informa healthcare

Topological descriptors in modelling antimalarial activity: N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine as prototype

SHREEKANT DESHPANDE, V. RAJA SOLOMON, SETU B. KATTI, & YENAMANDRA S. PRABHAKAR

Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow 226 001, India

Abstract
The QSAR of antimalarial activity of two distinct series of N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl) piperazine analogues are investigated with DRAGON descriptors in order to rationalize their activity. Of these two series of compounds, one has amide characteristics and the other has amine characteristics. Both the analogues have shared radial centric information (ICR) as common modelling descriptor with increased centricity in the molecules as preferred feature for antimalarial activity. Apart from this, the models of amide analogues suggested in favor of distantly placed nitrogen(s) and unfavorable nature of carbonyl moieties adjacent to nitrogen in the varying portion of the molecule for the activity. Moreover, for these analogues, the regression models have preferred the lone pair electrons on heteroatoms (N and O) for purposes other than H-bonds for better activity. In case of amine analogues, the models suggested in favor of compact structural moieties in the varying parts of the molecule for improved activity. Also, for these analogues, hydrophobicity of the compound is an important factor for influencing activity. The variations in the models of amide and amine analogues are attributed to the characteristic functional differences of these analogues.

Keywords: *N^1-(7-Chloro-4-quinolyl)-1, 4-bis(3-aminopropyl)piperazine derivatives, CP-MLR, antimalarial activity, topological descriptors, DRAGON software*

Introduction

The control of malaria is globally a high priority task as it is associated with the high morbidity and mortality [1]. In this task, chloroquine (CQ) and other aminoquinolines (AQ) are the frontline chemotherapeutic agents because of their therapeutic efficacy and low cost [2]. These compounds enter the food vacuole of the parasite and inhibit its growth by forming complex with hematin thereby preventing the formation of hemozoin [3]. However, the emergence and proliferation of multidrug-resistant strains of *Plasmodium* species necessitated the development of alternative antimalarial agents [4]. The mechanistic investigations underlying the drug resistance have indicated that the resistance is a consequence of decreased accumulation of the drug in the food vacuole of the parasite owing to the enhanced efflux and reduced uptake [5]. Moreover, the studies with CQ/AQ analogues have suggested that resistance mechanism does not involve any change in the target of this class of drugs [6].

The structure activity relationship studies in 4-aminoquinoline class have suggested that 7-chloro-4-aminoquinoline nucleus as obligatory for the antimalarial activity. It helps in maintaining the PK_a of quinoline nitrogen thereby the accumulation of the drug at the target and inhibition of the formation of hematin [7]. In an attempt to overcome the drug efflux mechanism of the parasite, Vennerstrom et al have investigated a series of bisquinolines (Figure 1) against CQ-sensitive and CQ resistant strains [8]. In this
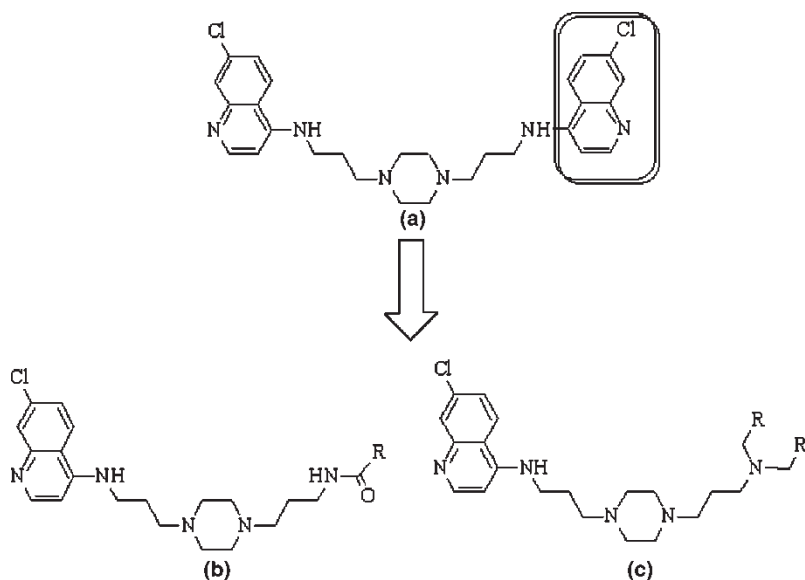
RIGHTSLINK

Figure 1. General structures of (a) bisquinoline (b) N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine amide derivatives and (c) N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine amine derivatives as antimalarial agents against the chloroquine-resistant *P. falciparum* FcB1.

scenario, Ryckebusch et al have prepared several N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine derivatives (Figure 1) by replacing one of the 7-chloro-4-aminoquinoline moieties of the bisquinolines with various amines and amides and evaluated them for antimalarial activity against the chloroquine-resistant *P. falciparum* FcB1 strain [9]. For these analogues, Ryckebusch et al have propounded CQ-like mechanism as well as the involvement of other mechanisms [9]. Our explorations with some modified 4-aminoquinoline antimalarials have suggested the role of lipophilicity and PK_a of the side chain on the activity [10].

The investigation of quantitative structure-activity relationships (QSARs) between the properties of the chemical entities and their activity will come to aid in advancing the understanding of the system. The early QSAR studies with tebuquine analogues proposed a correlation of the antimalarial activity with the size and electron donating property of the 4-anilino substituents of these molecules [11]. Our recent QSAR study with 7-chloro-4-(3′,5′-disubstituted anilino)quinolines suggested that the 4-anilino moiety of these compounds map a large domain in the activity space and also preferred electron rich substituent groups in this substructure space for better antimalarial activity [12].

In this milieu to rationalize and explore the scope of the antimalarial activity of aminoquinoline based agents we attempted a QSAR study on N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine derivatives (Figure 1) [9]. The physicochemical and topological features of the compounds are important in drug design and discovery studies [13]. They describe the hydrophobic, steric, electronic and topological characteristics of the molecules. In view of this, the chemical space of N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl) piperazine derivatives (Figure 1) [9]. has been parameterized in terms of different descriptor classes which include empirical, topological, functional, atom centered fragments and other molecular properties from DRAGON software [14]. to rationalize the antimalarial activity profile of these analogues. The QSAR models are developed using the variable selection procedure, combinatorial protocol in multiple linear regression (CP-MLR) [15]. The results are presented here.

Computation

Dataset

Several N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine analogues are reported in the recent literature along with their in vitro antimalarial activity (IC_{50}, inhibitory concentration in moles per liter against the chloroquine-resistant *P. falciparum* FcB1 strain) [9]. Among these analogues, one compound, cyclopropane carboxylic acid(3-[4-[3-(7-chloro-quinolin-4ylamino)-propyl]-piperazin-1-yl}-propyl)-cyclopropyl methyl amide, is embedded with a highly reactive cyclopropane carbonyl moiety. Due to this reason it has been kept out from the analogues selected for the QSAR study. The selected compounds are divided into two sets, as N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine amide derivatives (briefly referred as amides; Table I) and N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine amine derivatives (briefly referred as amines; Table II). For the purpose of QSAR study the antimalarial activity of the compounds has been transformed into logarithm of reciprocal inhibitory concentration and expressed as $-\log IC_{50}$. The structure databases of

Table I.   Observed and modeled in vitro antimalarial activity of N^1-(7-chloro-4-quinolyl)1,4-bis(3-aminoprpyl)piperazine- amide derivatives (Figure 1b) against the FcB1R strain of *P. falciparum*.

						$-\log IC_{50}$					
Comp No	R	Obs.[*]	Equation 1[†]	Set 1[‡]	Set 2[‡]	Equation 2	Set 1	Set 2	Equation 3	Set 1	Set 2
1	4-quinolinyl	7.70	7.32	7.41	*7.29*	7.08	7.11	*7.09*	7.32	7.35	6.99
2	1-naphthyl	6.81	7.12	7.18	7.04	7.18	7.25	7.17	7.10	7.13	6.98
3	Phenyl	7.09	7.21	*7.22*	*7.11*	7.15	*7.17*	*7.10*	7.16	*7.06*	*7.05*
4	3-thiophenyl	6.95	6.93	6.99	6.80	6.90	6.95	6.84	6.90	6.92	6.81
5	3-phenoxyphenyl	7.68	7.96	7.90	7.98	7.82	7.80	7.77	7.83	7.78	7.51
6	4-chlorophenyl	7.78	8.44	*8.38*	8.57	7.30	*7.42*	7.13	7.90	*7.84*	7.77
7	4-methoxyphenyl	7.70	7.56	7.54	*7.53*	7.66	7.73	*7.61*	7.60	7.69	*7.59*
8	4-fluorophenyl	7.31	7.11	7.14	7.01	7.31	7.26	7.27	7.20	7.32	7.21
9	4-nitrophenyl	7.91	7.88	8.06	7.83	8.34	7.94	8.37	8.03	8.07	8.02
10	4-hydroxyphenyl	6.38	6.92	*6.93*	6.88	6.95	6.96	6.94	6.87	7.09	6.72
11	Benzyl	7.17	6.96	6.98	*6.93*	7.10	7.10	*7.06*	7.01	6.98	6.96
12	phenethyl	7.09	7.43	7.42	7.42	7.54	7.52	7.52	7.40	7.38	7.31
13	cyclohexyl	6.67	6.97	7.01	6.93	6.90	6.90	6.89	6.90	6.85	6.86
14	cyclopropyl	6.54	6.66	*6.71*	6.59	6.64	*6.63*	6.61	6.62	*6.85*	6.67
15	Hexyl	7.46	7.43	7.39	*7.41*	7.31	7.22	*7.31*	7.37	7.21	*7.21*
16	Propyl	7.13	6.60	6.63	6.49	6.61	6.58	6.54	6.62	6.66	6.62
17	Ethyl	6.56	6.64	*6.69*	6.56	6.72	*6.74*	6.69	6.64	*6.73*	6.59
18	methyl	6.40	6.31	6.42	*6.23*	6.45	6.54	*6.40*	6.35	6.47	*6.28*
19	*tert*-butyl	6.79	6.61	6.67	*6.56*	6.89	6.98	*6.83*	6.71	6.64	*6.59*
20	isopropyl	6.83	6.64	*6.71*	6.58	6.77	*6.81*	6.73	6.68	*6.74*	6.63
21	N-Boc-Gly	6.41	6.71	*6.53*	6.86	6.71	*6.47*	6.87	6.66	*6.79*	6.63
22	N-Boc-L-Pro	6.37	6.57	6.48	*6.60*	6.59	6.35	*6.64*	6.56	6.37	6.27
23	N-Boc-(R)-1,2,3,4-tetrahydro-3-isoquinolinyl	6.41	6.44	*6.34*	6.53	6.62	*6.41*	6.80	6.50	*6.53*	6.28
24	N-Boc-(S)-1,2,3,4-tetrahydro-3-isoquinolinyl	6.40	6.45	6.28	*6.50*	6.62	6.42	*6.66*	6.50	6.53	*6.60*
25	Gly	6.50	6.38	6.64	*6.41*	6.54	6.64	*6.50*	6.46	6.70	6.87
26	L-Pro	6.44	6.75	6.99	6.77	6.57	6.62	6.54	6.66	6.77	6.80
27	(R)-1,2,3,4-tetrahydro-3-isoquinolinyl	7.72	7.23	7.33	7.18	7.37	7.39	7.29	7.37	7.43	7.47
28	(S)-1,2,3,4-tetrahydro-3-isoquinolinyl	7.56	7.27	7.38	7.25	7.39	7.42	7.33	7.37	7.43	7.47
29	1,2,3,4-tetrahydro-3-isoquinolinyl	7.42	7.30	7.28	7.26	7.48	7.46	7.43	7.36	7.40	7.34
30[¶]	1-propanol carbonyl	6.76	6.77	*6.81*	6.70	6.59	*6.61*	6.53	6.68	*6.64*	6.35
31[¶]	1-propanoic acid carbonyl	5.86	5.86	*5.68*	5.90	5.92	*5.67*	6.10	5.84	*6.49*	5.96
32[¶]	1-propanamide carbonyl	ND[§]	5.90			5.90			5.90		
33[¶]	N-Boc ethyl amine carbonyl	7.47	6.82	*6.83*	6.78	6.43	*6.41*	6.41	6.85	*7.25*	7.40
34[¶]	ethyl amine carbonyl	6.40	6.59	*6.74*	*6.52*	6.35	*6.39*	*6.32*	6.49	*6.56*	*6.54*
	Training r²			0.82	0.78		0.80	0.70		0.81	0.84
	Test r²			0.73	0.83		0.66	0.81		0.57	0.74

[*] Observed activity; ref. [9a] for compounds 1-29 and ref. [9b] for compounds 30–34; [†] For all equations 'set 1' corresponds to MACCS cluster and 'set 2' corresponds to random selection; unless otherwise stated, all predicted activities are from LOO cross-validation; [‡] Test set compounds are identified in italics and their activities are predicted with the corresponding model developed using the remaining compounds as the training group; [¶] For Compounds **30–34**, amide 'H' is replaced by methylene cyclopropane; [§] ND - not determined, see ref. [9b].

Table II. Observed and modeled in vitro antimalarial activity of N^1-(7-chloro-4-quinolyl)1,4-bis(3-aminoprpyl)piperazine-amine derivatives (Figure 1c) against the FcB1R strain of *P. falciparum*.

						$- \mathrm{logIC}_{50}$					
Comp.No*	R	Obs [†]	Equation 4 [‡]	Set 1 [¶]	Set 2 [¶]	Eq.5	Set 1	Set 2	Eq.6	Set 1	Set 2
35	4-quinolinyl	7.13	7.39	7.42	*7.38*	7.44	7.43	*7.45*	7.31	7.34	*7.35*
36	1-naphthyl	7.58	7.36	*7.38*	7.35	7.45	*7.42*	7.48	7.34	*7.35*	7.36
37	Phenyl	7.78	7.57	7.58	7.52	7.78	7.80	7.73	7.62	7.58	7.58
38	3-thiophenyl	7.94	7.61	7.62	*7.58*	7.56	7.55	*7.53*	7.56	7.61	*7.51*
39	3-phenoxyphenyl	7.70	7.12	*7.18*	7.14	7.54	*7.54*	7.58	7.28	*7.11*	7.36
40	4-chlorophenyl	7.78	7.97	8.06	*7.91*	7.86	7.84	*7.85*	7.91	7.96	*7.91*
41	4-methoxyphenyl	7.71	7.96	8.05	7.93	7.99	7.99	8.01	7.98	7.93	7.98
42	4-fluorophenyl	7.73	7.90	*7.95*	7.87	7.86	*7.83*	7.86	7.88	*7.94*	7.88
43	4-nitrophenyl	8.23	7.78	7.82	*7.81*	7.98	7.92	*8.04*	7.91	7.87	*7.95*
44	4-hydroxyphenyl	7.26	7.56	*7.61*	7.50	7.36	*7.30*	7.33	7.43	*7.55*	7.40
45	Benzyl	7.73	7.42	7.41	7.37	7.58	7.56	7.53	7.46	7.41	7.42
46	Phenethyl	7.59	7.49	*7.51*	7.45	7.67	*7.65*	*7.61*	7.57	*7.43*	*7.53*
47	Cyclohexyl	7.37	7.57	7.60	7.51	7.59	7.65	7.48	7.56	7.51	7.46
48	Cyclopropyl	7.99	7.70	7.72	7.61	7.48	7.48	7.33	7.65	7.69	7.54
49	Hexyl	8.02	7.60	*7.64*	*7.54*	7.76	*7.81*	*7.70*	7.72	*7.55*	*7.63*
50	Propyl	7.61	7.69	7.72	7.59	7.76	7.76	7.65	7.70	7.71	7.61
51	Ethyl	7.44	7.75	*7.76*	7.64	7.82	*7.78*	7.71	7.75	*7.79*	7.65
52	Methyl	7.24	7.67	7.72	7.56	7.75	7.77	7.64	7.64	7.74	7.52
53	tert-butyl	8.01	7.85	7.89	7.75	8.11	8.06	8.04	7.93	7.98	7.88
54	Isopropyl	7.73	7.79	7.83	7.68	7.97	7.95	7.88	7.84	7.87	7.76
55	4-quinolinyl	6.85	6.93	*6.84*	7.12	6.65	*6.55*	6.90	6.80	*6.85*	6.94
56	1-naphthyl	6.67	6.84	6.71	7.04	6.87	6.70	7.14	6.74	6.74	6.88
57	Phenyl	7.26	7.58	7.55	*7.59*	7.71	7.71	*7.71*	7.56	7.53	*7.60*
58	3-thiophenyl	7.20	7.45	7.39	7.46	7.36	7.30	7.38	7.34	7.33	7.34
59	3-phenoxyphenyl	6.58	6.67	6.53	*6.85*	6.86	6.69	*7.02*	6.73	6.61	*6.96*
60	4-chlorophenyl	ND§	7.90			7.59			7.82		
61	4-methoxyphenyl	8.02	7.73	7.69	7.71	7.66	7.49	7.75	7.83	7.77	7.92
62	4-fluorophenyl	ND§	7.74			7.68			7.76		
63	4-nitrophenyl	7.34	7.53	7.54	*7.60*	7.38	7.28	*7.52*	7.53	7.55	*7.66*
64	4-hydroxyphenyl	7.34	7.06	*7.20*	6.99	6.61	*6.67*	6.66	7.01	*7.17*	7.06
65	Benzyl	7.58	7.28	7.19	7.33	7.51	7.46	7.56	7.35	7.28	7.40
66	Phenethyl	7.40	7.44	*7.37*	7.48	7.36	*7.29*	7.41	7.47	*7.33*	7.50
67	Cyclohexyl	7.80	7.75	7.71	*7.77*	7.64	7.69	*7.63*	7.70	7.68	*7.67*
68	Cyclopropyl	8.06	8.56	*8.52*	8.55	8.10	*8.17*	8.03	8.36	*8.49*	8.29
69	Hexyl	7.42	7.75	7.71	7.71	7.88	7.98	7.89	7.80	7.58	7.75
70	propyl	8.13	8.09	8.09	8.05	8.28	8.31	8.23	8.14	8.14	8.11
71	Ethyl	8.66	8.31	*8.36*	*8.27*	8.28	*8.33*	8.66	8.37	*8.40*	*8.31*
72	methyl	8.30	7.93	7.84	7.75	7.94	7.79	7.78	8.05	8.06	7.96
73	tert-butyl	ND§	9.56			9.59			9.51		
74	isopropyl	9.05	8.80	8.93	8.70	8.86	8.97	8.83	8.91	9.07	8.93
75	hydrogen	7.99	7.72	7.75	*7.65*	7.56	7.57	*7.49*	7.68	7.75	*7.57*
76	1-pentane nitrile	8.00	8.23	8.25	8.19	7.78	7.78	7.75	8.11	8.11	8.07
77	N-Boc propyl amine carbonyl	7.90	7.89	7.89	*7.89*	7.97	7.88	*8.01*	7.99	7.84	*8.04*

Table II – *continued*

Comp.No*	R	Obs †	Equation 4 ‡	Set 1 ¶	Set 2 ¶	Eq.5	Set 1	Set 2	Eq.6	Set 1	Set 2
						$-\log IC_{50}$					
78	propyl amine carbonyl	6.82	7.40	7.38	7.35	7.18	7.08	7.12	7.16	7.32	7.09
79	pyrrolidinyl	7.71	7.88	7.85	7.80	7.84	7.86	7.73	7.86	7.89	7.75
80	piperidinyl	7.94	7.87	7.83	7.80	7.79	7.79	7.72	7.87	7.86	7.76
81	azepaninyl	7.87	7.87	7.84	7.78	7.71	7.72	7.57	7.85	7.83	7.74
	Training r²			0.79	0.73		0.76	0.77		0.79	0.79
	Test r²			0.57	0.67		0.69	0.62		0.53	0.65

* for compounds **55–74**, R = R₁; for compounds **75–81**, R₁ is cyclopropyl methyl; † observed activity; ref. [9a] for compounds **35–74** and ref. [9b] for compounds **75–81**; ‡ see the corresponding footnote 'b' and 'c' in Table I; ¶ ND - not determined, see ref.[9a].

these compounds were generated in ChemDraw [16]. In DRAGON software[14]. these structure databases have resulted in 474 and 483 molecular descriptors for the compounds of Tables I and II, respectively. They represent 0D- to 2D- characteristics of the molecular structures. According to the formalism of quantification of embedded structural information, all these indices belong to ten descriptor classes [14]. The QSAR studies with divergent descriptor classes provide scope to view and understand the activity from different perceptions embedded in the descriptors.

Model development

The QSAR model generation and validation was carried out using CP-MLR (combinatorial protocol in multiple linear regression) [15]. in conjunction with a three-stage descriptor classification protocol [12]. CP-MLR [15]. is a 'filter' based variable selection procedure for the model identification and development in QSAR and QSPR studies [17–18]. This involves a combinatorial strategy with appropriately placed 'filters' interfaced with MLR. It results in the extraction of diverse structure-activity models, each having unique combination of descriptors from the datasets under study. In this, the filters set the thresholds for the descriptors in terms of inter-parameter correlation cutoff limits in subset regressions (filter-1), t-values of the regression coefficients (filter-2), internal explanatory power (filter-3; square-root of adjusted multiple correlation coefficient of regression Equation, r-bar) and the external consistency (filter-4; Q^2 i.e. cross-validated R^2 from the leave-one-out procedure).

The three-stage descriptor classification protocol [12]. is implemented with the 3-descriptor combinations (baseline models) as they are the simplest ones obtained to explain the activity. In the first stage of classification protocol, the baseline models from the individual descriptor classes of the dataset were used to sort the descriptor classes into four categories. They are primary contributors (category I: a descriptor class forms model with its constituent descriptors), collective contributors (category II: a descriptor class unable to form model with its constituent descriptors, but forms model(s) in combination with a descriptor from another such descriptor class), secondary contributors (category III: a descriptor class forms model(s) only in combination with category I) and non-contributors (category IV: a descriptor class unable to form model(s) in any manner like category I, II or III; category IV are omitted from the study). The sorted descriptor classes were collated in the second stage to identify all the 3-descriptor models across the categories. In the last stage, the individual descriptors of all 3-descriptor models were pooled to discover the higher models for the activity.

Throughout this study, for the filters-1, 2, and 4 of CP-MLR the thresholds were assigned as 0.3, 2.0, and $0.3 \leq Q^2 \leq 1.0$, respectively. The filter-3 was assigned

an initial value of 0.71. In order to collect the descriptors with higher information content, 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. With these provisions CP-MLR was used in conjunction with the three-stage descriptor classification protocol [12].

All the identified models were reassessed for the chance correlations, if any, by repeated randomization of the biological response [18–19]. For this each identified model was subjected to one hundred simulation runs with scrambled activity. The emerging regression equations with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted to express the percent chance correlation of the model under examination. Additionally, the proposed models were verified by creating two divergent test sets, one emanating from the cluster analysis of the bit-packed version of the MACCS fingerprints (FP BIT MACCS) [20] of the compounds and the other from the random selection procedure, with each containing about one-third of the total compounds under analysis. As the total number of descriptors involved in this study is very large, only the descriptors significant to the models are addressed in the discussion. The complete descriptor file of each structure database is provided as supplementary material to the article.

Results and discussion

In multi-descriptor class environment, exploring for models along the descriptor class provides scope to understand the phenomenon under investigation in relation to the concepts embedded in them. With this view, N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine analogues [9] (Tables I and II) are analyzed using different 0D to 2D-descriptor classes from DRAGON software [14]. For these compounds all the descriptor classes have evolved as collective contributors (category II) to explain the antimalarial activity. At the end of a search for four parameter models, from a pool of equations the following are selected to explain the antimalarial activity of amide analogues (N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine amide derivatives; Table I).

$$-\log IC_{50} = -2.654 - 0.952(0.123)nDB$$
$$+ 2.989(0.438)ICR$$
$$+ 0.012(0.003)T(N..Cl)$$
$$- 0.236(0.063)nHDon$$
$$n = 33, \ r = 0.900, \ Q^2 = 0.707$$
$$s = 0.254, \ F = 29.97 \qquad (1)$$

$$-\log IC_{50} = 1.137 - 0.869(0.139)nDB$$
$$+ 3.311(0.503)ICR$$
$$- 2.160(1.004)GATS1e$$
$$- 3.062(1.293)GATS2p$$
$$n = 33, \ r = 0.864, \ Q^2 = 0.639,$$
$$s = 0.293, \ F = 20.57 \qquad (2)$$

In this and all other 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, s is the standard error of the estimate and F is the F-ratio between the variances of calculated and observed activities. The values given in the parentheses are the standard errors of the regression coefficients. In the randomization study, none of the identified models has shown any chance correlation. These are further validated through two test sets corresponding to the cluster analysis of the bit-packed version of the MACCS fingerprints (FP BIT MACCS) [20] of the compounds and the random selection procedure, with each one containing twelve out of thirty-three compounds of Table I. The test sets predictions are in agreement with their experimental values (Table I; Figure 2).

The Equations (1) and (2) have collectively shared six descriptors. They are from four different descriptor classes namely, constitutional (nDB), topological (ICR and T(N..Cl)), 2D-autocorrelations (GATS1e and GATS2p), and functional groups count (nHDon) [15]. In these equations, nDB represents the number of isolated double bonds in the molecule. In the compounds, these double bonds address the carbonyl moiety of amide and carboxyl groups present in the varying parts. Its negative regression coefficient suggests in favor of minimum number of such functions in the varying part of the amide analogues for improved activity. In terms of physicochemical properties, the presence of carbonyl moiety reduces the basic nature of amide nitrogen. The topological parameter ICR is radial centric information index. It represents the mean information content derived from atom eccentricities. Its regression coefficient suggests in favor of increased centricity (ICR) in the molecules for better activity. The fundamental concepts underlying the molecular centricity are explained in a recent review [21]. In simple terms, centricity explains the branching of edges/ vertices in a graph; e.g. a branched graph is more centric when compared to its linear homologue. In case of the compounds under study, other influencing factors remaining constant, highly branched substituents in the varying part are favorable for the activity. These two descriptors are common to Equations (1) and (2). The other topological descriptor T(N..Cl) in Equation (1) represents the sum of topological distances between N and Cl atoms in the molecules. As the distance between
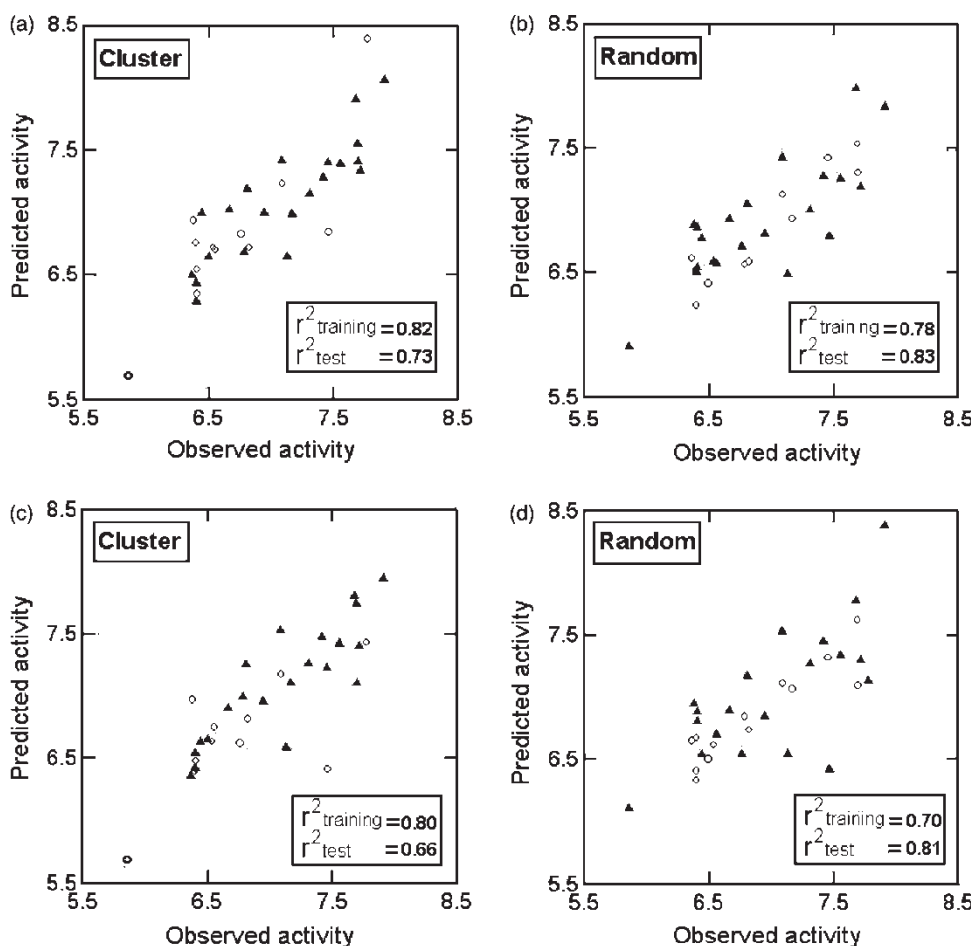
Figure 2.   Plots of training (▲) and test (o) sets predicted activities versus observed activity corresponding to Equations (1)(a, b) and (2) (c, d). Number of compounds in training set is 21 and in test set is 12.

the chloro group and quinoline nitrogen is constant and common to all the analogues, the T(N..Cl) practically refers to the separation of chloro group and rest of the nitrogens in the amide analogues. T(N..Cl)'s positive regression coefficient suggests in favor of increased separation between these two atoms. The functional group descriptor nHDon (Equation 1) represents the number of donor atoms for H-bonds (with N and O) in amide analogues. Its regression coefficient suggests in favor of minimum donor atoms for H-bonds for better activity. It may be viewed as the necessity of lone pair electrons on heteroatoms for purposes other than H-bonds for the activity. The 2D autocorrelation descriptors (Equation 2), GATS1e is Geary autocorrelation of lag 1 weighted by atomic Sanderson electro negativities (e) and GATS2p is Geary autocorrelation of lag 2 weighted by atomic polarizabilities, (p). These two descriptors represent influence of spatial autocorrelations of 1 and 2 path lengths (lags) of molecular graphs weighted by electro negativities and polarizabilities on the activity. These descriptors recommend in favor of decreasing one and two lag autocorrelation weighted by electronegativity and polarizabilities for the better activity.

The six descriptors of Equations (1) and (2) are furthermore evaluated using the PLS (partial least squares) analysis to come out with a 'single window' structure-activity model for the activity. The PLS cross-validation procedure [22] has suggested that three components as optimum for these descriptors to explain variance in the activity. Equation (3) is MLR like PLS equation resulted from the three components of six descriptors (Equations 1 and 2).

$$-\log IC_{50} = -1.132 - 0.939(0.343)\text{nDB}$$
$$+ 3.105(0.320)\text{ICR}$$
$$+ 0.010(0.135)\text{T(N..Cl)}$$
$$- 0.977(0.049)\text{GATS1e}$$
$$- 1.203(0.046)\text{GATS2p}$$
$$- 0.153(0.106)\text{nHDon}$$
$$n = 33,\ k = 6,\ r = 0.902,\ Q^2 = 0.763,$$
$$s = 0.247,\ F = 42.33 \tag{3}$$

In the PLS model, the values given in the parentheses following the regression coefficients are the fraction

contributions of corresponding descriptors in explaining the variance in the activity of the compounds. In the statistics of PLS equations, k is the number of explanatory descriptors used in the PLS analysis. The remaining statistical terms (n, r, Q^2, s and F) associated with the equations represent the same information as defined earlier. All the six descriptors have influenced the activity of amide analogues between 4-34 per cent. Among them, nDB and ICR have shown maximum influence (each one more than 30%) on the activity. Figure 3 shows a plot of the fraction contribution of normalized regression coefficients of these seven descriptors to the activity.

For the amine analogues (N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine amine derivatives; Table II) at the end of a search for 4- and 5-parameter models from all the descriptor classes together, the following models have emerged to explain their antimalarial activity.

$$-\text{logIC}_{50} = -18.903$$
$$- 0.000013(0.000001)\text{GMTIV}$$
$$+ 26.870(7.696)\text{MATS8}m$$
$$+ 10.479(1.936)\text{MATS6}e$$
$$- 0.536(0.116)\text{Hy}$$
$$n = 44,\ r = 0.854,\ Q^2 = 0.655,\ s = 0.264,$$
$$F = 26.32 \tag{4}$$

$$-\text{logIC}_{50} = -25.672 - 0.019(0.002)S0K$$
$$+ 1.344(0.398)\text{ICR}$$
$$+ 1.710(0.435)\text{IVDE}$$
$$+ 28.020(7.646)\text{MATS8}m$$
$$- 0.748(0.120)\text{Hy}$$
$$n = 44,\ r = 0.868,\ Q^2 = 0.659,\ s = 0.255,$$
$$F = 23.23 \tag{5}$$

The 483 descriptor dataset of amine analogues did not result in models with higher explained variance than Equations (4) and (5), and at the same time satisfying the CP-MLR filters criteria. Equations (4) and (5) are also validated through two test sets with each one containing fourteen out of forty four compounds of Table II and their predictions are in agreement with their experimental values (Table II; Figure 4). Collectively, seven descriptors have taken part in these two equations. They are from topological (GMTIV, S0K, ICR and IVDE), 2D-autocorrelations (MATS6e and MATS8m), and Empirical (Hy) descriptor classes.
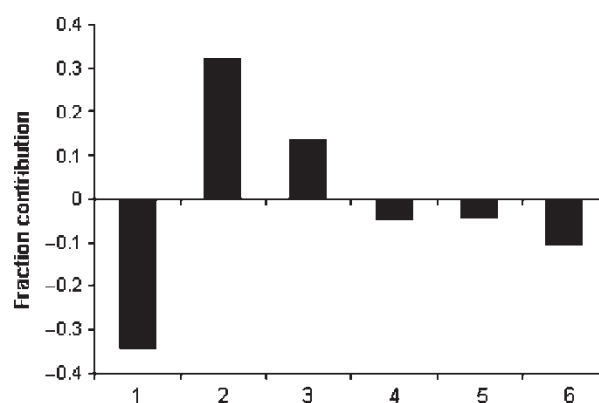

Figure 3. Plot of fraction contribution of MLR-like PLS coefficients (normalized) of the 6 descriptors from Equations (1) and (2) to the activity. The serial numbers 1 to 6 on the horizontal axis refer to the descriptors nDB, ICR, T(n..Cl), GATS1e, GATS2p, and nHDon, respectively.

Among the topological descriptors, ICR (radial centric information index) (Equation 5) has shown its significance in modelling the activity of amide (Table I) as well as amine (Table II) analogues. This descriptor has shown positive influence for both amide and amine analogues. The other topological descriptors identified for the amine analogues are GMTIV (Gutman Molecular Topological Index by valence vertex degrees) (Equation 4), S0K (Kier symmetry index) and IVDE (mean information content on the vertex degree equality) (Equation 5). GMTIV is the product summation of valence vertex degree of all of atoms for all topological distances. Its negative regression coefficient suggests in favor of closely placed vertices and compact structural scaffolds for improved activity. The Kier symmetry index, S0K is for the zero order paths (atoms) and accounts for the molecular symmetry in terms of atom topological uniqueness. In Equation (5) it is associated with a negative regression coefficient and points out for structural features with reduced topological equivalence for better activity. IVDE is based on the partition of vertices according to vertex degree equality. It is a measure of the lack of structural homogeneity or the diversity of a molecule and related to Shannon's entropy measure. Its positive regression coefficient in Equation (5) is in favor of increasing vertex degree equality for the activity. The 2D autocorrelation descriptors MATS6e and MATS8m (Equations 4 and 5) are Moran autocorrelations of lags 6 and 8 / weighted by atomic Sanderson electronegativities (e) and atomic mass (m) respectively. Their positive regression coefficient suggests in favor of increased autocorrelation contents of six- and eight-member structural graphs weighted by electronegativities and mass for the activity. In these equations Hy represents hydrophilic factor of the molecules. The negative coefficient of Hy suggests in favor of hydrophobic compounds for improved activity. MATS8m and Hy are common descriptors to Equations (4) and (5). Equation (6) is MLR like PLS
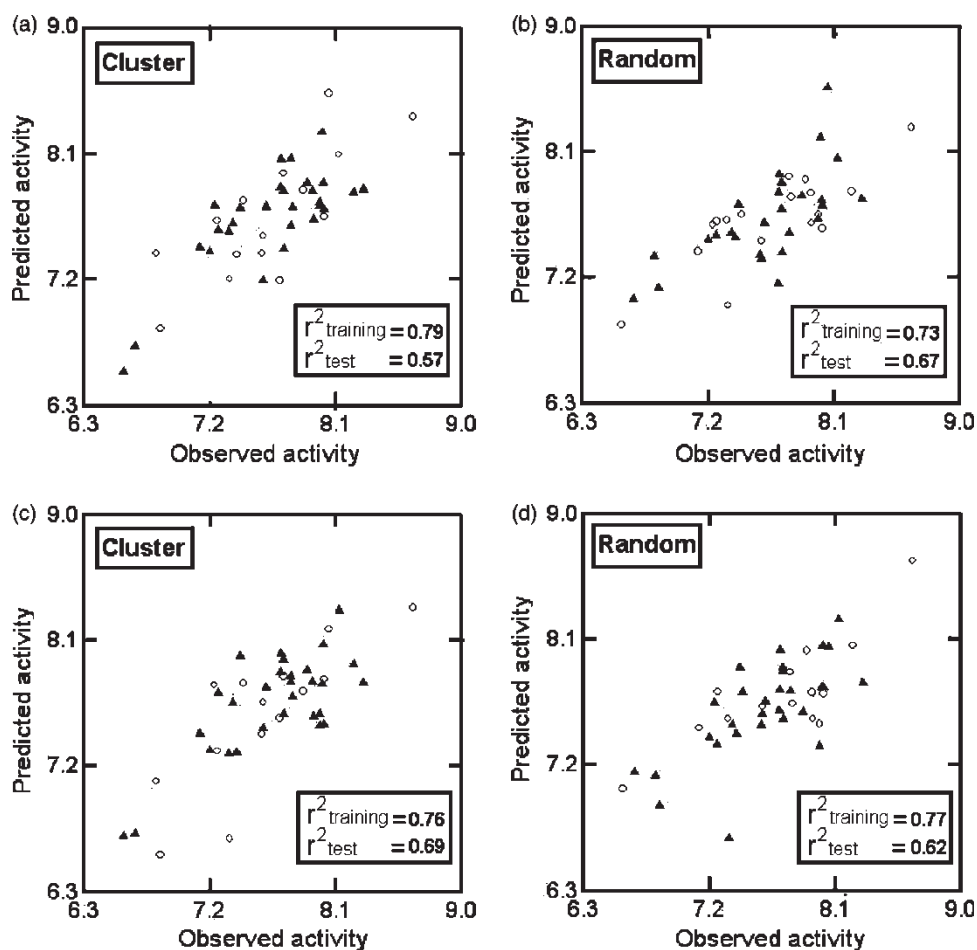
Figure 4.   Plots of training (▲) and test (o) sets predicted activities versus observed activity corresponding to Equations (4) (a, b) and (5) (c, d). Number of compounds in training set is 30 and in test set is 14.

equation for the activity from the three components of seven descriptors of Equations (4) and (5).

$$-\mathrm{logIC}_{50} = -18.964$$

$$- 0.000007(0.190)\mathrm{GMTIV}$$

$$- 0.008(0.192)S0K$$

$$+ 0.602(0.062)\mathrm{ICR}$$

$$+ 0.555(0.049)\mathrm{IVDE}$$

$$+ 24.677(0.128)\mathrm{MATS8}m$$

$$+ 8.202(0.157)\mathrm{MATS6e}$$

$$- 0.654(0.221)\mathrm{Hy}$$

$$n = 44,\ k = 7,\ r = 0.876,\ Q^2 = 0.715,$$

$$s = 0.241,\ F = 44.03 \tag{6}$$

In this equation, Hy is one descriptor having the maximum influence (~22%) on the activity of amine analogues (Table II). It suggests in favor of

hydrophobic compounds for better activity. In amine analogues (Table II), ICR has accounted for about 6 per cent variance in the activity. The remaining five descriptors have influenced the activity between 5 to 19 per cent. The fraction contributions of these seven descriptors to the activity are shown in Figure 5.

Conclusions

The descriptors from constitutional, topological, 2D-autocorrelation, functional and empirical classes have together formed models to explain the antimalarial activity of N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine analogues with amide and amine characteristics (Tables I and II). As the N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine moiety (Figure 1) is common for all the analogues, the attributions of the descriptors are due to the varying portions of the molecules. The amide (Table I) and amine (Table II) analogues have shared the radial centric information (ICR) as common modelling descriptor. It suggests in favor of increased centricity in the molecules for better antimalarial activity. This in terms of molecular features (other influencing factors
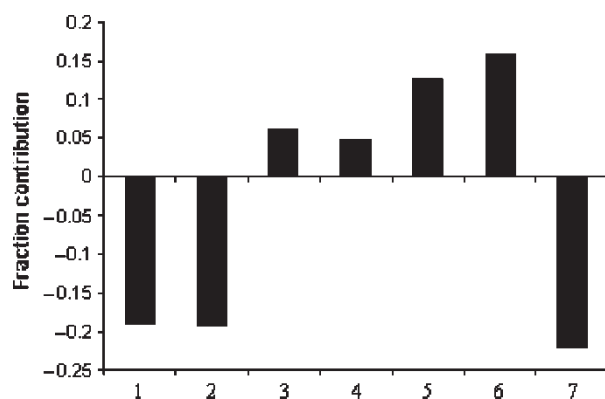
RIGHTSLINK

Figure 5. Plot of fraction contribution of MLR-like PLS coefficients (normalized) of the 7 descriptors from Equations (4) and (5) to the activity. The serial numbers 1 to 7 on the horizontal axis refer to the descriptors GMTIV, S0K, ICR, IVDE, MATS8m, MATS6e and Hy, respectively.

remaining constant) of the compounds under study, recommends highly branched substituents in the varying part for the activity. The remaining descriptors are different for both classes of compounds. This may be attributed to the characteristic difference between amides and amines.

The models derived for the amide analogues (Table I) have suggested in favor of minimum number of isolated double bonds (nDB) in the form of carbonyl moieties and distantly placed nitrogen(s) (from Cl group; T(N..Cl)) in the varying portion of the molecule for improved activity. In terms of physico-chemical properties, a carbonyl moiety adjacent to nitrogen alters the basic nature of the nitrogen which is essential for accumulation of compound in the food vacuole of the parasite. Moreover, for these analogues, the regression models have preferred the lone pair electrons on heteroatoms (N and O) for purposes other than H-bonds for better activity. The 2D-autocorrelation descriptors (GATS1e and GATS2p) in amide analogue models recommend in favor of decreasing one and two lag autocorrelation weighted by electro-negativity and polarizabilities for the better activity. The PLS model of amide analogues have suggested that decreased carbonyl moieties (nDB) and increased increased centricity (ICR) will be having maximum influence (each one more than 30%) on the activity.

For amine analogues (Table II), the regression coefficients of Gutman Molecular Topological Index by valence vertex degrees (GMTIV), Kier symmetry index (S0K) and Shannon's like entropy related measure (IVDE) have suggested in favor of compact structural moieties in the varying parts of the molecule for improved activity. The 2D-autocorrelation descriptors (MATS6e and MATS8m) of these analogues recommend in favor of increased auto-correlation contents of six- and eight-member structural graphs weighted by electronegativities and mass for the better activity. In these models, the

participation of hydrophilic factor (Hy) clearly suggested in favor of hydrophobicity for improved activity. The PLS model derived from these descriptors also suggested that Hy as the most influential descriptor ($\sim 22\%$) to determine the activity of amine analogues (Table II).

The modelling procedure adopted in this study being an 'indirect approach' involving closely related compounds with their 2D-descriptors, the central scaffold, N^1-(7-chloro-4-quinolyl)-1,4-bis(3-amino-propyl)piperazine moiety (Figure 1) may be viewed as the pharmacophore. The study suggests in favor of highly branched and compact groups in the varying part of the pharmacophore for the improvement of the activity. It also suggested in favor of minimum amide and/ or carboxyl moieties in the pharmacophore for the activity. The other preferred features in and around the pharmacophore are distantly placed nitrogen(s), lone pair electrons on heteroatoms (N and O) for purposes other than H-bonds and overall hydrophobicity. The study, while identifying the differences between amine and amide analogues in terms of their topological and other functional descriptors, may provide some insight to guide the design of new analogues.

Acknowledgements

References

[1] http://www.globalhealthfacts.org

[2] (a) Ridley RG. Medical need, scientific opportunity and the drive for antimalarial drugs. Nature 2002;415:686–693. (b) O'Neill PM, Bray PG, Hawley SR, Ward SA, Park BK. 4-Aminoquinolines–past, present, and future: A chemical perspective. Pharmacol Ther 1998; 7:29–58.

[3] (a) Pandey AV, Bisht H, Babbarwal VK, Srivastava J, Pandey KC, Chauhan VS. Mechanism of malarial haem detoxification inhibition by chloroquine. Biochem J 2001;355:333–338. (b) Chou AC, Chevli R, Fitch CD. Ferriprotoporphyrin IX fulfills the criteria for identification as the chloroquine receptor of malaria parasites. Biochemistry 1980;19:1543–1549. (c) Egan TJ, Marques HM. The role of haem in the activity of chloroquine and related antimalarial drugs. Coord Chem Rev 1999;190–192:493–517. (d) Dorn A, Stoffel R, Matile H, Bubendorf A, Ridley RG, Malarial hemozoin/β-hematin supports haem polymerization in the absence of protein. Nature 1995;374:269–271. (e) Sullivan DJ, Gluzman IY, Russell DG, Goldberg DE. On the molecular mechanism of chloroquine's antimalarial action. Proc Natl Acad Sci USA 1996;93:11865–11870.

[4] Wiesner J, Ortmann R, Jomaa H, Schlitzer M. New antimalarial drugs. Angew Chem Int Ed 2003;42:5274–5293.

[5] (a) Bray PG, Howells RE, Ward SA. Vacuolar acidification and chloroquine sensitivity in Plasmodium falciparum. Biochem Pharmacol 1992;43:1219–1227. (b) Bray PG, Howells RE, Ritchie GY, Ward SA. Rapid chloroquine efflux phenotype

in both chloroquine-sensitive and chloroquine-resistant Plasmodium falciparum. A correlation of chloroquine sensitivity with energy-dependent drug accumulation. Biochem Pharmacol 1992;44:1317–1324.

[6] (a) Ridley RG, Hofheinz H, Matile H, Jaquet C, Dorn A, Masciadri R, Jolidon S, Richter WF, Guenzi A, Girometta MA, Urwyler H, Huber W, Thiathong S, Peters W. 4-aminoquinoline analogues of CQ with shortened side chains retain activity against CQ resistant Plasmodium falciparum. Antimicrobial Chemother 1996;40:1846–1854. (b) De D, Krogstad FM, Byers LD, Krogstad DJ. Structure-activity relationships for antiplasmodial activity among 7-substituted 4-aminoquinolines. J Med Chem 1998;41:4918–4926.

[7] (a) Egan TJ, Hunter R, Kaschula CH, Marques HM, Misplon A, Walden JC. Structure-function relationships in aminoquinolines: Effect of amino and chloro groups on quinoline-hematin complex formation, inhibition of β-hematin formation, and antiplasmodial activity. J Med Chem 2000;43:283–291. (b) Kaschula CH, Egan TJ, Hunter R, Basilico N, Parapani S, Tarameli D, Pasini E, Monti D. Structure activity relationships in 4-aminoquinoline antiplasmodials. The role of the group at the 7-position. J Med Chem 2002;45:3531–3539.

[8] (a) Vennerstrom JL, Ellis WY, Ager AL, Andersen SL, Gerena L, Milhous WK. Bisquinolines. 1. *N,N*-Bis(7-chloroquinolin-4-yl)alkanediamines with Potential against chloroquine-resistant malaria. J Med Chem 1992;35:2129–2134. (b) Vennerstrom JL, Ager AL, Dorn A, Andersen SL, Gerena L, Ridley RG, Milhous WK. Bisquinolines. 2. *N,N*-Bis(7-chloroquinolin-4-yl)alkanediamines. J Med Chem 1998;4:4360–4364.

[9] (a) Ryckebusch A, Deprez-Poulain R, Maes L, Debreu-Fontaine MA, Mouray E, Grellier P, Sergheraert C. Synthesis and *in Vitro* and *in Vivo* antimalarial activity of N^1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine derivatives. J Med Chem 2003;46:542–557. (b) Ryckebusch A, Debreu-Fontaine MA, Mouray E, Grellier P, Sergheraert C, Melnyk P. Bioorg Med Chem Lett 2005;15: 297–302.

[10] (a) Solomon VR, Haq W, Srivastava K, Puri K, Katti SB. Synthesis and antimalarial activity of side chain modified 4-aminoquinoline derivatives. J Med Chem 2007;50:394–398. (b) Solomon VR, Puri SK, Srivastava K, Katti SB. Design and synthesis of new antimalarial agents from 4-aminoquinoline. Bioorg Med Chem 2005;13: 2157–2165.

[11] Werbel LM, Cook PD, Elslager EF, Hung JH, Johnson JL, Kesten SJ, McNamara DJ, Ortwine DF, Worth DF. Antimalarial activity, and quantitative structure-activity relationships of tebuquine and a series of related 5-[(7-chloro-4-quinolinyl)amino]-3-[(alky1amino)methyl][1,1'-biphenyl]-2-ols and N^ω-Oxides. J Med Chem 1986;29: 924–939.

[12] Gupta MK, Prabhakar YS. Topological descriptors in modeling the antimalarial activity of 4-(3',5'-disubstituted anilino)quinolines. J Chem Inf Model 2006;46:93–102.

[13] (a) Estrada E, Uriarte E. Recent advances on the role of topological indices in drug discovery research. Curr Med Chem 2001;8(13):1573–1588. (b) Votano JR. Recent uses of topological indices in the development of in silico ADMET models. Curr Opin Drug Discov Devel 2005;8(1):32–37. (c) Gozalbes R, Doucet JP, Derouin F. Application of topological descriptors in QSAR and drug design: History and new trends. Curr Drug Targets Infect Disord 2002;2(1):93–102. (d) Bagchi MC, Mills D, Basak SC. Quantitative structure-activity relationship (QSAR) studies of quinolone antibacterials against *M. fortuitum* and *M. smegmatis* using theoretical molecular descriptors. J Mol Model 2007;13:111–120. (e) González MP, Terán C, Teijeira M. A topological function based on spectral moments for predicting affinity toward A3 adenosine receptors. Bioorg Med Chem Lett 2006;16(5):1291–1296. (f) Gupta MK, Sagar R, Shaw AK, Prabhakar YS. CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols–topological descriptors in modeling the activity. Bioorg Med Chem 2005;13: 343–351.

[14] DRAGON software version 5.0-2005. By Todeschini R, Consonni V, Mauri A, Pavan M. Milano, Italy. http://disat.unimib.it/chm/Dragon.htm

[15] Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression analysis of Selwood et al. data set- a case study. QSAR Comb Sci 2003;22:583–595.

[16] ChemDraw Ultra 6.0 and Chem 3D Ultra, Cambridge Soft Corporation, Cambridge, USA.

[17] Prabhakar YS. A combinatorial protocol in multiple linear regression to model gas chromatographic response factor of organophosphonate esters. Internet Electron J Mol Des 2004; 3:150–162, http://www.biochempress.com.

[18] Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR /PLS Directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR Comb Sc 2004;23:234–244.

[19] So SS, Karplus M. Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks 2. Applications. J Med Chem 1997; 40:4347–4359.

[20] MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1255 University Street, Suite 1600, Montreal, Quebec, Canada H3B 3X3. (b) Brown RD, Martin YC. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. J Chem Inf Comput Sci 1996; 36:572–584.

[21] Bonchev D. My life–long journey in mathematical chemistry. Internet Electron J Mol Des 2005;4:434–490, http://www.biochempress.com

[22] Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 1978;20:397–405.

Supporting information

Complete dataset of molecular descriptors of structure databases corresponding Tables I and II, and PLS loadings, weights and sensitivity of independent and dependent descriptors of PLS-models will be provided as supplementary material by email upon request.