

Original article

QSAR study on tetrahydroquinoline analogues as plasmodium protein farnesyltransferase inhibitors: A comparison of rationales of malarial and mammalian enzyme inhibitory activities for selectivity

Manish K. Gupta, Yenamandra S. Prabhakar*

Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow 226001, India

Received 16 August 2007; received in revised form 14 January 2008; accepted 14 January 2008

Available online 1 February 2008

This paper is dedicated to Dr. C.M. Gupta, Director, Central Drug Research Institute, Lucknow 226 001, India.

Abstract

The quantitative structure–activity relationships of *Plasmodium falciparum* and Rat protein farnesyltransferase (PFT) inhibitory activities of 6-cyano-1-(3-methyl-3*H*-imidazolyl-4-ylmethyl)-3-substituted-1,2,3,4-tetrahydroquinoline (THQ) analogues are investigated in order to explore the similarities/deviations between the two enzymes for these analogues. The structure space of a ligand (BMS-214662) bound to Rat–PFT (PDB code 1SA5) has been used as the conformational space of the compounds under investigation. The study has been carried out using the combinatorial protocol in multiple linear regression with several 2D- and 3D-descriptors from molecular operating environment (MOE) representing the physicochemical and electronic features of the compounds. The molecular potential energy and partially charged van der Waals surface areas have taken part in the PFT models. They suggested in favor of molecular arrangement with minimum energy and low positively/negatively charged surfaces for optimum *Pf*–PFT inhibitory activity. Furthermore, less hydrophobic compounds are preferred for the activity. The Rat–PFT inhibitory activity models suggested in favor of more negatively as well as more positively charged surface area descriptors for the better activity. The PLS analysis carried out on the descriptors of the *Pf*–PFT and Rat–PFT models suggested that among the parameters, the partially charged surface areas in the range -0.20 to -0.15 (PEOE_VSA-3) and -0.30 to -0.25 (PEOE_VSA-5), hydrophobicity (a_{hyd} , $\log P(o/w)$ and $S \log P_{\text{VSA4}}$), and electronic energy (PM3_Eele) of the molecules hold promise for modulating the *Pf*–PFT/R–PFT inhibitory activities of the compounds. This suggested the possibility of modulating the *Pf*–PFT/R–PFT inhibitory activities and bringing about selectivity in the THQ analogues for the malarial parasite enzyme.

© 2008 Elsevier Masson SAS. All rights reserved.

Keywords: Protein farnesyltransferase; Tetrahydroquinolines; Malaria; Combinatorial protocol in multiple linear regression (CP-MLR); PLS; QSAR

1. Introduction

Malaria is associated with the high morbidity and mortality [1]. Its control is globally a high priority task. Chloroquine (CQ), a low-cost drug, is widely used for its treatment. However, the emergence of CQ resistant malarial parasite strains has prompted the search for alternative strategies to combat the disease. Therefore the identification and development of new antimalarial agents associated with different mechanism

of action is of great interest [2–7]. In eukaryotic cells, protein farnesyltransferase (PFT) transfers a 15-carbon farnesyl group from farnesyl pyrophosphate (FPP) to the C-terminals of selected proteins which include Ras GTPase, whereas protein geranylgeranyltransferase-1 (GGGT-1) transfers a 20-carbon geranylgeranyl moiety to the γ -subunits of heterotrimeric G-proteins [8]. The inhibition of these enzymes in mammalian cells is exploited in cancer chemotherapy [9]. The genome sequence data of *Plasmodium falciparum* has indicated the absence of GGGT-1 in the parasite [10]. This has led to the investigation of *P. falciparum* PFT (*Pf*–PFT) as potential target for the development of new antimalarials [11–18]. In this

* Corresponding author. Tel.: +91 522 2612411; fax: +91 522 2623405.

E-mail address: yenpra@yahoo.com (Y.S. Prabhakar).

background, Nallan et al. have evaluated different mammalian-cell-optimized PFT inhibitors as new class of antimalarials against *Pf*-PFT [19]. From these leads, Bristol-Myers Squibb's (BMS) tetrahydroquinoline (THQ) prototypes were adopted and explored as potential compounds (Fig. 1) for the inhibition of *Pf*-PFT [19].

The establishment of correlation between the structure and the associated activity will come to aid in advancing the understanding of the system under investigation. The earlier QSAR and modeling studies on PFT inhibitors are mostly focused on anticancer activity [20–25]. These investigations have involved a variety of structural types ranging from peptides, peptidomimetics to steroids, different heterocyclics etc. They have suggested the significance of molecular shape, size, hydrogen bonding, electronegativity, polarizability, and hydrophobic characteristics of the compounds in determining their PFT inhibitory activity [20–25]. Xie et al. have carried out CoMFA and CoMSIA analyses of some 2,5-diaminobenzophenones as antimalarial agents based on the PFT inhibitory activity [26]. It has proposed that steric, electrostatic, and hydrophobic properties of these analogues influence the activity [26].

In this milieu we have contemplated a comprehensive quantitative structure–activity relationship (QSAR) study of the antimalarial activity of the THQ analogues to rationalize their *Pf*-PFT inhibition and to analyze the ways to impart selectivity in them. As the crystal structure of Rat-PFT bound with a ligand (PDB code 1SA5) is known [27], we have adopted this receptor bound conformation of the ligand (BMS-214662; Fig. 1) as the structure space of the THQ analogues considered in the study. These conformations have been used to parameterize the chemical space in terms of different 2D- and 3D-features of the compounds. The QSAR models are developed using the variable selection procedure, combinatorial protocol in multiple linear regression (CP-MLR) [28]. The results are presented here.

2. Material and methods

2.1. Datasets

6-Cyano-1-(3-methyl-3*H*-imidazol-4-ylmethyl)-3-substituted-1,2,3,4-tetrahydroquinolines (briefly referred to as tetrahydroquinoline or THQ analogues) reported in the literature are

considered for the datasets (Tables 1 and 2) [19,29]. The protein farnesyltransferase (PFT) inhibitory concentrations (IC_{50}) against *P. falciparum* (*Pf*-PFT) and/or Rat (R-PFT), and effective inhibitory dose (ED_{50}) against *P. falciparum* 3D7 strain (*Pf*-3D7) were reported for these analogues [19,29]. They represent the inhibitor concentration, in moles per liter, to produce 50% inhibition of enzyme activity (IC_{50}) and that of parasite proliferation in culture (ED_{50}) [19]. These inhibitory activities have been considered for the modeling study after their transformation in the form of logarithm of inverse of inhibitory concentration ($-\log IC_{50}$ *Pf*-PFT, $-\log IC_{50}$ R-PFT and $-\log ED_{50}$ *Pf*-3D7).

2.2. Putative bioactive conformation (PBC)

For the QSAR study, all the structures of THQ analogues (Table 1) have been generated in the structure builder module of MOE (molecular operating environment) [30]. A good consensus has been reported between a benzodiazepine analogue (BMS-214662; Fig. 1) bound to Rat-PFT (PDB code 1SA5) and the predicted binding mode of THQ analogues to a homology model of *P. falciparum* PFT [31]. However, subtle difference in the binding conformation of a ligand to two different enzyme systems causes the variation in the activity of the ligand. To the best of our knowledge, co-crystal structure information of *P. falciparum* PFT with a ligand is not available. In light of this BMS-214662 (ligand) bound to Rat-PFT (PDB code 1SA5) has been adopted as a reference structure space (Fig. 2) for the compounds (Table 1). Fig. 1 intuitively suggests the likely spatial correspondence between BMS-214662 and THQ analogues. In this, the 'A', 'B', 'C', 'D' and 'E' rings (or regions) of THQ analogues may, respectively, fulfill the function of 'A', 'B', 'C', 'D' and 'E' rings (or regions) of BMS-214662. Among the compounds (Table 1), THQ analogue 5 is structurally close to BMS-214662 (Fig. 1). Hence, this analogue has been adopted as a core structure for conformation generation.

In MOE, using the flexible alignment module with default settings, the conformational space of the THQ analogue 5 is adjusted to align with the conformational space of the ligand (Fig. 2; BMS-214662) bound to Rat-PFT (PDB code 1SA5). In default mode this alignment procedure is a stochastic

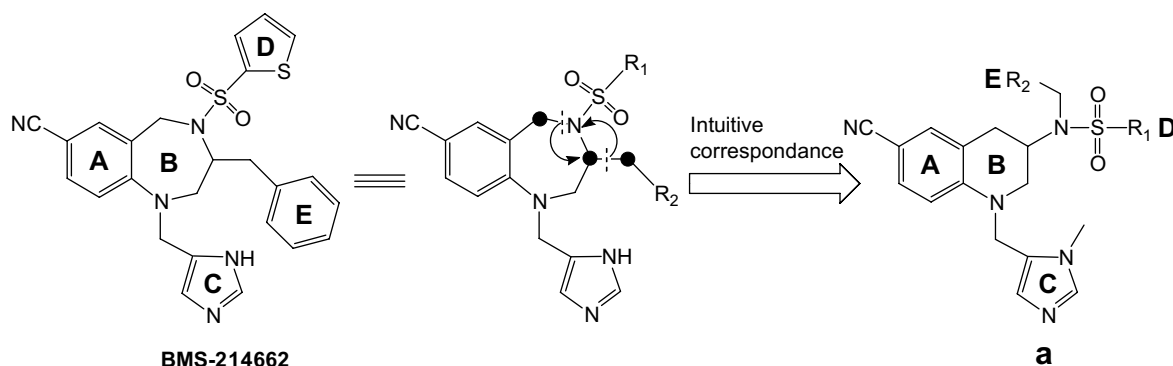
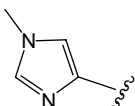
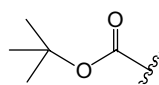
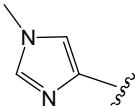

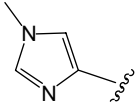

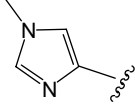
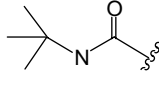
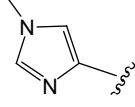
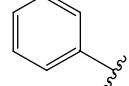
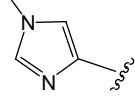
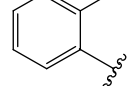
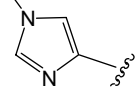
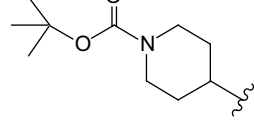
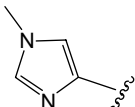
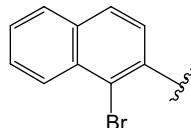
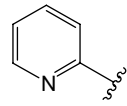
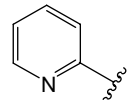


Fig. 1. BMS-214662; (a) general structure of tetrahydroquinoline (THQ) analogues associated with protein farnesyltransferase inhibitory and antimalarial activities. The correspondence between BMS-214662 and (a) are shown by means of reorganization of structural moieties of the former.

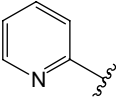
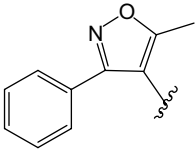
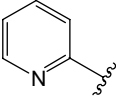
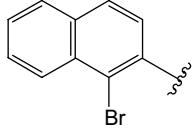
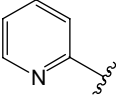
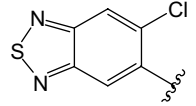
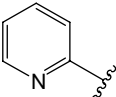
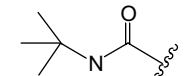
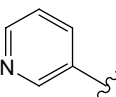
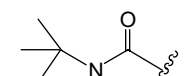
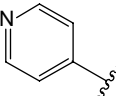
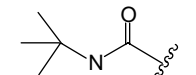
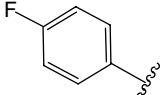
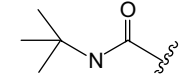
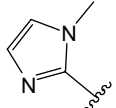
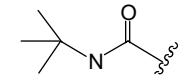
Table 1

The characteristic conformational energies and intramolecular distances of tetrahydroquinoline analogues (Fig. 1a) adopted for modeling the PFT inhibitory and antimalarial activities

Compound ^a	R ₁	R ₂	Energy ^b (kcal/mol)		Distance ^c (Å)	
			PBC	MEC	PBC	MEC
1			100.9	91.3	3.42	6.07
2			72.1	69.5	3.72	5.04
3			73.6	71.5	3.71	6.46
4			86.3	82.6	3.76	5.01
5			86.2	84.5	3.77	5.03
6			90.0	87.8	3.77	6.49
7			78.0	73.4	3.72	6.51
8			107.2	104.8	3.74	6.39
9			106.4	97.6	4.53	5.98

(continued on next page)

Table 1 (continued)

Compound ^a	R ₁	R ₂	Energy ^b (kcal/mol)		Distance ^c (Å)	
			PBC	MEC	PBC	MEC
10			95.3	92.1	3.58	6.27
11			113.7	111.9	3.61	5.56
12			113.1	98.7	4.32	6.61
13			92.3	90.3	3.61	5.55
14			79.2	68.1	3.45	6.46
15			77.8	68.2	3.53	5.81
16			85.2	73.0	3.5	6.67
17			88.0	71.1	3.45	6.23

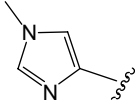
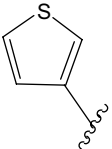
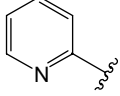
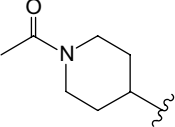
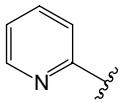
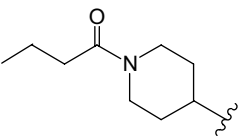
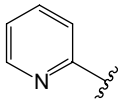
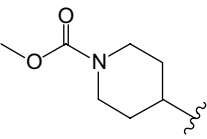
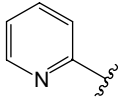
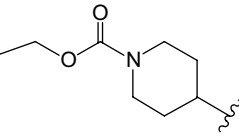
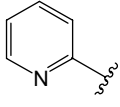
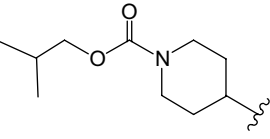
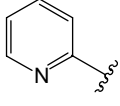
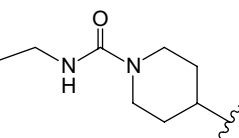
^a From Ref. [19].^b The forcefield involved in the potential energy is MMFF94. PBC represents putative bioactive conformation and MEC represents minimum energy conformation from systematic conformational search (see Ref. [30]).^c distance between the ring nitrogen of tetrahydroquinoline moiety (B-ring) and centroid of D-ring (R₁ substituent; Figs. 1a and 2a).

search approach. In this procedure alignments are “sampled” using a random incremental pulse search procedure with a quantitative measure of goodness of the alignment [30,32]. It involves generation of different conformations of the molecule by randomly rotating bonds and inverting unconstrained chiral centers followed by rigid-body optimization of a similarity

function. The goodness of each alignment is assessed in terms of its energy, similarity and the sum total of these two called as score. The energy term is calculated using the individual potential energies of the molecules in the alignment. It represents their average strain energy (kcal/mol). The forcefield MMFF94 [33] is applied to calculate the potential energy of

Table 2

The characteristic conformational energies and intramolecular distances of tetrahydroquinoline analogues (Fig. 1a) adopted as external test set for modeling the PFT inhibitory and antimalarial activities

Compound ^a	R ₁	R ₂	Energy (kcal/mol) ^b		Distance (Å) ^c	
			PBC	MEC	PBC	MEC
18			70.9	69.2	3.76	6.51
19			88.0	83.9	3.58	6.47
20 ^d			87.0	84.2	3.58	6.35
21			74.5	70.2	3.59	6.26
22 ^d			72.6	67.9	3.59	6.26
23 ^d			77.3	72.5	3.59	6.2
24 ^d			61.6	56.9	3.58	6.22

^a From Ref. [29].

^b The forcefield involved in the potential energy is MMFF94. PBC represents putative bioactive conformation and MEC represents minimum energy conformation from systematic conformational search (see Ref. [30]).

^c distance between the ring nitrogen of tetrahydroquinoline moiety (B-ring) and centroid of D-ring (R₁ substituent; Figs. 1a and 2a).

^d MEC is from stochastic conformational search approach implemented in MOE; Ref. [30]. The default systematic conformational search of the compound calls for the evaluation of more than 2 million conformers.

the aligning entities. In MOE, the default similarity measures include H-bond donor, H-bond acceptor, aromaticity and volume of the aligning chemical entities. Among the sampled alignments, those ones with smaller scores point to the better ones.

In the alignment of THQ analogue 5 and BMS-214662, the conformation of the latter was fixed to its Rat–PFT bound pose (PDB code 1SA5). From the alignments obtained, the conformer corresponding to the minimum score has been selected

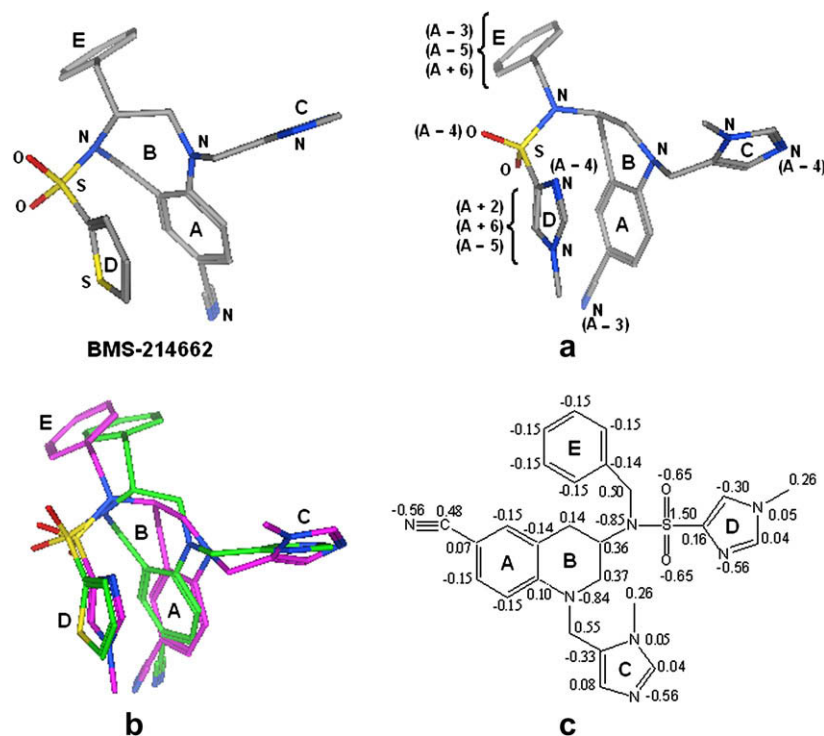


Fig. 2. BMS-214662 in the conformation of Rat-PFT bound ligand (PDB code 1SA5); (a) conformation of compound **5** aligned to BMS-214662; A + n/A – n ($n = 2, 3, 4, 5$ or 6) are abbreviated notation for PEOE_VSA + n/PEOE_VSA – n; they show the structural location or region of the van der Waals surface areas (VSA), computed using partial equalization of orbital electronegativities (PEOE), taken part in the *Plasmodium falciparum*- and Rat-PFT inhibitory activity predictive models; (b) superimposition pose of BMS-214662 (carbons: green) and compound **5** (carbons: magenta); in all, oxygens are in red, nitrogens are in blue and sulfurs are in yellow; (c) partial charge distribution on compound **5** was calculated by forcefield MMFF94; hydrogens are suppressed for clarity. For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.

and energy minimized (MMFF94; RMS gradient = 0.001) to generate the putative bioactive conformation (PBC) of THQ analogue **5** (Fig. 2a). In most of the small score alignment poses of BMS-214662 and THQ analogue **5**, the 'A', 'B', 'C', 'D' and 'E' regions of one entity are, respectively, in the vicinity of the 'A', 'B', 'C', 'D' and 'E' regions of the other entity (Fig. 2).

The PBC of THQ analogue **5** (Fig. 2a) has been used as template to generate the PBCs of all other THQ analogues (Table 1). The 'D' and 'E' regions (Fig. 1a; R₁ and R₂ substituents) of the template are appropriately modified using the MOE standard fragment library to result in the conformers of other THQ analogues (Table 1). These were energy minimized (MMFF94; RMSG = 0.001) to result in the PBCs used in the study. Also, the conformers of compounds listed in Table 2 have been generated from the PBC template of THQ analogue **5** using the same procedure as adopted in case of other THQ analogues (Table 1). No constraints were applied in generating these conformations of the THQ analogues from the template conformation (THQ analogue **5**). In all the conformations (PBCs), while 'A' and 'D' rings are almost parallel to each other, 'A' and 'C' rings are in near perpendicular planes. The 'A', 'B' and 'D' rings with the sulfonamide bridge (linker) formed a U-like shape. The 'E' ring is positioned at the bottom of the U-like shape from outside and located out of the way from 'A', 'B', 'C', and 'D' rings.

2.3. Minimum energy conformations (MECs)

A systematic conformational search of all the analogues (Tables 1 and 2) has been performed in order to compare the potential energy deviation between the minimum energy conformation (MEC) of each analogue and that of the PBC selected for the study. The systematic conformational search of THQ analogues has been carried out in MOE with default bond rotation parameters followed by the energy minimization (MMFF94; RMSG = 0.001) of the resulting conformations [30]. The potential energies of the PBCs of THQ analogues are within 20 kcal/mol from their respective MECs (Tables 1 and 2). The correlation between the potential energy (E) of the PBCs of THQ analogues and the biological activities under investigation is of the order of 0.45 (E vs. $-\log IC_{50}$ of Pf-PFT and R-PFT r is -0.45 and -0.48). There is a marked difference in the spatial orientation of structural moieties in PBCs and MECs. The MECs of the THQ analogues are more extended in space when compared to the PBCs. For example, in the MECs the average distance between the ring nitrogen of tetrahydroquinoline moiety and centroid of 'D' ring is 6.100 Å (S.D. = 0.506) while the same is 3.685 Å (S.D. = 0.253) in PBCs (Tables 1 and 2). In the MECs of the THQ analogues, the 'A', 'B', 'D' and 'E' rings are spread like 'propeller blades' around the sulfonamide nitrogen.

2.4. Descriptor generation and test set

The PBCs were used in MOE to compute various 2D- and 3D-descriptors of the chemical space. They include physico-chemical properties, surface areas, atom and bond counts, topological descriptors, H-bond donor/acceptor features, partial charges, and potential energies. It has resulted in 197 descriptors to qualify the chemical space of the PBCs of the compounds (Table 1). In addition to the compounds shown in Table 1, seven new THQ analogues (Table 2) have been chosen as external test set of the modeling study [29,34]. These compounds (Table 2) have been selected considering their descriptor span as well as structural proximity to the analogues listed in Table 1. Moreover, the compounds of Tables 1 and 2 have originated from the same laboratory [19,29]. The QSAR model generation and validation have been done using the combinatorial protocol in multiple linear regression (CP-MLR) [28] and partial least squares (PLS) analysis [35,36]. In the analysis the descriptors are not weighted with any bias. For the purpose of regression they are centered with their means ($XD = X - X_m$; where X is original descriptor, X_m is mean value of X and XD is deviation of X from X_m). As the total number of descriptors involved in this study is very large, only those descriptors identified in the models have been addressed in the discussion. The computational procedure is briefly described.

2.5. QSAR procedure

The CP-MLR is a ‘filter’ based variable selection procedure for model development [28]. Its procedural aspects and implementation are discussed in some of the recent publications [28,37–43]. The thrust of this procedure is in its embedded ‘filters’. They are briefly as follows: filter-1 seeds the variables by way of limiting inter-parameter correlations to predefined level (upper limit ≤ 0.79); filter-2 controls the variables entry to a regression equation through t -values of coefficients (threshold value ≥ 2.0); filter-3 provides comparability of equations with different number of variable in terms of square-root of adjusted multiple correlation coefficient of regression equation, r -bar; filter-4 estimates the consistency of the equation in terms of cross-validated R^2 or Q^2 with leave-one-out (LOO) cross-validation as default option (threshold value $0.3 \leq Q^2 \leq 1.0$). The filter-2 evaluates the significance of variables of each seed in terms of the t -values of regression coefficients. It involves a comparison of estimated regression coefficients of the variables and their standard errors; the seed is skipped if the ratio is below the threshold value. Successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient values. In light of this, filter-3 (r -bar value) compensates the increment in correlation coefficient due to the increasing number of explanatory variables in seeds and allows the comparison of different seeds.

Throughout this study, for the filters-1, 2, and 4 of CP-MLR the thresholds were assigned as 0.79, 2.0, and $0.3 \leq Q^2 \leq 1.0$, respectively. The filter-3 was assigned an initial value of 0.71 for two-variable models. In order to collect the descriptors with higher information content and explanatory power, the

threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r -bar value of the preceding optimum model as the new threshold for next generation. The identified models were reassessed for the chance correlations through 100 simulation runs with the randomized biological response [38,44]. In addition to this, the models were verified with two different internal test sets carved out from the 17 compounds considered in the analysis (Table 1) and an external test set of 7 compounds (Table 2). The internal test sets are from the descriptors’ clustering of the compounds and the random selection procedure, with each one containing 5 out of 17 compounds listed in Table 1. Furthermore, the descriptors of the CP-MLR models were evaluated in the PLS procedure in order to derive composite single-window MLR-like PLS models comprising all identified descriptors.

3. Result and discussion

In CP-MLR the Pf -PFT inhibitory activity of THQ analogues (Table 1) [19] has been analyzed for one to three parameter models in terms of physicochemical, topological and partial charge indices representing the 2D- and 3D-features of the compounds. Among large number of validated models, the following equations have best explained the Pf -PFT inhibitory activity of these analogues.

$$\begin{aligned} -\log IC_{50} Pf-PFT = & 0.356(0.071)(PEOE.VSA - 3) \\ & + 0.049(0.008)(PEOE.VSA + 2) \\ & - 0.484(0.108) \log P(o/w) + 0.323; \\ n = 17, r = 0.900, Q^2 = 0.623, \\ Q_{L30}^2 = 0.632, s = 0.485, \\ F = 18.56; r_{\text{rand}}(S.D.) = 0.395(0.153) \end{aligned} \quad (1)$$

$$\begin{aligned} -\log IC_{50} Pf-PFT = & -0.085(0.018)(PEOE.VSA - 5) \\ & + 0.043(0.013)vsa_other \\ & - 0.467(0.084)E_{\text{str}} + 10.033; \\ n = 17, r = 0.883, Q^2 = 0.636, \\ Q_{L30}^2 = 0.680, s = 0.523, \\ F = 15.35, r_{\text{rand}}(S.D.) = 0.394(0.136) \end{aligned} \quad (2)$$

$$\begin{aligned} -\log IC_{50} Pf-PFT = & -0.119(0.019)(PEOE.VSA - 5) \\ & - 0.448(0.078)a_{\text{hyd}} \\ & - 0.002(0.0006)PM3_E_{\text{ele}} + 13.896; \\ n = 17, r = 0.890, Q^2 = 0.657, \\ Q_{L30}^2 = 0.614, s = 0.509, \\ F = 16.46, r_{\text{rand}}(S.D.) = 0.375(0.143) \end{aligned} \quad (3)$$

The regression models outlined in this study are expressed for the variables in their original form. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. The magnitudes of regression coefficients of different descriptors are not for direct comparison to deduce their relative importance to the activity. In the latter part of the discussion, the relative importance of all

Table 3

Ten random correlation coefficients (r_{yrand}) from the activity (Y) randomization study of the Pf–PFT and R–PFT models

S.No. ^a	CP-MLR							PLS			
	–log IC ₅₀ Pf–PFT			–log IC ₅₀ R–PFT				–log IC ₅₀ Pf–PFT		–log IC ₅₀ R–PFT	
	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (4)	Eq. (5)	Eq. (6)	Eq. (7)	Table 10	Table 11	Table 10	Table 11
1	0.691	0.699	0.669	0.698	0.732	0.693	0.723	0.684	0.650	0.691	0.680
2	0.561	0.655	0.666	0.489	0.683	0.692	0.631	0.430	0.569	0.594	0.603
3	0.489	0.613	0.639	0.481	0.609	0.674	0.552	0.409	0.480	0.526	0.560
4	0.456	0.536	0.622	0.480	0.604	0.652	0.550	0.406	0.458	0.421	0.412
5	0.438	0.503	0.543	0.466	0.581	0.635	0.545	0.347	0.444	0.263	0.402
6	0.396	0.459	0.443	0.421	0.579	0.625	0.324	0.302	0.334	0.259	0.311
7	0.380	0.327	0.381	0.258	0.568	0.592	0.323	0.302	0.168	0.197	0.306
8	0.380	0.299	0.354	0.231	0.559	0.564	0.296	0.296	0.131	0.126	0.294
9	0.371	0.268	0.350	0.189	0.408	0.477	0.266	0.179	0.123	0.109	0.289
10	0.291	0.115	0.256	0.119	0.288	0.236	0.250	0.168	0.087	0.034	0.211

^a Values of 10 high r_{yrand} recorded from 100 Y-randomizations.

Table 4

Observed and CP-MLR/PLS predicted PFT inhibitory and antimalarial activities of the tetrahydroquinoline analogues (Table 1) analyzed as training and internal test (descriptor cluster) sets

Compound	–log IC ₅₀ Pf–PFT ^a				–log ED ₅₀ Pf–3D7 ^a			–log IC ₅₀ R–PFT ^a			
	Obsd ^b	Eq. (1) ^c	Eq. (2) ^c	PLS ^d	Obsd ^b	Model (1a) ^c	Model (3a) ^c	Obsd ^b	Eq. (5) ^c	Eq. (6) ^c	PLS ^e
1^f	9.05	9.18	8.69	8.93	8.30	7.97	7.35	8.92	9.03	8.85	8.73
2	9.22	9.32	9.59	9.58	8.16	8.11	9.43	9.16	9.41	9.06	8.94
3^f	9.40	8.96	9.51	9.51	8.22	7.66	8.76	9.10	9.44	9.01	8.93
4	8.92	9.66	8.01	8.83	8.30	8.33	6.78	8.82	8.59	8.74	8.61
5	9.22	8.76	8.97	9.14	8.30	7.30	7.45	8.92	8.26	8.49	8.60
6	8.82	8.71	8.61	8.95	7.26	7.39	7.39	8.47	8.01	8.03	8.59
7	8.96	8.84	9.30	9.06	7.00	7.68	8.04	8.50	7.98	8.47	8.73
8^f	7.55	8.05	7.33	8.19	5.52	6.40	5.97	6.00	6.69	6.63	7.17
9	8.55	7.79	8.43	7.94	7.29	6.50	7.46	8.42	7.97	8.29	7.92
10^f	9.10	8.30	9.36	8.46	6.95	7.47 ^g	7.17	8.52	8.17	7.92	8.01
11	6.78	7.13	7.33	6.90	5.56	4.82	5.18	6.36	6.88	6.59	6.40
12	7.26	7.44	7.07	7.40	5.52	5.93	6.58	8.00	8.09	7.74	8.13
13^f	8.85	8.34	7.84	7.57	7.07	7.15	6.79	8.75	8.52	8.66	7.88
14	7.60	7.78	6.50	7.26	5.57	6.44	5.72	8.22	7.51	7.72	7.82
15	6.00	7.46	7.52	6.83	<5.30	4.49	5.62	6.80	7.98	8.16	7.80
16	7.38	6.49	7.46	7.12	<5.30	4.39	5.57	8.12	8.57	7.70	7.78
17	8.10	6.91	8.55	7.81	<5.30	5.18	7.33	8.30	9.11	8.94	8.77
Regression statistics ^h											
Training set	12	12	12		10	9		12	12	12	
r	0.906	0.894	0.935		0.940	0.884		0.842	0.845	0.853	
Q^2	0.521	0.569	0.751		0.656	0.368		0.425	0.565	0.564	
Q^2_{L3O}	0.483	0.585	0.770		0.636	0.267		0.482	0.542	0.577	
s	0.524	0.553	0.413		0.462	0.697		0.520	0.516	0.474	
F	12.22	10.68	31.36		15.40	5.99		6.52	6.64	12.02	
r_{yrand}	0.512	0.522	0.358		0.511	0.593		0.469	0.482	0.383	
(S.D.)	(0.167)	(0.132)	(0.166)		(0.180)	(0.166)		(0.193)	(0.170)	(0.165)	
Test set	5	5	5		4	5		5	5	5	
r^2_{Pred}	0.710	0.723	0.464		0.777	0.716		0.883	0.881	0.627	

^a –log IC₅₀ against *P. falciparum* (Pf–PFT) and Rat (R–PFT) protein farnesyltransferase, and –log ED₅₀ against 3D7 strain of *P. falciparum* in culture.^b Observed activity from Ref. [19].^c Predictions are from the matching training set equation (of the identified equation) derived after partitioning the compounds into training and test sets.^d Predictions are from the matching training set Pf–PFT PLS equation (of the PLS model in Table 10) derived after partitioning the compounds into training and test sets.^e Predictions are from the matching training set R–PFT PLS equation (of the PLS model in Table 11) derived after partitioning the compounds into training and test sets.^f Test set compound; the predictions are from corresponding training set equation.^g Included in training set to overcome the colinearity in the partitioned (subset) data.^h The notations in the regression statistics stand for the same as outlined following the Eqs. (1)–(3).

Table 5

Observed and CP-MLR/PLS predicted PFT inhibitory and antimalarial activities of the tetrahydroquinoline analogues (Table 1) analyzed as training and internal test (random selection) sets

Compound	–log IC ₅₀ Pf–PFT ^a				–log ED ₅₀ Pf–3D7 ^a			–log IC ₅₀ R–PFT ^a			
	Obsd ^b	Eq. (1) ^c	Eq. (2) ^c	PLS ^d	Obsd ^b	Model (1a) ^c	Model (3a) ^c	Obsd ^b	Eq. (5) ^c	Eq. (6) ^c	PLS ^e
1^f	9.05	9.17	8.66	8.94	8.30	7.95	7.59	8.92	8.81	8.80	8.82
2	9.22	9.32	9.40	9.52	8.16	8.17	8.57	9.16	9.11	9.08	9.15
3	9.40	8.79	9.37	9.45	8.22	7.34	8.46	9.10	9.33	9.02	9.17
4	8.92	9.73	8.07	8.72	8.30	8.54	7.08	8.82	8.28	8.58	8.58
5^f	9.22	8.79	8.96	8.84	8.30	7.32	7.08	8.92	8.07	8.46	8.53
6^f	8.82	8.65	8.62	8.51	7.26	7.09	6.79	8.47	7.78	7.94	8.38
7	8.96	8.76	9.15	8.72	7.00	7.42	7.33	8.50	8.01	8.17	8.61
8	7.55	8.15	6.97	7.20	5.52	6.13	3.92	6.00	6.61	6.56	6.37
9^f	8.55	8.75	8.45	8.81	7.29	7.10	7.07	8.42	8.32	8.82	8.38
10	9.10	8.41	9.33	9.09	6.95	6.93	6.94	8.52	8.25	7.95	8.35
11^f	6.78	7.21	7.09	6.91	5.56	4.53	4.60	6.36	6.59	6.50	6.09
12	7.26	8.03	7.67	7.89	5.52	5.23	6.68	8.00	8.32	7.73	8.33
13	8.85	8.81	7.80	8.49	7.07	7.44	6.75	8.75	8.70	9.23	8.37
14	7.60	7.72	6.85	7.15	5.57	5.53	5.69	8.22	7.13	7.54	7.50
15	6.00	7.37	7.68	6.93	<5.30	4.28	5.60	6.80	7.45	7.97	7.48
16	7.38	6.32	7.61	7.00	<5.30	3.89	5.34	8.12	8.54	7.52	7.55
17	8.10	7.06	8.56	8.18	<5.30	5.25	7.55	8.30	8.80	8.82	8.82
Regression statistics ^g											
Training set	12	12	12		9	9		12	12	12	
<i>r</i>	0.879	0.850	0.910		0.965	0.921		0.909	0.887	0.884	
<i>Q</i> ²	0.451	0.490	0.659		0.853	0.464		0.665	0.618	0.617	
<i>Q</i> _{L3O} ²	0.441	0.503	0.639		0.815	0.513		0.654	0.578	0.584	
<i>s</i>	0.583	0.645	0.479		0.385	0.576		0.453	0.502	0.480	
<i>F</i>	9.07	6.94	21.64		22.83	9.26		12.66	9.85	16.04	
<i>r</i> _{rand}	0.511	0.515	0.402		0.589	0.577		0.501	0.523	0.426	
(S.D.)	(0.163)	(0.172)	(0.193)		(0.221)	(0.168)		(0.145)	(0.160)	(0.176)	
Test set	5	5	5		5	5		5	5	5	
<i>r</i> _{Pred} ²	0.894	0.916	0.921		0.625	0.460		0.723	0.850	0.946	

^a –log IC₅₀ against *P. falciparum* (Pf–PFT) and Rat (R–PFT) protein farnesyltransferase, and –log ED₅₀ against 3D7 strain of *P. falciparum* in culture.

^b Observed activity from Ref. [19].

^c Predictions are from the matching training set equation (of the identified equation) derived after partitioning the compounds into training and test sets.

^d Predictions are from the matching training set Pf–PFT PLS equation (of the PLS model in Table 10) derived after partitioning the compounds into training and test sets.

^e Predictions are from the matching training set R–PFT PLS equation (of the PLS model in Table 11) derived after partitioning the compounds into training and test sets.

^f Test set compound; the predictions are from corresponding training set equation.

^g The notations in the regression statistics stand for the same as outlined following the Eqs. (1)–(3).

Table 6

Observed and CP-MLR/PLS predicted Pf–PFT inhibitory and antimalarial activities of the tetrahydroquinoline analogues (Table 2) analyzed as external test set compounds

Compound	–log IC ₅₀ Pf–PFT ^a					–log ED ₅₀ Pf–3D7 ^a			
	Obsd ^b	Eq. (1) ^c	Eq. (2) ^c	Eq. (3) ^c	PLS ^d	Obsd ^b	Model (1a) ^c	Model (2a) ^c	Model (3a) ^c
18	9.40	9.02	9.48	9.03	9.30	7.92	7.74	8.35	7.87
19	8.62	8.76	7.86	7.64	8.01	7.43	7.37	6.85	6.26
20	8.52	8.32	7.59	7.22	7.60	7.10	6.75	6.51	5.59
21	9.24	8.62	9.49	9.51	9.20	7.80	7.17	7.62	7.86
22	9.30	8.46	9.23	8.99	8.89	7.26	6.94	7.47	7.27
23	8.85	7.99	9.00	8.61	8.49	7.21	6.28	7.19	6.63
24	9.19	8.77	9.21	9.29	8.94	7.17	7.38	7.41	7.53
<i>r</i> _{Pred} ^f		0.508	0.657	0.327	0.648		–0.077	0.306	–1.871
<i>r</i> _{Pred} ^g		0.511	0.848	0.699	0.836		0.008	0.557	–0.296

^a –log IC₅₀ against *P. falciparum* (Pf–PFT) protein farnesyltransferase, and –log ED₅₀ against 3D7 strain of *P. falciparum* in culture.

^b Observed activity from Ref. [29].

^c Predictions are from the Eqs. (1), (2), or (3) derived using 17 THQ analogues (Table 1).

^d Predictions are from the Pf–PFT PLS equation (Table 10) derived using 17 THQ analogues (Table 1).

^e Predictions are from the models 1a, 2a, or 3a (Table 7) derived using 14 THQ analogues (Table 1).

^f For 7 compounds (18–24).

^g For 6 compounds (excluded compound 20).

descriptors identified from the CP-MLR models are explained in the form of their fraction contributions from normalized regression coefficients of PLS models. Also, Ref. [45] gives a detailed procedure to compare the regression coefficients of different descriptors. In all regression equations, n is the number of compounds, r is the correlation coefficient, Q^2 is cross-validated R^2 from leave-one-out (LOO) procedure, Q^2_{L3O} is cross-validated R^2 from leave three compounds out (leave-many-out) procedure (where a group of compounds are randomly kept outside the analysis each time in such a way that all compounds are in the predictive groups for once), s is the standard error of the estimate and F is the F -ratio between the variances of calculated and observed activities. The $r_{y\text{rand}}$ (S.D.) is the mean correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. The values given in the parentheses (in regression equation) are the standard errors of the regression coefficients. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. For these and other PFT models, 10 high $r_{y\text{rand}}$ values from 100 Y -randomizations are shown Table 3. The models 1–3 are validated with two test sets (internal), with each one containing 5 compounds out of the 17 analogues listed in Table 1. The predictions of these test sets (5 compounds each) emanating from the remaining 12 training compounds are shown in Tables 4 and 5.

In addition to the internal test sets, the models 1–3 are validated using an external test set of 7 compounds as shown in Table 2. These THQ analogues are part of a recent report from the same laboratory as that of compounds of Table 1 [19,29]. The parameter and structure space of the THQ analogues listed in Table 1 are considered in the selection of these compounds (Table 2). Among the compounds listed in Table 1, the THQ analogue 7 is in the neighborhood of these test set analogues (Table 2). For these analogues, the predicted activities from Eqs. (1)–(3) and the resultant predictive r^2 values are shown in Table 6. In Eq. (3), the predicted activity of THQ analogue 20 (Table 6) has shown a large deviation from its observed value. We do not have any reasoning for this deviation. However, on exclusion of this compound the predictive r^2 values have improved significantly. These results give significance to the identified models.

The Pf –PFT inhibitory activity of THQ analogues is not directly correlated to the potential energy (E) of the conformers. However, the potential energy (E) together with electrostatic component of the potential energy (E_{ele}) have moderately explained the activity ($r^2 = 0.55$; equation is shown in Supplementary data). The energy related parameters involved in the Pf –PFT inhibitory models are E_{str} (the bond stretching potential energy in Eq. (2) and $PM3_E_{\text{ele}}$ (electronic energy of the molecules calculated using the $PM3$ Hamiltonian in Eq. (3)). These descriptors suggest in favor of molecular arrangement with minimum energy for optimum activity.

Different descriptors measuring the van der Waals surface areas (VSA), computed using partial equalization of orbital electronegativities (PEOE) have taken part in the models

Table 7
The scope of descriptors of CP-MLR models specific to Pf –PFT inhibitory activity (Eqs. (1)–(3)) of THQ analogues (Table 1) in explaining their Pf –3D7 and R–PFT inhibitory activities

Descriptors	−log ED ₅₀ Pf−3D7											−log IC ₅₀ R−PFT												
	Model					Regression statistics						Model					Regression statistics							
						<i>n</i>	<i>r</i>	<i>Q</i> ²	<i>Q</i> ² _{L3O}	<i>s</i>	<i>F</i>	<i>r</i> _{yrand} (S.D.)						<i>n</i>	<i>r</i>	<i>Q</i> ²	<i>Q</i> ² _{L3O}	<i>s</i>	<i>F</i>	<i>r</i> _{yrand} (S.D.)
PEOE_VSA − 3, PEOE_VSA + 2, log P(o/w)	1a					14	0.930	0.71, 0.67	0.47		21.26	0.48 (0.16)	1b					17	0.797	0.25, 0.17	0.63		7.56	0.39 (0.14)
PEOE_VSA − 5, vsa_other, E_str	2a					14	0.904	0.51, 0.55	0.55		14.86	0.44 (0.14)	2b					17	0.878	0.53, 0.50	0.50		14.61	0.38 (0.14)
PEOE_VSA − 5, a_hyd, PM3_Eele	3a					14	0.887	0.61, 0.64	0.59		12.34	0.43 (0.16)	3b					17	0.853	0.54, 0.58	0.54		11.61	0.40 (0.13)

1–3. Among these descriptors, PEOE_VSA–3 (Eq. (1)) and PEOE_VSA–5 (Eqs. (2) and (3)) represent surface areas with partial negative charges in the ranges of –0.20 to –0.15 and –0.30 to –0.25, respectively. In these compounds, PEOE_VSA–3 is due to the N of nitrile group of A-ring and pyridyl nitrogen of E-region (Fig. 2a). The PEOE_VSA–5 is due to the N and/or O of D- and E-regions of the molecules (Fig. 2a). The regression coefficients of these descriptors suggest that less negatively charged surface areas are favorable for the activity. In Eq. (1), PEOE_VSA+2 represents surface area with partial positive charge in the range of 0.10 to 0.15. In these analogues, the –CH= of C- and D-rings form the PEOE_VSA+2 (Fig. 2a). The positive regression coefficient of this descriptor suggests its favorable nature for the activity. In Eq. (2), vsa_other has shown positive influence on the activity. It accounts for the van der Waals surface area of the pharmacophore. The log P(o/w) in Eq. (1) suggests in favor of less hydrophobic molecules for better activity. The parameter a_hyd in Eq. (3) accounts for the number of hydrophobic atoms in the molecule. Its negative regression coefficient in Eq. (3) is in agreement with the coefficient of log P(o/w) in Eq. (1). It also suggests in favor of decreasing hydrophobicity for better activity.

Among the 17 THQ analogues listed in Table 1, the inhibitory doses of *P. falciparum* 3D7 strain in culture are available only for 14 compounds. The *Pf*–3D7 inhibitory doses of the THQ analogues 15–17 were reported as more than 5000 nM (in –log ED₅₀*Pf*–3D7 scale, it is equivalent to less than 5.30) [19]. The *Pf*–3D7 and *Pf*–PFT inhibitory activities of the remaining 14 THQ analogues (Table 1) are highly intercorrelated ($n = 14$; $r^2 = 0.815$). Since *P. falciparum* is a single cell eukaryote, this agreement between the compounds' inhibitory activity profiles against *Pf*–3D7 parasite in culture and *Pf*–PFT enzyme is not unusual. In light of this, the descriptors of *Pf*–PFT inhibitory activity models (Eqs. (1)–(3)) are extended to explain the inhibition of *Pf*–3D7 strain of these analogues as well.

The *Pf*–3D7 inhibitory activity models 1a–3a shown in Table 7 have been derived using 14 THQ analogues (Table 1). These models have well explained the inhibition of the *Pf*–3D7 parasite in culture. In terms of descriptor composition each one of these (models 1a–3a; Table 7) are counterparts of *Pf*–PFT inhibitory activity Eqs. (1)–(3). Similar to *Pf*–PFT models, the *Pf*–3D7 models are also validated with two internal test sets of 5 compounds each (descriptor cluster and random selections from Table 1) and one external test set of 7 compounds (Table 2). The predictions of the internal test sets (5 compounds each) from the remaining 9

compounds are shown in Tables 4 and 5. Table 6 shows the predicted *Pf*–3D7 inhibitory activities of the external test set THQ analogues (Table 2) from the models 1a–3a (Table 7) derived using the 14 THQ analogues (Table 1). Five out of seven test set THQ analogues are predicted well (Table 6) by these models (Table 7). Among them, model 2a has predicted the activities more satisfactorily than others. Similar to the case of *Pf*–PFT, the predicted *Pf*–3D7 inhibitory activity of THQ analogue 20 has shown large deviation from its observed one (Table 6). Moreover, in case of test set THQ analogues (Table 2), their observed *Pf*–PFT and *Pf*–3D7 inhibitory activities (Table 6) are intercorrelated only to a limited extent ($n = 7$; $r^2 = 0.307$). In general terms, the differences in the quality equations (also predictions) from within or between two or more activities can be attributed to various factors; one reason could be insufficient reflection of the properties of selected analogues towards the activities. Barring these differences, the *Pf*–PFT and *Pf*–3D7 equations (models 1–3 and 1a–3a) are in agreement with each other and it further strengthened the former as a potential target for the antimalarial agents.

The PFT is originally identified as a target of anticancer agents and subsequently explored as a target of antimalarial agents. For the THQ analogues (Table 1), their *Pf*–PFT and R–PFT inhibitory activities are intercorrelated to the extent of 67% ($r^2 = 0.669$). In order to ascertain the similarities/deviations between the *Pf*–PFT and R–PFT systems for THQ analogues, the models specific to the *Pf*–PFT inhibitory activity (Eqs. (1)–(3)) are examined for the R–PFT inhibitory activity also (Table 7; models 1b–3b). Although these R–PFT inhibitory activity models (Table 7; models 1b–3b) are statistically significant, they are inferior when compared to those derived for the *Pf*–PFT inhibitory activity (Eqs. (1)–(3)). This suggests the possible deviations between these two enzyme systems. The equations shown below (Eqs. (4)–(7)) are specific to the R–PFT inhibitory activity of the THQ analogues (Table 1).

$$\begin{aligned}
 -\log \text{IC}_{50}\text{R-PFT} &= 0.051(0.015)(\text{PEOE_VSA} + 6) \\
 &\quad - 0.539(0.083)\text{E_str} + 11.028; \\
 n &= 17, r = 0.868, Q^2 = 0.600, \\
 Q_{\text{L30}}^2 &= 0.575, s = 0.497, \\
 F &= 21.37, r_{\text{yand}}(\text{S.D.}) = 0.328(0.162) \quad (4)
 \end{aligned}$$

Table 8

The scope of descriptors of CP-MLR models specific to R–PFT inhibitory activity (Eqs. (4)–(7)) of THQ analogues (Table 1) in explaining their *Pf*–PFT inhibitory activity

Descriptors	Model	–log IC ₅₀ <i>Pf</i> –PFT					
		<i>n</i>	<i>r</i>	Q^2, Q_{L30}^2	<i>s</i>	<i>F</i>	<i>r</i> _{yand} (S.D.)
PEOE_VSA+6, E_str	4a	17	0.730	0.23, 0.28	0.73	8.01	0.31 (0.14)
PEOE_VSA–3, PEOE_VSA–4, E_vdw	5a	17	0.689	–0.03, –0.16	0.81	3.91	0.39 (0.13)
PEOE_VSA–3, Q_VSA_FPPOS, E_str	6a	17	0.705	0.04, –0.23	0.79	4.30	0.39 (0.16)
PEOE_VSA–4, S log P_VSA4, a_hyd	7a	17	0.710	0.27, 0.32	0.78	4.40	0.40 (0.16)

$$\begin{aligned}
 -\log \text{IC}_{50}\text{R-PFT} &= 0.232(0.057)(\text{PEOE_VSA} - 3) \\
 &\quad + 0.070(0.028)(\text{PEOE_VSA} - 4) \\
 &\quad - 0.104(0.014)E_{\text{vdw}} + 6.667; \\
 n &= 17, r = 0.901, Q^2 = 0.703, \\
 Q_{\text{L3O}}^2 &= 0.725, s = 0.450, \\
 F &= 18.73, r_{\text{yrand}}(\text{S.D.}) = 0.415(0.131)
 \end{aligned} \quad (5)$$

$$\begin{aligned}
 -\log \text{IC}_{50}\text{R-PFT} &= 0.158(0.049)(\text{PEOE_VSA} - 3) \\
 &\quad + 17.501(5.068)(Q_{\text{VSA_FPPOS}}) \\
 &\quad - 0.491(0.081)E_{\text{str}} + 4.151; \\
 n &= 17, r = 0.905, Q^2 = 0.726, \\
 Q_{\text{L3O}}^2 &= 0.655, s = 0.441, \\
 F &= 19.64, r_{\text{yrand}}(\text{S.D.}) = 0.368(0.144)
 \end{aligned} \quad (6)$$

$$\begin{aligned}
 -\log \text{IC}_{50}\text{R-PFT} &= 0.085(0.028)(\text{PEOE_VSA} - 4) \\
 &\quad + 0.193(0.053)(S \log P_{\text{VSA4}}) \\
 &\quad - 0.409(0.055)a_{\text{hyd}} + 12.282; \\
 n &= 17, r = 0.900, Q^2 = 0.699, \\
 Q_{\text{L3O}}^2 &= 0.726, s = 0.453, \\
 F &= 18.38, r_{\text{yrand}}(\text{S.D.}) = 0.375(0.123)
 \end{aligned} \quad (7)$$

These equations have satisfied the validation measures and optimally explained the R-PFT inhibitory activity of the compounds (Tables 4 and 5). The R-PFT inhibitory activity is not reported for the external test set THQ analogues (Table 2). Hence, they are not used for R-PFT model evaluation. There are only three common descriptors (E_{str} , PEOE_VSA-3 and a_{hyd}) between Eqs. (1)–(3) and Eqs. (4)–(7). The R-PFT specific models (Eqs. (4)–(7)) are composed of more negatively (PEOE_VSA-4) as well as more positively (PEOE_VSA+6) charged surface area descriptors for the inhibitory activity. In these compounds, the oxygens of sulfonamide and -N= of imidazole rings (C- and D-regions) contribute to the PEOE_VSA-4 whereas highly polarized carbons of D- and E-regions contribute to the PEOE_VSA+6 (Fig. 2a). The descriptors $Q_{\text{VSA_FPPOS}}$ (fractional positive polar van der Waals surface area) in Eq. (6) and $S \log P_{\text{VSA4}}$ (sum of atomic $\log P(\text{o/w})$ with polarities in the range 0.1 to 0.15) in Eq. (7) also suggest in favor of positively polarized surface area for the R-PFT inhibitory activity. Similar to the examination of Pf -PFT specific models for R-PFT (Table 7; models 1b–3b), the scope of R-PFT specific models (Eqs. (4)–(7)) in explaining the Pf -PFT inhibitory activity is shown in Table 8 (models 4a–7a). All these Pf -PFT models (Table 8) are far inferior to the R-PFT specific models (Eqs. (4)–(7)). For the given set of compounds, the variations

Table 9
Correlation matrix of the 13 descriptors took part in the Pf -PFT specific and R-PFT specific CP-MLR models (Eqs. (1)–(7)) and their individual correlations with the PFT inhibitory activities of the THQ analogues (Table 1)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 PEOE_VSA-3	1.0														
2 E_{str}	0.531	1.0													
3 a_{hyd}	0.261	0.756	1.0												
4 PEOE_VSA+2	-0.639	-0.409	-0.131	1.0											
5 PEOE_VSA-5	-0.052	-0.001	-0.340	-0.490	1.0										
6 $v_{\text{sa_other}}$	0.392	0.294	-0.074	-0.283	0.385	1.0									
7 $\log P(\text{o/w})$	0.626	0.864	0.799	-0.355	-0.282	0.119	1.0								
8 PM3_Fele	-0.210	-0.559	-0.690	0.157	-0.074	-0.531	-0.437	1.0							
9 PEOE_VSA+6	0.280	0.404	0.056	0.006	-0.044	0.769	0.218	-0.387	1.0						
10 PEOE_VSA-4	-0.235	0.010	0.307	0.083	-0.001	0.133	0.035	-0.492	0.191	1.0					
11 $Q_{\text{VSA_FPPOS}}$	-0.298	-0.293	-0.609	0.278	0.278	0.551	-0.542	0.087	0.629	0.134	1.0				
12 $S \log P_{\text{VSA4}}$	0.378	0.373	0.582	0.092	-0.584	0.189	0.542	-0.468	0.301	0.042	-0.299	1.0			
13 E_{vdw}	0.611	0.881	0.854	-0.494	-0.079	0.076	0.850	-0.567	0.128	0.094	-0.598	0.457	1.0		
14 $-\log \text{IC}_{50}Pf\text{-PFT}$	-0.113	-0.617	-0.440	0.620	-0.467	0.010	-0.448	0.253	0.109	-0.023	0.356	0.163	-0.592	1.0	
15 $-\log \text{IC}_{50}\text{R-PFT}$	-0.134	-0.747	-0.733	0.339	-0.014	0.179	-0.620	0.381	0.103	0.049	0.550	-0.120	-0.749	0.818	1.0

in the *Pf*-PFT specific models (Eqs. (1)–(3)) and R-PFT specific models (Eqs. (4)–(7)) may be viewed as dissimilarity in the requirements of these enzyme systems. This suggests the possibility of modulating the *Pf*-PFT and R-PFT inhibitory activities as well as bringing about selectivity in the THQ analogues for the malarial parasite enzyme.

A PLS analysis has been carried out on the descriptors participated in *Pf*-PFT specific and R-PFT specific CP-MLR models to facilitate the development of single-window structure–activity models comprising all those descriptors and to identify their (descriptors) potential in modulating the PFT inhibitory activities of the THQ analogues (Table 1). It also gives an opportunity to make a comparison of relative significance among the descriptors. The fraction contributions obtainable from the normalized regression coefficients of descriptors allow this comparison within each modeled activity as well as between the modeled activities. For this, the 8 descriptors (dataset-1) of *Pf*-PFT specific models (Eqs. (1)–(3)), 8 descriptors (dataset-2) of R-PFT specific models (Eqs. (4)–(7)), and 13 descriptors

(dataset-3) of both of them together (Eqs. (1)–(7)) have been analyzed in the PLS. The intercorrelations between all 13 descriptors (corresponding to Eqs. (1)–(7) of datasets 1–3) as well as their correlation with the *Pf*-PFT and R-PFT inhibitory activities of the analogues (Table 1) are shown in Table 9. Only 4 descriptor-pairs consisting of E_str, E_vdw, a_hyd and log P(o/w) have shown intercorrelation above 0.80 (Table 9; r is between 0.854 and 0.887). Among the descriptors, only PEOE_VSA–4 has shown a correlation less than 0.10 with both the activities (Table 9; PEOE_VSA–4 vs. *Pf*-PFT and R-PFT, $r < |0.10|$). As all 13 descriptors are from the prior discussed CP-MLR models, they are opted for the PLS analysis without any omissions. The descriptors have been autoscaled (zero mean and unit S.D.) to give each one of them equal weight in the PLS analysis. Both the inhibitory activities (*Pf*-PFT and R-PFT) are concurrently analyzed (both X and Y descriptor blocks are in multivariate mode) with each dataset to identify the relative individual contributions of the descriptors to the modeled activities. In the cross-validation procedure of the PLS analysis [35,36] of these datasets, two components for each one of them are found to be the optimum to explain the activities.

Table 10

PLS and MLR-like PLS models of THQ analogues (Table 1) from the 8 descriptors of CP-MLR models specific to *Pf*-PFT activity (Eqs. (1)–(3)) for *Pf*-PFT and R-PFT inhibitory activities

A	PLS equation	–log IC ₅₀ <i>Pf</i> -PFT	–log IC ₅₀ R-PFT
	PLS factors	PLS coeff ^a (s.e)	PLS coeff ^a (s.e)
	Factor-1	0.352 (0.052)	0.370 (0.072)
	Factor-2	0.688 (0.102)	0.380 (0.141)
	Constant	8.280	8.198
B	MLR-like PLS equation	–log IC ₅₀ <i>Pf</i> -PFT	–log IC ₅₀ R-PFT
	S. No. Descriptor	MLR-like coeff ^b (f.c)	MLR-like coeff ^b (f.c)
1	PEOE_VSA–3	0.123 (0.133)	0.061 (0.092)
2	E_str	–0.239 (–0.158)	–0.196 (–0.183)
3	a_hyd	–0.106 (–0.114)	–0.093 (–0.141)
4	PEOE_VSA+2	0.020 (0.166)	0.015 (0.170)
5	PEOE_VSA–5	–0.063 (–0.196)	–0.039 (–0.172)
6	vsa_other	0.035 (0.158)	0.021 (0.131)
7	log P(o/w)	–0.104 (–0.060)	–0.115 (–0.094)
8	PM3_Eele	–0.0001 (–0.014)	0.0001 (0.017)
	Constant	8.076	9.785
Regression statistics			
	n	17	17
	r	0.931	0.840
	Q^2	0.778	0.561
	Q^2_{L3O}	0.795	0.542
	s	0.392	0.543
	F	45.66	16.74
	r_{yrand} (S.D.)	0.314 (0.156)	0.316 (0.174)
	External test set		
	$r^2_{\text{Pred}}^c$	0.648	—
	$r^2_{\text{Pred}}^d$	0.836	—

^a Regression coefficient of PLS factor and its standard error.

^b Coefficients of MLR-like PLS equation in terms of descriptors for their original values; f.c is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit S.D.) data.

^c For 7 compounds (18–24 of Table 6).

^d For 6 compounds (18, 19, and 21–24 of Table 6).

Table 11

PLS and MLR-like PLS models of THQ analogues (Table 1) from the 8 descriptors of CP-MLR models specific to R-PFT activity (Eqs. (4)–(7)) for *Pf*-PFT and R-PFT inhibitory activities

A	PLS equation	–log IC ₅₀ <i>Pf</i> -PFT	–log IC ₅₀ R-PFT
	PLS factors	PLS coeff ^a (s.e)	PLS coeff ^a (s.e)
	Factor-1	0.311 (0.087)	0.395 (0.058)
	Factor-2	–0.502 (0.154)	–0.368 (0.103)
	Constant	8.280	8.198
B	MLR-like PLS equation	–log IC ₅₀ <i>Pf</i> -PFT	–log IC ₅₀ R-PFT
	S. No. Descriptor	MLR-like coeff (f.c) ^b	MLR-like coeff (f.c) ^b
1	PEOE_VSA–3	0.085 (0.116)	0.056 (0.083)
2	E_str	–0.243 (–0.204)	–0.238 (–0.219)
3	a_hyd	–0.075 (–0.103)	–0.088 (–0.132)
9	PEOE_VSA+6	0.021 (0.093)	0.017 (0.083)
10	PEOE_VSA–4	0.015 (0.033)	0.012 (0.028)
11	Q_VSA_FPPOS	6.857 (0.081)	7.906 (0.103)
12	S log P_VSA4	0.164 (0.221)	0.120 (0.178)
13	E_vdw	–0.027 (–0.150)	–0.029 (–0.175)
	Constant	7.224	8.491
Regression Statistics			
	n	17	17
	r	0.790	0.899
	Q^2	0.460	0.718
	Q^2_{L3O}	0.412	0.691
	s	0.659	0.438
	F	11.59	29.55
	r_{yrand} (S.D.)	0.319 (0.139)	0.292 (0.136)

^a Regression coefficient of PLS factor and its standard error.

^b Coefficients of MLR-like PLS equation in terms of descriptors for their original values; f.c is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit S.D.) data.

The PLS model of the dataset-1 (*Pf*–PFT specific models) has explained 86.7% variance ($r = 0.931$, $s = 0.392$, $F = 45.66$) in the *Pf*–PFT inhibitory activity of the compounds. For the R–PFT inhibitory activity of the compounds, this PLS model (dataset-1) has explained about 70.5% variance ($r = 0.840$, $s = 0.543$, $F = 16.74$) in the activity. On the contrary, the PLS model of the dataset-2 (R–PFT specific models) has explained only 62.4% variance ($r = 0.790$, $s = 0.659$, $F = 11.60$) in the *Pf*–PFT inhibitory activity of the compounds. But, this PLS model (dataset-2) has explained higher variance, about 80.8% ($r = 0.899$, $s = 0.438$, $F = 29.55$) in the R–PFT inhibitory activity of the compounds. The two-component PLS equations and the corresponding MLR-like PLS equations of these two datasets are shown in Tables 10 and 11. The values in the parentheses following the coefficients of PLS factors are their standard errors. They suggest that the PLS factors are significant at more than 95% level. In MLR-like PLS equations, the values in the parentheses following the MLR-like regression coefficients are the fraction contributions of descriptors to the modeled activity. They are computed using the normalized regression

coefficients (not shown). They are used for the comparison of contribution of each descriptor with every other descriptor within and between the modeled activities. The *Pf*–PFT specific PLS model from datasets-1 (Table 10) has well predicted the *Pf*–PFT inhibitory activities of the external test set compounds (Tables 2 and 6). The predictions of internal test sets of the *Pf*–PFT specific and R–PFT specific inhibitory activities from appropriate PLS models of datasets-1 and -2 are shown in Tables 4 and 5. Fig. 3 shows the agreement between the observed versus predicted activities (*Pf*–PFT and R–PFT) of the compounds from the PLS analysis.

The dataset-3 includes all descriptors of datasets-1 and -2. The 2-component PLS model of this dataset has explained 78.5% variance ($r = 0.886$, $s = 0.498$, $F = 25.62$) in the *Pf*–PFT inhibitory activity and 68.6% variance ($r = 0.828$, $s = 0.561$, $F = 15.28$) in the R–PFT inhibitory activity of the compounds (Supplementary data). This analysis has been carried out to draw a comparison all 13 descriptors' contribution to each inhibitory activity. Fig. 4 shows the plots of comparative fraction contributions of the normalized regression coefficients of the PLS models' descriptors to the inhibitory

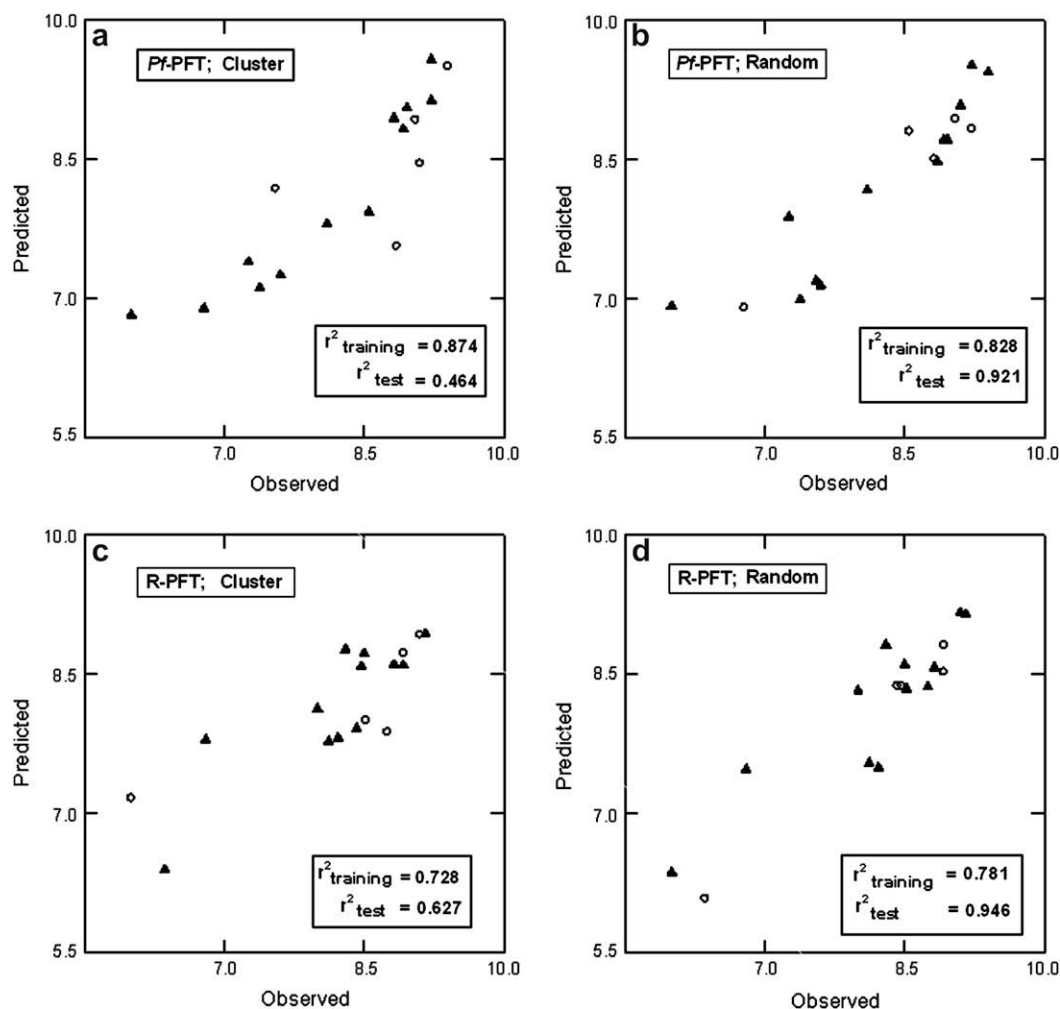


Fig. 3. Plots of training (▲) and test sets (○) predicted activities versus observed activity corresponding to PLS equations of *Pf*–PFT specific (a, b) (Table 10) and R–PFT specific (c, d) (Table 11) descriptors. Number of compounds in training set is 12 and test set is 5.

activities. The PLS model from the dataset-1 (*Pf*–PFT specific descriptors) indicates that PEOE_VSA+2 has made equal contribution to the *Pf*–PFT and R–PFT inhibitory activities of the compounds (Table 10; Fig. 4a). The coefficients of remaining 7 *Pf*–PFT specific descriptors suggest the possibility of modulating the *Pf*–PFT/R–PFT inhibitory activities of the compounds (Table 10; Fig. 4a). Among these 7 descriptors, PEOE_VSA–5, PEOE_VSA–3, *a*_{hyd}, log P(o/w) and PM3_Eele hold promise to impart selectivity to these compounds. For example, *a*_{hyd} and log P(o/w) of the compounds influenced the R–PFT inhibitory activity more pronouncedly when compared to that of *Pf*–PFT. The PLS model from the dataset-2 (R–PFT specific descriptors) has also shown some preferences for *Pf*–PFT and R–PFT inhibitory activities of the compounds (Table 11; Fig. 4b). It also suggests that PEOE_VSA–3 and *a*_{hyd} as potential descriptors to modulate the activities. The descriptor S log P_VSA4 specific to dataset-2 has displayed maximum deviation in its fraction contributions towards the activities (Table 11, Fig. 4b), thereby suggested its relevance in activity modulation. The preferences of PLS models of datasets-1 and -2 have clearly reflected in the PLS model of dataset-3 (Fig. 4c). It also clearly infers that PEOE_VSA+2 has matching influence on the *Pf*–PFT and R–PFT inhibitory activities of the analogues. Furthermore, it suggests that among the properties, partially negatively charged surfaces (PEOE_VSA–5 and PEOE_VSA–3), *a*_{hyd}, log P(o/w), electronic energy (PM3_Eele) and S log P_VSA4 hold promise for modulating the *Pf*–PFT/R–PFT inhibitory activities of the compounds (Fig. 4c). These results clearly differentiate the THQ analogues' requirements for *Pf*–PFT and R–PFT inhibitory activities and may point towards achieving the desired selectivity.

4. Conclusions

The molecular conformations of THQ analogues considered in the study have similarity with BMS-214662 (Fig. 2) bound to Rat–PFT (PDB code 1SA5). The high order of correspondence between *Pf*–PFT models (Eqs. (1)–(3)) and *Pf*–3D7 models (Table 7; models 1a–3a) signify the importance of the *Pf*–PFT specific descriptors in modeling the antimalarial activity of THQ analogues. The molecular energy parameters (*E*_{str} and PM3_Eele) present in the models (Eqs. (2) and (3)) suggest in favor of molecular arrangement with minimum energy levels for optimum activity. The van der Waals surface area (VSA) descriptors taken part in the models (Eqs. (1)–(3)) suggest in favor of low positively (PEOE_VSA+2) and negatively (PEOE_VSA–3) charged surfaces for better *Pf*–PFT inhibitory activity. In these compounds PEOE_VSA–3 represents the surface area composed of the N of nitrile group of A-ring and pyridyl nitrogen of E-region, whereas PEOE_VSA+2 represents the surface area composed of –CH= fragments of C- and D-rings (Fig. 2a). The participation descriptors log P(o/w) (Eq. (1)) and *a*_{hyd} (Eq. (3)) recommend decreasing hydrophobicity for better activity.

The *Pf*–PFT and R–PFT inhibitory activities of the THQ analogues are correlated to the extent of 67%. For *Pf*–PFT

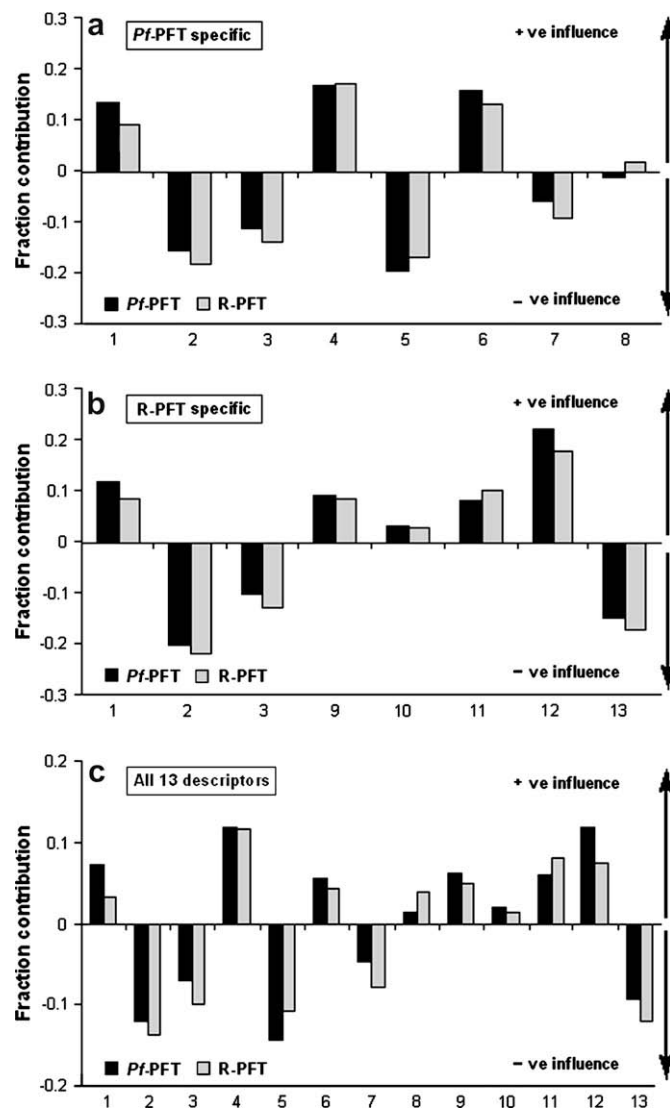


Fig. 4. Plots of fraction contribution of MLR-like PLS coefficients (normalized) of the (a) *Pf*–PFT specific (Table 10), (b) R–PFT specific (Table 11), and (c) jointly all *Pf*–PFT and R–PFT specific descriptors to the activities. The horizontal axis refers to the descriptors' numbers (Tables 10 and 11).

and *Pf*–3D7 inhibitory activities of the THQ analogues, this correlation is 81.5%. This deviation between the *Pf*–PFT and R–PFT systems for THQ analogues is well reflected in the models developed for the respective ones (Eqs. (1)–(3) and (4)–(7)). They share only three common descriptors (*E*_{str}, PEOE_VSA–3 and *a*_{hyd}) between them. Interestingly, the R–PFT specific models (Eqs. (4)–(7)) are in favor of more negatively (PEOE_VSA–4) as well as more positively (PEOE_VSA+6) charged surface areas for the inhibitory activity (Fig. 2a). The other charge/polar descriptors took part in R–PFT models also favor more polar hydrophilic surfaces for the activity. This suggested the possibility of modulating the *Pf*–PFT/R–PFT inhibitory activities and bringing about selectivity in the THQ analogues for the malarial parasite enzyme. Fig. 2a identifies the structural location and region of the van der Waals surface areas (PEOE_VSA

descriptors) of THQ analogues that have taken part in the PFT inhibitory activity equations.

The PLS analysis carried out on the descriptors of all the models suggested that PEOE_VSA+2 equally contribute to the *Pf*–PFT and R–PFT inhibitory activities. It also suggests that among the properties, the partially charged surface areas in the range -0.20 to -0.15 (PEOE_VSA–3) and -0.30 to -0.25 (PEOE_VSA–5), a_{hyd} , $\log P(o/w)$, electronic energy (PM3_Eele) and $S \log P_{\text{VSA4}}$ of the molecules hold promise for modulating the *Pf*–PFT/R–PFT inhibitory activities of the compounds (Fig. 4). For example, in these compounds a_{hyd} and $\log P(o/w)$ influence the R–PFT inhibitory activity more dominantly when compared to that of *Pf*–PFT. These results clearly differentiated the THQ analogues' requirements for *Pf*–PFT and R–PFT inhibitory activities and set direction for achieving the desired selectivity.

Acknowledgment

MKG thanks CSIR, New Delhi India for the financial support in the form of Senior Research Fellowship. CDRI Communication No. 7329.

Appendix A. Supplementary data

Complete dataset of molecular descriptors of structure databases of Tables 1 and 2; regression equations of Tables 7 and 8; PLS factor scores, loadings, weights and sensitivity of independent and dependent descriptors of the PLS models. Supplementary data associated with this article can be found in the online version, at [doi:10.1016/j.ejmech.2008.01.025](https://doi.org/10.1016/j.ejmech.2008.01.025).

References

- <http://www.globalhealthfacts.org>.
- G.A. McConkey, Antimicrob. Agents. Chemother. 43 (1999) 175–177.
- M. Calas, M.L. Ancelin, G. Cordina, P. Portefaix, G. Piquet, V. Vidal-Sailhan, H. Vial, J. Med. Chem. 43 (2000) 505–516.
- R. McLeod, S.P. Muench, J.B. Rafferty, R. McLeod, S.P. Muench, J.B. Rafferty, D.E. Kyle, E.J. Mui, M.J. Kirisits, D.G. Mack, C.W. Roberts, B.U. Samuel, R.E. Lyons, M. Dorris, W.K. Milhous, D.W. Rice, Int. J. Parasitol. 31 (2001) 109–113.
- T. Joet, U. Eckstein-Ludwig, C. Morin, S. Krishna, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 7476–7479.
- M.H. Gelb, W.C. Van Voorhis, F.S. Buckner, K. Yokoyama, R. Eastman, E.P. Carpenter, C. Panethymitaki, K.A. Brown, D.F. Smith, Mol. Biochem. Parasitol. 126 (2003) 155–163.
- E. Nduati, S. Hunt, E.M. Kamau, A. Nzila, Antimicrob. Agents Chemother. 49 (2005) 3652–3657.
- K. Yokoyama, G.W. Goodwin, F. Ghomashchi, J. Glomset, M.H. Gelb, Biochem. Soc. Trans. 20 (1992) 489–494.
- R.B. Lobell, C.A. Omer, M.T. Abrams, H.G. Bhimnathwala, M.J. Brucker, C.A. Buser, J.P. Davide, S.J. deSolms, C.J. Dinsmore, M.S. Ellis-Hutchings, A.M. Kral, D. Liu, W.C. Lumma, S.V. Machotka, E. Rands, T.M. Williams, S.L. Graham, G.D. Hartman, A.I. Oliff, D.C. Heimbrook, N.E. Kohl, Cancer Res. 61 (2001) 8758–8768.
- www.plasmodb.org.
- H. Ohkanda, J.W. Lockman, K. Yokoyama, M.H. Gelb, S.L. Croft, H. Kendrick, M.I. Harrell, J.E. Feagin, M.A. Blaskovich, S.M. Sehti, A.D. Hamilton, Bioorg. Med. Chem. Lett. 11 (2001) 761–764.
- D. Chakrabarti, T. Da Silva, J. Barger, S. Paquette, H. Patel, S. Patterson, C.M. Allen, J. Biol. Chem. 277 (2002) 42066–42073.
- J. Wiesner, K. Kettler, J. Sakowski, R. Ortmann, A.M. Katzin, E.A. Kimura, K. Silber, G. Klebe, H. Jomaa, M. Schlitzer, Angew. Chem., Int. Ed. Engl. 43 (2004) 251–254.
- D. Carrico, J. Ohkanda, H. Kendrick, K. Yokoyama, M.A. Blaskovich, C.J. Bucher, F.S. Buckner, W.C. Van Voorhis, D. Chakrabarti, S.L. Croft, M.H. Gelb, S.M. Sehti, A.D. Hamilton, Bioorg. Med. Chem. 12 (2004) 6517–6526.
- M.P. Glenn, S.Y. Chang, O. Huckle, C.L. Verlinde, K. Rivas, C. Hornéy, K. Yokoyama, F.S. Buckner, P.R. Pendyala, D. Chakrabarti, M. Gelb, W.C. Van Voorhis, S.M. Sehti, A.D. Hamilton, Angew. Chem., Int. Ed. Engl. 44 (2005) 4903–4906.
- A. Ryckebusch, P. Gilleron, R. Millet, R. Houssin, A. Lemoine, N. Pommery, P. Grellier, C. Sergheraert, J.P. Hénichart, Chem. Pharm. Bull. 53 (2005) 1324–1326.
- K. Kettler, J. Wiesner, K. Silber, P. Haebel, R. Ortmann, I. Sattler, H.M. Dahse, H. Jomaa, G. Klebe, M. Schlitzer, Eur. J. Med. Chem. 40 (2005) 93–101.
- M.P. Glenn, S.Y. Chang, C. Hornéy, K. Rivas, K. Yokoyama, E.E. Pusateri, S. Fletcher, C.G. Cummings, F.S. Buckner, P.R. Pendyala, D. Chakrabarti, S.M. Sehti, M. Gelb, W.C. Van Voorhis, A.D. Hamilton, J. Med. Chem. 49 (2006) 5710–5727.
- L. Nallan, K.D. Bauer, P. Bendale, K. Rivas, K. Yokoyama, C.P. Horney, P.R. Pendyala, D. Floyd, L.J. Lombardo, D.K. Williams, A. Hamilton, S. Sehti, W.T. Windsor, P.C. Weber, F.S. Buckner, D. Chakrabarti, M.H. Gelb, W.C. Van Voorhis, J. Med. Chem. 48 (2005) 3704–3713.
- M.J. Polley, D.A. Winkler, F.R. Burden, J. Med. Chem. 47 (2004) 6230–6238.
- M.P. Gonzalez, J. Caballero, A. Tundidor-Camba, A.M. Helguera, M. Fernandez, Bioorg. Med. Chem. 14 (2006) 200–213.
- D.S. Puntambekar, R. Giridhar, M.R. Yadav, Bioorg. Med. Chem. Lett. 16 (2006) 1821–1827.
- D.S. Puntambekar, R. Giridhar, M.R. Yadav, Eur. J. Med. Chem. 41 (2006) 1279–1292.
- A. Lu, J. Zhang, X. Yin, X. Luo, H. Jiang, Bioorg. Med. Chem. Lett. 17 (2007) 243–249.
- T. Equbal, O. Silakari, G. Rambabu, M. Ravikumar, Bioorg. Med. Chem. Lett. 17 (2007) 1594–1600.
- A. Xie, P. Sivaprakasam, R.J. Doerksen, Bioorg. Med. Chem. 14 (2006) 7311–7323.
- T.S. Reid, L.S. Beese, Biochemistry 43 (2004) 6877–6884.
- Y.S. Prabhakar, QSAR Comb. Sci. 22 (2003) 583–595.
- P. Bendale, S. Olepu, P.K. Suryadevara, V. Bulbule, K. Rivas, L. Nallan, B. Smart, K. Yokoyama, S. Ankala, P.R. Pendyala, D. Floyd, L.J. Lombardo, D.K. Williams, F.S. Buckner, D. Chakrabarti, C.L. Verlinde, W.C. Van Voorhis, M.H. Gelb, J. Med. Chem. 50 (2007) 4585–4605.
- MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1255 University Street, Suite 1600, Montreal, Quebec, Canada H3B 3X3. <http://www.chemcomp.com>.
- R.T. Eastman, J. White, O. Huckle, K. Bauer, K. Yokoyama, L. Nallan, D. Chakrabarti, C.L. Verlinde, M.H. Gelb, P.K. Rathod, W.C. Van Voorhis, J. Biol. Chem. 280 (2005) 13554–13559.
- D.M. Ferguson, D.J. Raber, J. Am. Chem. Soc. 111 (1989) 4371–4378.
- T.A. Halgren, J. Comput. Chem. 17 (1996) 490–519.
- The compounds of Table 2 have been included as external test set of the study while revising the manuscript after the referees' suggestions on the initial submission.
- S. Wold, Technometrics 20 (1978) 397–406.
- L. Stahle, S. Wold, in: G.P. Ellis, W.B. West (Eds.), Progress in Medicinal Chemistry, vol. 25, Elsevier Science Publishers, B.V. Amsterdam, 1988, pp. 291–338 (Chapter 6).
- Y.S. Prabhakar, Internet Electron. J. Mol. Des. 3 (2004) 150–162. <http://www.biochempress.com>.
- Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta, S.B. Katti, QSAR Comb. Sci. 23 (2004) 234–244.

- [39] M.K. Gupta, R. Sagar, A.K. Shaw, Y.S. Prabhakar, *Bioorg. Med. Chem.* 13 (2005) 343–351.
- [40] Y.S. Prabhakar, R.K. Rawal, M.K. Gupta, V.R. Solomon, S.B. Katti, *Comb. Chem. High Throughput Screen* 5 (2005) 431–437.
- [41] Y.S. Prabhakar, M.K. Gupta, N. Roy, Y. Venkateswarlu, *J. Chem. Inform. Model.* 46 (2006) 86–92.
- [42] M.K. Gupta, Y.S. Prabhakar, *J. Chem. Inform. Model.* 46 (2006) 93–102.
- [43] M. Saquib, M.K. Gupta, R. Sagar, Y.S. Prabhakar, A.K. Shaw, R. Kumar, P.R. Maulik, A.N. Gaikwad, S. Sinha, A.K. Srivastava, V. Chaturvedi, R. Srivastava, B.S. Srivastava, *J. Med. Chem.* 50 (2007) 2942–2950.
- [44] S.S. So, M. Karplus, *J. Med. Chem.* 40 (1997) 4347–4359.
- [45] J.B. Bhonsle, A.K. Bhattacharjee, R.K. Gupta, *J. Mol. Model.* 13 (2007) 179–208.