2D Chemometrics Analyses of Tetrahydroquinoline and Ethylenediamine Derivatives with Antimalarial Activity

Humberto Fonseca de Freitas and Marcelo Santos Castilho*

Faculdade de Farmácia, Universidade Federal da Bahia (UFBA), Salvador, Bahia, 40170-115

Abstract: Malaria, one of the most widespread and deadly infectious diseases continues to kill over 1 million people every year. This scenario is getting even worse as *P. falciparum* develops resistance to existing drugs. Thus, there is an imperative need for novel and more effective antimalarials. Farnesyltransferase (PFT) appears to be a promising therapeutic target to development of antimalarial drugs and many analogs of PFT inhibitors have proved active against *P. falciparum*. In order to shed some light on the structure-activity relationships of 192 tetrahydroquinoline and ethylenediamine derivatives that are active against *P.falciparum*, exploratory analysis as well as classical and hologram QSAR strategies were employed. No global QSAR could be developed for the whole dataset, instead local QSAR models were developed for 118 compounds (classical QSAR r^2=0.78, q^2=0.75, r^2_{pred}= 0.77 with 2 PCs; HQSAR r^2=0.82, q^2=0.72, r^2_{pred}= 0.79 with 3 PCs) and 74 compounds (r^2=0.79, q^2=0.74, r^2_{pred}= 0.57 with 2PCs; r^2=0.86, q^2=0.77, r^2_{pred}= 0.75 with 4 PCs) using partial least square (PLS) regression. Furthermore, the careful and integrated analysis of contribution maps and regression vector suggest that these inhibitors might have dissimilar requirements to their biological activity.

Keywords: Antimalarial activity, chemometric analysis, classical 2D QSAR, ethylenediamine, hologram QSAR, tetrahydroquinoline.

INTRODUCTION

Malaria, one of the most widespread and deadly infectious diseases [1] kills over 1 million people every year and nearly 50% of the world population is at stake [2]. This scenario is getting even worse as *P. falciparum*, the protozoan parasite responsible for the majority of malaria mortality [3], develops resistance to existing drugs (chloroquine, mefloquine, sulfadoxime/pyrimethamine) [4], the vector eliminating campaigns are limited by *Anopheles* mosquito resistance to the pesticide DDT and the global warming contributes to malaria spread [5]. Thus, there is an imperative need for novel and more effective antimalarials [6]. In order to accomplish this goal, attention has been directed toward parasite biochemical pathways that are absent in the human host or that allow selective inhibition [7-12]. Among several macromolecular targets for antimalarial drug development [14-16], the protein farnesyltransferase (PFT) appears to be a promising therapeutic target, once it plays a pivotal role in a posttranslational modification of proteins (protein prenylation) that facilitates membrane association as well as protein-protein interactions [13]. In the human host, the protein prenylation is mediated by three different enzymes: PFT, protein geranylgeranyltransferase type I (PGGT-I) and protein geranylgeranyltransferase type II (PGGT-II). However, the analysis of *P. falciparum* genome indicates an apparent lack of PGGT-I [17], hence suggesting that the parasite would rely on PFT only for its prenylation requirements. In fact, a number of PFT inhibitors, previously developed as anticancer agents, were found to be highly toxic to *P. falciparum* [11].

Drug design efforts are made easier when the selected biochemical target has already been explored for the treatment of other diseases, inasmuch as pharmacological information and initial lead compounds can be readily available [10, 11, 13]. For that reason many groups have synthesized analogs of PFT inhibitors [10, 11, 13, 18-27] (Fig. **1**) and studied their structure-activity requirements for the selective inhibition of parasite's enzyme [28].

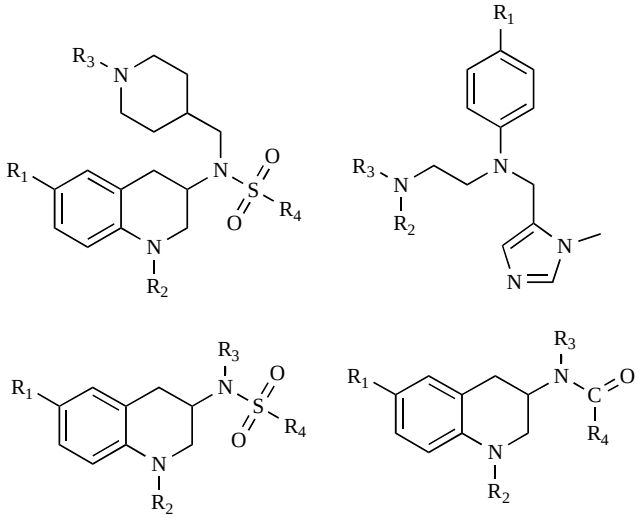

Fig. (1). Representative scaffold of PFT inhibitors analogs with antimalarial activity.R_1, R_2, R_3 and R_4 stand for the different moieties that can be found in the inhibitors scaffold (see supplementary material S1).

*Address correspondence to this author at the Faculdade de Farmácia, Universidade Federal da Bahia (UFBA), Salvador, Bahia, 40170-115; Tel: +55-071-3283-6930; Fax: +55-017-3283-6949; E-mail: castilho@ufba.br

1875-6638/12 $58.00+.00

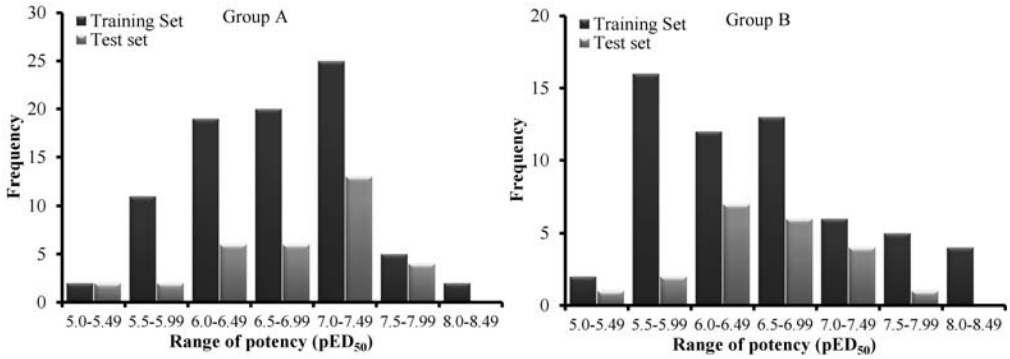

Fig. (2). Histogram of potency distribution of group A and B.

Nevertheless, as far as we are aware, no effort was made to investigate this matter from a quantitative stand point of view.Considering that QSAR studies have played an important role in the design of many antimalarial compounds [29-31], it would also be useful to employ this approach in the development of second generation PFT inhibitors. Accordingly, descriptor and fragment based QSAR models were developed, which not only show good predictive power, but also underscore the dissimilar chemical and structural requirements for biological activity among this series of inhibitors.

MATERIALS AND METHODS

Data Set

A data set of 192 tetrahydroquinoline (THQ) and ethylenediamine (EDA) derivatives, that were designed as PFT inhibitors [11, 18, 23], was employed for the QSAR studies. All structures were drawn in SYBYL 8.0 platform and single point optimized using AM1 method (Keywords: 1SCF XYZ ESP NOINTER SCALE=1.4 NSURF=2 SCINCR=0.4 NOMM). The biological activity (ED_{50}– the concentration required to inhibit 50% of parasite growth) of these compounds was converted to pED_{50} ($-\log ED_{50}$) and used as dependent variable in the QSAR modeling. The chemical structures and biological data are listed in Tables **S1** and **S2** (Supplementary material). A Hierarchical Cluster Analysis (HCA), using Euclidean distances and autoscaled variables, guided the division of the complete data into two smaller data sets (A-118 compounds and B-74 compounds) that were further split into training (compounds A 1–89, B 119–179) and test (compounds A 90–118 and B 180–192) set for external validation purpose. Special attention was paid to guarantee that test sets represent the structural diversity of the training sets and show a similar distribution of potency (Fig. **2**).

Classical QSAR Models

2D molecular descriptors such as topological descriptors, connectivity index, 2D correlation descriptors and so forth were computed with DRAGON5.5 software (Taletesrl., Milano, Italy) and used as independent variables in QSAR modeling.

2489 calculated descriptors were then subjected to the following selection criteria; descriptors with constant values, pair correlation of 99% or superior and found to have poor correlation to biological property ($r^2 < 0.10$) were discarded. This strategy afforded 1008 descriptors that were employed to develop multiple linear regression (MLR) models with up to 4 variables as described below.

Variable selection was carried out by genetic algorithm (GA) [32], as available in MOBYDIGS 1.0 software (Taletesrl., Milano, Italy) using the following parameters: QUIK rule (0.005), asymptotic Q2 rule (-0.005), redundancy RP rule (0.1) and overfitting RN rule (0.01) [33, 34]. These criteria were employed to discard highly correlated variables or those completely uncorrelated to the biological activity during MLR model development. Due to the stochastic nature of genetic algorithm, the search was performed in ten independent populations of 100 models. Each population evolved for 1000 generations or at least until 10^6 (one million) models were evaluated. The variables from the best 5 models of each population were polled together, autoscaled and used for Principal Components Analyses (PCA) and Partial Least Square (PLS) analysis in the PIROUETTE 4.0 software.

HQSAR Models

Statistical HQSAR modeling was carried out as previously described [34, 35]. Briefly, each molecule is broken down into several unique structural fragments which are arranged to form a molecular hologram, an extended form of fingerprint that encodes all possible molecular fragments (e.g., linear, branched, cyclic, overlapping). With the transformation of the chemical representation of a molecule into its corresponding molecular hologram, this method requires no explicit 3D information for the ligands (e.g. determination of 3D structure, putative binding conformations, and molecular alignment). Furthermore, the molecular hologram accounts for the frequency of each molecular fragment type in length fixed bins (hologram length), according to a set of rules, such as size and type of fragment substructures. For that reason, several combinations of fragment distinction and size were considered during the QSAR modeling runs and holograms were generated using 7 distinct fragment sizes over the 12 default series of hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins). The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using PLS analysis. All models generated in our studies were investigated using the full cross-validated r^2 (q^2) Partial Least Squares (PLS) Leave-One-Out (LOO) method. The models internal consistency was also evaluated by means of Leave-Group-Out cross validation.

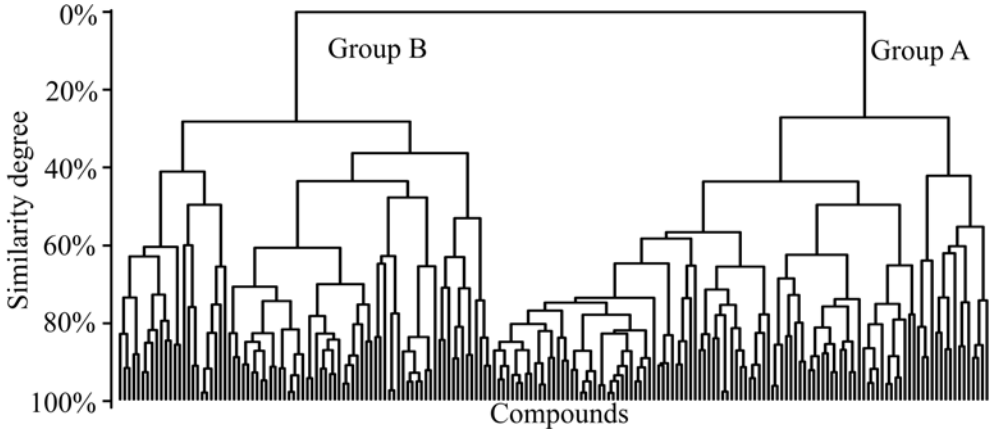

Fig. (3). Hierarchical Cluster Analysis of 192 tetrahydroquinoline and ethylenediamine derivatives with antimalarial activity. The dendogram highlights the structural diversity of the dataset, that splits apart into group A and B at 25% similarity degree.

QSAR Models Validation

The ability to predict compounds not employed during QSAR model development was evaluated by external validation using a test set of 29 compounds for model A and 13 compounds for model B. The test set compounds were selected randomly and none of these compounds were employed during model development. The QSAR models´ predictive power was calculated as described by Schuurmann and coworkers [36]:

$$r^2_{pred} = 1 - \frac{\Sigma_{i=1}^{N}(y_i^{pred} - y_i)^2}{\Sigma_{i=1}^{N}(y_i - y_{test\ mean})^2} \quad (1)$$

where N is the number of compounds; y_i is the experimental value of potency for the i-th compound; y_i^{pred} is the predicted value of potency for the i-th compound; $y_{test\ mean}$ is the mean potency of the test set compounds.

And further evaluated by the $r^2_{m\ (test)}$ values [37-39]:

$$r^2_{m(test)} = r^2 * (1 - \sqrt{r^2 - r_0^2}) \quad (2)$$

were r^2 and r_0^2 are the correlation coefficient values between the observed and predicted values of the test set compounds with and without zero as the intercept, respectively.$r^2_{m(test)}$ values greater than 0.5 are expected in QSAR model with good external predictability.

RESULTS AND DISCUSSION

QSAR models have been widely used to further understand the structural and physicochemical features that are crucial for the biological property of a series of compounds, as well as to guide their potency and selectivity optimization [40-43]. However, QSAR models are best fitted to describe compounds potency when no activity cliff is present. To some extent, the sudden changes in potency, following minor modification in the compounds´ structure, explain why global QSAR models show lower predictive ability than local ones. Yuan *et al.* argued that global models are too coarse to capture detailed and subtle structure-activity relationships found in the activity landscape [44]. Besides it has been pointed out that structurally diverse inhibitors might bind differently to the biological target thus clouding the predictive ability of QSAR models [45, 46]. In order to investigate the impact of these points in the present work, a hierarchical cluster analysis of the whole dataset, using 1008 2D descriptors, was carried out before any supervised modeling step was pursued. The HCA clearly shows that the 192 compounds, previously described as PFT inhibitors [11, 18, 23], cluster in two different families, group A and B (Fig. **3**). No clear structural or biological feature that might suggest the presence of outliers was evident in each group, but a global QSAR model using the whole dataset provided poor statistical values (data not shown). Therefore, it was decided to investigate if local QSAR models would provide better results.

Thus, the 118 compounds from group A were randomly split into training (1–89) and test set (90-118), as described in the experimental section. Then, MOBYDIGS 1.0 software was employed to search for multiple linear regression (MLR) models, with up to 4 variables, through genetic algorithm.

This strategy aimed at selecting a subset of uncorrelated descriptors that could be useful for QSAR model development. The best MLR model shows moderate statistical parameters (n=89, r^2=0.72, q^2=0.70), with poor predictive ability (r^2_{pred}=0.65).

Aiming at increasing the QSAR model´s predictive ability, we resorted to more powerful statistical tools such as PLS and PCA available in PIROUETTE 4.0 software. For that purpose, the descriptors found in the 5 best models from each population (42 descriptors from 27 best models – Supplementary material Table **S3**) were gathered and used for further QSAR model development.

PCA shows that the 3 first principal components (PCs) are responsible for 73.5% of accumulated variance and that PC2 broadly accounts for inhibitors potency: All weak inhibitors (pED_{50}< 6.0) have positive PC2 values, 95% of moderate inhibitors (6.0 < pED_{50} < 7.0) have PC2 values ranging from 0.5 to -1.6 and >99% of potent inhibitors (pED_{50} > 7.0) have negative PC2 values (Fig. **4**).

This preliminary result suggests that selected descriptors somehow explain inhibitors different biological profile towards *P. falciparum* and hence deserve further investigation. Analysis of the PCA loading plot reveals that nTB (–0.45), nArCN (–0.45), N-074 (–0.45), Hy (0.34) and H-050 (0.36)

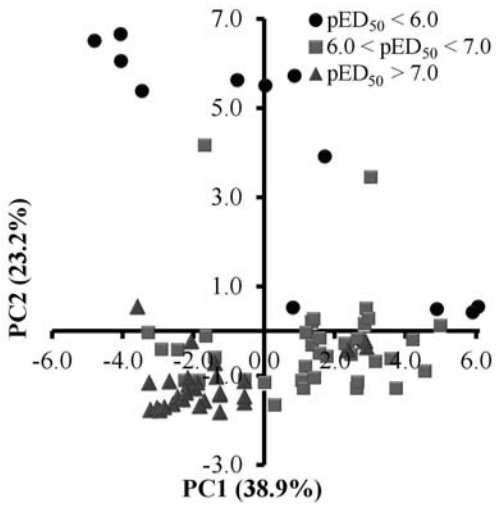

Fig. (4). Principal component analysis of group A compounds.Score plot shows the positioning of potent, moderate and weak inhibitors according to PC1 and PC2.

have a major contribution to PC2 and thus can be useful to explain the potency variation among this series.

The number of triple bonds (nTB) and of nitriles bonded to aromatic ring (nArCN) highlight the importance of such fragments to increase potency. Likewise, an optimal balance of hydrophobic and hydrophilic contributions must be critical to activity, as evaluated by the importance of Hy (hydrophilic index) [47] to the second principal component. Hy accounts for molecule hydrophilicity through the following ratio: sum of number of hydrophilic groups (eg. number of OH, NH, SH, etc.) plus carbons atoms divided by the number of atoms, except for Hydrogen. Although log P measures a somewhat related property these variables are poorly correlated (r≈ 0.55).

In order to further explore the information contained in these descriptors, PLS analysis was carried out next. The initial QSAR model shows a decline in statistical parameters (r^2= 0.52 and q^2= 0.45, 2PCs), however the further iterative exclusion of descriptors that have minor contribution to the regression vector significantly improved the QSAR model.

The final QSAR model, containing 19 descriptors (Table **1**), shows higher internal consistency (r^2= 0.78, q^2= 0.75, 2 PCs) and predictive ability (r^2_{pred}= 0.77and $r^2_{m(test)}$=0.67) than those of preliminary MLR models (Fig. **5** and Table **2**).

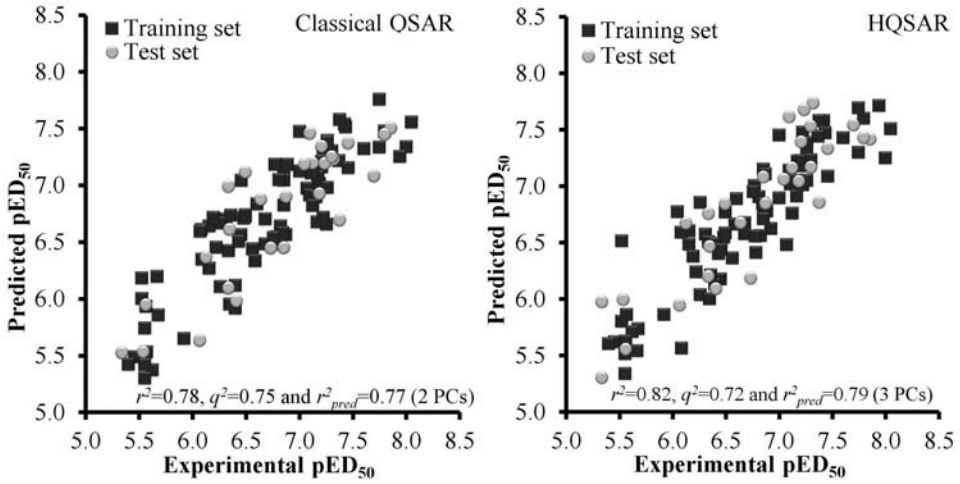

Fig. (5). Plot of predicted versus experimental values of pED_{50} for training and test sets of group A, according to 2D classical QSAR models (left panel) and hologram QSAR (right panel).

Table 1. Descriptors Found in the Final Classical QSAR Model for Group A Compounds

Variables	Description
BEHm8	Highest eigenvalue n. 8 of Burden matrix / weighted by atomic masses
C-002	Ghose-Crippen atom-centred fragments / CH_3R or CH_4
C-025	Ghose-Crippen atom-centred fragments / R– –CR– –R[a]
H-050	Ghose-Crippen atom-centred fragments / H attached to heteroatom
N-074	Ghose-Crippen atom-centred fragments / R#N[b] or R=N–
EEig04r	Eigenvalue 04 from edge adj. matrix weighted by resonance integrals
EEig13r	Eigenvalue 13 from edge adj. matrix weighted by resonance integrals
EEig04x	Eigenvalue 04 from edge adj. matrix weighted by edge degrees
ESpm01x	Spectral moment 01 from edge adj. matrix weighted by edge degrees
F06[C-N]	Frequency of C-N at topological distance 6

Table 1. contd….

Variables	Description
Hy	Hydrophilic index
nBM	Number of multiple bonds
nTB	Number of triple bonds
nArCN	Number of nitriles (aromatic)
nBnz	Number of benzene-like rings
nCb-	Number of substituted benzene C(sp2)
piPC05	Molecular multiple path count of order 05
piPC07	Molecular multiple path count of order 07
piPC08	Molecular multiple path count of order 08

[a]R stands for any groups linked through carbon and "– –" for aromatic bonds as in benzene or delocalized bond / [b] "#" stands for the triple bond

Table 2. Predictive Power of Classic QSAR (r^2_{pred}=0.77 and $r^2_{m(test)}$=0.67) for Group A Test Set Compounds*

ID	ED_{50} (nM)	pED_{50}		Residue	ID	ED_{50} (nM)	pED_{50}		Residue
		Exp.	Pred.				Exp.	Pred.	
90	230	6.64	6.88	-0.24	**105**	16	7.80	7.45	0.35
91	133	6.88	6.90	-0.02	**106**	50	7.30	7.25	0.05
92	320	6.50	7.12	-0.62	**107**	65	7.19	6.93	0.26
93	450	6.35	6.61	-0.26	**108**	75	7.13	7.19	-0.06
94	2750	5.56	5.95	-0.39	**109**	90	7.05	7.18	-0.13
95	460	6.34	6.99	-0.66	**110**	140	6.85	6.45	0.40
96	14	7.85	7.50	0.35	**111**	62	7.21	7.34	-0.13
97	42	7.38	6.69	0.68	**112**	185	6.73	6.50	0.23
98	390	6.41	6.00	0.43	**113**	35	7.46	7.38	0.08
99	48	7.32	7.23	0.09	**114**	860	6.07	5.64	0.43
100	20	7.70	7.09	0.62	**115**	460	6.34	6.10	0.24
101	80	7.10	7.46	-0.36	**116**	4600	5.34	5.53	-0.19
102	58	7.24	7.19	0.05	**117**	2900	5.54	5.53	0.01
103	50	7.30	7.25	0.05	**118**	4600	5.34	4.94	0.40
104	750	6.13	6.37	-0.24					

*"Exp." stands for experimental values and "Pred." stands for predicted values

Table 3. Influence of Fragment Distinction on the Statistical Parameters of HQSAR Models for Group A Compounds

Model	Fragment Distinction	q^2	r^2	SEE	HL	N	Fragments' Size
1	A/B/C	0.58	0.76	0.440	307	3	4-7
2	A/B/C/H	0.58	0.75	0.441	97	4	4-7
3	A/B/C/Ch	0.58	0.76	0.438	257	3	4-7
4	A/B/C/H/Ch	0.58	0.76	0.441	97	4	4-7
5	A/B/C/H/Ch/DA	0.51	0.75	0.478	97	4	4-7

Table 3. contd….

Model	Fragment Distinction	q^2	r^2	SEE	HL	N	Fragments' Size
6	A/B/Ch	0.62	0.79	0.421	59	4	4-7
7	A/C/Ch	0.62	0.75	0.416	61	3	4-7
8	A/B/H/CH	0.58	0.70	0.455	83	4	4-7
9	A/C/DA	0.52	0.68	0.481	275	3	4-7
10	A/B/Ch	0.71	0.83	0.367	97	4	2-5
11	A/B/Ch	0.72	0.82	0.358	97	3	3-6
12	A/B/Ch	0.56	0.72	0.450	401	3	5-8
13	A/C/Ch	0.66	0.78	0.396	71	3	2-5
14	A/C/Ch	0.66	0.80	0.397	59	4	3-6
15	A/C/Ch	0.60	0.78	0.429	151	4	5-8

q^2, cross-validated correlation coefficient; r^2, noncross-validated correlation coefficient; SEE, noncross-validated standard error; HL, hologram length; *N*, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

Table 4. Predictive Power of HQSAR Model (r^2_{pred}=0.79 and $r^2_{m(test)}$=0.64) For Group A Test Set Compounds*

ID	ED_{50} (nM)	pED_{50}		Residue	ID	ED_{50} (nM)	pED_{50}		Residue
		Exp.	Pred.				Exp.	Pred.	
90	230	6.64	6.68	-0.04	**105**	16	7.80	7.43	0.37
91	133	6.88	6.85	0.03	**106**	50	7.30	7.17	0.13
92	320	6.50	6.84	-0.34	**107**	65	7.19	7.04	0.15
93	450	6.35	6.47	-0.12	**108**	75	7.16	7.20	-0.04
94	2750	5.56	5.56	0.00	**109**	90	7.05	7.06	-0.01
95	460	6.34	6.76	-0.42	**110**	140	6.85	7.08	-0.23
96	14	7.85	7.42	0.43	**111**	62	7.21	7.39	-0.18
97	42	7.38	6.86	0.52	**112**	185	6.73	6.19	0.54
98	390	6.41	6.09	0.32	**113**	35	7.46	7.33	0.12
99	48	7.32	7.74	-0.42	**114**	860	6.07	5.95	0.12
100	20	7.70	7.54	0.16	**115**	460	6.34	6.20	0.14
101	80	7.10	7.62	-0.52	**116**	4600	5.34	5.98	-0.64
102	58	7.24	7.68	-0.44	**117**	2900	5.54	6.00	-0.46
103	50	7.30	7.53	-0.23	**118**	4600	5.34	5.31	0.03
104	750	6.13	6.67	-0.54					

*"Exp." stands for experimental values and "Pred." stands for predicted values

It has been shown that hologram QSAR models generally show statistical parameters and predictive ability comparable to those of more complex 3D QSAR approaches [35]. Thus, aiming at developing more statistically sound 2D QSAR models, the compounds from training set A were broken into several overlapping fragments of 4-7 atoms that were employed to generate molecular holograms which were then used as independent variable in PLS analysis. During QSAR model development the influence of fragment distinction over the statistical parameters was investigated (Table **3**). Adding Hydrogen (H) or Chirality (Ch) to fragment distinction provided no improvement in the model (**2** and **3** vs **1**), whilst the further addition of donor and acceptor atoms decreased the quality of the model (**5** vs **1**). On the other hand, the exclusion of either Bond (B) or Connectivity (C), while keeping Ch in fragment distinction increased the internal consistency of the model (**6** and **7** vs **1**). These results underscore the importance of chirality towards the biological activity in group A compounds.

Next, the influence of fragment size over statistical parameter was investigated for the 2 best HQSAR models (**6** and **7**). Smaller fragment sized models show improved statistical parameters (**10** and **11**vs **6**) as well as good predictive ability (r^2_{pred}= 0.79and $r^2_{m(test)}$=0.64) (Table **4** and Fig. **5**). A similar trend was observed for model 7, though to a lower extent.

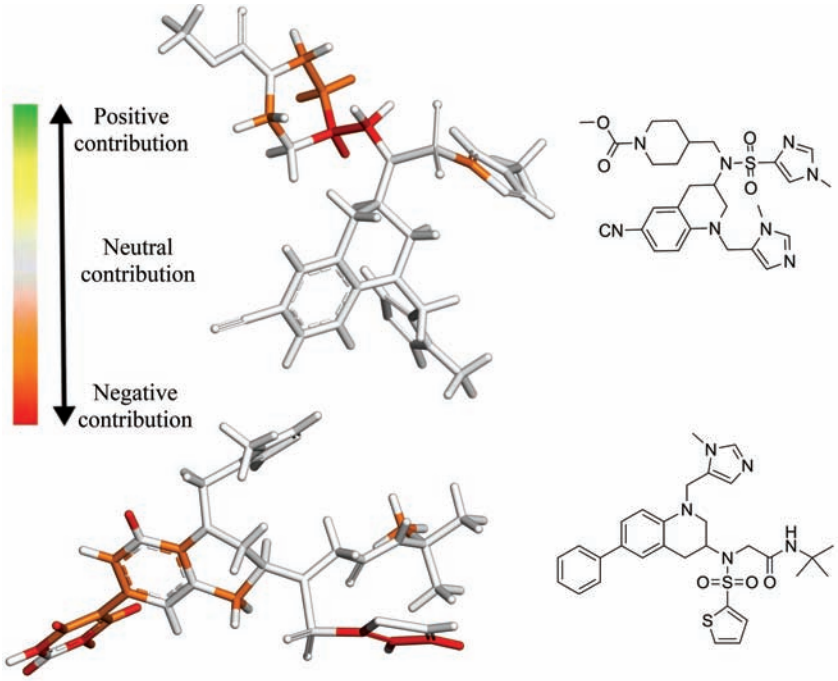

Fig. (6). HQSAR contribution map and the corresponding 2D chemical structure for molecules 20 (ED_{50} = 9 nM) – upper panel – and 62 (ED_{50} = 3500 nM) – lower panel. Both potent and weak inhibitors in group A have contribution maps that highlight mainly detrimental contributions (orange/red) to potency. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

Besides good predictive power, a useful QSAR model should also provide insights into the structure-activity relationships for this series of compounds. This task can be achieved by the analysis of contribution maps from the HQSAR models, which highlight fragments that have positive (green and yellow) or negative (orange and red) contribution to the biological property. Fragments that do not contribute significantly to differentiate compounds are colored white (Fig. **6**).

The contribution map underscores that even in the most potent compound (ED_{50}= 9nM) negative contributions towards the biological property are apparent, suggesting that improved potency can be achieved by careful modification of the inhibitors' scaffold.

Next, it was decided to develop QSAR models for the remaining 74 compounds (group B) using the same steps applied to build group A QSAR models.Thus 74 compounds, randomly split into training (119-179) and test (180-192) set, were employed to build multiple linear regression (MLR) models as available in MOBYDIGS 1.0 software. Once again, MLR model with moderate statistical parameters (n=61, r^2=0.74, q^2=0.71 and r^2_{pred}=0.69) were obtained. Hence, PLS and PCA tools, available in PIROUETTE 4.0 software were employed to build statistically sound QSAR models. Accordingly 37 descriptors (Supplementary material Table **S4**) were pooled together and used for further QSAR model development. Unsupervised exploratory analysis shows that the 3 first principal components (PCs) of PCA account for 69.4% of accumulated variance and the analysis of score plot shows that although all potent *Pf*PFT inhibitors (pED_{50}>7.0) show positive PC2 values, weak and moderate inhibitors are scattered along this axis (Fig. **7**).

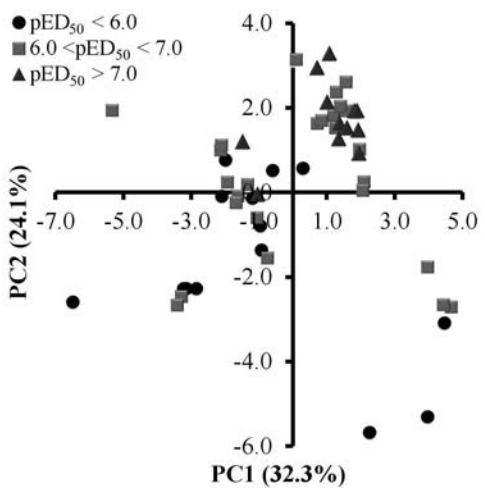

Fig. (7). Principal component analysis of group B compounds.Score plot shows the positioning of potent, moderate and weak inhibitors according to PC1 and PC2.

Hoping that supervised chemometric tools would provide more detailed information on the structure-activity relationship of group B compounds, PLS analysis was carried out.

A similar iterative optimization protocol, that discards descriptors with minor contribution to the regression vector, allowed the selection of 20 descriptors (Table **5**), which lead to a QSAR model (r^2=0.79, q^2=0.74, 2 PCs) statistically equivalent to the one found for group A compounds, but whose descriptors contribution to the final QSAR model differ significantly from those seen in the group A QSAR model (See the comparison of group A and B QSAR models below).

Table 5. Descriptors Found in the Final Classical QSAR Model for Group B Compounds

Variable	Description
C-033	Ghose-Crippen atom-centred fragments / R– –CH..X[a]
N-073	Ghose-Crippenatom-centredfragments /Ar_2NH / Ar_3N / Ar_2N-Al / R..N..R
N-074	Ghose-Crippen atom-centred fragments / R#N[b] or R=N–
EEig04r	Eigenvalue 04 from edge adj. matrix weighted by resonance integrals
EEig12d	Eigenvalue 12 from edge adj. matrix weighted by dipole moments
ESpm03d	Spectral moment 03 from edge adj. matrix weighted by dipole moments
ESpm05d	Spectral moment 05 from edge adj. matrix weighted by dipole moments
F06[C-N]	Frequency of C-N at topological distance 6
LP1	Lovasz-Pelikan index (leading eigenvalue)
MAXDP	Maximal electrotopological positive variation
MPC08	Molecular path count of order 08
nArCN	Number of nitriles (aromatic)
nBnz	Number of benzene-like rings
nTB	Number of triple bonds
VRe2	Average eigenvector coefficient sum from electronegativity weighted distance matrix
VRm2	Average Randic-type eigenvector-based index from mass weighted distance matrix
VRp2	Average eigenvector coefficient sum from polarizability weighted distance matrix
X4A	Average connectivity index chi-4

[a]R stands for any groups linked through carbon and "– –" for aromatic bonds as in benzene or delocalized bond / [b]"#" stands for the triple bond

Table 6. Predictive Power of Classical QSAR Model (r^2_{pred}=0.57 and $r^2_{m(test)}$=0.56)For Group B Test Set Compounds

ID	ED_{50} (nM)	pED_{50}		Residue	ID	ED_{50} (nM)	pED_{50}		Residue
		Exp.	Pred.				Exp.	Pred.	
180	510	6.29	6.36	-0.07	**187**	175	6.76	7.07	-0.31
181	230	6.64	6.55	0.09	**188**	120	6.92	6.71	0.21
182	450	6.38	6.89	-0.49	**189**	750	6.13	6.37	-0.24
183	2100	5.68	5.99	-0.31	**190**	230	6.64	6.58	0.06
184	65	7.19	7.53	-0.34	**191**	700	6.16	5.85	0.31
185	21	7.68	7.83	-0.15	**192**	260	6.59	6.37	0.21
186	4100	5.39	5.90	-0.51					

*"Exp." stands for experimental values and "Pred." stands for predicted values

The predictive power of the group B QSAR model (r^2_{pred}=0.57and $r^2_{m(test)}$=0.56), evaluated by means of external validation as described previously (Fig. **8** and Table **6**) would be considered unacceptable for most drug design efforts, so we decided to focus on hologram QSAR models in the search for improved statistical parameters and further insight into the reasons that rendered global QSAR modeling unsuccessful.

As a consequence, HQSAR method was applied to investigate if fragments that are important for group B compounds activity differ from those identified for group A compounds. Initially, models with standard 4-7 fragment size and different fragment distinctions were explored, as shown in Table **7**. The same trend observed for group A HQSAR model was identified in group B HQSAR models: use of Chirality (Ch) in fragment distinction improved the model (**18**vs**16**) where-

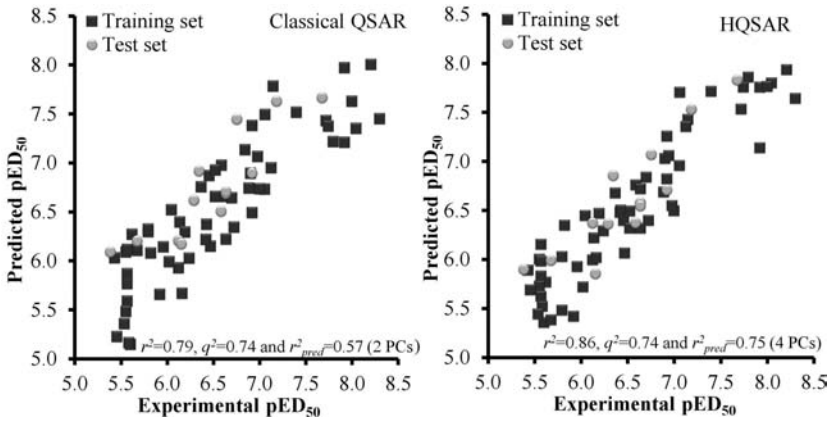

Fig. (8). Predicted versus experimental values of pED_{50} for training and test sets of group B, according to 2D classical QSAR models (left panel) and hologram QSAR (right panel).

Table 7. Influence of Fragment Distinction on the Statistical Parameters of HQSAR Models for Group B Compounds

Model	Fragment Distinction	q^2	r^2	SEE	HL	N	Fragments' Size
16	A/B/C	0.69	0.82	0.473	151	4	4-7
17	A/B/C/H	0.68	0.83	0.476	61	4	4-7
18	**A/B/C/Ch**	**0.71**	**0.83**	**0.451**	**151**	**3**	**4-7**
19	A/B/C/H/Ch	0.65	0.82	0.503	61	4	4-7
20	A/B/C/H/Ch/DA	0.64	0.87	0.511	151	4	4-7
21	A/B/Ch	0.69	0.84	0.471	151	4	4-7
22	A/C/Ch	0.66	0.77	0.487	151	3	4-7
23	A/C/DA	0.66	0.83	0.486	151	4	4-7
24	A/B/C/DA	0.64	0.82	0.512	151	4	4-7
25	A/B/C/Ch	0.67	0.84	0.483	151	4	2-5
26	**A/B/C/Ch**	**0.74**	**0.86**	**0.430**	**71**	**4**	**3-6**
27	A/B/C/Ch	0.70	0.84	0.466	71	4	5-8

as the addition of Hydrogen (H) or donor and acceptor atoms (DA) to fragment distinction decreased the quality of the model (**17**, **20** and **24**vs**16**) (Table **7**).

Smaller sized fragments produced HQSAR models with improved statistical parameters (**26**vs**18**), whose predictive ability was evaluated by means of external validation (Fig. **8** and Table **8**). This result might suggest that HQSAR models are interchangeable, but predicting the biological activity of test set A (90-118) with group B HQSAR models, and vice-versa, proves otherwise (supplementary material Table **S5** and **S6**) and suggests that each group has unique chemical features that are crucial for their biological activity.

Overall, the best HQSAR model for group B compounds (26) shows good predictive ability (r^2_{pred}=0.75 and $r^2_{m(test)}$=0.69) and thus can be useful to shed some light on the structural fragments that are important for activity. This goal can be accomplished by visual analysis of contribution maps, which highlight that both weak and potent compounds show positive contributions to biological activity (Fig. **9**).

Other interesting feature is that moieties with negative influence in group A model, such as the phenyl ring attached to tetrahydroquinoline (THQ) scaffold, have a positive contribution in group B HQSAR model (compare lower panels of Fig. (**6**) and (**9**). In order to further investigate this apparent paradoxical result a detailed comparison of all QSAR models was carried out.

It has been shown that synergic interpretation of hologram QSAR and classical QSAR results afford new insights into the structure activity relationships [43, 48, 49]. For that reason, comparing and contrasting previous models can be very informative for future drug development efforts. The graphical analysis of HQSAR contribution maps highlights that even the most potent compounds of group A suffer from the negative contribution of several fragments. Therefore, it is tempting to assume that more potent inhibitors could be designed, as long as the structural ballast could be either removed or reduced.

Table 8. Predictive Power of HQSAR Model (r^2_{pred}=0.75 and $r^2_{m(test)}$=0.69) for Group B Test Set Compounds*

ID	ED_{50} (nM)	pED_{50}		Residue	ID	ED_{50} (nM)	pED_{50}		Residue
		Exp.	Pred.				Exp.	Pred.	
180	510	6.29	6.62	-0.33	187	175	6.76	180	6.29
181	230	6.64	6.70	-0.06	188	120	6.92	182	6.35
182	450	6.35	6.92	-0.57	189	750	6.13	183	6.64
183	2100	6.64	6.69	-0.05	190	230	6.64	184	7.19
184	65	7.19	7.63	-0.44	191	700	6.16	185	7.68
185	21	7.68	7.66	0.02	192	260	6.59	181	6.64
186	4100	5.39	6.09	-0.70					

*"Exp." stands for experimental values and "Pred." stands for predicted values

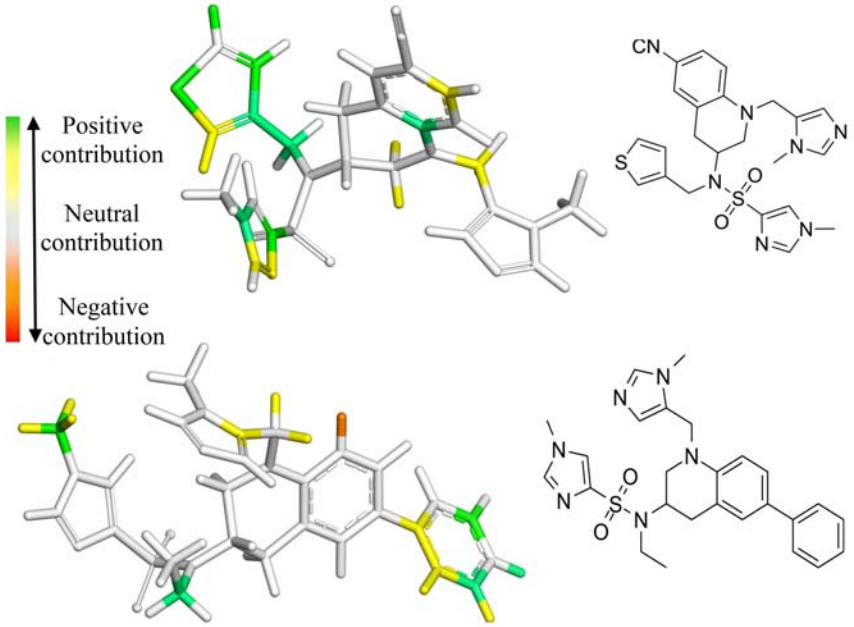

Fig. (9). HQSAR contribution maps and the corresponding 2D chemical structure for molecule 139 (ED_{50}= 12 nM) – upper panel – and 156 (ED_{50} = 5000 nM) – lower panel. Positive contribution (green/yellow) can be found in potent and weak inhibitors throughout this group. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

Unfortunately, this is not an easy task as red/orange colored fragments can be crucial to maintaining the overall bioactive conformation. Besides, the final color scheme results from the contribution of many overlapping fragments, whose individual contributions to activity cannot be easily deconvoluted. However, it is interesting to note that a correlation seems to exist among contribution maps and regression vectors: Most descriptors found in group A regression vector plot show negative contribution and a predominance of red/orange fragments can be found in the contribution maps from group A HQSAR models. The opposite behavior is found in group B regression vector, which shows a larger number of descriptors that positively contribute to potency. This data is in good agreement with hologram QSAR contribution map for group B, inasmuch as the hologram QSAR contribution map also highlights positive contribution to activity. These results underscore the fact that both fragment-based and descriptor-based QSAR models are capturing the same information, but from a different stand point of view. Taking this assumption into consideration we made a careful analysis of descriptors contribution to the regression vector of the final classical QSAR models. The reasoning is that descriptors with large influence to the regression vector (important to the biological activity), might help to explain the chemical and structural features that contribute to the final coloring scheme of contribution maps. The comparison of regression vector plots for group A and B compounds reveals that 6 descriptors (EEig04r, F06[C-N], N-074, nArCN, nBnz and nTB) can be found in both plots (Fig. **10**). This result merely indicates that certain structural [50, 51] and chemical features are essential for antimalarial activity along all dataset.

However, it should be emphasized that descriptors´ load vary from one model to another, for instance nTB, nArCN,

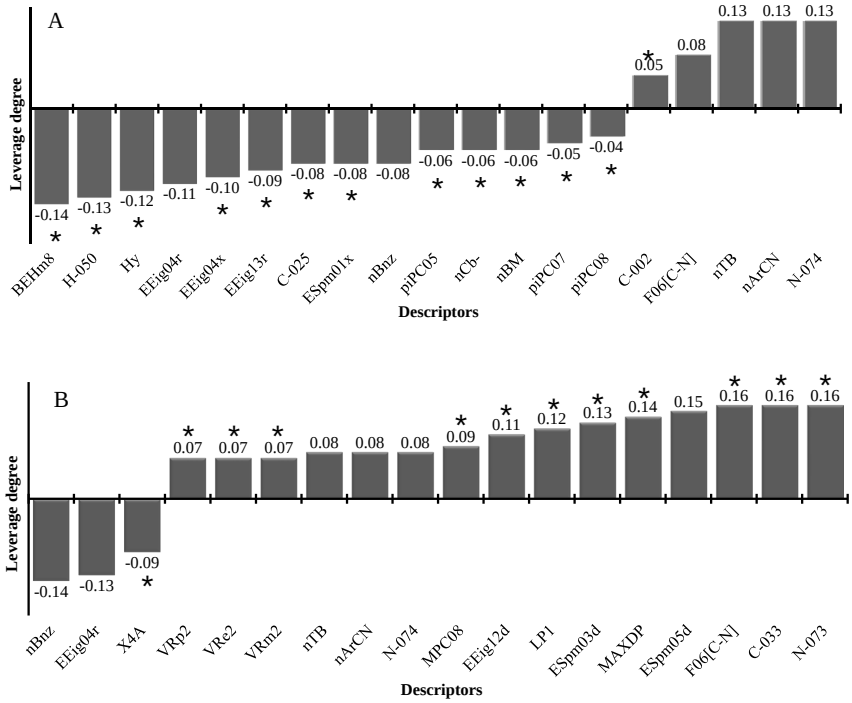

Fig. (10). Regression vector of classical QSAR models for group A compounds (upper panel) and group B compounds (lower panel). Descriptors that are found in one but not in the model are marked with *.

N-074 and F06[C-N] show pronounced contribution to group A regression vector, but only moderate importance to group B regression vector. This data might be a consequence of dissimilar chemical diversity between groups, thus pointing that the number of triple bonds(nTB), the number of nitriles(nArCN), and so forth vary less in group B than in group A compounds. On the other hand, this result could also be rationalized by different chemical requirements for biological activity. Aiming at further explore this hypothesis, our attention was turned to those descriptors that are found in group A QSAR model but not group B QSAR model and vice-versa (highlighted in Fig. **10**).

A straight inspection of regression vector plots clearly shows a predominance of steric-related descriptors in group A QSAR model (BEHm8, piPC05, piPC07, piPC05, nCb, nBm, C-025, C-002) [52-58] whereas group B QSAR model seems to be more influenced by electronic features (E-Eig12d, ESpm03d, MAXDP, ESpm05d) [50, 58]. Thus it is possible that the negative contributions found in group A HQSAR model are the consequence of steric clashes, whereas the positive contributions seen in group B HQSAR model are due to electrostatic interactions.However it must be pointed out that both QSAR models underscore the importance of steric related descriptors to potency, as described below.

A detailed analysis of the descriptors with great positive (N-074) and negative (BEHm8) leverage to the Group A QSAR model shows that the nitrile moiety presence (N-074) [51, 59, 60] increases PFT inhibitors potency, although it is not clear if this is a consequence of electronic or steric properties. On the other hand, BEHm8, which accounts for the highest eigenvalues of the Burden connectivity matrix adjusted by relative atomic mass at the topological distance of 8 bonds [54, 61], clearly indicates that steric features play a crucial role towards PFT inhibition. As BEHm8 values increase with molecular volume it is reasonable to assume that the red-orange observed in HQSAR contribution maps are a consequence of poor complementarity fit inside their binding sites.

A similar approach was employed to investigate the chemical meaning of the most important descriptors for Group B QSAR model. The large negative contribution associated with the number of benzene-like rings (nBnz) [47] emphasizes that increasing molecular volume adversely affects potency. This hypothesis is in good agreement with our interpretation og Group A QSAR model and highlights that steric ballast is counter balanced by the positive contribution of electronic factors in highly potent inhibitors.

N-073, a Ghose-Crippen atom-centered descriptor that accounts for nitrogen heterocyclic aromatic rings (i.e. imidazole ring) [51, 59, 60] has the most positive contribution in Group B QSAR model, hinting at the importance of its interaction to potency. In fact, this feature was also captured in HQSAR contribution maps (see Fig. **9**). Nevertheless, as pointed out before, it is not completely clear whether this is a consequence of steric or electronic characteristics.

Overall, steric complementarity seems to be the most important singular feature underscored by classical QSAR models that shall be optimized in order to develop more potent *P. falciparum* PFT inhibitors.

CONCLUSION

The careful usage of unsupervised analysis tools such as HCA proved crucial to shed some light on the structure-activity relationships of PFT inhibitors once it guided the separation of the original dataset into two smaller families

(Group A and B) that were investigated separately. Next, robust and predictive QSAR models were developed and analyzed independently for each dataset. This strategy afforded an interesting result as both fragment-based QSAR models (HQSAR) and descriptor-based QSAR models underscore that steric and electronic features which are pivotal to the biological activity of group A compounds are not the same as those from group B. A striking difference was observed in HQSAR contribution maps that display contrasting contributions to potency. Whereas molecules from Group A have red/orange colored fragments that negatively impact potency, Group B compounds display contribution maps with a predominance of yellow/green colored fragments, which have positive contribution to potency. The predictive ability of the final models, along with the fact that QSAR models for group A cannot be used to predict Group B-compounds potency and vice-versa, suggest that no statistical artifact can be blamed for this result, leading us to conclude that two separate SAR pathways should be followed in the design of novel and more potent antimalarial compounds. Moreover, the integrated analysis of contribution maps and the chemical interpretation of topological descriptors pointed out that steric, rather than electronic features should be optimized in Group A compounds, whereas electronic properties can be held responsible for the potency of Group B compounds. Nevertheless, the analysis of descriptor-based QSAR models alone suggests that steric complementarity towards their binding site is the main global feature that should be investigated in the pursue for more potent *P. falciparum* PFTinhibitors.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from FAPESB (The State of Bahia Research Foundation) and from CNPq (The National Council for Scientific and Technological Development).

SUPPLEMENTARY MATERIAL

Supplementary data including chemical structure and biological activity of data set compounds (group A - Table **S1**, group B – Table **S2**); Complete set of descriptor employed for PLS model development (group A compounds - Table **S3**, group B compounds - Table **S4**) is available on the publisher's web site along with the published article.

REFERENCES

[1] Sahu, N. K.; Sahu, S.; Kohli, D. V. Novel molecular targets for antimalarial drug development. *Chem. Biol. Drug Des.* **2008**, *71*(4), 287-297.

[2] WHO. *World malaria report 2008*; World Health Organization: Geneva, **2008**.

[3] Snow, R. W.; Guerra, C. A.; Noor, A. M.; Myint, H. Y.; Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, **2005**, *434*(7030), 214-217.

[4] Hyde, J. E.Drug-resistant malaria - an insight. *FEBS*, **2007**, *274*(18), 4688-4698.

[5] Greenwood, B.; Mutabingwa, T. Malaria in 2002. *Nature*, **2002**, *415*(6872), 670-672.

[6] Singh, V.; Tyagi, L.; Singhal, M.; Sharma, C.; Kori, M. A new generation of 7-Chloro-4-Aminoquinoline antimalarials. *Syst. Rev. Pharm.* **2010**, *1*(2), 182-189.

[7] Clough, B.; Wilson, R. J. M. In *Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery;* Rosenthal, P. J., Ed.; Humana Press Inc.: Totowa, **2001**, pp. 265-286.

[8] Loyevsky, M.; Gordeuk, V. R. In *Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery;* Rosenthal, P. J., Ed.; Humana Press Inc: Totowa, **2001**, pp. 307-324.

[9] Chakrabarti, D.; Azam, T.; DelVecchio, C.; Qiu, L.; Park, Y. I.; Allen, C. M. Protein prenyl transferase activities of *Plasmodium-falciparum*. *Mol. Biochem. Parasitol.* **1998**, *94*(2), 175-184.

[10] Gelb, M. H.; Van Voorhis, W. C.; Buckner, F. S.; Yokoyama, K.; Eastman, R.; Carpenter, E. P.; Panethymitaki, C.; Brown, K. A.; Smith, D. F. Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of anti-trypanosomatid and antimalarial therapeutics. *Mol. Biochem. Parasitol.* **2003**, *126*(2), 155-163.

[11] Nallan, L.; Bauer, K. D.; Bendale, P.; Rivas, K.; Yokoyama, K.; Horney, C. P.; Pendyala, P. R.; Floyd, D.; Lombardo, L. J.; Williams, D. K.; Hamilton, A.; Sebti, S.; Windsor, W. T.; Weber, P. C.; Buckner, F. S.; Chakrabarti, D.; Gelb, M. H.; Van Voorhis, W. C.Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. *J. Med. Chem.* **2005**, *48*(11), 3704-3713.

[12] Nwaka, S.; Riopel, L.; Ubben, D.; Craft, J. C. Medicines for Malaria Venture new developments in antimalarials. *Travel. Med. Infect. Dis.* **2004**, *2*(3-4), 161-170.

[13] Eastman, R. T.; Buckner, F. S.; Yokoyama, K.; Gelb, M. H.; Van Voorhis, W. C. Thematic review series: lipid posttranslational modifications. Fighting parasitic disease by blocking protein farnesylation. *J. Lipid. Res.* **2006**, *47*(2), 233-240.

[14] Krungkrai, S. R.; Krungkrai, J. Malaria parasite carbonic anhydrase: inhibition of aromatic/heterocyclic sulfonamides and its therapeutic potential. *Asian Pac. J. Trop. Biomed.* **2011**, *1*(3), 233-242.

[15] Crowther, G. J.; Napuli, A. J.; Gilligan, J. H.; Gagaring, K.; Borboa, R.; Francek, C.; Chen, Z.; Dagostino, E. F.; Stockmyer, J. B.; Wang, Y.; Rodenbough, P. P.; Castaneda, L. J.; Leibly, D. J.; Bhandari, J.; Gelb, M. H.; Brinker, A.; Engels, I. H.; Taylor, J.; Chatterjee, A. K.; Fantauzzi, P.; Glynne, R. J.; Van Voorhis, W. C.; Kuhen, K. L. Identification of inhibitors for putative malaria drug targets among novel antimalarial compounds. *Mol. Biochem. Parasitol.* **2011**, *175*(1), 21-29.

[16] Aparicio, I. M.; Marin-Menendez, A.; Bell, A.; Engel, P. C. Susceptibility of Plasmodium falciparum to glutamate dehydrogenase inhibitors--a possible new antimalarial target. *Mol. Biochem. Parasitol.* **2010**, *172*(2), 152-155.

[17] Aurrecoechea, C.; Brestelli, J.; Brunk, B. P.; Dommer, J.; Fischer, S.; Gajria, B.; Gao, X.; Gingle, A.; Grant, G.; Harb, O. S.; Heiges, M.; Innamorato, F.; Iodice, J.; Kissinger, J. C.; Kraemer, E.; Li, W.; Miller, J. A.; Nayak, V.; Pennington, C.; Pinney, D. F.; Roos, D. S.; Ross, C.; Stoeckert, C. J., Jr.; Treatman, C.; Wang, H. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic. Acids Res.* **2009**, *37*, 539-543.

[18] Bendale, P.; Olepu, S.; Suryadevara, P. K.; Bulbule, V.; Rivas, K.; Nallan, L.; Smart, B.; Yokoyama, K.; Ankala, S.; Pendyala, P. R.; Floyd, D.; Lombardo, L. J.; Williams, D. K.; Buckner, F. S.; Chakrabarti, D.; Verlinde, C. L.; Van Voorhis, W. C.; Gelb, M. H. Second generation tetrahydroquinoline-based protein farnesyltransferase inhibitors as antimalarials. *J. Med. Chem.* **2007**, *50*(19), 4585-4605.

[19] Carrico, D.; Ohkanda, J.; Kendrick, H.; Yokoyama, K.; Blaskovich, M. A.; Bucher, C. J.; Buckner, F. S.; Van Voorhis, W. C.; Chakrabarti, D.; Croft, S. L.; Gelb, M. H.; Sebti, S. M.; Hamilton, A. D. *In vitro* and *in vivo* antimalarial activity of peptidomimetic protein farnesyltransferase inhibitors with improved membrane permeability. *Bioorg. Med. Chem.* **2004**, *12*(24), 6517-6526.

[20] Eastman, R. T.; White, J.; Hucke, O.; Bauer, K.; Yokoyama, K.; Nallan, L.; Chakrabarti, D.; Verlinde, C. L.; Gelb, M. H.; Rathod, P. K.; Van Voorhis, W. C. Resistance to a protein farnesyltransferase inhibitor *Plasmodium falciparum*. *J. Biol. Chem.* **2005**, *280*(14), 13554-13559.

[21] Eastman, R. T.; White, J.; Hucke, O.; Yokoyama, K.; Verlinde, C. L.; Hast, M. A.; Beese, L. S.; Gelb, M. H.; Rathod, P. K.; Van Voorhis, W. C. Resistance mutations at the lipid substrate binding site of *Plasmodium falciparum* protein farnesyltransferase. *Mol. Biochem. Parasitol.* **2007**, *152*(1), 66-71.

[22] Fletcher, S.; Cummings, C. G.; Rivas, K.; Katt, W. P.; Horney, C.; Buckner, F. S.; Chakrabarti, D.; Sebti, S. M.; Gelb, M. H.; Van

Voorhis, W. C.; Hamilton, A. D. Potent, Plasmodium-selective farnesyltransferase inhibitors that arrest the growth of malaria parasites: structure-activity relationships of ethylenediamine-analogue scaffolds and homology model validation. *J. Med. Chem.* **2008**, *51*(17), 5176-5197.

[23] Glenn, M. P.; Chang, S. Y.; Horney, C.; Rivas, K.; Yokoyama, K.; Pusateri, E. E.; Fletcher, S.; Cummings, C. G.; Buckner, F. S.; Pendyala, P. R.; Chakrabarti, D.; Sebti, S. M.; Gelb, M.; Van Voorhis, W. C.; Hamilton, A. D. Structurally simple, potent, *Plasmodium* selective farnesyltransferase inhibitors that arrest the growth of malaria parasites. *J. Med. Chem.* **2006**, *49*(19), 5710-5727.

[24] Hast, M. A.; Fletcher, S.; Cummings, C. G.; Pusateri, E. E.; Blaskovich, M. A.; Rivas, K.; Gelb, M. H.; Van Voorhis, W. C.; Sebti, S. M.; Hamilton, A. D.; Beese, L. S. Structural basis for binding and selectivity of antimalarial and anticancer ethylenediamine inhibitors to protein farnesyltransferase. *Chem. Biol.* **2009**, *16*(2), 181-192.

[25] Ohkanda, J.; Lockman, J. W.; Yokoyama, K.; Gelb, M. H.; Croft, S. L.; Kendrick, H.; Harrell, M. I.; Feagin, J. E.; Blaskovich, M. A.; Sebti, S. M.; Hamilton, A. D. Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity. *Bioorg. Med. Chem. Lett.* **2001**, *11*(6), 761-764.

[26] Olepu, S.; Suryadevara, P. K.; Rivas, K.; Yokoyama, K.; Verlinde, C. L.; Chakrabarti, D.; Van Voorhis, W. C.; Gelb, M. H. 2-Oxo-tetrahydro-1,8-naphthyridines as selective inhibitors of malarial protein farnesyltransferase and as anti-malarials. *Bioorg. Med. Chem. Lett.* **2008**, *18*(2), 494-497.

[27] Van Voorhis, W. C.; Rivas, K. L.; Bendale, P.; Nallan, L.; Horney, C.; Barrett, L. K.; Bauer, K. D.; Smart, B. P.; Ankala, S.; Hucke, O.; Verlinde, C. L.; Chakrabarti, D.; Strickland, C.; Yokoyama, K.; Buckner, F. S.; Hamilton, A. D.; Williams, D. K.; Lombardo, L. J.; Floyd, D.; Gelb, M. H. Efficacy, pharmacokinetics, and metabolism of tetrahydroquinoline inhibitors of *Plasmodium falciparum* protein farnesyltransferase. *Antimicrob. Agents Chemother.* **2007**, *51*(10), 3659-3671.

[28] Xie, A.; Sivaprakasam, P.; Doerksen, R. J. 3D-QSAR analysis of antimalarial farnesyltransferase inhibitors based on a 2,5-diaminobenzophenone scaffold. *Bioorg. Med. Chem.* **2006**, *14*(21), 7311-7323.

[29] Sahu, N. K.; Sharma, M. C.; Mourya, V.; Kohli, D. V. QSAR studies of some side chain modified 7-chloro-4-aminoquinolines as antimalarial agents. *Arabian J. Chem.* **2010**, *in press.*

[30] Bringmann, G.; Bischof, S. K.; Muller, S.; Gulder, T.; Winter, C.; Stich, A.; Moll, H.; Kaiser, M.; Brun, R.; Dreher, J.; Baumann, K. QSAR guided synthesis of simplified antiplasmodial analogs of naphthylisoquinoline alkaloids. *Eur. J. Med. Chem.* **2010**, *45*(11), 5370-5383.

[31] Ojha, P. K.; Roy, K. Chemometric modeling, docking and in silico design of triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors as antimalarials. *Eur. J. Med. Chem.* **2010**, *45*(10), 4645-4656.

[32] Gramatica, P.; Corradi, M.; Consonni, V. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere*, **2000**, *41*(5), 763-777.

[33] Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. In *Data Handling in Science and Technology;* Leardi, R., Ed.; Elsevier, **2003**; 23, pp. 141-167.

[34] Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Detecting "bad" regression models: multicriteria fitness functions in regression analysis. *Anal. Chim. Acta*, **2004**, *515*(1), 199-208.

[35] Castilho, M. S.; Postigo, M. P.; de Paula, C. B.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D. Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg. Med. Chem.* **2006**, *14*(2), 516-527.

[36] Schuurmann, G.; Ebert, R. U.; Chen, J.; Wang, B.; Kuhne, R. External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *J. Chem. Infor. Model.* **2008**, *48*(11), 2140-2145.

[37] Ojha, P. K.; Mitra, I.; Das, R. N.; Roy, K. Further exploring rm2 metrics for validation of QSPR models. *Chemometr. Intell. Lab.* **2011**, *107*(1), 194-205.

[38] Pratim Roy, P.; Paul, S.; Mitra, I.; Roy, K. On two novel parameters for validation of predictive QSAR models. *Molecules* **2009**, *14*(5), 1660-1701.

[39] Roy, P. P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Combinatorial Sci.* **2008**, *27*(3), 302-313.

[40] Puntambekar, D. S.; Giridhar, R.; Yadav, M. R. Insights into the structural requirements of farnesyltransferase inhibitors as potential anti-tumor agents based on 3D-QSAR CoMFA and CoMSIA models. *Eur. J. Med. Chem.* **2008**, *43*(1), 142-154.

[41] Freitas, H. F.; Paz, O. S.; Castilho, M. S. Qsar 3d Studies of a Series of Human Butyrylcholinesterase Inhibitors. *Quim. Nova*, **2009**, *32*(8), 2114-2153.

[42] Castilho, M. S.; Postigo, M. P.; Pereira, H. M.; Oliva, G.; Andricopulo, A. D. Structural basis for selective inhibition of purine nucleoside phosphorylase from Schistosoma mansoni: Kinetic and structural studies. *Bioorg. Med. Chem.* **2010**, *18*(4), 1421-1427.

[43] Castilho, M. S.; C. Guido, R. V.; Andricopulo, A. D. Classical and Hologram QSAR Studies on a Series of Tacrine Derivatives as Butyrylcholinesterase Inhibitors. *Lett. Drug Des. Disc.* **2007**, *4*, 106-113.

[44] Yuan, H.; Wang, Y.; Cheng, Y. Local and global quantitative structure-activity relationship modeling and prediction for the baseline toxicity. *J. Chem. Inf. Model.* **2007**, *47*(1), 159-169.

[45] Peltason, L.; Hu, Y.; Bajorath, J.From structure-activity to structure-selectivity relationships: quantitative assessment, selectivity cliffs, and key compounds. *Chem. Med. Chem.* **2009**, *4*(11), 1864-1873.

[46] Sisay, M. T.; Peltason, L.; Bajorath, J. Structural interpretation of activity cliffs revealed by systematic analysis of structure-activity relationships in analog series. *J. Chem. Inf. Model.* **2009**, *49*(10), 2179-2189.

[47] Todeschini, R.; Consonni, V.; Mannhold, R.; Kubinyi, H.; Timmerman, H., *Handbook of Molecular Descriptors*. Wiley-VCH, **2000**.

[48] Guido, R.; Castilho, M.; Mota, S.; Oliva, G.; Andricopulo, A. Classical and hologram QSAR studies on a series of inhibitors of trypanosomatid glyceraldehyde-3-Phosphate dehydrogenase. *QSAR Comb. Sci.* **2008**, *27*(6), 768-781.

[49] Mota, S. G. R.; Barros, T. F.; Castilho, M. S. 2D QSAR studies on a series of bifonazole derivatives with antifungal activity. *J. Brazil. Chem. Soc.* **2009**, *20*(3), 451-459.

[50] Estrada, E.; Ramírez, A. Edge Adjacency relationships and molecular topographic descriptors. definition and QSAR applications. *J. Chem. Inf. Comput. Sci.* **1996**, *36*(4), 837-843.

[51] Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*(3), 163-172.

[52] Diudea, M. V.; Topan, M.; Graovac, A. Molecular Topology. 17. Layer Matrixes of Walk Degrees. *J. Chem. Inf. Comput. Sci.* **1994**, *34*(5), 1072-1078.

[53] Benigni, R.; Passerini, L.; Pino, A.; Giuliani, A. The information content of the eigenvalues from modified adjacency matrices: Large scale and small scale correlations. *Quant. Struct.-Act. Rel.* **1999**, *18*(5), 449-455.

[54] Burden, F. R. A chemically intuitive molecular index based on the eigenvalues of a modified adjacency matrix. *Quant. Struct.-Act. Rel.* **1997**, *16* (4), 309-314.

[55] Rucker, C.; Rucker, G. Mathematical relation between extended connectivity and eigenvector coefficients. *J. Chem. Inf. Comput. Sci.* **1994**, *34*(3), 534-538.

[56] Rucker, G.; Rucker, C. Counts of all walks as atomic and molecular descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*(5), 683-695.

[57] Rucker, G.; Rucker, C. Walk counts, labyrinthicity, and complexity of acyclic and cyclic graphs and molecules. *J. Chem. Inf. Comput. Sci.* **2000**, *40*(1), 99-106.

[58] Randić, M.; Wilkins, C. L. A Procedure for characterization of the rings of a molecule. *J. Chem. Inf. Comput. Sci.* **1980**, *20*(1), 36-46.

[59] Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships I. partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**. *7*, 565-577.

[60] Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions. *J. Comput. Chem.* **1988**. *9*, 80-90.

[61] Pearlman, R. S.; Smith, K. M. Metric validation and the receptor-relevant subspace concept. *J. Chem. Inform. Comput. Sci.* **1999**, *39*(1), 28-35.

Received: January 25, 2011 Revised: June 11, 2011 Accepted: June 19, 2011