Inter*net* **Electronic** Journal of 𝔐𝔬𝔩𝔢𝔠𝔲𝔩𝔞𝔯 𝔇𝔢𝔰𝔦𝔤𝔫

Quantitative Structure–Activity Relationships of a Series of Chalcone Derivatives (1,3–Diphenyl–2–propen–1–one) as Anti *Plasmodium falciparum* Agents (Anti Malaria Agents)

Luiz Frederico Motta,[1,2] Anderson Coser Gaudio,[3] and Yuji Takahata [2]

[1] Laboratory of Computational Chemistry and Chemometrics, Department of Chemistry, Federal Center of Technological Education, 56314–512 Petrolina, PE, Brazil
[2] Institute of Chemistry, State University of Campinas, 13083–970 Campinas, SP, Brazil
[3] Physics Department, Federal University of Espírito Santo, 29075–910 Vitória, ES, Brazil

Citation of the article:
L. F. Motta, A. C. Gaudio, and Y. Takahata, Quantitative Structure–Activity Relationships of a Series of Chalcone Derivatives (1,3–Diphenyl–2–propen–1–one) as Anti *Plasmodium falciparum* Agents (Anti Malaria Agents), *Internet Electron. J. Mol. Des.* **2006**, *5*, 555–569, http://www.biochempress.com.

Quantitative Structure–Activity Relationships of a Series of Chalcone Derivatives (1,3–Diphenyl–2–propen–1–one) as Anti *Plasmodium falciparum* Agents (Anti Malaria Agents) [#]

Luiz Frederico Motta,[1,2,]* Anderson Coser Gaudio,[3] and Yuji Takahata[2]

[1] Laboratory of Computational Chemistry and Chemometrics, Department of Chemistry, Federal Center of Technological Education, 56314–512 Petrolina, PE, Brazil
[2] Institute of Chemistry, State University of Campinas, 13083–970 Campinas, SP, Brazil
[3] Physics Department, Federal University of Espírito Santo, 29075–910 Vitória, ES, Brazil

Abstract

Motivation. Malaria is globally recognized as serious problem of public health, mainly in the tropical and subtropical regions of the world. The increase of resistant malarial parasite strains represents the largest obstacle to antimalarial chemotherapy. In this work, we studied chalcone derivatives, because they present wide range of biological activity and a supposed mechanism of action: inhibition of malarial cysteine protease.

Method. QSAR analysis was performed on the series of chalcone derivatives. Various physical-chemical parameters such as hydrophobic, electronic, steric, thermodynamic and structural were calculated using computational package Molecular Modeling Pro 4.0, ChemSite Pro 5.0 and Arguslab 4.0 programs. QSAR models with up to four variables were generated employing multiple linear regression method using BuildQSAR program.

Results. Statistically significant models with *R*–values 0.931 and 0.958 were obtained. Results obtained show that hydrophobic and steric properties seem to play an important role in the explanation of the activity. Molar refractivity and molecular length have positive contribution to the activity against chloroquine–resistant (W2) *Plasmodium falciparum* strains, while molecular weight against mefloquine–resistant (D6) strains. The results indicated that the activity against W2 and D6 strains is favored if ring A is a width–limited chemical substituent and the limited molecular width of these derivatives can be related with the activity against D6 strain.

Conclusions. The present QSAR study reveal descriptors that may be important in the inhibitory activity of chalcone derivatives on *P. falciparum* cysteine protease. We obtained two different models against *Plasmodium falciparum* strains. The models have good capacity to explain the observed values of biological activity, good adjustment level, statistical significance and good predictive capacity.

Keywords. Chalcone derivatives; quantitative structure–activity relationships; QSAR; cysteine protease; Falcipain; Hansch analysis; *Plasmodium falciparum*.

1 INTRODUCTION

Malaria is an infection caused by the protozoa of the genus *Plasmodium*. The man is commonly infected by four species of the parasite, *i.e.*, *Plasmodium ovale*, *P. vivax*, *P. malariae* and *P. falciparum*, the last being the most virulent to humans. These parasites can be introduced in the human organism through the bite of a female *Anopheles* mosquito, the injection or transfusion of infected blood and through the hypodermic syringes. According to the World Health Organization (WHO), malaria is globally recognized as serious problem of public health, mainly in the tropical and subtropical regions of the world. It reaches 40% of the population of more than a hundred countries and it is considered as one of the diseases that caused already great damage to millions of people. More than 300 million cases and at least one million consequent deaths are estimated to occur annually. Previous researches point that *P. falciparum* is responsible for more than 40% of malaria cases registered in the globe and for more than 90% of all the deaths caused by this disease [1].

Some drugs such as chloroquine, mefloquine, pyrimethamine, dapsone, and cycloguanil have been used for years against malaria. However, the increase of resistant malarial parasite strains represents the largest obstacle to antimalarial chemotherapy [2]. The massive use of classical antimalarials promoted fast selection of drug–resistant strains of *P. falciparum*, which requires an urgent development of new antimalarial drugs. As malaria is a disease that most affects poor people from poor countries, the biggest pharmaceutical industries are generally not interested in spending efforts in research in this area. The WHO recently established the Roll Back Malaria initiative that attempts to increase the research on antimalarials through effective global cooperation. As a result, many vaccines and new drugs are currently being developed [3].
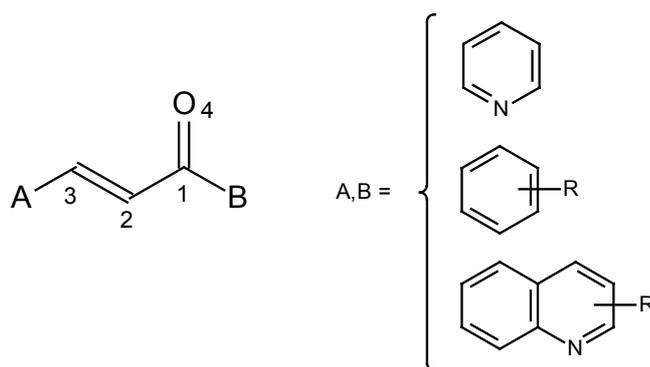

I – Chalcone derivatives

Among many promising substances, chalcones of varied chemical structures with antimalarial activity have been investigated (Structure **I**) [4–6]. Chalcones are biosynthetic precursory of flavonoids. Licochalcone A was the first compound of this family to present antimalarial activity

[5]. The mechanism of action of chalcone derivatives appears to be based on the competitive inhibition of malarial cysteine protease (falcipain) [7]. This is a key parasitic enzyme responsible for the degradation of hemoglobin, which generates the essential amino acids needed for the *P. falciparum* to growth. The first structure–activity relationship study on chalcone derivatives as falcipain inhibitors was performed by Li and co–workers [7]. They synthesized a series of chalcones derived from structure **I** and tested *in vitro* against *P. falciparum* strains CDC/Indochina III (W2) and CDC/Sierra Leone I (D6). The W2 strain is resistant to chloroquine, quinine, and pyrimethamine and susceptible to mefloquine. D6 is resistant to mefloquine and susceptible to chloroquine, quinine, and pyrimethamine. In structure **I**, A and B are pyridinyl, phenyl, and quinolinyl substituents.

The main conclusions of this work were the following: (*a*) The C_2–C_3 double bond is essential for high inhibitory activity. It is not only a conjugated linker between A and B aromatic substituents, but it keeps extended the molecular conformation. In this way, the drug molecule seems to bind much better the active site, which resembles a cleft on the surface of falcipain; (*b*) Substitutions on the bridge portion of the chalcone series cause a pronounced decrease in the inhibitory activity, probably due to steric interactions; (*c*) Chloro or fluoro substitution on the A ring and electron–donating substitution on the B ring increase the antimalarial activity; (*d*) Quinolinyl group in the A ring results in increase of the activity. Under acidic conditions nitrogen–containing heterocycles can be protonated, which may favor the interaction with the protease at His67 of falcipain.


II – 1,3–disubstituted–phenyl–2–propen–1–one
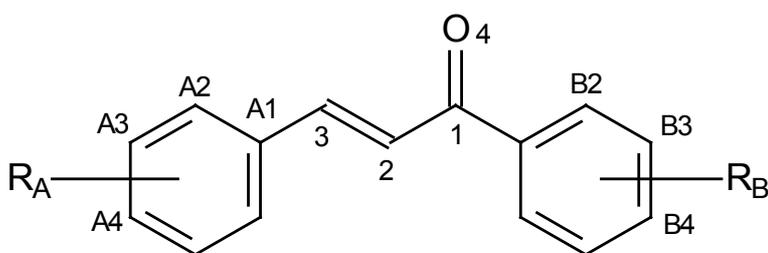
More recently, Liu and co–workers [8] performed a chemometric QSAR study on 93 chalcones derived from structure **II** against a strain of chloroquine–resistent human malarial parasite, *P. falciparum*, equivalent to W2 strain.

The results from the multivariate analysis indicated that: (*a*) Factors that contribute to high activity are different depending on the substitution pattern on the B ring; (*b*) A separate analysis involving the 19 most active compounds suggested that the influence of the B ring on the activity is related to size considerations, while the A ring may be more important in influencing

hidrophobicity; (*c*) Presence of chloro and fluoro substitution on ring A does not necessarily increase the antimalarial activity. The influence of these atoms on ring A seems to be dependent on the kind of substitution on ring B. It is interesting to note that not a single Hansch model could be derived from this dataset by using classical and quantum chemical descriptors.

At the time of the first QSAR study involving chalcone derivatives [7] only one parasitic papain–like cysteine protease was known. Falcipain–1 was extracted, isolated and identified by Salas and co–workers [9] who thought it was the main enzyme involved in the hemoglobin degradation. Studies involving Falcipain–1 have proved not to be promising due to its low abundance. More recently, two other cysteine proteases have been identified. It was shown that the highest concentration of cysteine protease in the parasitic food vacuole is due to Falcipain–2 [10]. The last member of this family, Falcipain–3 [11] was shown to be approximately 1.8 times less concentrated and twice as active as falcipain–2, which give these enzymes approximately the same importance in the parasitic hemoglobin degradation process.

Unfortunately, there is no experimental crystal structure of any of the three falcipain available so far. Avery and co–workers built homology models of falcipain–2 [12] and falcipain–3 [13], which is being proven to be very useful at this moment. Although the coordinates of these models are not available to the scientific community, some important qualitative information may be taken from them. The most important are the general topography and the amino acid configuration of the falcipain binding site. Figure 4 from Ref. [13] shows a schematic comparison of the binding sites of falcipain–2 and falcipain–3, with good residue conservation.
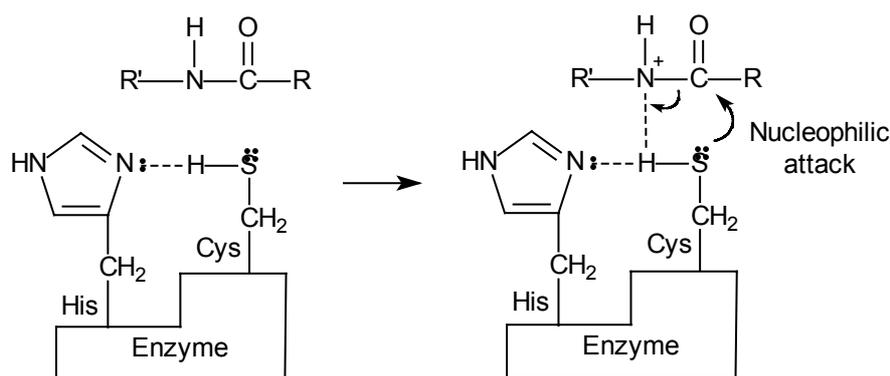
Figure 1. Initial mechanism of the cleavage of a peptide bond by cysteine protease.

Of course the most important residues are cysteine (Cys42 in falcipain–2, and Cys51 in falcipain–3) and histidine (His183 in falcipain–2, and His174 in falcipain–3), which constitute the catalytic center of the enzyme responsible for the cleavage of the peptide bonds. Glutamine and

triptophan residues form what is called *oxyanion hole*, which stabilizes the tetrahedral transition state intermediate of the bond cleavage. The other residues provide a region of extra stability to the ligand, through the formation of a variety of intermolecular bonds. It may be important to review briefly the initial mechanism of peptide bond cleavage by these enzymes, shown in Figure 1. Amino acid cysteine acts as nucleophile and donates an electron pair to the carbon atom of carbonyl. Histidine residue plays the role of a base and temporally accepts a proton from cysteine in order to place it on the correct position for the reaction with the nitrogen from the peptide bond. This mechanism is very important in the analysis of the present QSAR study.

The purpose of the present work is to perform a quantum chemical QSAR study on the chalcone derivatives in order to investigate the binding mode of these compounds and the properties that are relevant for their activity.

2 MATERIALS AND METHODS

Biological data of the activity of chalcone derivatives was extracted from the paper of Li and co–workers [7] (Table 1). The activity data have been given as IC_{50} values, where IC_{50} refers to the experimentally determined molar concentration of drug required for 50% inhibition (IC_{50}) of [^{3}H] hypoxanthine uptake into *P. falciparum* (W2 and D6 strains). Chloroquine and mefloquine were used as controls in the assays [7].
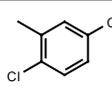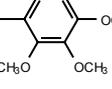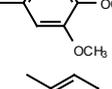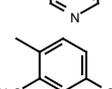
2.1 Chemical Data

The biological activity values [IC_{50} (μM)] reported in the literature were converted to molar units and then further to –log scale (Table 1). The molecular structures of the dataset were sketched using ACD/ChemSketch from Advanced Chemistry Development.

2.2 Computer Software

The first step consisted in obtaining the molecular geometry of all derivatives from the dataset of Li and co–workers [7] (Table 1). We initially minimized molecular energy in the Molecular Modeling Pro 4.0 computational package (MMP) [14] for each derivative using Allinger's MM2 [15] force field by fixing Root Mean Square Gradient (RMS) to 0.1 Kcal/molÅ. Further, geometry optimization was done using semi–empirical PM3 (Parametric Method) Hamiltonian method in the Hyperchem 6.0 program [16]. Finally, we minimized molecular energy with Simplex Method using dielectric constant equals to 78.54 (aqueous environment), Lennard–Jones potential 6–12 and Hydrogen bond function in the MMP package [17]. The low energy conformers obtained were used to quantum chemical QSAR analysis.

Table 1. Dataset Used in the Quantum Chemical QSAR Study [7] (Structure **I**)

Compd.	A	B	IC$_{50}$ W2 (µM)	– log IC$_{50}$ W2	IC$_{50}$ D6 (µM)	– log IC$_{50}$ D6
7			0.23	6.638	0.19	6.721
8			0.36	6.443	0.43	6.366
9			0.51	6.292	0.72	6.142
10			0.71	6.148	2.53	5.596
11			0.75	6.124	0.68	6.167
12			0.75	6.124	0.92	6.036
13			0.77	6.113	1.24	5.906
14			0.80	6.096	1.69	5.772
15			1.00	6.000	1.29	5.889
16			1.00	6.000	0.98	6.008
17			0.90	6.045	0.77	6.113
18			0.92	6.036	1.10	5.958
19			1.14	5.943	1.96	5.707
20			1.18	5.928	0.90	6.045
21			1.38	5.860	1.29	5.889
22			1.90	5.721	1.95	5.709
23			2.09	5.679	2.10	5.677
24			3.80	5.420	3.40	5.468
25			11.1	4.954	19.1	4.718
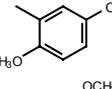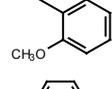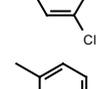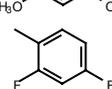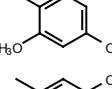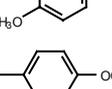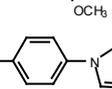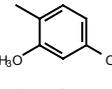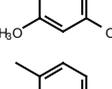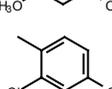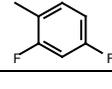
2.2.1 Quantum chemistry QSAR analysis

In the quantum chemical analysis we calculated seventy molecular properties of all the compounds. The calculated physical–chemical parameters are the types of hydrophobic, electronic, steric, thermodynamic, and structural. All the molecular properties were calculated by the MMP computational package, except the van der Waals surface area and van der Waals surface area accessible to the solvent, that was obtained using the ChemSite Pro 5.0 program [18]. The eletronic properties of quantum nature were calculated in the Arguslab program [19] except hydrogen bond donor and hydrogen bond acceptor that were obtained using MMP package.

The calculated hydrophobic parameters are: octanol/water partition coefficient (Hansch log P), Ghose and Crippen log P, hydrophilic–lipophilic balance, solubility parameter, dispersion of Hansen 3D, polarity of Hansen 3D, hydrogen bond Hansen 3D, hydrophilic surface area, % hydrophilic surface area, Klopman water solubility, hydration number and molar refractivity (MR).

The electronic parameters were subdivided into properties of empirical and quantum nature. The empirical electronic parameters (Hammett substituent constant) calculated are: σ_p, σ_m, σ^* and σ_i for the substituents A and B (Structure **I**). We used the semi–empirical PM3 method [20] to calculate quantum electronic parameters: dipole moment (μ), hydrogen bond donor, hydrogen bond acceptor, frontier orbital energy (E_{HOMO} and E_{LUMO}), hardness, heat of formation (ΔH_f) and partial charges qC_1, qC_2, qC_3, qO_4, qC_{A1} and qC_{B1} (Structure **II**).

The steric parameters calculated are: molecular volume (VM), molecular density, molecular length (L_x), molecular width (L_y), molecular depth (L_z), Verloop parameters (L_1, B_1, B_2, B_3 and B_4) for the substituents A and B, van der Waals surface area, van der Waals surface area accessible to the solvent, molecular weight (MW) and polar surface area.

The thermodynamic parameters evaluated are: boiling point, parachor, effective number of torsional bonds, number of hydrogen bonds (HBN), critical temperature, and critical pressure. Finally we complemented the list of molecular properties calculating the structural properties for each derivative of the optimized molecular geometry: interatomic distances (C_1–C_2, C_1–O_4, C_2–C_3, C_3–C_{A1}, and C_1–C_{B1}), bond angles (C_2–C_1–O_4, C_2–C_1–C_{B1}, C_3–C_2–C_1, C_2–C_3–C_{A1}), and dihedral angles (C_{A1}–C_3–C_2–C_1, C_3–C_2–C_1–O_4 and C_3–C_2–C_1–C_{B1}).

2.2.2 QSAR model selection

In order to estimate multidimensional linear Hansch [21] model, we based on the Topliss criterion [22], the low correlation among the variables (Tables 2 and 4), the adjustment degree (*R* and *s*), the degree of statistical significance (*F*–values), the level of confidence (*p*) and the previsibility degree (Q^2 and S_{PRESS}) of the model.

BuildQSAR program [23] was used to select the best combinations of the variables and to establish multidimensional QSAR models through the multiple linear regression. Initially we

executed the linear regression, one by one, for the seventy molecular descriptors and the biological activity. The values of the correlation coefficient (R), squared correlation coefficient (R^2), standard deviation (s), Fischer's test (F) and the level of confidence (p) were evaluated. Then, we built a correlation matrix, discarding combinations among variables with correlation coefficient greater than 0.6.

The selection of molecular properties was done by systematic search for models with up to four descriptors. The equations were analyzed using the following statistical data: correlation coefficient (R), squared correlation coefficient (R^2), standard deviation (s), Fischer's test (F), level of confidence (p), squared cross– correlation coefficient (Q^2) and standard deviation of sum of square of difference between predicted and observed values (S_{PRESS}). We also used graphical analyses to determine if there exists relation between observed and predicted properties (Figures 2 and 3).

3 RESULTS AND DISCUSSION

The QSAR analysis included the 19 compounds presented in Table 1, which let us to investigate models with up to four variables. The descriptors selected for modeling antimalarial activity of chalcone derivatives are summarized in Tables 3 and 5. The best model generated for W2 strain is Eq. (1), denoted as model 1:

$$pIC_{50} (W2) = + 0.033 (\pm 0.014) MR - 10.693 (\pm 5.517) d (C_3-C_{A1}) + 0.231 (\pm 0.087) Lx$$
$$- 0.220 (\pm 0.092) B2(A) + 17.709 (\pm 8.428) \tag{1}$$
$$n = 19; R = 0.931; R^2 = 0.866; s = 0.151; F = 22.886; p < 0.0001 \; Q^2 = 0.752; S_{PRESS} = 0.206$$

The data that generated this model are shown in Table 3. The model 1 has good capacity to explain the observed values of biological activity because it possesses high explanation coefficient (R^2). It explains 86.6% of the variance in the antimalarial activity against W2 strains, which can also be confirmed by the low standard deviation. The model has good adjustment level: high correlation coefficient ($R = 0.931$) and low standard deviation ($s = 0.151$). The evaluation of the degree of statistical significance was accomplished by the Fischer test (F) and level of confidence ($p < 0.0001$). The validation parameters ($Q^2 = 0.752$ and $S_{PRESS} = 0.206$), reflects the good predictive power of the model 1. Table 2 shows the correlation matrix among the four descriptors that appear in Eq. (1). The correlation coefficients between any two out of the four descriptors are sufficiently low.

Table 2. Correlation matrix for model 1

|	MR	d (C_3–C_{A1})	Lx	B2 (A)
MR	1.000	0.026	0.011	0.062
d (C_3–C_{A1})		1.000	0.046	0.030
Lx			1.000	0.001
B2 (A)				1.000

http://www.biochempress.com

Table 3. Numerical values of the four selected descriptors, Observed Activity, Predicted Activity and Residual for QSAR model 1

Derivatives	Observed Activity	Predicted Activity	Residual	MR	d (C_3–C_{A1})	Lx	B2 (A)
7	6.638	6.398	0.240	90.2262	1.522	11.55320	3.197927
8	6.443	6.312	0.131	95.8762	1.489	10.07700	4.482778
9	6.292	6.291	0.001	99.5157	1.493	9.655001	4.487854
10	6.148	6.116	0.033	84.2943	1.465	9.788700	4.490911
11	6.124	6.304	− 0.179	85.7771	1.515	11.39000	3.125317
12	6.124	6.073	0.052	89.4130	1.494	9.002701	3.228987
13	6.113	6.221	− 0.108	84.6082	1.500	10.57440	3.192143
14	6.096	6.124	− 0.027	90.2262	1.497	11.61950	5.717386
15	6.000	6.220	− 0.220	99.5157	1.501	9.718800	4.492193
16	6.000	5.880	0.120	87.0221	1.496	9.833000	4.515000
17	6.045	6.109	− 0.063	89.4130	1.497	9.269899	3.199387
18	6.036	5.810	0.226	89.4130	1.488	8.811399	4.507043
19	5.943	5.931	0.012	89.4130	1.490	9.443800	4.526188
20	5.928	5.836	0.092	93.8589	1.510	9.302600	4.506215
21	5.860	5.914	− 0.054	89.4130	1.511	9.122200	3.249254
22	5.721	5.714	0.007	80.2362	1.501	9.078300	3.212538
23	5.679	5.723	− 0.043	80.2362	1.499	9.098499	3.292225
24	5.420	5.671	− 0.250	89.4130	1.491	9.087100	5.281165
25	4.954	4.925	0.029	80.2362	1.532	9.141100	5.349347

A QSAR model that employed only MR, pIC_{50} (W2) = f [MR], resulted in following statistics: R = 0.514; s = 0.322; F = 6.100; Q^2 = 0.049; S_{PRESS} = 0.366. This means that the hydrophobic parameter, MR, alone can explain 26,4% of the variance in the biological data. A QSAR model that employed only d (C_3–C_{A1}), pIC_{50} (W2) = f [d (C_3–C_{A1})], resulted in following statistics: R = 0.295; s = 0.359; F = 1.617; Q^2 = 0.446; S_{PRESS} = 0.451. This means that the structural parameter, d (C_3–C_{A1}), alone can explain 8.7% of the variance in the biological data. A QSAR model that employed only Lx, pIC_{50} (W2) = f [Lx], resulted in following statistics: R = 0.541; s = 0.316; F = 7.040; Q^2 = 0.106; S_{PRESS} = 0.355. This means that the topological parameter, Lx, alone can explain 29.2% of the variance in the biological data. Similarly, the steric parameter, Verloop B2(A), resulted in following statistics: R = 0.330; s = 0.354; F = 2.075; Q^2 = 0.193; S_{PRESS} = 0.410, therefore, alone can explain 10.9% of the variance in the biological data. The parameters MR and Lx have large contribution in terms of biological activity. Although each variable alone has a small prediction capacity of the biological activity, the prediction capacity increases considerably (Q^2 = 0.752) when the four descriptors were combined in a QSAR model.

A QSAR model that employed only the descriptors MR, Lx and Verloop B2(A), pIC_{50} (W2) = f [MR, Lx, B2(A)] resulted in following statistics: R = 0.840; s = 0.217; F = 11.992; Q^2 = 0.506; S_{PRESS} = 0.281. This means that the steric and hydrophobic parameters can explain approximately 70.5% of the variance in the biological data. This shows the great importance of these parameters.

Figure 2 shows the observed activity versus the predicted activity. The regression model has small residuals (Table 3). Out of the nineteen compounds, fourteen are fitted very well. Only five lie in the confidence limit area and no derivative is considered to be outlier.
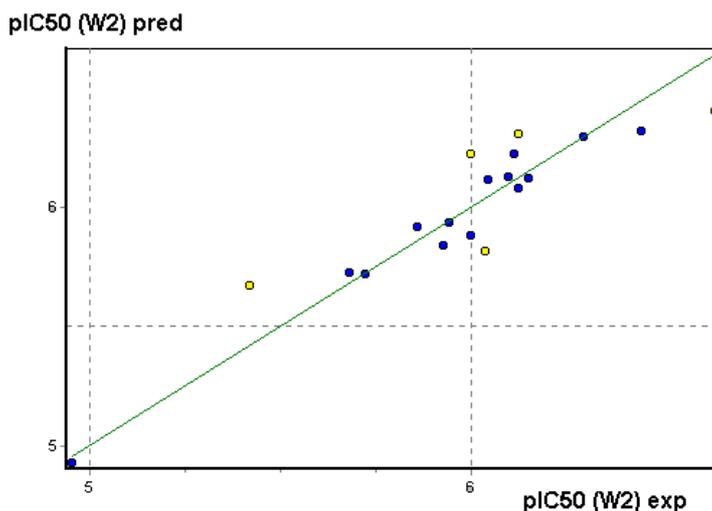
Figure 2. The linear relation between observed and predicted descriptors for the compounds of the dataset.

The model 1 indicates that inhibitory activity of chalcone derivatives against cysteine protease W2 strain depends upon steric and hydrophobic factors. The steric parameter, Verloop B2(A), indicates that substituent A to be width–limited in order to favor the activity. The positive sign of Lx indicates that chalcone molecule should have large length. The positive sign of hydrophobic parameter MR, shows that the compound should present as large volume as possible to increase the activity, because molar refractivity is related to the molecular volume. The negative sign in the d $(C_3–C_{A1})$ parameter indicates that the smaller the distance $C_3–C_{A1}$, the greater will be the biological activity.

The multivariate study of Liu and co–workers [8] detected the influence of the charge on the carbonyl oxygen and the LUMO energy on the chalcone activity against *P falciparum*. These observations indicate that chalcone–cysteine protease interaction may be dependent on a nucleophilic interaction with one or more amino acid residues of the active site of the parasitic enzyme.

The initial step of the peptide bond cleavage mechanism shown in Figure 1 suggests that the cysteine present in the active site could be the nucleophile agent. This interaction can be physicochemical, mediated by intermolecular forces, or it may result of chemical reaction between drug and receptor. In fact, the $C_2–C_3$ π–bonds of the chalcone derivatives are subject to a classical chemical reaction known as Michael addition. From the drug–receptor point of view, addition to chalcone molecule would result in irreversible (suicide) inhibition [24]. While this would be desirable for the chalcone–falcipain interaction, the use of such drugs in humans would result in irreversible damage to the human cysteine protease. In the present case, it appears that there is no chemical reaction between chalcone derivatives and the active site. A study involving the parasitic cysteine protease from *Trypanosoma cruzi*, called Cruzain, which shares 33% of homology with falcipain, has demonstrated to have a reversible interaction with similar chalcones [7]. The d (C_3–

C_{A1}) parameter may be related with C_2–C_3 π–bond of the chalcone derivatives. An important fact is that the studied compounds are capable of taking resonance structures. They can be good Michael's acceptors. The nucleophilicity in the carbon C_3 and C_1 along the α–β unsaturated chain can be explained by resonance mechanism. One can observe the positive partial charge in the carbon C_3 [qC_3] and in the carbon C_1 [qC_1]. There exists high electronic density in the oxygen of the carbonyl evidenced by negative partial charge [qO_4].

Another important point is the fact that the internal environment of the digestive vacuole of *Plasmodium* presents character of weak acid (pH = 5.0–5.5) [7]. Chalcones in acid environment are subject to the protonation in the oxygen of the carbonyl, intensifying the positive partial charge in the carbons C_1 and C_3 and favoring the nucleophilic addition in these carbons. Nucleophiles attacks in carbon C_3 take place through orbital interaction between LUMO of the carbon C_3 with HOMO of the cysteine protease.

The activity of chalcones on D6 strain of *P. falciparum* was best modeled by Eq. (2), denoted as model 2:

$$pIC_{50} (D6) = + 0.016 (\pm 0.003) \text{ MW} - 172.709 (\pm 68.313) \text{ HBN} - 0.143 (\pm 0.07553) \text{ Ly}$$
$$- 0.290 (\pm 0.085) \text{ B2 (A)} + 4.889 (\pm 1.265984) \tag{2}$$
$$n = 19 \quad R = 0.958 \quad R^2 = 0.917 \quad s = 0.131 \quad F = 38.857 \quad p < 0.0001 \quad Q^2 = 0.843 \quad S_{PRESS} = 0.181$$

The data that generated this model are shown in Table 5. The model 2 has good capacity of explaining the observed values of biological activity. It explains 91.7% of the variance in the antimalarial activity against D6 strains, which can also be confirmed by the low standard deviation. The model possesses excellent adjustment level: high correlation coefficient ($R = 0.958$) and low standard deviation ($s = 0.131$). The model also has good predictive capacity ($Q^2 = 0.843$ and $S_{PRESS} = 0.181$). The Table 4 shows correlation matrix among the four descriptors that appears in Eq. (2).

Table 4. Correlation Matrix for Model 2

	MW	HBN	Ly	B2 (A)
MW	1.000	0.002	0.242	0.026
HBN		1.000	0.004	0.089
Ly			1.000	0.090
B2 (A)				1.000

A QSAR model that employed only MW, $pIC50$ (D6) = f [MW], resulted in following statistics: R = 0.523; s = 0.354; F = 6.395; Q2 = 0.097; SPRESS = 0.394. This means that the steric parameter, MW, alone can explain 27.3% of the variance in the biological data. A QSAR model that employed only HBN, $pIC50$ (D6) = f [HBN], resulted in following statistics: R = 0.316; s = 0.394; F = 1.881; Q2 = 0.158; SPRESS = 0.446. This means that the thermodynamic parameter, HBN, alone can explain 9,9% of the variance in the biological data. A QSAR model that employed only Ly, $pIC50$ (D6) = f [Ly], resulted in following statistics: R = 0.201; s = 0.406; F = 0.712; Q2 = 0.269; SPRESS = 0.467. This means that the topological parameter, Ly, alone can explain 4% of the

variance in the biological data. Similarly, the steric parameter, Verloop B2(A), resulted in following statistics: R = 0.472; s = 0.366; F = 4.874; Q2 = 0.020; SPRESS = 0.419. Therefore, B2(A) alone can explain 22.2% of the variance in the biological data.

Table 5. Numerical Values of the Four Selected Descriptors, Observed Activity, Predicted Activity and Residual for QSAR Model 2
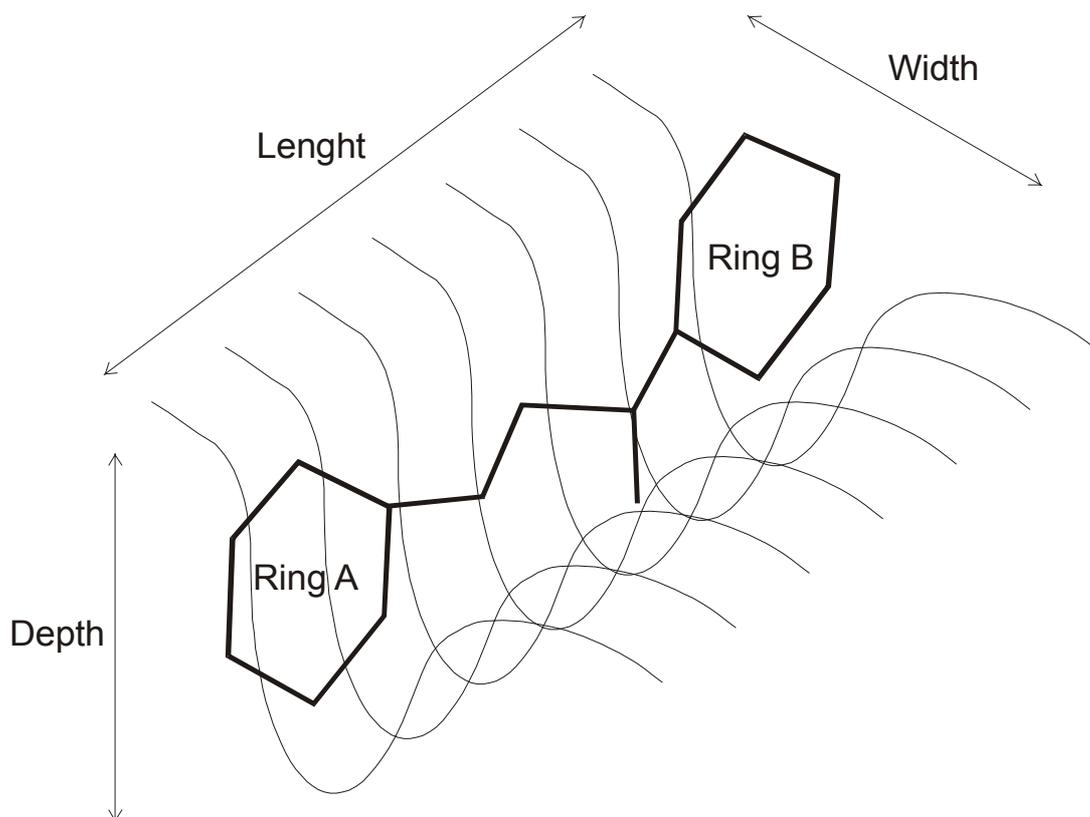
Derivatives	Observed Activity	Predicted Activity	Residual	MW	HBN	Ly	B2 (A)
7	6.721	6.736	−0.015	328.1967	0.003046953	13.7285	3.197927
8	6.366	6.272	0.094	367.2281	0.005446205	15.8380	4.482778
9	6.142	5.938	0.205	353.8047	0.004895499	17.3179	4.487854
10	5.596	5.491	0.106	294.7397	0.003392824	15.6241	4.490911
11	6.167	6.139	0.028	336.3105	0.005150153	16.4127	3.125317
12	6.008	6.178	−0.169	337.2018	0.005136542	16.0490	3.228987
13	5.906	5.801	0.106	302.7570	0.005720926	14.1902	3.192143
14	5.772	5.876	−0.104	328.1967	0.003046953	14.6143	5.717386
15	5.889	5.948	−0.059	353.8047	0.004895499	17.2375	4.492193
16	6.008	6.007	0.002	329.7328	0.003032758	16.3316	4.515000
17	6.113	6.170	−0.057	337.2018	0.005136542	16.1618	3.199387
18	5.958	5.843	0.116	337.2018	0.005136542	15.7934	4.507043
19	5.707	5.712	−0.004	337.2018	0.005136542	16.6691	4.526188
20	6.045	6.007	0.039	343.2115	0.002913655	18.0037	4.506215
21	5.889	6.141	−0.251	337.2018	0.005136542	16.2655	3.249254
22	5.709	5.655	0.055	304.2932	0.005692047	15.3741	3.212538
23	5.677	5.598	0.079	304.2932	0.005692047	15.6064	3.292225
24	5.468	5.457	0.011	337.2018	0.005136542	16.9126	5.281165
25	4.718	4.903	−0.184	304.2932	0.005692047	16.2847	5.349347
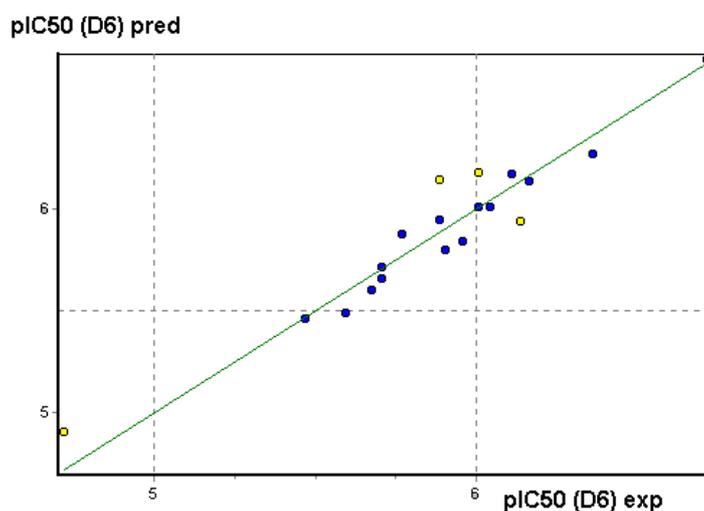
Figure 3. The linear relation between observed and predicted descriptors for the compounds of the dataset.

A QSAR model that only employed of the descriptors MW, HBN and Verloop B2(A), pIC_{50} (D6) = f [MW, HBN, B2(A)] resulted in following statistics: $R = 0.906$; $s = 0.187$; $F = 22.900$; $Q^2 = 0.679$; $S_{PRESS} = 0.250$. This means that, the steric and thermodynamic parameters can explain approximately 82.0% of the variance in the biological data. The steric parameter MW presents large contribution in terms of biological activity. When we combined the four descriptors in a linear

http://www.biochempress.com

QSAR model, the prediction capacity increases to $Q^2 = 0.843$.

Figure 4. The active site of cysteine protease where chalcone derivatives bind to possess a cleft on the enzyme surface. The width of the molecule matches the depth of the active site and vice–versa.

Figure 3 shows the observed activity versus the predicted activity. The regression model proposed has no compounds with high residuals (Table 4). Out of the nineteen compounds participants of the model, fifteen are fitted very well, four lie in the confidence limit area and no derivative was considered to be outlier.

Eq. (2) has two steric parameters. B2(A) has already appeared in Eq. (1). L_y is a molecular parameter and it refers to the maximum width of the molecule when its highest dimension is aligned to the *x*–direction. According to Eq. (2), the molecular width of chalcone derivatives should be small in order to increase activity. The interpretation of L_y at the molecular level requires some knowledge of active site. There are evidences that chalcones interact with a surface cleft of cysteine protease [7], in which the width of drug molecule matches the depth of the active site (Figure 4). The limited molecular width indicated by parameter L_y can be related to a limited depth in the active site depth. A wide molecular structure may not have an ideal interaction with the site, because part of its width would be out of the cleft.

The positive sign of thermodynamic parameter MW, shows that the chalcone derivatives should have as large volume as possible to increase the activity. This property can be related with the molar

refractivity and with the molecular volume. The molecular weight is also considered as bioavailability parameter. In terms of central nervous system, the acceptable molecular weight should be smaller than 600 a.u. In terms of oral bioavailability, the acceptable molecular weight should be smaller than 1000 a.u. All derivatives studied present admissible oral and central nervous system bioavailability. The presence of MW in Eq. (2) could be related to the molecular transport across biological membranes.

The previous structure–activity studies on chalcone derivatives [7–8] concluded that the presence of methoxy groups on substituent B favors drug activity. This may also give some meaning to the presence of molecular weight property (MW) in Eq. (2), whose positive coefficient sign is in agreement with an increasing number of methoxy groups in the molecular structure. Another model can be derived to help explaining the activity of chalcones on D6 *P. falciparum* strain. It is shown in condensed form just for qualitative purposes in Eq. (3).

$$pIC_{50} \, (D6) = f \, (+ \, VM; - \, HBN; - \, L_Y; - \, B_{2,A})$$
$$n = 19 \quad R = 0.954 \quad s = 0.138 \quad F = 35.044 \quad Q^2 = 0.823 \quad S_{PRESS} = 0.192 \tag{3}$$

This model is similar to structure of Eq. (2), except the presence of parameter VM. Parameter VM also is a topological measure and possess same signal in the equation. Although slightly different, Eqs. (1) and (2) have two important features in common. The activity on strains W2 and D6 is dependent on the properties of the whole molecule (MR and Lx for W2 strain, and MW and L_y for D6 strain) and on the property of the substituents attached to ring A – Verloop B2(A).

4 CONCLUSIONS

The present QSAR study investigated the factors that may be important in the inhibitory activity of chalcone derivatives on *P. falciparum* cysteine protease. The obtained models presented good capacity to explain the observed values of biological activity, high adjustment level, statistical significance and good predictive capacity. Hydrophobic and steric properties seem to play an important role in the explanation of the activity of the dataset. The results indicated that the activity on W2 and D6 strains is favored if ring A is a width–limited chemical substituent. The limited molecular width of these derivatives can be related with the activity against D6 strain. The molecular weight, which is related to molecular volume, appears to influence only the activity of D6 strain. The results also indicated that molar refractivity and molecular length have positive contribution to the activity against chloroquine–resistant (W2) *Plasmodium falciparum* strains, while molecular weight against mefloquine–resistant (D6) strains. The crystal structure of the parasitic cysteine protease remains undetermined. The availability of the enzyme structure would help researchers to go further in understanding the interactions that dominate chalcone–receptor binding.

Acknowledgment

5 REFERENCES

[1] K. N. Suh,, K. C. Kain, J. S. Keystone, Malaria. *Canadian Med. Assoc. J.* **2004**, *170*, 1693–1702.

[2] N. J. White, Antimalarial Drug Resistance. *J. Clin. Investigation* **2004**, *113*, 1084–1092.

[3] M. L. Go, Novel Antiplasmodial Agents. *Med. Res. Rev.* **2003**, *23*, 456–487.

[4] M. Chen, S. B. Christensen, L. Zhai, M. H. Rasmussen, T. G. Theander, S. Frokjaer, B. Steffansen, J. Davidsen, A. Kharazmi, The Novel Oxygenated Chalcone, 2,4–Dimethoxy–4'–butoxychalcone, Exhibits Potent Activity Against Human Malaria Parasite *Plasmodium falciparum In Vitro* and Rodent Parasites *Plasmodium berghei* and *Plasmodium yoelii In Vivo*. *J. Infectious Diseases* **1997**, *176*, 1327–1333.

[5] M. Chen, T. G. Theander, S. B. Christensen, L. Hviid, L. Zhai, A. Kharazmi, Licochalcone–A, A New Antimalarial Agent, Inhibits *In–Vitro* Growth Of The Human Malaria Parasite *Plasmodium falciparum* And Protects Mice From *P. Yoelii* Infection. *Antimicrob. Agents Chemother.* **1994**, *38*, 1470–1475.

[6] A. Kharazmi, M. Chen, T. Theander, S. B. Christensen, Discovery of Oxygentade Chalcones as Novel Antimalarial Agents. *Annal. Trop. Med. Parasitol.* **1997**, *91*, S91–S95.

[7] R. S. Li, G. L. Kenyon, F. E. Cohen, X. W. Chen, B. Q. Gong, J. N. Dominguez, E. Davidson, G. Kurzban, R. E. Miller, E. O. Nuzum, P. J. Rosenthal, J. H. McKerrow, *In Vitro* Antimalarial Activity of Chalcones and Their Derivatives. *J. Med. Chem.* **1995**, *38*, 5031–5037.

[8] M. Liu, P. Wilairat, M. L. Go, Antimalarial Alkoxylated and Hydroxylated Chalones: Structure–Activity Relationship Analysis. *J. Med. Chem.* **2001**, *44*, 4443–4452.

[9] F. Salas, J. Fichmann, G. K. Lee, M. D. Scott, P. J. Rosenthal, Functional Expression of Falcipain, a *Plasmodium falciparum* Cysteine Proteinase, Supports Its Role as a Malarial Hemoglobinase. *Infection Immunity* **1995**, *63*, 2120–2125.

[10] B. R. Shenai, P. S. Sijwali, A. Singh, P. J. Rosenthal, Characterization of Native and Recombinant Falcipain–2, a Principal Trophozoite Cysteine Protease and Essential Hemoglobinase of *Plasmodium falciparum*. *J. Biol. Chem.* **2000**, *275*, 29000–29010.

[11] P. S. Sijwali, B. R. Shenai, J. Gut, A. Singh, P. J. Rosenthal, Expression and Characterization of the *Plasmodium falciparum* Haemoglobinase Falcipain–3. *Biochem. J.* **2001**, *360*, 481–489.

[12] Y. Sabnis, P. J. Rosenthal, P. Desai, M. A. Avery, Homology Modeling of Falcipain–2: Validation, De Novo Ligand Design and Synthesis of Novel Inhibitors. *J. Biomol. Struct. Dynamics* **2002**, *19*, 765–774.

[13] Y. A. Sabnis, P. V. Desai, P. J. Rosenthal, M. A. Avery, Probing the Structure of Falcipain–3, a Cysteine Protease from *Plasmodium falciparum*: Comparative Protein Modeling and Docking Studies. *Protein Sci.* **2003**, *12*, 501–509.

[14] *Molecular Modeling Pro*, Computational Chemistry Program, 4, ChemSW, Inc.: Fairfield, **2001**.

[15] N. L. Allinger, Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.

[16] *Hyperchem: Molecular Modeling System*, 6, Hypercube, Inc.: Florida, **2000**.

[17] T. M. Dyott, A. J. Stuper, G. S. Zander, Moly – Interactive System For Molecular Analysis. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 28–35.

[18] *Interactive 3D Molecular Modeling*, 5, ChemSW, Inc.: Fairfield, **2001**.

[19] *Arguslab 4.0*. Mark Thompson and Planaria Software LLC, **2004**.

[20] J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods II. Applications, *J. Comput. Chem.* **1989**, *10*, 221–264.

[21] S. H. Unger, and C. Hansch, On Model Building in Structure–Activity Relashionships. A Reexamination of Adrenergic Blocking Activity of β–halo–β–arilalkylamines. *J. Med. Chem.* **1973**, *16*, 745–749

[22] J. G. Topliss, and R. J. Costello, Chance Correlations in Structure–Activity Studies using Multiple Regression Analysis. *J. Med. Chem.* **1972**, *15*, 1066–1068.

[23] D. B. De Oliveira, A. C. Gaudio, BuildQSAR: A New Computer Program for QSAR Studies. *Quant. Struct.–Act. Relat.* **2000**, *19*, 599–601.

[24] G. Thomas, Química Medicinal – Uma introdução. Ed. *Guanabara Koogan* **2003**, 187.