

Electronic index	Description
E_T	Total energy
E_P	Potential energy
E_K	Kinetic energy
E_{XC}	Exchange-Correlation energy
E_{HOMO-1}	Energy of the level just below the highest occupied molecular orbital (HOMO-1)
E_{HOMO}	Energy of the highest occupied molecular orbital (HOMO), or vertical ionization potential
E_{LUMO}	Energy of the lowest unoccupied molecular orbital (LUMO)
E_{LUMO+1}	Energy of the level just above the lowest unoccupied molecular orbital (LUMO+1)
Δ_{L-H}	Energy difference between LUMO and HOMO levels
Δ_{H-H1}	Energy difference between HOMO and HOMO-1 levels
Δ_{L1-L}	Energy difference between LUMO+1 and LUMO levels
Δ_{L1-H1}	Energy difference between LUMO+1 and HOMO-1 levels
Dip_x, Dip_y, Dip_z	Components of the electric dipole moment
Dip_T	Total electric dipole moment
BO_{i-j}	Bond order associated to the molecular bond involving atoms i and j
$CHAR^{MP}_i$	Electric charge associated to the i-th atom of the compound structure (Mulliken partition)
$CHAR^{LP}_i$	Electric charge associated to the i-th atom of the compound structure (Lowdin partition)
VAL_i	Total valence associated to the i-th atom of the compound structure
$BVAL_i$	Bond valence associated to the i-th atom of the compound structure

All the SP calculations were performed through a DFT approach, employing three distinct XC functionals: B3LYP, X3LYP,[21] and PBE0;[22] 6-31G(1p,1d) basis set was adopted in all the cases. All these functionals are hybrid ones, with slight differences between them: *(i)* X3LYP is a B3LYP-like functional, adjusted to better describe unbounded interactions, and *(ii)* PBE0 presents a distinct XC functional in relation to the B3LYP. The results presented here relate mainly to the DFT/B3LYP/6-31G(1p,1d) approach. X3LYP and PBE0 functionals were considered so as to check the robustness of the B3LYP findings.

Correlation studies

Distinct multivariate methods were employed to evaluate correlations between the electronic structure and the biological activity of the THP derivatives.

Simple and multiple linear regressions were first performed with the aim of obtaining predictive equations capable of estimating the biological activity based on linear combinations of a few electronic descriptors. To

avoid overfitting effects, a maximum of three independent variables (electronic descriptors) was employed in MLR analyses.[23] In this sense, all the combinations with up to three variables of the whole set of descriptors (195 electronic indexes) were considered. The quality of the regressions were evaluated by analyzing the correlation parameter between the predicted and the experimental values of the dependent variable for each regression. Since *%ISPf* parameters are not good quantitative descriptors of a compound's activity, IC_{50} values were considered as the dependent variable in linear regressions.

Subsequently, PCA and LDA were employed for pattern recognition and compound classification, respectively. For this purpose, *%ISPf$_{1.25}$* indexes were employed to define active and non-active compounds.

In PCA, a new orthogonal coordinate system is generated by linear combinations of the original electronic descriptors (independent variables), which are a convenient means of analyzing similarity between compounds.[24,25] The resulting principal axes, called principal components (PCs), are ordered according to the maximum variance of the data set: PC_1 presents more statistical information than PC_2; PC_2 presents more statistical information than PC_3; and so on.

Finally, in LDA, a discriminant function (DF) is obtained through a linear combination of electronic descriptors.[26] DF is constructed in such a way that it is able to promote the highest distinction between active and non-active subgroups of compounds. By defining a delimiting parameter (the DF cutoff), the molecules can be classified as active or not, according to their DF score. A "stepwise" procedure, based on Mahalanobis distance criterion (validated by F statistic), was adopted so as to minimize the number of descriptors necessary for DF construction.[26,27]

LDA calculations were performed employing the commercial software SPSS.[27] SLR, MLR, and PCA studies were performed using our homemade statistical package.

Results and Discussion

In this section, only the results relating to electronic descriptors derived from the DFT/B3LYP/6-31G(1p,1d) approach are presented. The same trends and conclusions were obtained for PBE0 and X3LYP functionals.

Following the similarity criteria outlined for RMSD-AP analysis, only one conformer was obtained for each derivative (for details see Supplementary Information).

Simple and multiple linear regressions

Linear regressions were performed employing all the combinations of up to three independent variables

(SLR: 195, MLR$_2$: 18,915, and MLR$_3$: 1,216,865 combinations). As the dependent variable, we considered distinct functional forms of IC_{50}: IC_{50}, $log(IC_{50})$, and $(1/IC_{50})$. Equation 3 illustrates the more representative regression equation (with correlation parameter of $R_{corr} = 0.91$) obtained by considering $1/IC_{50}$ as the dependent variable. Figure 2 shows a comparison between experimental data and values predicted by equation 3.

$$\frac{1}{IC_{50}} = -331.756 + (191.282 \times BO_{28\text{-}29}) - (29.286 \times CHAR^{MP}_{15}) \quad (3)$$
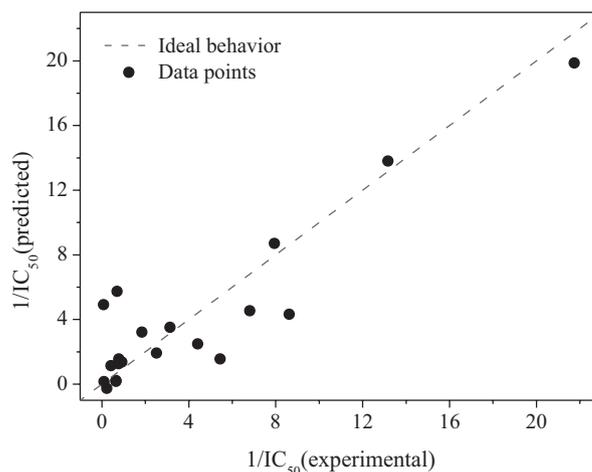$$- (606.482 \times CHAR^{MP}_{29})$$

Figure 2. Comparison between $1/IC_{50}$ values predicted by equation 3 and experimental values.

The descriptors involved are the bond order of atoms 28 and 29, and the net charge on atoms 15 and 29 of the basic structure of the THP molecule. A quite similar regression equation was also obtained for the descriptors $BO_{28\text{-}29}$, $CHAR^{MP}_{29}$, and $CHAR^{MP}_{32}$; however, since atoms 15 and 32 are located at equivalent positions of rings B and B', it contains the same physical information as equation 3 (see Supplementary Information for details).

The *BO* descriptor illustrates the strength of the chemical bond between two atoms. In general, high bond orders indicate an excess of electrons in the region between the atoms, and thus a strong chemical bond; low values of *BO* indicate a lack of electrons in the bond region, indicating a weak linkage of the two atoms. The *CHAR* descriptor is defined as the difference between the number of electrons on an isolated atom and the calculated fraction of the molecular electronic population located on it. Negative net charges are associated with an excess of electrons and positive charges with a lack of electrons.

The presence of the descriptors $CHAR^{MP}_{15}$ and $CHAR^{MP}_{32}$ in the most representative linear equations suggests that rings B and B' can play an important role in the inhibitory mechanism of the drugs. The electronic

indexes $CHAR^{MP}_{29}$ and BO_{28-29}, in turn, are associated with the ring A', which marks another active site of the molecule.

In general, the regression equation obtained indicates that active derivatives must present the following features: (*i*) Strong chemical bond between atoms 28 and 29 and (*ii*) Excess of electrons (or low positive charge) on atoms 29 and 15 (or 32).

These trends can be adjusted by choosing appropriate R^1 and R^2 substituents. For instance, an electron donating group attached to position 16 (or 33) of the ring B (or B') could induce a high electronic population at carbon 15 (or 32). On the other hand, the addition of electron withdrawing groups on position 27 of the ring A could promote an excess of negative charge on atom 29. However, the substituent in this case should not be a strong electron withdrawing specie, for fear of weakening the chemical bond between atoms 28 and 29.

Despite the useful information provided by equation 3, it is important to note that $1/IC_{50}$ values do not promote a good dispersion of the experimental data. As can be seen in Figure 2, there is some data agglomeration at low values of $1/IC_{50}$. This feature, in addition to the absence of significant regressions for IC_{50} and $log\ (IC_{50})$ functional forms, suggests that classificatory and/or pattern recognition methods could be appropriated. Such analyses are performed in the following sections.

Principal component analysis

PCA studies were performed considering the entire data set (21 cases × 195 descriptors), with a view to better evaluating similarities between the electronic structures of the compounds. Figure 3 shows a plot of the first and second components (PC_1 and PC_2) derived from this PCA study (full PCA).
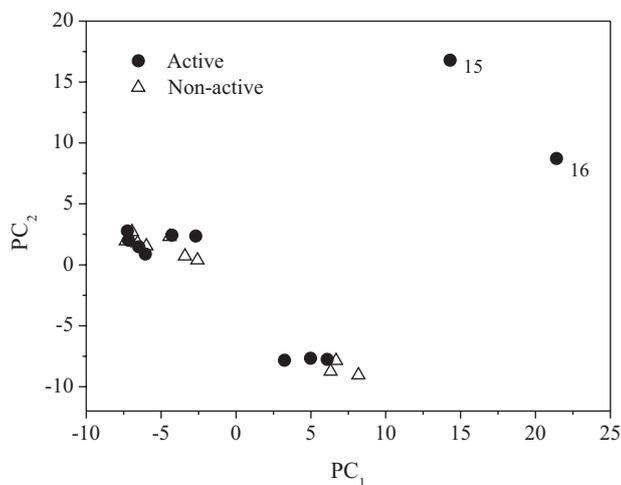

Figure 3. PC_1 *versus* PC_2 from a PCA study involving the whole data set.

Compounds **15** and **16** present a quite different electronic structure from the others. These compounds are the only ones that present thiophene-2-carboxaldehyde as rings B and B', being very distinct from the other derivatives. Figure 4 shows a plot of the first and second components (PC_1 and PC_2) of a full PCA performed after removing compounds **15** and **16** from the data set.
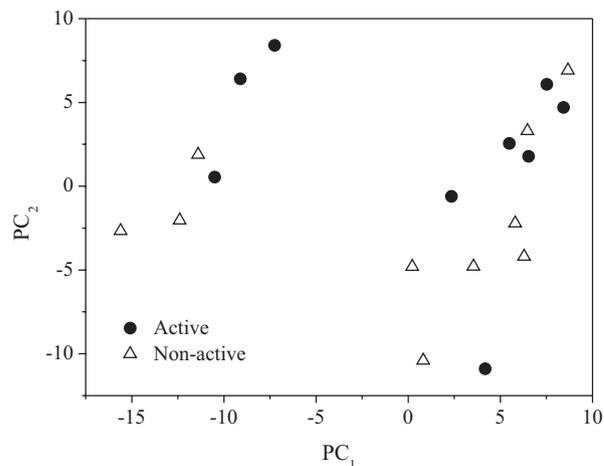

Figure 4. PC_1 *versus* PC_2 from full PCA performed excluding molecules **15** and **16** from the data set.

Note that two distinct groups can be observed. However, there appears to be no direct relationship between the observed clusters and the biological activity of the compounds. Indeed, these groups are mainly associated with R^1 substituents. In general, compounds with PC_1 scores lower than −2.5 present R^1 = 4-methoxy, while higher values are observed for the other structures (a same trend was observed even for compounds **15** and **16**, see Figure 3).

This result suggests that changes in the rings A and A' can induce significant alterations in the electronic structure of the compounds; however, it has no direct influence on their biological activity. Indeed, this information is complementary to MLR results. Despite the presence of BO_{28-29} and $CHAR^{MP}_{29}$ descriptors in the equation 3, which are indexes relating to ring A', it is necessary the presence of one electronic descriptor associated with ring B (or B') to obtain significant correlations ($CHAR^{MP}_{15}$ or $CHAR^{MP}_{32}$). This result suggests that the rings B and B' play an important role in the antimalarial activity of THPs.

Linear discriminant analysis

LDA was performed with the aim of identifying which electronic descriptors better discriminate between active and non-active compounds. The *%ISPf* observed at 1.25 µg mL⁻¹ (*%ISPf*$_{1.25}$) was employed to identify

two groups of compounds: *(i)* active compounds with %ISPf$_{1.25}$ ≥ 86 and *(ii)* non-active compounds with %ISPf$_{1.25}$ < 86.

Equation 4 presents the best DF obtained, which promotes the most significant separation between these two groups.

$$DF = -3.139 + (4.695 \times 10^{-4})E_T - (3.724 \times 10^{+1})\text{CHAR}^{LP}_{17} \quad (4)$$

The descriptors involved are the total energy (E_T) and Löwdin's net charge on atom 17 ($CHAR^{LP}_{17}$) (similar results were also obtained from considering atoms 15, 32, and 34, which are not shown for simplicity). E_T represents the total energy of the compound and can be correlated with the stability of the molecule. The *CHAR* descriptor has the same definition as presented before. Table 3 summarizes some statistical information about the values obtained for these descriptors for each subset of molecules (active, non active and the whole set).

In general, higher total energy and a more negative charge on site 17 (by Löwdin's partition) are observed in active molecules. Non-active compounds are also observed to present larger dispersion of the descriptors, mainly regarding E_T values, which indicates that the non-active subgroup is much more heterogeneous than the active one.

The DF obtained (equation 4) shows a statistical significance higher than 99.84% (Wilk's Lambda = 0.489, χ^2 = 12.867 with 2 degrees of freedom, and p < 0.0016) and correctly classifies 95.2% of the molecules. The same hit percentage was obtained by leave-one-out cross validation (in which each molecule is tested with a model derived from all the other molecules).

Figure 5 shows the DF scores plot. The centroids for active and non-active molecules are respectively 0.927 and −1.019. The obtained cutoff is −0.046, such that: *(i)* if *DF* > −0.046, the compound is predicted to be active and *(ii)* if *DF* < −0.046 the compound is predicted to be non-active.

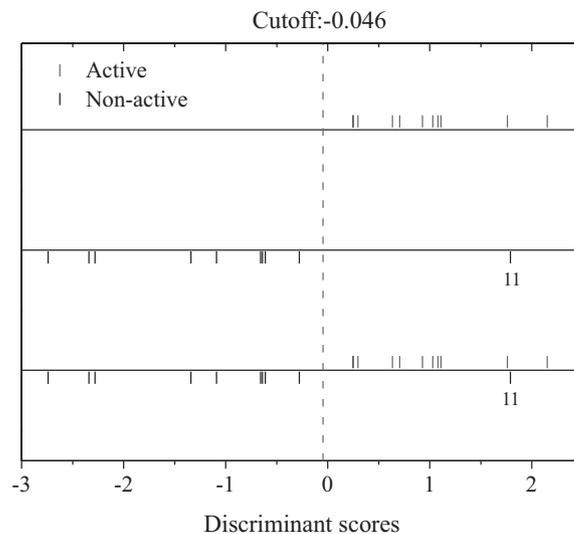Note that just one of the non-active THP derivatives (compound **11**) is misclassified by this rule.

Figure 5. Discriminant function scores of active and non-active compounds. Dotted line indicates the cutoff value (–0.046).

According to the trends presented in Table 3, the DF obtained indicates that two essential features must be observed in active compounds: *(i)* high total energy and *(ii)* high negative (or low positive) net charge on site 17 (similar results were obtained by considering sites 15, 32, and 34, not shown).

The energy component suggests that active compounds are supposed to be more (thermodynamically) unstable molecules. In our systems such high values of E_T can be also related to compounds with a reduced electronic structure, i.e., molecules a lower number of electrons in their structure (see Figure S9 in Supplementary Information). Since the number of electrons are frequently associated with the molecule volume, this result can suggest that derivatives with reduced volume (smaller ligands) tend to present higher antimalarial activity.

Such consideration, indicates the possibility that not just topological, but also lipophilic properties could be associated with THP's activity. As a matter of fact, lipophilic parameters (such as octanol-water partition coefficient, log P) usually provide valuable information about the interaction of the molecules with cell membranes,

Table 3. Statistical information regarding the relevant descriptors identified by LDA

Descriptor	Classification	Maximum	Minimum	Average	Standard deviation
	Active	−1457.254	−3057.068	−2037.957	522.995
E_T	Non active	−2148.003	−7522.598	−4331.543	2283.648
	Total	−1457.254	−7522.598	−3130.141	1965.017
	Active	−0.109	−0.164	−0.135	0.017
$CHAR^{LP}_{17}$	Non active	−0.097	−0.159	−0.112	0.020
	Total	−0.097	−0.164	−0.124	0.022

being a very important descriptor in QSAR studies.[28] Since such parameters often present some dependence with molecular volume of the compounds (in general larger molecules present higher hydrophobicity[29-31]), the presence of E_T descriptor in DF equation, can suggest the relevance of hydrophobicity in the biological activity of the compounds.

Indeed, in our case it is possible to observe a linear dependence between the hydrophobic parameter, log P, and IC_{50} values of THP's (see Figure S10 in Supplementary Information), what reinforces that E_T parameter carries more information about molecule size than stability. Nevertheless, since the evaluation of lipophilic/topological parameters is not in the scope of the present work, more profound studies are still necessary in this subject.

Despite of the apparent relevance of the hydrophobicity discussed above, the presence of the descriptor *CHAR* in equation 4 indicates that electrostatic interactions can play an important role in the compounds' activity, suggesting that site 17 can be linked to inhibitory mechanisms of THP derivatives. In fact, very accurate DFs can be also obtained from considering sites 15, 32, or 34, which represent equivalent positions in relation to site 17 in the THP basic structure. Figure 6 illustrates these molecular sites. It is important to note that sites 15 and 32 were also identified as relevant sites by MLR studies, suggesting that the biological activity of the compounds is mainly associated with R^2 substituents. In particular, all the relevant sites are related to meta positions (in relation to the main structure), and are prone to be tuned by the appropriate choice of R^2.
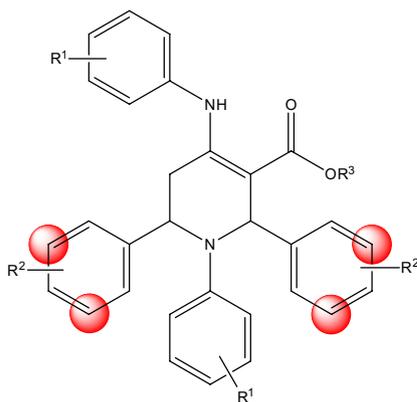

Figure 6. Relevant sites for the discrimination of biological activity in THP derivatives.

As can be seen in Figure 5, only one of the compounds is misclassified by the LDA model. Compound **11** is defined as a non-active derivative in the experimental work,[4] although it is classified as an active molecule by its DF score. The optimized structure of this compound is presented in Figure 7. Note that this derivative has a large group attached to the rings B and B', which is composed of

a single bounded ring. At room temperature, these structures can rotate very easily around the bond axis, hindering, for example, the possible coupling of this compound to some biological target. In this sense, although this molecule presents a $CHAR^{LP}_{17}$ value typical of an active compound, the steric hindrance disturbs specific couplings that can be relevant to the antimalarial activity of these molecules, turning it into a non-active compound.
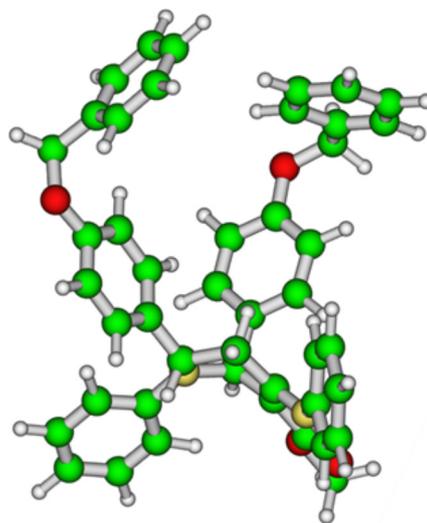

Figure 7. Structure of compound **11** after geometry optimization.

In this context, the misclassification of compound **11** reinforces our hypothesis that the rings B and B' have an important role in compound activity, demonstrating that the active sites of THP derivatives can in fact be located on it. It is also consistent with the relevance of hydrophobicity in the activity of the compounds, since the lateral groups of this molecule make it a very hydrophobic molecule. These considerations confirm that the antimalarial properties of the compounds can be truly enhanced by an appropriate choice of R^2 substituents, indicating, however, that large attached groups must be avoided in these rings, as already stated from the evaluation of E_T parameter.

In addition to the substitutions proposed in the SLR and MLR section, the presence of halogens, hydroxylated, or alkoxylated groups attached to sites 15, 17, 32, and 34 could also lead to a high negative net charge on these positions. As a matter of fact, the presence of such substituents has already been associated with pyridine-based compounds with improved antimalarial activity.[32–35] Nevertheless, as observed for compound **11**, we need to consider that the presence of such groups on R^2 position could impede the action of the molecule, deactivating it. Thus, we believe that small electron donating groups attached on position 16 (or 33) of the rings B and B' and weak electron withdrawing groups attached on position 27 of the rings A and A' could

Table 4. Changes induced by substitutions on rings A, A', B and B' in relation to compound **12**

Substitution	R[1]	R[2]	Percentage change in relation to compound **12** (+) parameter increase; (−) parameter reduction				
			$\Delta_{\%}BO_{28\text{-}29}$	$\Delta_{\%}CHAR^{MP}_{15}$	$\Delta_{\%}CHAR^{MP}_{29}$	$\Delta_{\%}E_T$	$\Delta_{\%}CHAR^{LP}_{17}$
1a	NO$_2$	H	0.279 (+)	1.093 (+)	24.775 (−)	28.055 (−)	4.853 (+)
2a	F	H	1.882 (−)	2.312 (−)	8.009 (+)	13.614 (−)	1.827 (−)
3a	CF$_3$	H	0.453 (−)	0.369 (+)	11.326 (−)	46.238 (−)	3.269 (+)
4a	CH$_3$C=O	H	0.697 (−)	0.378 (+)	18.921 (−)	31.259 (−)	2.816 (+)
1b	H	NH$_2$	0.279 (−)	32.658 (−)	4.148 (+)	7.593 (−)	41.854 (−)
2b	H	F	0.383 (−)	64.601 (−)	6.341 (+)	13.614 (−)	31.614 (−)
3b	H	CH$_3$	0.279 (−)	37.075 (−)	3.808 (+)	5.392 (−)	12.874 (−)
4b	H	CCH	0.348 (−)	0.868 (+)	5.533 (+)	10.443 (−)	11.151 (+)

be an interesting alternative means of obtaining THP derivatives with improved biological activity.

In order to test these hypotheses, additional calculations were performed for new derivatives. Two types of substitutions were tested: *(i)* R^2 = H and R^1 = NO$_2$, F, CF$_3$ and CH$_3$C=O (small electron withdrawing groups) and *(ii)* R^1 = H and R^2 = NH$_2$, F, CH$_3$ and CCH (small electron donating groups); all the substitutions were performed on *para* positions. Fluorine were inserted in R^1 and R^2 because this element can promote both the effects: electron withdrawing by induction and electron releasing by resonance.[36] The effect of each substitution was evaluated by comparing the changes induced on E_T, $CHAR^{LP}_{17}$, $BO_{28\text{-}29}$, $CHAR^{MP}_{15}$ and $CHAR^{MP}_{29}$ descriptors, in relation to those observed in compound **12** (R^1 = R^2 = H). Table 4 presents the percentage change for each index: (+) means an increase in the parameter while (−) represents a reduction of the value.

All the substitutions result in lower E_T values than those obtained for compound **12**, which is an undesired effect for obtaining active derivatives. This result is associated with the increase of the number of electrons in the resulting structures. In particular smaller changes are observed for **1a** and **3a**.

Substitutions performed at the rings A and A' (type *a*) promote a desired increase in the electron density on atom 29 (more negative charge), except for **2a**; however, undesired small effects are concomitantly induced on $CHAR^{MP}_{15}$, $CHAR^{LP}_{17}$ and $BO_{28\text{-}29}$ descriptors. For substitutions promoted in R^2, we can observe that **1b**, **2b** and **3b** promote a desired increase in the electron density on atoms 15 and 17 (more negative charges) as expected. However, small undesired changes on $BO_{28\text{-}29}$ and $CHAR^{MP}_{29}$ parameters are also observed for all the cases, including **4a**. These results evidence the relevance of substitutions on both the rings, A (A') and B (B').

Balancing both, favorable and unfavorable changes in the parameters presented in Table 4 (accordingly to LDA

and MLR results), it is possible to suggest some promising substitutions to achieve active compounds. For instance, R^1 = (NO$_2$) / R^2 = (NH$_2$) (**1a-1b**) and R^1 = (CH$_3$C=O) / R^2 = (NH$_2$) (**4a-1b**) may be cited as the two most promising compounds in the evaluated set. Individually these groups have shown the desired effect on the rings were they were attached, with small negative interference on others relevant parameters.

Complementary calculations, employing the same optimization and single point methodology described before, indicate that the compounds obtained from **1a-1b** and **4a-1b** substitutions are indeed active derivatives. DF scores of 1.558 and 1.581 and IC$_{50}$ values of 0.052 and 0.078 were respectively obtained for **1a-1b** and **4a-1b** substitutions, pointing out these compounds as active ones (DF > −0.046), with very low values of IC$_{50}$; and then good alternatives for further *in vitro / in vivo* tests.

Conclusions

The correlation between the electronic structure and antimalarial activity of a set of 21 tetrahydropyridines derivatives was evaluated by distinct multivariate methods.

Predictive equations obtained by multiple linear regressions indicate positions 15, 32, and 29 as important sites for drug activity. The results demonstrate that active derivatives must present a strong chemical bond between atoms 28 and 29 and a high net negative charge (or low positive charge) on atoms 29 and 15 (or 32).

PCA results show that much of the electronic structure of THP derivatives is defined by R^1 substituents. However, it was also observed that they are not directly associated with the biological response of the compounds.

By LDA, it was possible to define a discrimination function with high statistical significance. The obtained function combines just two electronic indexes and is able to correctly predict around 95% of the compounds, suggesting

that active derivatives present high total energy and a high negative (or low positive) net charge on meta positions of rings B and B' (atoms 15, 17, 32. and 34 of the main structure), which is compatible with MLR results. The results also suggest that large substituents must be avoided, since steric interactions could disturb the effective interaction between active sites of the molecules and their biological environment. The relevance of lipophilicity was also observed, but more studies are still necessary on this subject.

Based on these findings, we recommend the addition of small electron donating groups on position 16 (or 33) and electron withdrawing groups on position 27, as alternative means of obtaining THP derivatives with enhanced antimalarial activity.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgements

References

1. World Health Organization (WHO); *World Malaria Report*, France, 2013; p. 255.

2. Miller, L. H.; Baruch, D. I.; Marsh, K.; Doumbo, O. K.; *Nature* **2002**, *415*, 673.

3. Rahman, M. A.; *Chem. Sci. J.* **2011**, *2011*.

4. Misra, M.; Pandey, S. K.; Pandey, V. P.; Pandey, J.; Tripathi, R.; Tripathi, R. P.; *Bioorg. Med. Chem.* **2009**, *17*, 625.

5. Ramin, G.-V.; Hajar, S.; *Comptes Rendus Chim.* **2013**, *16*, 1047.

6. Lavarda, F. C. ; *Int. J. Quantum Chem.* **2003**, *95*, 219.

7. Ghiotto, R. C. T.; Lavarda, F. C.; Ferreira, F. J. B.; *Int. J. Quantum Chem.* **2004**, *97*, 949.

8. Autreto, P. A. da S.; Lavarda, F. C.; *Rev. Inst. Med. Trop. São Paulo* **2008**, *50*, 21.

9. Saghaie, L.; Sakhi, H.; Sabzyan, H.; Shahlaei, M.; Shamshirian, D.; *Med. Chem. Res.* **2013**, *22*, 1679.

10. Batagin-Neto, A.; Lavarda, F. C.; *Med. Chem. Res.* **2014**, *23*, 580.

11. Allouche, A.-R.; *J. Comput. Chem.* **2011**, *32*, 174.

12. Batagin-Neto, A.; Oliveira, E. F.; Graeff, C. F. O.; Lavarda, F. C.; *Mol. Simul.* **2013**, *39*, 309.

13. Stewart, J. J. P.; *J. Mol. Model.* **2007**, *13*, 1173.

14. Stewart, J. J. P.; *J. Comput. Aided Mol. Des.* **1990**, *4*, 1.

15. Stewart, J. J. P. *MOPAC2012: Molecular Orbital Package*; Stewart Computational Chemistry, 2012.

16. Lu, H.; Li, H.; Banu B. S. Rashid, S.; Leow, W. K.; Liou, Y.-C. In *Pattern Recognition in Bioinformatics*; Kadirkamanathan, V.; Sanguinetti, G.; Girolami, M.; Niranjan, M.; Noirel, J., eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009, Vol. 5780, pp. 175.

17. Gans, J. D.; Shalloway, D.; *J. Mol. Graphics Modell.* **2001**, *19*, 557.

18. Becke, A. D.; *J. Chem. Phys.* **1993**, *98*, 5648.

19. Lee, C.; Yang, W.; Parr, R. G.; *Phys. Rev. B* **1988**, *37*, 785.

20. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A.; *J. Comput. Chem.* **1993**, *14*, 1347.

21. Xu, X.; Goddard, W. A.; *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673.

22. Adamo, C.; Barone, V.; *J. Chem. Phys.* **1999**, *110*, 6158.

23. Ferreira, M. M. C.; Montanari, C. A.; Gaudio, A. C.; *Quim. Nova* **2002**, *25*, 439.

24. Brown, S. D.; *J. Chemom.* **1988**, *2*, 298.

25. Ferreira, M. M. C.; *J. Braz. Chem. Soc.* **2002**, *13*, 742.

26. Hair, J. F.; Black, W. C.; Babin, B. J.; Anderson, R. E. In *Multivariate Data Analysis*; Prentice Hall: Englewood Cliffs, 2010.

27. *SPSS for Windows*; SPSS Inc.: Chicago, 2001.

28. Kubinyi, H. *QSAR Hansch Analysis and Related Approaches*; VCH: Weinheim; New York, 1993.

29. Sedov, I. A.; Solomonov, B. N.; *J. Chem. Thermodyn.* **2010**, *42*, 1126.

30. Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V.; *J. Pharm. Sci.* **2009**, *98*, 861.

31. Leo, A.; Hansch, C.; Jow, P. Y.; *J. Med. Chem.* **1976**, *19*, 611.

32. Iwasa, K.; Nishiyama, Y.; Ichimaru, M.; Moriyasu, M.; Kim, H.-S.; Wataya, Y.; Yamori, T.; Takashi, T.; Lee, D.-U.; *Eur. J. Med. Chem.* **1999**, *34*, 1077.

33. Miroshnikova, O. V.; Hudson, T. H.; Gerena, L.; Kyle, D. E.; Lin, A. J.; *J. Med. Chem.* **2007**, *50*, 889.

34. O'Neill, P. M.; Shone, A. E.; Stanford, D.; Nixon, G.; Asadollahy, E.; Park, B. K.; Maggs, J. L.; Roberts, P.; Stocks, P. A.; Biagini, G.; Bray, P. G.; Davies, J.; Berry, N.; Hall, C.; Rimmer, K.; Winstanley, P. A.; Hindley, S.; Bambal, R. B.; Davis, C. B.; Bates, M.; Gresham, S. L.; Brigandi, R. A.; Gomez-de-las-Heras, F. M.; Gargallo, D. V.; Parapini, S.;

Vivas, L.; Lander, H.; Taramelli, D.; Ward, S. A.; *J. Med. Chem.* **2009**, *52*, 1828.

35. Păunescu, E.; Susplugas, S.; Boll, E.; Varga, R.; Mouray, E.; Grosu, I.; Grellier, P.; Melnyk, P.; *ChemMedChem* **2009**, *4*, 549.

36. Carey, F. A.; Sundberg, R. J.; *Advanced Organic Chemistry: Part A: Structure and Mechanisms*; Springer Science & Business Media: New York, 2007.