Hindawi Publishing Corporation
Journal of Chemistry
Volume 2013, Article ID 154629, 14 pages
http://dx.doi.org/10.1155/2013/154629

Research Article

Topological Features in Profiling the Antimalarial Activity Landscape of Anilinoquinolines: A Multipronged QSAR Study

Shreekant Deshpande,[1] Mohammad Goodarzi,[2] Seturam B. Katti,[1] and Yenamandra S. Prabhakar[1]

[1] *Medicinal and Process Chemistry Division, CSIR-Central Drug Research Institute, Lucknow 226001, India*
[2] *Young Researchers Club, Islamic Azad University, Arak Branch, Markazi, Arak, P.O. Box 38135-567, Iran*

Correspondence should be addressed to Yenamandra S. Prabhakar; yenpra@yahoo.com

Received 27 June 2012; Revised 30 August 2012; Accepted 30 August 2012

Academic Editor: Marjana Novic

The antimalarial activity of a series of 4-anilinoquinolines was modeled with topological and other functional descriptors using feature selection approaches CP-MLR and GA. Five models were identified from each approach to explain the activity of the compounds. They jointly shared eighteen descriptors. Among them five descriptors, namely, H-052, MATS4m, MATS7e, Mor30p, and R7m, were common to both approaches. In PLS analysis the eighteen descriptors have led to a three-component model ($r^2 = 0.731$, $Q^2 = 0.688$, $r_t^2 = 0.676$). and the common descriptors were among the most influential ones to modulate the activity. Among them, MATS7e indicated the favorability of nonlinear and branched molecular topology for higher activity. MATS4m has also advocated in favor of branching/nonlinearity in the molecule for the activity. The H-052 argued that R'CH2-CHX-CH2R fragments (X is halogen) in the scaffold enhance the activity. In BP-ANN these descriptors led to very good predictive models (training $r^2 > 0.81$; validation $r^2 > 0.81$; test $r^2 > 0.75$). The study has offered direction to understand the patterns of the antimalarial activity of anilinoquinolines for exploring potential prototype compounds.

1. Introduction

Malaria is a vector-born parasitic infection (vector: female mosquitoes of the *Anopheles* genus; parasite: protozoa, genus *Plasmodium*) of the tropical regions with serious health and economic implications. The interventional measures of the last decade have resulted in the form of some relief to the incidents of deaths due to malaria. However, these efforts did not decline the manifestation of drug resistance cases [1]. In fact, until the recognition of drug-resistant strains of *Plasmodium falciparum*, the treatment of malaria has heavily relied on chloroquine as first-line drug [2, 3]. Also, in clinical practice chloroquine suffers due to several limitations/side effects which include gastrointestinal, stomach, neural, and blurring of vision [4, 5]. The mechanistic investigations on the antimalarial activity of this (quinoline) class have indicated that chloroquine and other analogues follow similar pathway in the expression of the activity [6, 7]. The drug resistance of parasite is compound centric and not due to altered mechanism of action [8–10]. This has renewed the research interest to explore alternative quinolines as potential antimalarial agents. Moreover, existence of large preclinical and clinical information and low cost/ease of preparation of alternative/new drugs or drug-like molecules encouraged the researchers to venture into this chemical class [11–16].

In quinoline class of compounds, amodiaquine (Figure 1) is a clinically practiced antimalarial agent [17]. The chloroquine-resistant *Plasmodium* parasites are not automatically cross-resistant to it [18]. However, amodiaquine is reported to cause agranulocytosis and hepatitis [19, 20]. The side effects are attributed to the 4-hydroxyanilino moiety of amodiaquine. In biological system it undergoes enzymatic oxidation to the quinoneimine form and makes nucleophilic addition to proteins [21, 22]. In this scenario, to overcome the undesirable side effects of amodiaquine, different 7-chloro-4-(3′,5′-disubstituted anilino)quinolines were explored as alternative antimalarial agents [23–26]. These compounds structurally resemble

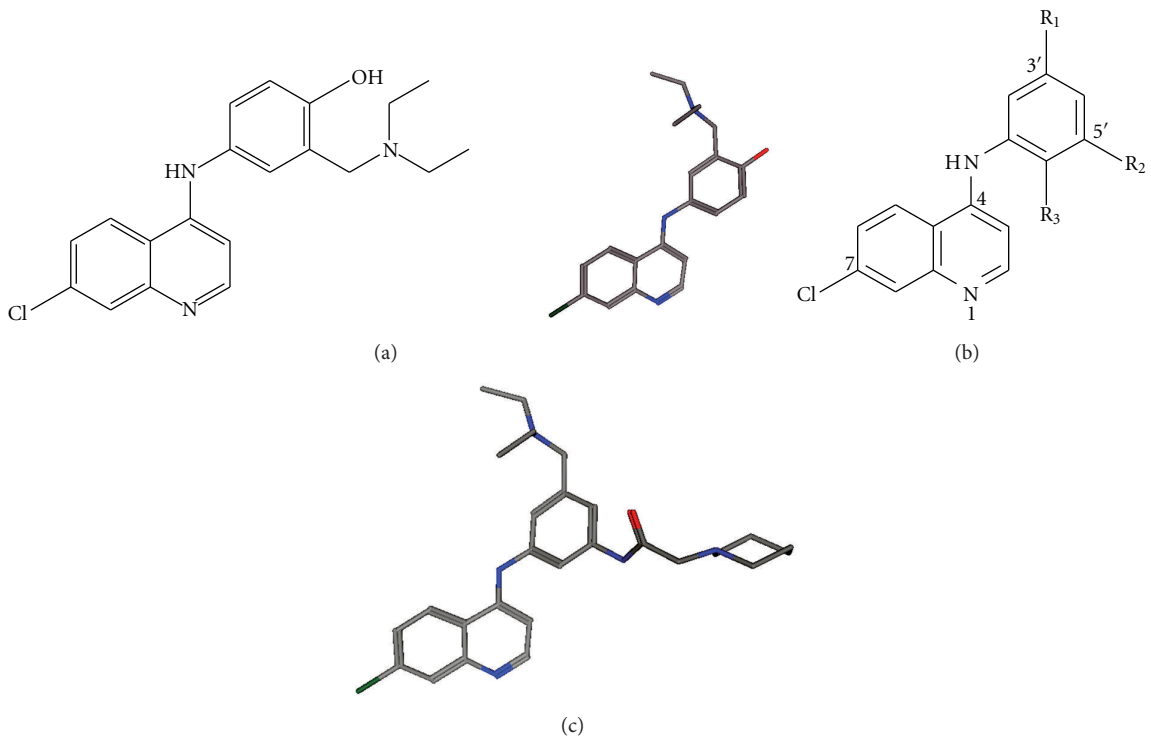

Figure 1: Structure of amodiaquine (a) X-ray crystal structure of amodiaquine from CCDC (626826). (b) General structure of anilinoquinolines in the database. (c) Putative bioactive conformation of compound **40**. Hydrogens are suppressed for clarity.

amodiaquine but are devoid of amodiaquine's 4-hydroxyl on the anilino moiety which is attributed for the side effects.

In medicinal chemistry paradigm the rational drug design approaches, which include quantitative structure-activity relationship (QSAR) and molecular modeling protocols, cull out structural and functional information of chemical entities desirable for biological response. This may come handy to modulate/design the biological response of intended compounds. Here structure-activity elucidation of the compounds is attempted taking into account the correlation between the chemical structure space indices and their biological response landscape. The earlier QSAR study [27] on some 7-chloro-4-(3′,5′-disubstituted anilino)quinolines, involving 2D molecular features, has denoted that 3′- and 5′-substituents of the anilino moiety map different domains with substructure preferences in the activity space. It also gave indication in favor of the electron rich centers in the aniline substituent groups for better antimalarial activity. In this background, the QSAR analysis of the antimalarial activity of an enlarged dataset of 4-anilinoquinolines has been undertaken with a perspective to broaden the structural information relevant to the activity space. The results are presented hereunder.

2. Materials and Methods

2.1. Chemical Structure Database and Biological Activity. A dataset of 90 anilinoquinolines (Figure 1(b)) along with their antimalarial activity (IC_{50}, inhibitory concentration or dose in micromoles of compound to reduce 50% FcB1R strain of *P. falciparum*) reported in the literature was considered for this study [23–26]. The substitution positions in these compounds are briefly summarized in Table 1. The antimalarial activity (IC_{50}) of all these compounds was reported using the same experimental protocol [23–26]. The compounds exhibited good variation (~2.6 orders) in their antimalarial activity. For the purpose of modeling study the activity has been transformed in the form of logarithm of inverse of inhibitory concentration and expressed as pIC_{50} (Table 2).

The structure database of the compounds under investigation has been generated using the X-ray crystal structure of amodiaquine (Figure 1(a)) [28] to impart 3D characteristics to the chemical space of the agents. Accordingly, in SYBYL [29] by making use of the procedure implemented therein the 3D structures of the compounds (Figure 1(b)) were generated from the X-ray crystal structure of amodiaquine (Figure 1(a)). In Dragon software [30] these conformations have resulted in 490 and 686 descriptors, respectively, to profile the 0D to 2D and 3D characteristics of the molecules. Prior to the QSAR study, all those descriptors showing a correlation of less than 0.1 with the dependent variable (descriptor versus activity $r < 0.1$) and descriptors showing intercorrelation greater than or equal to 0.9 ($r \geq 0.9$) were excluded. This has reduced the 0D to 2D and 3D descriptors to 101 and 131, respectively, for correlating with the activity.

For QSAR study, using the fingerprints of BIT-packed version of Molecular ACCess System (FP-BIT-MACCS) of the compounds, the dataset was divided into two mutually exclusive groups as training and test sets. The concepts

Table 1: The variations in substitution positions of anilinoquinolines (Figure 1(b)).

Comp no.	R_1 [a]	R_2	R_3
1	R''	$-CH_2OH$	H
2–19	$-NHC(O)-(CH_2)_n-NRR'$	$-CH_2OH$	H
20–31	$-NHC(O)-CH_2-N(CH_2)_5$	$-CH_2-OC(O)-(CH_2)_n-R''$	H
32	$-NHC(O)-CH_2-N(CH_2)_5$	R''	H
33–61	$-NHC(O)-CH_2-N(CH_2)_5$	$-CH_2-R''$	H
62	R''	$-CH_2NEt_2$	–OH
63	R''	H	–OH
64–67	R''	$-CH_2NEt_2$	H
68–73	$-NHC(O)-CH_2-N(CH_2)_5$	$-CH_2-OC(O)-(CH_2)_n-R''$	H
74–80	$-NHC(O)-CH_2-N(CH_2)_5$	$-CH_2-OC(O)-NH-(CH_2)_n-R''$	H
81–90	$-NHC(O)-CH_2-N(CH_2)_5$	$-C(O)NRR'$	H

[a] The NRR' and R'' in R_1/R_2 groups represent the extended structural moieties (functionalized alkyl, aryl, and other functional units) attached to (in place of) them.

of molecular finger-prints were originally introduced by Molecular Design Limited, Inc (MDL) as a part of informatics services to the life sciences and chemical industry [31]. In molecular operating environment (MOE) software [32], the cluster analysis the MACCS fingerprints of the compounds was carried out at 85% similarity to segregate them (compounds) into training and test sets. All compounds were arbitrarily put into training set (50 compounds) and test set (40 compounds) in such a way that members of the clusters were distributed in both the sets. Furthermore, to facilitate the comparison of significance of descriptors with one another in the derived models, all the descriptor values are scaled between "0" and "1" (inclusive of both values). For this the original descriptor values have been scaled using the following transformation:

$$X_S = \frac{X - X_{\mathrm{MIN}}}{X_{\mathrm{MAX}} - X_{\mathrm{MIN}}}, \quad (1)$$

where X, X_{MIN}, X_{MAX}, and X_S are the training set feature X's original, minimum, maximum, and transformed descriptor values, respectively.

The significance of obtained molecular features in explaining the antimalarial activity of the compounds has been investigated using the combinatorial protocol in multiple linear regression (CP-MLR) [33], genetic algorithm (GA) [34], partial least squares (PLS) [35, 36], and artificial neural networks (ANN) [37] methods. Only the training set compounds were used for deriving the models and the test set compounds were used for the external validation of the derived models. Purposefully a large test set was created to facilitate a follow-up study of derived models in back-propagation artificial neural networks (BP-ANN). The modeling procedures and the computations are briefly described below.

2.2. CP-MLR. Combinatorial protocol in multiple linear regression (CP-MLR) is filter-based variable selection procedure [33]. The procedural aspects are discussed in some of the recent publications [38, 39]. This operates through a combinatorial seeding strategy followed by predefined filters to assess the significance of seeds and finally employs MLR to develop the models from the significant seeds. A unique combination of descriptors (variables) is referred to as a seed. Here, filter-2 controls the seeds through t-values (default t-value ≥ 2.0) of the coefficients of individual descriptors of the seed in regression; filter-2 controls the seeds through t-values of variables' coefficients in regression which is set as greater than or equal to two; filter-3 provides a comparison of seeds in different equations in terms of square root of adjusted multiple correlation coefficient of the regression, 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 ($0.3 \leq Q^2 \leq 1.0$). In CP-MLR, for the selection of features from datasets the initial threshold of filter-1 was assigned as 0.3 and subsequently liberated to 0.79 to boost the formation of different seeds. The search was started with two-variable seeds and with an initial filter-3 value of 0.74. The information rich descriptors were collected by successively incrementing the number of variables per seed as well as the threshold of filter-3 to the optimum r-bar value of the preceding generation.

2.3. GA. The genetic algorithm variable subset selection (GA-VSS) routine as implemented in MOBY DIGS [40, 41] was used for the selection of GA features. It has proceeded with an initial population of one hundred solutions (chromosomes) with maximum allowed variables in a solution as five. The fitness for each chromosome was calculated based on leave-one-out (LOO) cross-validation (Q^2). The reproduction/mutation trade-off (T) value was set to 0.5. Based on the T value, the crossover and mutation values of GA were automatically fixed *in situ* in the computation. The optimum solutions were identified at the end of one hundred generations of GA evolution process (selection, crossover, and mutation).

The models emerged from the CP-MLR and GA approaches are further regressed for the chance correlation through one hundred simulation runs with repeated randomization of biological response [42, 43]. The correlation coefficients of simulated regressions have been used to determine the average correlation coefficient of Y-randomization

Table 2: Anilinoquinolines (Figure 1(b); Table 1) and their observed and predicted antimalarial activities.

Compd no.[a]	n	NRR'/R''[b]	Obs	pIC_{50}[c] CP Eq. (1)	GA Eq. (6)	PLS All	BP-ANN dEq. (1)	dEq. (6)	Com
AQ01(V)		NH_2	6.55	6.00	6.92	6.56	6.32	6.59	6.70
AQ02(V)	1	4-Methyl-piperidine	7.11	7.04	6.80	6.87	7.11	6.97	7.35
AQ03(T)	2	4-Methyl-piperidine	6.87	6.96	7.00	6.89	6.98	7.00	7.49
AQ04(T)	4	4-Methyl-piperidine	6.33	6.63	6.60	6.50	6.50	6.45	6.93
AQ05(Tr)	5	4-Methyl-piperidine	6.52	6.57	6.60	6.40	6.51	6.55	6.75
AQ06(V)	7	4-Methyl-piperidine	6.51	6.89	7.11	6.83	6.75	6.89	6.55
AQ07(Tr)	11	4-Methyl-piperidine	7.07	6.58	7.29	6.65	6.77	7.15	6.79
AQ08(Tr)	1	Piperidine	7.35	6.94	6.86	6.89	7.13	7.00	7.40
AQ09(T)	1	N-Methyl-piperazine	6.46	6.58	6.69	6.74	6.44	6.50	6.55
AQ10(Tr)	1	4-Hydroxy-piperidine	5.98	6.13	6.20	6.13	6.26	6.06	6.22
AQ11(T)	1	Pyrrolidine	7.36	6.47	6.53	6.50	6.97	6.95	7.16
AQ12(Tr)	1	Thiazolidine	6.23	6.20	6.53	6.31	6.35	6.46	6.21
AQ13(T)	1	$NHC(CH_3)_3$	7.29	7.53	6.85	6.83	7.36	7.14	7.66
AQ14(V)	1	NEt_2	7.38	6.53	6.25	6.50	6.82	7.10	7.02
AQ15(V)	1	$NHCH_2C_6H_5$	6.72	6.14	6.29	6.32	6.39	6.67	6.35
AQ16(Tr)	1	$NHCH_2C_6H_4Cl$ (para)	6.07	6.41	6.19	6.43	6.51	6.21	6.07
AQ17(Tr)	1	1,2,3,4-Tetrahydroisoquinoline	5.84	6.62	6.26	6.42	6.55	6.16	6.42
AQ18(Tr)	1	3-Aminopyrazole	6.09	6.59	6.32	6.53	6.74	6.05	6.59
AQ19(Tr)	0	H	6.44	6.10	6.35	6.22	6.26	6.41	6.44
AQ20(Tr)	1	Piperidine	7.27	7.89	7.61	7.78	7.76	7.41	7.61
AQ21(T)	2	Piperidine	7.11	7.90	8.09	7.91	7.55	7.31	7.27
AQ22(V)	4	Piperidine	7.81	7.64	7.41	7.43	7.69	7.38	7.40
AQ23(Tr)	5	Piperidine	7.56	7.44	7.51	7.39	7.66	7.54	7.36
AQ24(V)	7	Piperidine	6.78	7.31	7.94	7.40	7.24	7.29	6.87
AQ25(Tr)	4	Pyrrolidine	7.64	7.47	7.39	7.38	7.35	7.57	7.38
AQ26(Tr)	4	Morpholine	7.85	7.22	7.60	7.34	7.35	7.52	7.27
AQ27(T)	4	N-methyl-piperazine	7.19	7.23	7.27	7.27	7.20	7.21	7.26
AQ28(Tr)	4	NEt_2	6.85	7.39	7.16	7.29	7.05	6.96	7.38
AQ29(Tr)	4	Br	6.82	6.87	6.99	6.80	6.81	7.04	7.04
AQ30(T)	0	Phenyl	7.12	7.33	6.69	7.20	7.31	6.64	7.39
AQ31(Tr)	0	Quinol-4-yl	7.59	7.15	6.97	7.20	7.64	7.41	7.30
AQ32(T)		CHO	6.55	7.38	6.95	7.00	7.22	6.87	6.75
AQ33(T)		$NHCH_2CH_2NMe_2$	7.81	7.81	7.95	7.95	7.42	7.88	7.88
AQ34(Tr)		$NHCH_2CH_2CH_2NMe_2$	6.87	7.69	7.94	7.76	7.65	7.26	7.22
AQ35(V)		$NMeCH_2CH_2NMe_2$	8.30	7.81	8.26	8.10	7.89	8.23	7.88
AQ36(T)		$NHCH_2CH_2$-pyrrolidine	7.99	7.95	8.12	8.02	7.94	8.04	8.05
AQ37(Tr)		$NHCH_2CH_2$-piperidine	8.18	7.96	8.01	7.99	7.74	8.00	8.04
AQ38(V)		$NHCH_2CH_2$-morpholine	8.10	7.55	7.79	7.77	7.81	7.95	7.78
AQ39(Tr)		NEt_2	8.09	7.95	7.92	7.89	7.93	7.86	8.23
AQ40(Tr)		NH-tBu	8.43	8.56	7.97	7.98	8.35	8.14	8.22
AQ41(V)		Piperidine	7.99	8.09	7.94	7.93	8.14	7.99	8.36
AQ42(V)		Pyrrolidine	8.16	8.06	7.97	7.91	8.09	7.99	8.34
AQ43(T)		N-methyl-piperazine	8.12	7.78	8.20	7.93	7.97	8.11	8.01
AQ44(Tr)		Morpholine	8.02	7.58	7.86	7.75	7.77	7.96	8.04
AQ45(Tr)		4-Hydroxy-piperidine	8.19	8.14	8.19	8.05	8.19	8.30	8.43
AQ46(V)		N-(2-hydroxyethyl)piperazine	7.94	7.47	7.76	7.88	7.87	7.87	7.78
AQ47(T)		N-phenylpiperazine	8.26	7.44	8.00	8.12	7.80	7.98	7.63
AQ48(Tr)		N-benzylpiperazine	7.89	7.78	7.83	8.23	7.87	7.72	7.84
AQ49(Tr)		N-(diphenylmethyl)piperazine	7.90	7.58	7.33	7.92	7.80	7.49	7.52

Table 2: Continued.

Compd no.[a]	n	NRR'/R''[b]	Obs	pIC_{50}[c]					
				CP	GA	PLS	BP-ANN		
				Eq. (1)	Eq. (6)	All	dEq. (1)	dEq. (6)	Com
AQ50(T)		N-(4-chlorobenzyl)piperazine	7.94	7.74	7.63	7.96	7.90	7.88	7.70
AQ51(T)		N-(4-methoxybenzyl)piperazine	7.96	7.68	7.69	8.18	7.79	7.87	7.72
AQ52(Tr)		N-(4-nitrobenzyl)piperazine	7.88	7.54	7.65	7.84	7.64	7.58	7.44
AQ53(Tr)		N-(4-diethylaminobenzyl)piperazine	7.92	7.71	7.64	8.16	7.77	7.64	7.82
AQ54(V)		N-(4-cyanobenzyl)piperazine	7.88	7.86	8.10	8.29	7.85	7.72	7.81
AQ55(T)		N-piperonylpiperazine	8.16	8.01	8.46	8.53	8.10	8.25	7.98
AQ56(V)		NHPh	7.67	7.81	7.82	7.94	7.62	7.74	7.87
AQ57(Tr)		$NHCH_2Ph$	8.07	8.13	8.42	8.32	8.18	8.14	8.09
AQ58(T)		$NHCHPh_2$	8.04	7.60	7.93	7.97	7.73	7.88	7.78
AQ59(Tr)		$NHCH_2C_6H_4Cl$ (para)	8.34	8.36	8.57	8.46	8.31	8.35	8.19
AQ60(V)		$NHCH_2C_6H_4OMe$ (para)	8.36	7.94	8.19	8.15	8.02	8.35	8.03
AQ61(T)		$NHCH_2C_6H_4CF_3$ (para)	8.26	7.84	8.10	7.97	8.25	7.91	7.79
AQ62(Tr)		H	7.02	7.45	7.42	7.39	7.47	7.14	7.58
AQ63(Tr)		CH_2NEt_2	7.81	7.14	7.35	7.20	7.25	7.33	7.65
AQ64(Tr)		H	7.80	7.78	7.94	7.74	7.84	7.89	8.02
AQ65(T)		OH	7.45	7.42	7.86	7.67	7.48	7.76	7.73
AQ66(Tr)		NH_2	7.74	7.62	7.79	7.65	7.75	7.75	7.89
AQ67(Tr)		$NH(CH_2)_2$-piperidine	7.72	8.10	8.05	8.07	8.02	7.82	8.00
AQ68(V)	2	1,4-dioxy-1,4-dihydronaphthalen-2-yl	6.84	6.95	6.80	6.93	6.89	6.82	6.85
AQ69(Tr)	4	1,4-Dioxy-1,4-dihydronaphthalen-2-yl	7.33	7.22	6.97	7.19	7.31	7.19	7.34
AQ70(Tr)	5	1,4-Dioxy-1,4-dihydronaphthalen-2-yl	7.64	7.14	7.11	7.06	7.62	7.85	7.10
AQ71(V)	4	8-OH-1,4-dioxy-1,4-dihydronaphthalen-2-yl	7.54	7.06	7.01	7.09	7.53	7.45	7.24
AQ72(Tr)	5	8-OH-1,4-dioxy-1,4-dihydronaphthalen-2-yl	7.25	7.25	7.40	7.22	7.26	7.70	7.19
AQ73(Tr)	5	Phenyl	6.43	6.93	7.10	6.90	6.42	6.88	6.82
AQ74(T)	5	1,4-Dioxy-1,4-dihydro-3-methyl-naphthalen-2-yl	6.84	7.32	6.82	7.05	7.00	6.95	7.31
AQ75(Tr)	2	$N(Me)_2$	7.99	7.72	7.74	7.76	7.79	7.93	7.85
AQ76(Tr)	2	Piperidine	7.95	7.76	7.59	7.66	7.74	7.69	7.65
AQ77(Tr)	2	Morpholine	7.32	7.30	7.41	7.44	7.34	7.46	7.41
AQ78(Tr)	2	$N(Et)_2$	7.84	7.60	7.37	7.64	7.67	7.76	7.55
AQ79(V)	3	N-Methylpiperazine	7.36	6.88	7.21	7.13	6.99	7.26	6.79
AQ80(Tr)	1	Ph	7.44	7.20	7.26	7.22	7.36	7.25	7.26
AQ81(Tr)		$NH(CH_2)_2N(Me)_2$	7.50	7.68	7.33	7.43	7.58	7.53	7.68
AQ82(Tr)		$NH(CH_2)_3N(Me)_2$	6.74	7.17	7.46	7.16	7.10	7.21	7.17
AQ83(V)		$NMe(CH_2)_2N(Me)_2$	7.73	7.58	7.71	7.57	7.67	7.74	7.69
AQ84(Tr)		$NH(CH_2)_2$-pyrrolidine	7.78	7.78	7.46	7.53	7.77	7.57	7.73
AQ85(Tr)		$NH(CH_2)_2$-morpholine	7.23	7.15	7.19	7.11	7.25	7.25	6.92
AQ86(Tr)		Piperidine	6.54	7.46	7.36	7.29	6.83	7.01	6.84
AQ87(Tr)		Morpholine	6.97	7.15	6.96	7.01	7.07	6.87	7.16
AQ88(Tr)		N-methylpiperizine	7.72	7.42	7.29	7.32	7.31	7.69	7.52
AQ89(Tr)		N-Phenylpiperazine	6.82	6.94	6.85	7.03	7.00	6.75	6.81
AQ90(V)		N-Benzylpiperazine	7.30	7.30	7.20	7.55	7.35	7.27	7.19

[a] Compounds: 1–61, [23]; 62–67, [24]; 68–74, [25]; 75–90, [26]. The training, validation, and test sets are specified by suffixing the compound numbers with (Tr), (V), and (T), respectively.

[b] See Table 1 for positioning of the NRR'/R'' variations in the compounds.

[c] CP, GA, PLS, and BP-ANN represent CP-MLR, GA PLS, and BP-ANN regressions; Eq. (1) and Eq. (6), respectively, correspond to Eq. (1) and Eq. (6) of Table 3; "All" corresponds to 18-descriptor PLS regression model (Table 5); dEq. (1), dEq. (6), and "com" represent BP-ANN models from the descriptors of Eq. (1), Eq. (6) (Table 3), and common feature of CP-MLR and GA (Table 4), respectively.

as well as percent chance correlation of the model under scrutiny. Also, the derived models are externally validated by predicting the activities of test set compounds which are not a part of the model generation exercise. The test set predictions are used for computing the test set r-square statistics (r_t^2) of the model in question. Normally, models with r_t^2 value greater than 0.5 are treated as reliable. Finally, the descriptors identified in CP-MLR and GA have been further subjected to the partial least squares (PLS) [35, 36] analysis to present single-window QSAR models comprising all identified descriptors.

2.4. Applicability Domain. The usefulness of a model may be declared based on its ability to predict new compounds. In this context, applicability domain defines the predictive space of a model. The training set data, when projected in the model's multivariate parameter space, demarcates the plotting regions as populated with data and empty ones. Here, the populated regions define the applicability domain of the model and indicate that the space is suitable for the predictions. Computationally, the applicability domain of the models is evaluated through the plot of standardized residuals versus leverage values (h) for each compound [44]. It is also known as Williams plot and is useful for the detection of both the response outliers (Y-outliers) and structurally influential chemicals (X-outliers) in the model. In this plot, the applicability domain is determined inside squared area within $\pm x$ times standard deviations (where x may be given a value between 2 to 3) and leverage threshold (h^*) which is typically fixed at $3(k + 1)/n$ (where n is the number of compounds in the training set and k is the number of parameters in the model). In this plot, if a compound's leverage value (h) is smaller than the h^*, the probability of its prediction come true may be as high as that of the training set compounds. Making use of these settings, the applicability domain of the models from the CP-MLR, GA, and PLS have been scrutinized for their predictive capability.

2.5. BP-ANN. In ANN modeling a training set was used for the model generation while a validation set was applied to stop the overfitting of the network. Additionally a test set was used to verify the predictivity of the generated model. In computation, the CP-MLR/GA training set (50 compounds) was considered as such for the training the network of ANN. However, the test set (40 compounds) of the CP-MLR/GA was randomly divided into ANN's validation (20 compounds) and test (20 compounds) sets. Coinciding with the number of descriptors in individual feature selection models, for ANN also five descriptors were considered in the input. Before training the networks, the input and output values were normalized with autoscaling of all data. The initial weights were selected randomly between (−0.3) and (0.3). Using the standard evaluation procedure with different numbers of hidden layer nodes, the optimum number of nodes for the hidden layer was assessed. The goal of training the network is to minimize the output errors by changing the weights between the layers [37]. Equation (2) gives the changes in the values of the weights in the network in the optimization of the output, as follows:

$$\Delta w_{ij,n} = F_n + \alpha \Delta w_{ij,n-1}. \tag{2}$$

In this, Δw_{ij} is the change in the weight factor for each network node, α is the momentum factor, and F is a weight update function, which indicates how weights are changed during the learning process. The weights of hidden layer were optimized using the second derivative optimization method, namely, Levenberg-Marquardt algorithm [45, 46].

2.6. Levenberg-Marquardt Algorithm. In this algorithm, the update function, F_n, is calculated the following using equations:

$$F_0 = -g_0, \tag{3}$$

$$g = J^T e, \tag{4}$$

$$F_n = -\left[J^T \times J + \mu I\right]^{-1} \times J^T \times e, \tag{5}$$

where g is gradient, J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights, and e is a vector of network errors. The parameter μ is multiplied by some factor (λ) whenever a step would result in an increased e and when a step reduces e, μ is divided by λ.

2.7. Statistical Parameters. In training the network, the overfitting of data was controlled by comparing the root-mean-square errors (RMSEs) of training and validation sets. It measures the goodness of the output and is useful for the comparison of the target values. The training of the network for the prediction of target value was stopped when the RMSE of the validation set began to increase while that of training set continues to decrease. The goodness of fit of activity of the test set compounds was used to further validate the developed models. The predictive ability of the constructed models was assessed using different statistical measures, namely, the training, validation, and test sets' correlation coefficients (r^2), and corresponding root-mean-square error of prediction (RMSEP), relative standard error of prediction (RSEP), and mean absolute error (MAE) values. More information on the statistical parameters can be found in applied statistics handbook [47]. The statistical parameters used in the study are calculated using the following equations:

$$r^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_{\text{pred}} - y_{\text{obs}}\right)^2}{\sum_{i=1}^{n} \left(y_{\text{obs}} - y_{\text{mean}}\right)^2}, \tag{6}$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_{\text{pred}} - y_{\text{obs}}\right)^2}{n}}, \tag{7}$$

Table 3: The CP-MLR and GA five-parameter models for antimalarial activity of anilinoquinolines (Table 2) along with statistics[a].

Eq.	Constant	Descriptor and scaled regression coefficient					r^2	s	Q^2	Q^2_{G5}	F	r_t^2
					CP-MLR							
(1)		MATS4m	MATS7e	H-052	Mor30p	RTe+						
	7.382	+0.816	−0.560	+1.440	−1.098	−0.905	0.692	0.395	0.604	0.588	19.76	0.562
(2)		MATS7e	H-052	RDF085p	Mor30p	R7m						
	7.802	−0.545	+1.162	+0.990	−1.073	−0.918	0.687	0.399	0.599	0.577	19.34	0.611
(3)		MATS7e	H-052	Mor30p	E1m	R7m						
	8.734	−0.633	+1.114	−0.994	−0.980	−0.885	0.684	0.400	0.6	0.607	19.08	0.632
(4)		MATS7e	H-052	Mor30p	E2m	R7m+						
	7.749	−0.548	+1.309	−1.043	+0.574	−0.948	0.666	0.411	0.565	0.564	17.58	0.610
(5)		MATS7e	H-052	Mor30p	E2m	R6m						
	7.805	−0.574	+1.349	−1.026	+0.951	−0.850	0.660	0.415	0.565	0.522	17.09	0.628
					GA							
(6)		MATS4m	MATS7e	Mor17p	R7m	R6e+						
	8.238	+1.096	−0.955	−2.267	−0.777	+0.786	0.692	0.395	0.604	0.588	19.76	0.562
(7)		MATS4m	MATS7e	H-047	H-052	Mor30p						
	6.635	+0.813	−0.570	+0.959	+1.890	−1.118	0.692	0.395	0.615	0.618	19.73	0.607
(8)		MATS7e	H-047	H-052	Mor30p	R7m						
	7.505	−0.592	+0.922	+1.357	−0.923	−0.705	0.682	0.401	0.584	0.572	18.84	0.65
(9)		MATS7e	$n\mathrm{NR}_2$	H-052	Mor15m	Mor30p						
	6.943	−0.596	+0.834	+1.614	+0.691	−1.090	0.673	0.407	0.576	0.581	18.08	0.634
(10)		MATS8m	MATS5e	MATS7e	Mor17p	R6e+						
	9.247	−0.951	−0.579	−1.042	−2.257	+0.705	0.668	0.410	0.588	0.567	17.68	0.579

[a] Number of compounds used in regression (training set) is 50 and number of compounds used in test set is 40. The regression coefficients of the descriptors shown are from the scaled descriptor values (scaled between "0" and "1", inclusive of both values). All regression coefficients are significant at higher than 95% confidence level. In Y-randomization study (100 simulations each time) no model has shown average $r^2_{Y\mathrm{rand}}$ greater than 0.107 and maximum $r^2_{Y\mathrm{rand}}$ greater than 0.385. For all regression equations, r is the multiple correlation coefficient, Q^2 and Q^2_{G5} are cross-validated r^2 from leave-one-out (LOO) and leave group of five out, respectively, s is the standard error of the estimate, F is the F-ratio between the variances of calculated and observed activities, and r_t^2 is test set r^2 value.

$$\mathrm{RSEP}\ (\%) = 100\sqrt{\frac{\sum_{i=1}^{n}\left(y_{\mathrm{pred}} - y_{\mathrm{obs}}\right)^2}{\sum_{i=1}^{n}\left(y_{\mathrm{obs}}\right)^2}}, \tag{8}$$

$$\mathrm{MAE}\ (\%) = \frac{100}{n}\sqrt{\sum_{i=1}^{n}\left|\left(y_{\mathrm{pred}} - y_{\mathrm{obs}}\right)\right|}, \tag{9}$$

where y_{obs} is the observed activity, y_{mean} is the mean of observed activity values, y_{pred} is the predicted activity of the compound in the sample, and n is the number of samples in the concerned set. The ANN computations were carried out using the MATLAB 7.6 for Windows [48].

3. Results and Discussion

The QSAR analysis of the antimalarial activity of anilinoquinolines has been carried out in CP-MLR and GA approaches using the 0D to 3D features of the molecules from Dragon software. At the end of the analysis, from each approach, five 5-parameter equations were identified as significant ones to model the activity of the compounds. The models identified from each approach are shown in Table 3. There are no common models between CP-MLR and GA approaches. However, several descriptors are common to models from both approaches. The models have predicted the activities of training and test set compounds within the reasonable limits of their actual values. Statistically, they have explained between 66% to 69% variance ($r^2 \approx 0.66$ to 0.69) in the activity of training set compounds and also predicted higher than 50% variance ($r_t^2 \geq 0.50$) in the activity of test set compounds (Table 3). For selected CP-MLR and GA models the training and test set predictions are shown in Table 2.

The equations from CP-MLR have jointly shared eleven descriptors and likewise the GA equations have shared twelve descriptors (Table 4). Together, these models have led to 18 descriptors (Table 4) as information rich features to model the antimalarial activity of the compounds. All these descriptors belong to seven different classes, namely, functional groups (nCOOR, $n\mathrm{NR}_2$), atom-centered fragments (H-047, H-052), 2D autocorrelations (MATS4m, MATS8m, MATS5e, MATS7e), radial distribution function (RDF085p), 3D molecule representation of structures based on electron diffraction signals (Mor15m, Mor28m, Mor17p, Mor30p), Weighted Holistic Invariant Molecular descriptors (E1m, E2m), and GEometry, Topology, and Atom-Weights AssemblY (R6m, R7m, R7m+, R6e+, RTe+) descriptors (Table 4). A brief physical meaning of these descriptors in terms of structural features is described in Table 4.

Among the identified variables (Table 4), 5 descriptors (H-052, MATS4m, MATS7e, Mor30p, and R7m) are common

Table 4: Information content of descriptors identified from CP-MLR and GA approaches.

S. no.	Descriptor	Class[a]	FS[b]		Information content[c]
1	MATS4m	2D-Auto	C	G	Moran autocorrelation—lag 4, weighted by atomic masses
2	MATS8m			G	Moran autocorrelation—lag 8, weighted by atomic masses
3	MATS5e			G	Moran autocorrelation—lag 5, weighted by atomic Sanderson electronegativities
4	MATS7e		C	G	Moran autocorrelation—lag 7, weighted by atomic Sanderson electronegativities
5	$n\mathrm{NR}_2$	Funct		G	Number of tertiary amines (aliphatic)
6	H-047	ACF		G	H attached to C_1(sp3), C_0(sp2)
7	H-052		C	G	H attached to C_0(sp3) with 1X attached to next C
8	RDF085p	RDF	C		Radial distribution function—8.5, weighted by atomic polarizabilities
9	Mor15m	3D-Morse		G	3D-MoRSE—signal 15, weighted by atomic masses
10	Mor17p			G	3D-MoRSE—signal 17, weighted by atomic polarizabilities
11	Mor30p		C	G	3D-MoRSE—signal 30, weighted by atomic polarizabilities
12	E1m	WHIM	C		1st component accessibility directional WHIM index, weighted by atomic masses
13	E2m		C		2nd component accessibility directional WHIM index, weighted by atomic masses
14	R6m	GETAWAY	C		R autocorrelation of lag 6, weighted by atomic masses
15	R7m		C	G	R autocorrelation of lag 7, weighted by atomic masses
16	R7m+		C		R maximal autocorrelation of lag 7, weighted by atomic masses
17	R6e+			G	R maximal autocorrelation of lag 6, weighted by atomic Sanderson electronegativities
18	RTe+		C		R maximal index, weighted by atomic Sanderson electronegativities

[a] Descriptor class: 2D-Auto: 2D autocorrelations; Funct: Functional; ACF: atom-centered fragments; RDF: radial distribution function; 3D-Morse: 3D molecule representation of structures based on electron diffraction signals; WHIM: Weighted Holistic Invariant Molecular descriptors (E1m, E2m); GETAWAY: GEometry, Topology, and Atom-Weights AssemblY.

[b] Feature selection approach involved in descriptor identification, C for CP-MLR and G for GA.

[c] See [30] for more details.

to both CP-MLR and GA approaches. Of these, MATS7e (Moran autocorrelation of lag 7 weighted by atomic Sanderson electronegativities) has appeared in all models with negative regression coefficient (Table 3). This has pointed that molecular topology leading to a reduced autocorrelation of lag 7 weighted by atomic electronegativities improves activity. This in turn explains that nonlinear and/or branched molecular topology leads to higher activity. The descriptors H-052 (with a positive regression coefficient) and Mor30p (with a negative regression coefficient) are part of all CP-MLR models as well as present in some GA models (Table 3). The H-052 argues in favor of $\mathrm{R}'\mathrm{CH}_2$–CHX–$\mathrm{CH}_2\mathrm{R}$ fragments (X is halogen atom) in the scaffold for the activity. Mor30p is 3D molecule representation of structure based on specific electron diffraction weighted by atomic polarizability. It describes the mutual arrangement of atoms in molecule leading to the 3D distribution of chosen property, that is, polarizability. The negative regression coefficient of Mor30p recommends typical arrangement of atoms in molecule leading to small descriptor values for high activity. The descriptors MATS4m (with a positive regression coefficient) and R7m (with a negative regression coefficient) have appeared in selected models of both CP-MLR and GA approaches (Table 3). The positive regression coefficient of MATS4m shows that small path lengths and branching in the molecule (lag 4 weighted by atomic mass) contribute to higher activity. The R7m is also a kind of autocorrelation of lag 7 weighted by atomic mass derived from the molecular leverage matrix. The negative regression coefficient of R7m argues that similar or almost similar atomic leverages (of lag 7) raise the activity (Table 3). Apart from the foregoing features, RDF085p, E1m, E2m, RTe+, R6m, and R7m+ are exclusive to the models from CP-MLR and MATS8m, MATS5e, H-047, $n\mathrm{NR}_2$, Mor15m, Mor17p, and R6e+ are exclusive to those from GA approach.

The descriptor RDF085p (Table 3; (2)) measures the probability of finding molecular constituents in a spherical volume of radius 8.5 Å weighted by atomic polarizability. Its positive regression coefficient argues in favor of this for improvement in antimalarial activity. The descriptors E1m and E2m (Table 3; (3)–(5)) represent 3D molecular information of atomic densities along principal axes 1 and 2 weighted by atomic mass. Principal axes of a molecule are from the eigenvalues and eigenvectors of weighted covariance matrix of its centered Cartesian coordinates. They are derived from the projections of the atoms (of the molecule) along each individual principal axis and convey information related to molecular size, shape, symmetry, and atom distribution. In the regression equations (Table 3; (3)–(5)) E1m and E2m are associated with negative and positive coefficients, respectively. This argues in favor of an atomic arrangement to maximize the 2nd principle axis of the molecule for high activity. Similar to R7m, the other GETAWAY class descriptors RTe+, R7m+, and R6m (Table 3; (1), (4), (5)) are associated with negative regression coefficients. In the molecules, while RTe+ accounts for the maximal molecular leverage autocorrelation index weighted by atomic Sanderson electronegativities, R7m+ accounts for the maximal molecular leverage autocorrelation of lag 7 weighted by atomic

Table 5: MLR-like PLS models from the combined as well as common descriptors of CP-MLR and GA approaches (Table 4) for the antimalarial activity (pIC_{50}) of anilinoquinolines (Table 2).

S. no.	Descriptor	MLR-like coeff (f.c)[a]	
		(CP-MLR)∪(GA)[b]	(CP-MLR)∩(GA)[c]
1	MATS4m	12.491 (0.058)	25.075 (0.143)
2	MATS8m	−1.965 (−0.015)	
3	MATS5e	−0.708 (−0.027)	
4	MATS7e	−2.605 (−0.143)	−2.520 (−0.169)
5	$n\mathrm{NR}_2$	0.079 (0.043)	
6	H-047	−0.001 (−0.003)	
7	H-052	0.049 (0.109)	0.110 (0.297)
8	RDF085p	0.005 (0.026)	
9	Mor15m	0.093 (0.018)	
10	Mor17p	−0.518 (−0.112)	
11	Mor30p	−1.395 (−0.151)	−1.601 (−0.212)
12	E1m	−0.390 (−0.033)	
13	E2m	0.623 (0.054)	
14	R6m	−0.819 (−0.025)	
15	R7m	−2.268 (−0.067)	−4.961 (−0.180)
16	R7m+	−2.749 (−0.018)	
17	R6e+	19.530 (0.059)	
18	RTe+	−1.080 (−0.037)	
	Constant	−2.821	−16.535
	Statistics		
n		50	50
r^2		0.731	0.638
s		0.361	0.418
F		41.68	27.04
Q^2		0.688	0.582
Q^2_{G5}		0.687	0.571
SPRESS		0.388	0.450
SEDP		0.373	0.431
r^2_t		0.676	0.510
$r^2_{Y\mathrm{rand}}$ (max)		0.115 (0.374)	0.089 (0.236)

[a]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.

[b]Combined descriptors of CP-MLR and GA.

[c]Descriptors common to CP-MLR and GA.

mass. The R6m is the molecular leverage autocorrelation of lag 6 weighted by atomic mass. All these descriptors advocate similar or almost similar leverages for high activity.

Concerning the descriptors exclusive to equations from GA, the functional groups descriptor $n\mathrm{NR}_2$ (Table 3; (9)) accounts for number of tertiary aliphatic amines in the molecule. Its positive regression coefficient speaks in favor of tertiary aliphatic amines for high activity. The H-047 (Table 3; (7), (8)) has appeared in GA equations with positive regression coefficient. In these analogues, it argues that unsubstituted methylenes lead to activity improvement. The 2D autocorrelation descriptors MATS8m, and MATS5e (Table 3; (10)) have appeared with negative regression coefficient. Both these descriptors infer in favor of molecular topology leading to a reduced autocorrelation of lag 8 weighted by atomic mass and lag 5 weighted by atomic electronegativities for improved activity. They further illustrate that nonlinear and/or branched molecular topology increases the activity. The descriptors Mor15m (Table 3; (9); positive regression coefficient) and Mor17p (Table 3; (6), (10); negative regression coefficient), similar to Mor30p, most probably show the influence of the specific distribution of atoms in the molecule on its activity. The GETAWAY class descriptor R6e+ has appeared in (Table 3; (6) and (10)) with positive regression coefficient. It represents maximal molecular leverage autocorrelation of lag 6 weighted by atomic Sanderson electronegativities. This suggests that increasing divergence in the leverage of lag 6 contributes to higher activity.

As a followup of feature identification, PLS analysis has been carried out on the eighteen descriptors of CP-MLR and

Table 6: Architecture and goodness of fit of BP-ANN models for antimalarial activity of anilinoquinolines (Table 2) from selected feature sets of CP-MLR, GA, and their common five descriptors.

BP-ANN architecture and parameters			
Layer	Nodes	Training parameters	
Input	5 + 1 (bias)	Learning rate (μ)	0.57–0.66
Hidden	6	Momentum (α)	0.55–0.77
Output	1	Transfer function	Sigmoid
		Optimization algorithm	Levenberg-Marquardt
		Iterations (λ)	17–40

Model statistics														
Feature set[a]	μ	α	r^{2b}			RMSEP			RSEP (%)			MAE (%)		
			Train	Valid	Test	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
CP-MLR	0.601	0.558	0.813	0.814	0.818	0.289	0.256	0.267	3.914	3.404	3.564	6.457	9.859	9.879
GA	0.577	0.657	0.883	0.886	0.878	0.229	0.200	0.219	3.099	2.662	2.925	6.019	8.349	9.287
Common	0.66	0.77	0.823	0.810	0.757	0.284	0.279	0.333	3.852	3.710	4.448	6.819	10.832	11.810

[a] The ANN input features of CP-MLR and GA sets, respectively, correspond to (1) and (6) given in Table 3. The ANN input features common sets are features common to CP-MLR and GA.
[b] r^2: squared correlation coefficient; RMSEP: root-mean-square error of prediction; RSEP: relative standard error of prediction; MAE: mean absolute error.

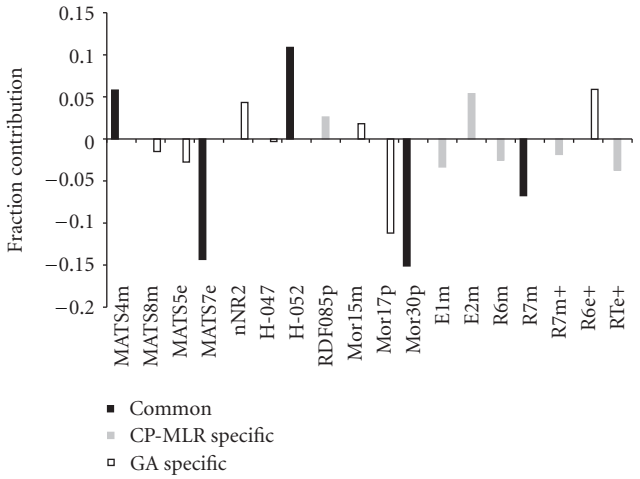

Figure 2: Plot of fraction contributions of MLR-like PLS coefficients (normalized) from the combined descriptors of CP-MLR and GA for the antimalarial activity of anilinoquinolines.

GA and the five common descriptors of both the approaches to facilitate the development of single-window structure-activity models. For PLS analysis, the descriptors have been autoscaled (zero mean and unit s.d) to give each one of them equal weight in the study. In the cross-validation procedure of the PLS analysis [35, 36], three components are found to be the optimum to explain the activity of the compounds. The PLS model from the eighteen descriptors of CP-MLR cum GA has explained 73.1% variance ($r^2 = 0.731, Q^2 = 0.688, s = 0.361, F = 41.68$) in the antimalarial activity of the training set compounds and showed a test set r^2 value 0.676. Figure 2 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity. Of the eighteen descriptors, the fraction contributions of five common descriptors of both approaches are found amongst the most significant ones to modulate the activity of the compounds. Also, the PLS model from these five common descriptors of CP-MLR and GA has explained 63.8% variance ($r^2 = 0.638, Q^2 = 0.582, s = 0.418, F = 27.04$) in the antimalarial activity of the training set compounds and showed a test set r^2 value 0.510. The MLR-like PLS coefficients of these two feature sets are shown in Table 5. All descriptors have conveyed the same meaning as in the case of regression equations from CP-MLR and GA.

The predictive ability of regression models derived from the CP-MLR, GA, and PLS approaches is assessed using applicability domain (AD) analysis. The AD plots for Eq. (1) and Eq. (6) and the PLS model are shown in Figure 3. They are from the models involving all the compounds, that is, training and test sets together. In the plots, the Y-outliers (response outliers) limits were set to 2.5 times the standard deviation units. In the AD plot of Eq. (6) (Figure 3(b)), two test set compounds are marginally outside the allowed region. Of these two, one compound (AQ14) is response outlier (observed residual value is 1.061; limiting residual value is ±0.993) and the other compound (AQ01) is leverage outlier (observed leverage is 0.366; limiting leverage value is 0.36). Except for these minor deviations, the AD plots argue in support of the predictive power of the presented models. Also the models are free from serious or influential outliers (Figure 3).

The models discussed so far could explain up to 73% variance in the activity. Prevalence of some degree of non-linearity in the activity in relation to the structural features is among the main reasons for this kind of situation. Often the biological activity landscape of chemical entities is far more nonlinear when compared to their physicochemical (also other properties) arena. In modeling studies artificial neural networks (ANNs) have a special place to address these situations. In ANN, involving of descriptors from feature selection approaches is a desirable option as they provide direction for the modification of chemical space to carry

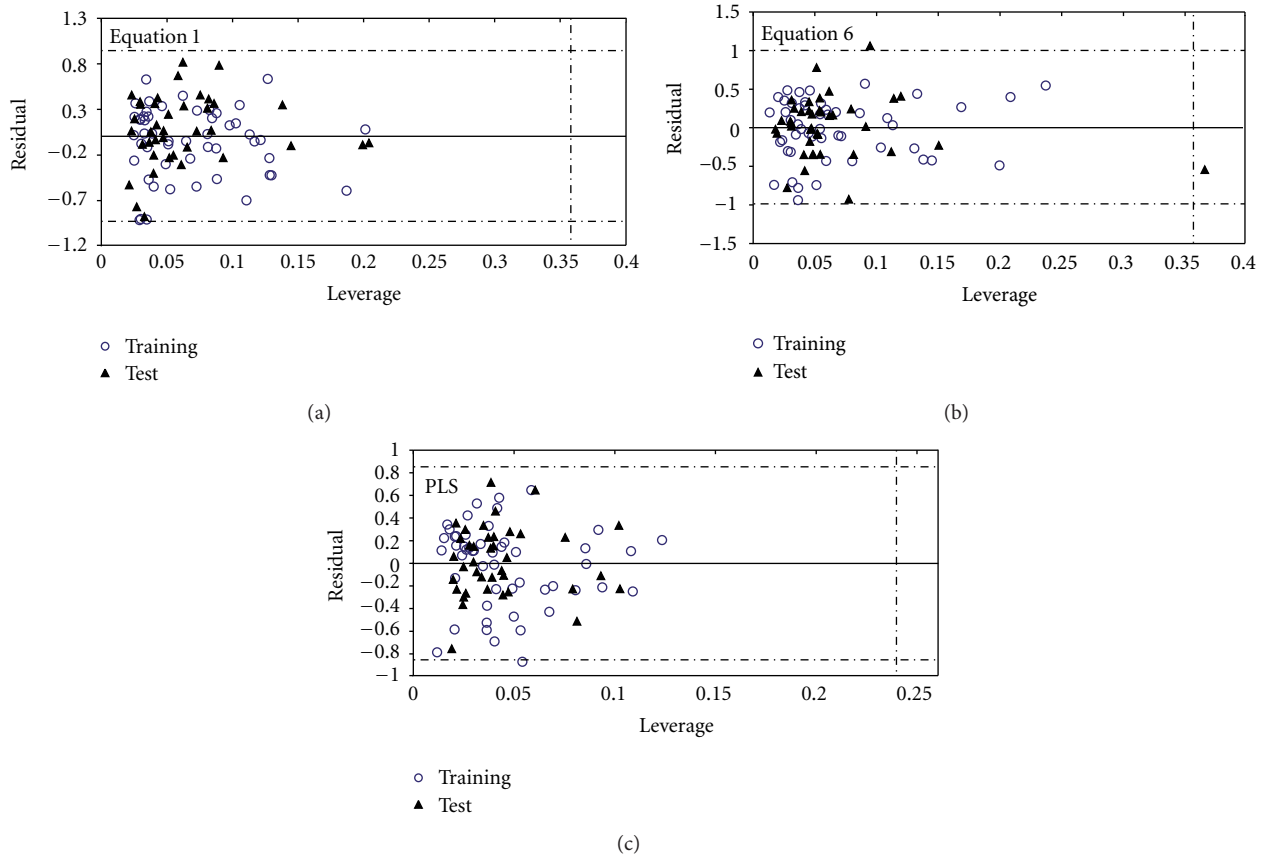

Figure 3: Williams plot (applicability domain) of the training compounds and external prediction set compounds for Eq. (1) and Eq. (6) and 3-component PLS equation of 18 descriptors. Cut-off value of leverage for Eq. (1) and Eq. (6) is 0.366 and for PLS model is 0.24; for residual cut-off value is taken as 2.5 of standard deviation.

out activity modulation [49]. In view of this the features of selected models of CP-MLR and GA (Table 3; (1) and (6)) and the five common descriptors of CP-MLR and GA (MATS4m, MATS7e, H-052, Mor30p, and R7m) have been used separately for the development of three BP-ANN models for the activity. The ANN architecture with network parameters and the predictive statistics of the emerged models are shown in Table 6. In ANN models, these descriptors have well explained the antimalarial activity of the compounds ($r^2 \geq 0.81$). Also they gave satisfactory predictions for the test set compounds (test set $r^2 \geq 0.75$). The plots of observed versus ANN predicted activities are shown in Figure 4. In ANN models also the features of CP-MLR, GA, and common sets infer the same meaning as discussed in previous paragraphs. The results clearly demonstrated that these descriptors have the ability to identify the patterns in the data and predict the activity of potential analogues.

4. Conclusions

The antimalarial activity of a series of anilinoquinolines was modeled with the feature selection approaches CP-MLR and GA. This has led to the identification of eighteen descriptors to model the activity of the compounds. Among the identified descriptors, five (H-052, MATS4m, MATS7e, Mor30p, and R7m) are common to both CP-MLR and GA approaches. For the development of the single-window structure-activity model, all eighteen features were analyzed in PLS. In PLS analysis, the common descriptors of CP-MLR and GA are found among the most influential ones to modulate the activity of the anilinoquinolines. In regression as well as PLS models the negative coefficient of MATS7e argued that nonlinear and/or branched molecular topology leads to higher activity. H-052 represents the hydrogen(s) attached to sp3 carbon which is next to the carbon anchoring halogens. Its regression coefficient advocated in favor of such fragments for higher activity. The regression coefficient of H-052 advocated for the groups containing hydrogen of sp3 carbon attached to next carbon containing halogens in the substituents for higher activity. In BP-ANN, the descriptors from the selected equations of both feature selection approaches and the five most significant descriptors of PLS analysis (MATS4m, MATS7e, H-052, Mor30p, and R7m) have explained higher than 81% variance in the antimalarial activity of the training set compounds and showed a test set r^2 value greater than 0.75. These results offered direction to understand the patterns of the antimalarial activity of anilinoquinolines and may serve to predict the activity of potential prototype compounds. The values of the eighteen

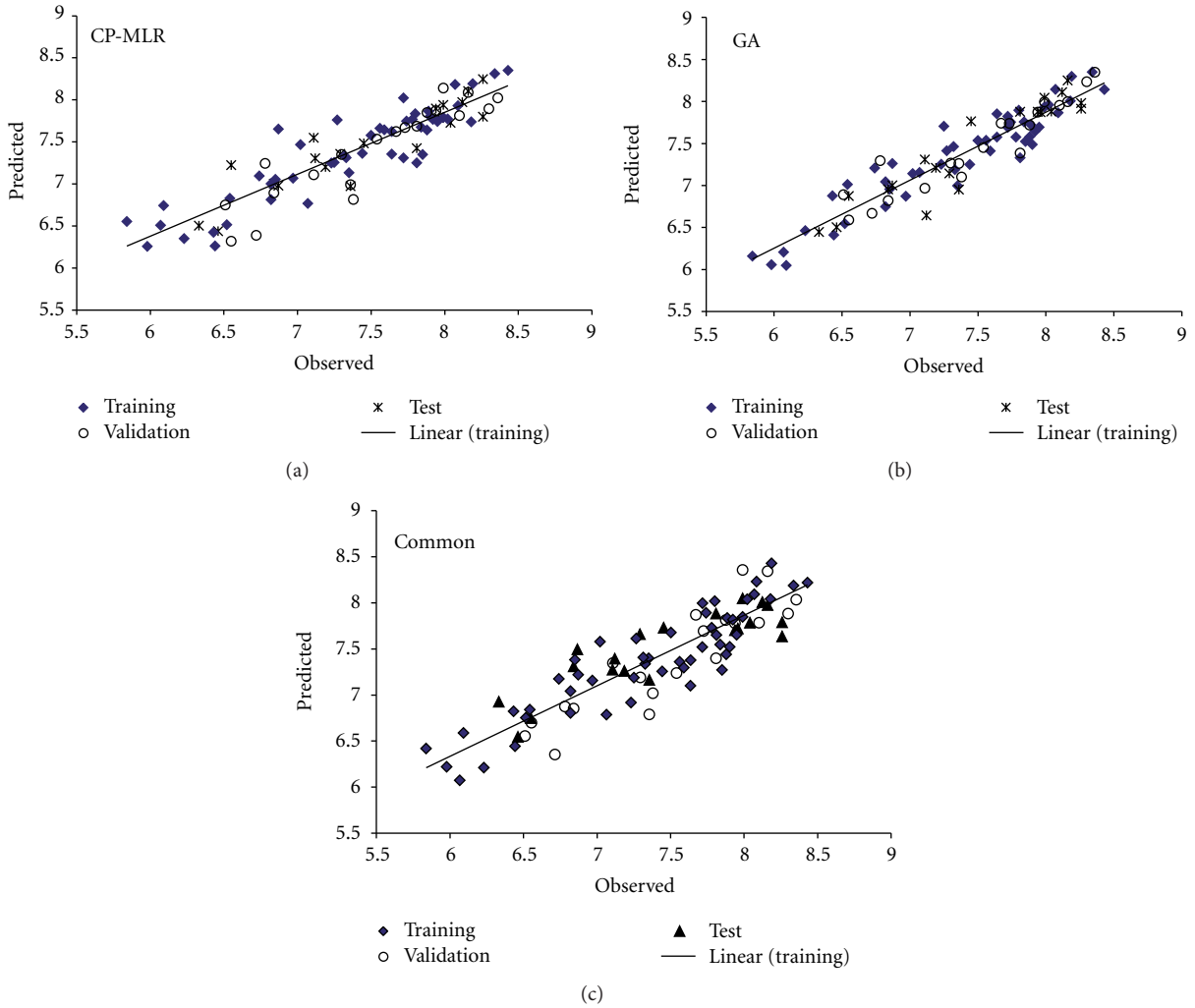

Figure 4: The plots of observed versus predicted activities of anilinoquinolines' training, validation, and test sets from BP-ANN models derived using the descriptors of Eq. (1) (CP-MLR), Eq. (6) (GA) (Table 3), and common descriptors of CP-MLR and GA (Table 4). The solid line indicates the best fit.

descriptors involved in the regression equations are provided as supplementary material to facilitate likely structural exploration (Supplementary material will be available online at http://dx.doi.org/10.1155/2013/154629).

Acknowledgment

This work is supported by CDRI Communication no. 8310.

References

[1] WHO report on communicable diseases (World Malaria Report 2011), June 2012, http://www.searo.who.int/en/Section10/Section21.htm.

[2] C. V. Plowe, "Antimalarial drug resistance in Africa: strategies for monitoring and deterrence," *Current Topics in Microbiology and Immunology*, vol. 295, pp. 55–79, 2005.

[3] A. C. Uhlemann and S. Krishna, "Antimalarial multi-drug resistance in Asia: mechanisms and assessment," *Current Topics in Microbiology and Immunology*, vol. 295, pp. 39–53, 2005.

[4] M. Foley and L. Tilley, "Quinoline antimalarials: mechanisms of action and resistance," *International Journal for Parasitology*, vol. 27, no. 2, pp. 231–240, 1997.

[5] S. Leecharoen, S. Wangkaew, and W. Louthrenoo, "Ocular side effects of chloroquine in patients with rheumatoid arthritis, systemic lupus erythematosus and scleroderma," *Journal of The Medical Association of Thailand*, vol. 90, no. 1, pp. 52–58, 2007.

[6] A. C. Chou and C. D. Fitch, "Control of heme polymerase by chloroquine and other quinoline derivatives," *Biochemical and Biophysical Research Communications*, vol. 195, no. 1, pp. 422–427, 1993.

[7] A. F. G. Slater, "Chloroquine: mechanism of drug action and resistance in *Plasmodium falciparum*," *Pharmacology & Therapeutics*, vol. 57, no. 2-3, pp. 203–235, 1993.

[8] C. Biot, G. Glorian, L. A. Maciejewski et al., "Synthesis and antimalarial activity *in vitro* and *in vivo* of a new ferrocene-chloroquine analogue," *Journal of Medicinal Chemistry*, vol. 40, no. 23, pp. 3715–3718, 1997.

[9] D. De, F. M. Krogstad, L. D. Byers, and D. Krogstad, "Structure-activity relationships for antiplasmodial activity among 7-substituted 4-aminoquinolines," *Journal of Medicinal Chemistry*, vol. 41, no. 2, pp. 4918–4926, 1998.

[10] P. A. Stocks, K. J. Raynes, P. G. Bray, B. K. Park, P. M. O'Neill, and S. A. Ward, "Novel short chain chloroquine analogues retain activity against chloroquine resistant K1 *Plasmodium falciparum*," *Journal of Medicinal Chemistry*, vol. 45, no. 23, pp. 4975–4983, 2002.

[11] L. M. Werbel, P. D. Cook, E. F. Elslager, J. H. Hung, J. L. Johnson, S. J. Kesten et al., "Antimalarial drugs. 60. Synthesis, antimalarial activity, and quantitative structure-activity relationships of tebuquine and a series of related 5-[(7-chloro-4-quinolinyl)amino]-3-[(alkylamino)methyl][1,1′-biphenyl]-2-ols and N.omega.-oxides," *Journal of Medicinal Chemistry*, vol. 29, no. 6, pp. 924–939, 1986.

[12] H. L. Koh, M. L. Go, T. L. Ngiam, and J. W. Mak, "Conformational and structural features determining *in vitro* antimalarial activity in some indolo[3,2-*c*]quinolines, anilinoquinolines and tetrahydroindolo[3,2-*d*]benzazepines," *European Journal of Medicinal Chemistry*, vol. 29, pp. 107–113, 1994.

[13] P. M. O'Neill, D. J. Willock, S. R. Hawley et al., "Synthesis, antimalarial activity, and molecular modeling of tebuquine analogues," *Journal of Medicinal Chemistry*, vol. 40, no. 4, pp. 437–448, 1997.

[14] S. R. Hawley, P. G. Bray, M. Mungthin, J. D. Atkinson, P. M. O'Neill, and S. A. Ward, "Relationship between antimalarial drug activity, accumulation, and inhibition of heme polymerization in *Plasmodium falciparumin vitro*," *Antimicrobial Agents and Chemotherapy*, vol. 42, no. 3, pp. 682–686, 1998.

[15] P. M. O'Neill, P. G. Bray, S. R. Hawley, S. A. Ward, and B. K. Park, "4-Aminoquinolines—past, present, and future; a chemical perspective," *Pharmacology & Therapeutics*, vol. 77, no. 1, pp. 29–58, 1998.

[16] T. J. Egan, "Structure-function relationships in chloroquine and related 4-aminoquinoline antimalarials," *Mini Reviews in Medicinal Chemistry*, vol. 1, no. 1, pp. 113–123, 2001.

[17] S. R. Meshnick and A. P. Alker, "Amodiaquine and combination chemotherapy for malaria," *American Journal of Tropical Medicine and Hygiene*, vol. 73, no. 5, pp. 821–823, 2005.

[18] K. H. Rieckmann, "Determination of the drug sensitivity of *Plasmodium falciparum*," *Journal of the American Medical Association*, vol. 217, no. 5, pp. 573–578, 1971.

[19] C. S. R. Hatton, C. Bunch, T. E. A. Peto et al., "Frequency of severe neutropenia associated with amodiaquine prophylaxis against malaria," *The Lancet*, vol. 327, no. 8478, pp. 411–414, 1986.

[20] K. A. Neftel, W. Woodtly, M. Schmid, P. G. Frick, and J. Fehr, "Amodiaquine induced agranulocytosis and liver damage," *British Medical Journal*, vol. 292, no. 6522, pp. 721–723, 1986.

[21] J. L. Maggs, N. R. Kitteringham, and B. K. Park, "Drug-protein conjugates—XIV: mechanisms of formation of protein-arylating intermediates from amodiaquine, a myelotoxin and hepatotoxin in man," *Biochemical Pharmacology*, vol. 37, no. 2, pp. 303–311, 1988.

[22] A. C. Harrison, N. R. Kitteringham, J. B. Clarke, and B. K. Park, "The mechanism of bioactivation and antigen formation of amodiaquine in the rat," *Biochemical Pharmacology*, vol. 43, no. 7, pp. 1421–1430, 1992.

[23] S. Delarue, S. Girault, L. Maes et al., "Synthesis and *in vitro* and *in vivo* antimalarial activity of new 4-anilinoquinolines," *Journal of Medicinal Chemistry*, vol. 44, no. 17, pp. 2827–2833, 2001.

[24] S. Delarue-Cochin, E. Paunescu, L. Maes et al., "Synthesis and antimalarial activity of new analogues of amodiaquine," *European Journal of Medicinal Chemistry*, vol. 43, no. 2, pp. 252–260, 2008.

[25] E. Davioud-Charvet, S. Delarue, C. Biot et al., "A prodrug form of a *Plasmodium falciparum* glutathione reductase inhibitor conjugated with a 4-anilinoquinoline," *Journal of Medicinal Chemistry*, vol. 44, no. 24, pp. 4268–4276, 2001.

[26] S. Delarue-Cochin, P. Grellier, L. Maes, E. Mouray, C. Sergheraert, and P. Melnyk, "Synthesis and antimalarial activity of carbamate and amide derivatives of 4-anilinoquinoline," *European Journal of Medicinal Chemistry*, vol. 43, no. 10, pp. 2045–2055, 2008.

[27] M. K. Gupta and Y. S. Prabhakar, "Topological descriptors in modeling the antimalarial activity of 4-(3′,5′-disubstituted anilino)quinolines," *Journal of Chemical Information and Modeling*, vol. 46, no. 1, pp. 93–102, 2006.

[28] A. Semeniuk, A. Niedospial, J. Kalinowska-Tluscik, W. Nitek, and B. J. Oleksyn, "Molecular geometry of antimalarial amodiaquine in different crystalline environments," *Journal of Molecular Structure*, vol. 875, no. 1–3, pp. 32–41, 2008.

[29] SYBYL, version 7.3, Tripos associates, St. Louis, Mo, USA, 2006.

[30] R. Todeschini, V. Consonni, A. Mauri, and M. Pavan, DRAGON software version 5. 0-2005, Milano, Italy, http://michem.disat.unimib.it/chm/.

[31] M. Waldrop, "Unique writing course for student engineers," *Chemical & Engineering News*, vol. 57, no. 3, pp. 28–29, 1979.

[32] MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1255 University Street, Suite 1600, Montreal, Quebec, Canada H3B 3X3.

[33] Y. S. Prabhakar, "A combinatorial approach to the variable selection in multiple linear regression: analysis of selwood *et al.* Data set—a case study," *QSAR & Combinatorial Science*, vol. 22, no. 6, pp. 583–595, 2003.

[34] D. R. Rogers and A. J. Hopfinger, "Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships," *Journal of Chemical Information and Computer Sciences*, vol. 34, no. 4, pp. 854–866, 1994.

[35] S. Wold, "Cross-validatory estimation of the number of components in factor and principal components models," *Technometrics*, vol. 20, no. 4, pp. 397–405, 1978.

[36] L. Stahle and S. Wold, "Multivariate data analysis and experimental design in biomedical research," in *Progress in Medicinal Chemistry*, G. P. Eillis and W. B. West, Eds., vol. 25, pp. 291–338, Elsevier Science, Amsterdam, The Netherlands, 1988.

[37] D. Graupe, in *Principles of Artificial Neural Networks*, pp. 59–111, World Scientific, Singapore, 2nd edition, 2007.

[38] S. Deshpande, V. R. Solomon, S. B. Katti, and Y. S. Prabhakar, "Topological descriptors in modelling antimalarial activity: N1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine as prototype," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 24, no. 1, pp. 94–104, 2009.

[39] S. Deshpande, R. Singh, M. Goodarzi, S. B. Katti, and Y. S. Prabhakar, "Consensus features of CP-MLR and GA in modeling HIV-1 RT inhibitory activity of 4-benzyl/benzoylpyridin-2-one analogues," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 26, no. 5, pp. 696–705, 2011.

[40] R. Todeschini, V. Consonni, and M. Pavan, MOBYDIGS software, Version 1.2 for Windows, Talete Srl, Milan, Italy, 2002, http://www.ccl.net/qsar/archives/0407/0516.html.

[41] M. Pavan, A. Mauri, and R. Todeschini, "Total ranking models by the genetic algorithm variable subset selection (GA–VSS) approach for environmental priority settings," *Analytical and Bioanalytical Chemistry*, vol. 380, no. 3, pp. 430–444, 2004.

[42] S.-S. So and M. Karplus, "Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations," *Journal of Medicinal Chemistry*, vol. 40, no. 26, pp. 4347–4359, 1997.

[43] Y. S. Prabhakar, V. R. Solomon, R. K. Rawal, M. K. Gupta, and S. B. Katti, "CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones," *QSAR & Combinatorial Science*, vol. 23, no. 4, pp. 234–244, 2004.

[44] P. Gramatica, "Principles of QSAR models validation: internal and external," *QSAR & Combinatorial Science*, vol. 26, no. 5, pp. 694–701, 2007.

[45] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[46] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[47] L. Sachs, *Applied Statistics: A Handbook of Techniques*, Springer, Berlin, Germany, 1982.

[48] MATLAB, Version 7.6, http://www.mathworks.com/products/matlab/.

[49] M. Goodarzi, S. Deshpande, V. Murugesan, S. B. Katti, and Y. S. Prabhakar, "Is feature selection essential for ANN modeling?" *QSAR & Combinatorial Science*, vol. 28, no. 11-12, pp. 1487–1499, 2009.