Topological Descriptors in Modeling the Antimalarial Activity of 4-(3′,5′-Disubstituted anilino)quinolines†

Manish K. Gupta and Yenamandra S. Prabhakar*

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

Received April 9, 2005

Two series of closely related antimalarial agents, 7-chloro-4-(3′,5′-disubstituted anilino)quinolines, have been analyzed using Combinatorial Protocol in Multiple Linear Regression (CP-MLR) for the structure−activity relations with more than 450 topological descriptors for each set. The study clearly suggested that 3′- and 5′-substituents of the anilino moiety map different domains in the activity space. While one domain favors the compact structural frames having aromatic, heterocyclic ring(s) substituted with closely spaced F, NO_2, and O functional groups, the other prefers structural frames enriched with unsaturation, loops, branches, electronic content, and devoid of carbonyl function. Also, this study gives an indication in favor of the electron rich centers in the aniline substituent groups for better antimalarial activity, an observation in line with several of the previous reports too. The models developed, and the participating descriptors suggest that the substituent groups of the 4-anilino moiety of the 4-(3′,5′-disubstituted anilino)quinolines hold scope for further modification in the optimization of the antimalarial activity.

INTRODUCTION

In drug research, the chemical structure and its biological response, activity, toxicity, physiological concentrations (accumulation/retention/excretion), etc. are conjoined systems. A study of these systems with respect to each other will come to aid in promoting the insight and in tuning them for optimum response. In this, optimization of the biological response of the chemical structure is a cherished goal. Here, understanding of the structure−activity relations comes to aid this process. One of the contemporary chemotherapeutic demands is the development of new and alternative antimalarial agents to address the proliferation of multidrug-resistant strains of Plasmodium parasites. Research interest have been renewed in quinoline based antimalarial agents due to the existence of large information base and the low cost involved in the preparation of alternative/new drugs or druglike molecules in this chemical class.[1−7] In this scenario amodiaquine, a 4-aminoquinoline analogue, has emerged as potential lead for modification because of the fact that the chloroquine-resistant Plasmodium parasites are not automatically cross resistant to it.[8] Also, the mechanistic investigations of the antimalarial activity of chloroquine and amodiaquine have suggested that they follow a similar pathway in the expression of the activity.[9,10] However, amodiaquine has been reported to cause agranulocytosis and hepatitis.[11,12] This has been attributed to the 4-hydroxyanilino moiety of amodiaquine, which is suggested to undergo enzymatic oxidation to its quinoneimine form for subsequent nucleophilic addition to proteins.[13,14] This knowledge has prompted for the synthesis and evaluation of some 7-chloro-4-(3′,5′-disubstituted anilino)quinolines (hereafter referred to as 4-(3′,5′-disubstituted anilino)quinolines or briefly as 4-anilinoquinolines) having resemblance with amodiaquine but devoid of a hydroxyl group at the para position in the anilino moiety as alternative antimalarial agents.[15] In this background, a quantitative structure−activity relationship (QSAR) study of the antimalarial activity of these 4-anilinoquinolines has been undertaken to rationalize their activity profile. In QSAR studies, transformation of a chemical structure into numerical descriptors (variables) plays a pivotal role. A meaningful inter- and intravariables communication results in the evolution of a variety of models with different predictive and diagnostic values. In this each model may address different substructural regions and attributes in explaining the chosen phenomenon. This provides scope to understand the predictive and diagnostic aspects of different substructural regions and in averaging them beyond the individual models. For this, it is necessary to characterize the molecules and their structural fragments from different perspectives for the generation of a large number of diverse descriptors. For this several computer programs based on graph theory are available to compute characteristic descriptors of the molecules and their structural fragments.[16−20] Moreover, when dealing with a large number of descriptors, for the optimum utilization of information content of the generated data sets it is necessary to adopt typical protocol(s) to identify the best models as well as information rich descriptors corresponding to the phenomenon under investigation. The Genetic Function Approximation (GFA),[21] MUtation and SElection Uncover Models (MUSEUM),[22] and Combinatorial Protocol in Multiple Linear Regression (CP-MLR)[23] are a few approaches among many others to address the model evolution in high-dimensional QSAR studies. In this milieu, the antimalarial activity of 4-anilinoquinolines (Figure 1)[15] has been analyzed with the molecular descriptors from

* Corresponding author phone: +91-522-2612411; fax: +91-522-2624305; e-mail: yenpra@yahoo.com.

† This article is based on material presented in the Fourth Indo-US Workshop on Mathematical Chemistry, January 8-12, 2005, University of Pune, Pune, India organized by Natural Resources Research Institute, University of Minnesota Duluth, Duluth, Minnesota, U.S.A. and University of Pune, Pune, India. C.D.R.I. Communication no. 6763.

10.1021/ci0501140 CCC: $33.50 © 2006 American Chemical Society
Published on Web 07/12/2005

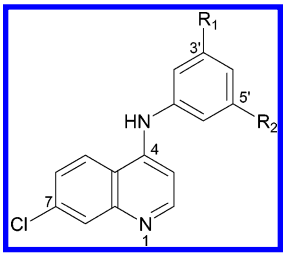

Figure 1. General structure of the 7-chloro-4-(3′,5′-disubstituted anilino)quinolines (Tables 1 and 2) associated with in vitro antimalarial activity against the FcB1R strain of *P. falciparum*. For the compounds of Table 1, R_2 is $-CH_2OH$ and R_1 is a varying group of $-NHC(O)-(CH_2)_n$-NRR′. And for the compounds of Table 2, R_1 is $-NHC(O)-CH_2-N(CH_2)_5$ where $N(CH_2)_5$ corresponds to piperidinyl and R_2 may be $-CH_2-X$ or $-CH_2-OC(O)-(CH_2)_n-X$.

Table 1. Observed and Modeled in Vitro Antimalarial Activity of 7-Chloro-4-(3′,5′-disubstituted anilino)quinolines (Figure 1)[a] against the FcB1R Strain of *P. falciparum*

compd no.	*n*	NRR′	−logIC50 obs[b]	eq 1[c]	eq 2
1	1	4-methylpiperidine	7.11	6.57	7.08
2	2	4-methylpiperidine	6.87	6.56	6.87
3	4	4-methylpiperidine	6.33	6.65	6.58
4	5	4-methylpiperidine	6.52	6.65	6.50
5	7	4-methylpiperidine	6.51	6.71	6.53
6	11	4-methylpiperidine	7.07	7.12	6.93
7	1	piperidine	7.35	6.83	6.99
8	1	*N*-methylpiperazine	6.46	6.74	6.79
9[d]	1	morpholine	<6.00	6.67	6.07
10	1	4-hydroxypiperidine	5.98	6.32	6.03
11	1	pyrrolidine	7.36	7.29	7.18
12	1	thiazolidine	6.23	6.27	6.32
13	1	$NHC(CH_3)_3$	7.29	7.38	7.48
14	1	NEt_2	7.38	7.37	7.26
15	1	$NHCH_2C_6H_5$	6.72	6.70	6.89
16	1	$NHCH_2C_6H_4Cl$ (*para*)	6.07	6.01	5.83
17	1	1,2,3,4-tetrahydroisoquinoline	5.84	5.79	5.89
18	1	3-aminopyrazole	6.09	6.09	6.05
19[d]	1	Cl	<6.00	5.79	5.83
20[d]	1	OH	<6.00	6.52	5.73
21[d]	0	phenyl	<6.00	6.68	7.71
22	0	H	6.44	6.56	6.45

[a] R_1 is $-NHC(O)-(CH_2)_n$-NRR′ and R_2 is $-CH_2OH$. [b] From ref 15. [c] Activity from a corresponding equation. [d] Compound reported as active at a dose more than 1 μM.

DRAGON software[19] using a typical descriptor search protocol involving the CP-MLR approach.[23−26] The procedural aspects and the results are presented here.

MATERIALS AND METHODS

Data Set. In this study two prototypes of 4-(3′,5′-disubstituted anilino)quinoline analogues (Tables 1 and 2) have been considered from the literature report[15] along with their in vitro antimalarial activity in the form of logarithm of the inverse of inhibitory concentration ($-\log IC_{50}$ where IC_{50} is in moles per liter against the chloroquine-resistant *P. falciparum* FcB1R strain). DRAGON software[19] has been used for the parametrization of the compounds of this study (Figure 1) (Tables 1 and 2)[15] as it offers several hundreds of descriptors from different perspectives corresponding to empirical, constitutional, and topological indices characteristic to the molecules and their structural fragments under multidescriptor class environment. Since 4-anilinoquinoline

Table 2. Observed and Modeled in Vitro Antimalarial Activity of 7-Chloro-4-(3′,5′-disubstituted anilino)quinolines (Figure 1)[a] against the FcB1R Strain of *P. falciparum*

compd no.	*n*	X	−logIC50 obs[b]	eq 3[c]	eq 4	eq 5
7			7.35	7.41	7.23	7.32
23	1	piperidine	7.27	7.46	7.48	7.70
24	2	piperidine	7.11	7.42	7.47	7.21
25	4	piperidine	7.81	7.31	7.33	7.44
26	5	piperidine	7.56	7.23	7.22	7.28
27	7	piperidine	6.78	7.06	7.01	6.96
28	4	pyrrolidine	7.64	7.30	7.39	7.54
29	4	morpholine	7.85	7.31	7.34	7.45
30	4	*N*-methylpiperazine	7.19	7.31	7.32	7.42
31	4	NEt_2	6.85	7.07	7.09	7.05
32	4	Br	6.82	6.94	6.93	6.93
33	0	phenyl	7.12	7.46	7.41	7.24
34	0	quinol-4-yl	7.59	7.57	7.49	7.35
35[a]			6.55	6.71	6.66	6.57
36		$NHCH_2CH_2NMe_2$	7.81	7.83	7.93	7.83
37[d]		$NHCH_2CH_2CH_2NMe_2$	6.87	7.84	7.92	7.79
38		$NMeCH_2CH_2NMe_2$	8.30	8.04	8.18	8.14
39		$NHCH_2CH_2$-pyrrolidine	7.99	8.05	7.92	7.90
40		$NHCH_2CH_2$-piperidine	8.18	8.04	8.05	8.01
41		$NHCH_2CH_2$-morpholine	8.10	8.01	7.98	7.94
42		NEt_2	8.09	7.88	8.18	8.12
43		NH-*t*Bu	8.43	8.01	8.34	8.36
44		piperidine	7.99	7.99	8.04	8.08
45		pyrrolidine	8.16	7.98	8.06	8.13
46		*N*-methylpiperazine	8.12	8.10	8.15	8.20
47		morpholine	8.02	7.99	8.04	8.08
48		4-hydroxypiperidine	8.19	8.04	8.06	8.10
49		*N*-(2-hydroxyethyl)piperazine	7.94	8.05	8.04	8.03
50		*N*-phenylpiperazine	8.26	8.19	8.11	8.15
51		*N*-benzylpiperazine	7.89	8.17	8.15	8.17
52		*N*-(diphenylmethyl)piperazine	7.90	8.11	8.03	8.00
53		*N*-(4-chlorobenzyl)piperazine	7.94	8.13	8.08	8.06
54		*N*-(4-methoxybenzyl)piperazine	7.96	8.13	8.12	8.09
55		*N*-(4-nitrobenzyl)piperazine	7.88	8.14	8.12	8.10
56		*N*-(4-diethylaminobenzyl)-piperazine	7.92	8.05	7.98	7.88
57		*N*-(4-cyanobenzyl)piperazine	7.88	8.17	8.14	8.13
58		*N*-piperonylpiperazine	8.16	8.22	8.19	8.22
59		NHPh	7.67	7.96	7.96	8.00
60		$NHCH_2Ph$	8.07	8.07	8.04	8.06
61		$NHCHPh_2$	8.04	8.13	8.02	8.02
62		$NHCH_2C_6H_4Cl$ (*para*)	8.34	8.03	7.94	7.93
63		$NHCH_2C_6H_4OMe$ (*para*)	8.36	8.14	8.09	8.09
64		$NHCH_2C_6H_4CF_3$ (*para*)	8.26	8.13	8.06	8.07

[a] For compounds 23−34 R_2 is $-CH_2-OC(O)-(CH_2)_n-X$, for compound 35 R_2 is $-CHO$, and for compounds 36−64 is R_2 is CH_2-X; R_1 is $-NHC(O)-CH_2-N(CH_2)_5$ where $N(CH_2)_5$ corresponds to piperidinyl. [b] From ref 15. [c] Activity from a corresponding equation. [d] Outlier for the analysis.

skeletons are common to all the compounds (Figure 1), only the R_1 and R_2 fragments attached to this skeleton have been considered for the parametrization to represent the changing structural features of these analogues. Here, for the purpose of computation of the descriptors, the valence satisfied (free valence was satisfied with hydrogen) structural fragments of the R_1 and R_2 groups of the compounds (Figure 1) have been drawn in ChemDraw[27] using the standard procedure. All these structures of the varying R_1 and R_2 groups of respective compounds have been ported to DRAGON software[19] for computing the parameters corresponding to 0D-, 1D-, and 2D- descriptor classes. The descriptor classes considered in the study along with their definitions and scope in addressing the structural features have been presented in Table 3. As the total number of descriptors involved in this

Table 3. Descriptor Classes Used for the Analysis of Antimalarial Activity of 7-Chloro-4-(3′,5′-disubstituted anilino)quinolines (Tables 1 and 2) and the Identified Categories in Modeling the Activity

descriptor class (acronyms)[a]	definition and scope	descriptors' category[b] Table 1	descriptors' category[b] Table 2
constitutional (CONST)	dimensionless or 0D descriptors; independent from molecular connectivity and conformations	I	I
topological (TOPO)	2D-descriptors from molecular graphs and independent of conformations	I	I
molecular walk counts (MWC)	2D-descriptors representing self-returning walks counts of different lengths	II	
modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements & atoms	III	II
Galvez topological charge indices (GVZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix	II	III
2D-autocorrelations (2DAUTO)	molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)	II	II
functional groups (FUN)	molecular descriptors based on the counting of the chemical functional groups		II
atom centered fragments (ACF)	molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen	II	I
empirical (EMP)	1D-descriptors represent the counts of nonsingle bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule		

[a] Reference 19. [b] Descriptor categories identified at the end of second stage of the flowchart (Figure 2); in this the filter values in CP-MLR are as follows: filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.71, and filter-4 as $0.3 \leq Q^2 \leq 1.0$; the number of compounds in the study from Tables 1 and 2 are 18 and 42, respectively.

study is more than 450 for each set of compounds, only the names of descriptor classes and the actual descriptor involved in the models have been listed.[28] The computational procedure of model development is briefly described below.

Model Development. The strategy shown in the flowchart (Figure 2) has been used for the identification of information rich descriptors corresponding to the phenomenon, the biological activity, under investigation. This has three stages in it—the first stage sorts the descriptor classes into different categories depending on their ability to form any model to explain the variance in the activity; the second stage collates the information rich descriptor classes to select the individual descriptors significant to the activity; and the last stage reuses the selected individual descriptors to discover higher models and/or to explain the phenomenon in a comprehensive manner. In this process, the first stage has been developed based on the philosophy of elimination through selection. This has three iterations in it. It has been devised to address the multiple descriptor classes' environment in high-dimensional modeling studies. It operates by way of categorized treatment of descriptor classes. In this the contributing descriptor classes will be identified using simple models called 'baseline models'. Here, a baseline model represents any entry-level cross-validated regression equation with minimum variables (for example one- to three-descriptors) and capable of explaining at least 50% of variance in the dependent variable. This has been considered with the view that among multidescriptor models, the two- or at the most three-descriptor equations are the simplest to understand and explain the chosen phenomenon. Also, at this stage the level of importance of the descriptors to the phenomenon of study can be seen clearly. Moreover, the 'baseline model' concept helps in efficiently handling a large number of variables in each descriptor class and in identifying the information rich descriptors of all classes corresponding to the phenomenon. For the identification of the baseline models, the CP-MLR (combinatorial protocol in multiple linear regression)[23]—a 'filter' based variable selection procedure for model development—has been used in its simplest form with predefined filter thresholds. The procedural aspects of the CP-MLR and its implementation are available in some of the recent publications.[23−26] The thrust of this procedure is in its embedded 'filters'. They are briefly explained as follows: filter-1 limits interparameter correlation of the variables to a predefined cutoff value (default acceptable value ≤ 0.3); filter-2 controls the variables entry to a regression equation through *t*-values of coefficients (default acceptable value ≥ 2.0); filter-3 provides comparability of equations with a different number of variables in terms of square-root of adjusted multiple correlation coefficient of regression equation, r-bar (for 'baseline model' the minimum value is 0.71); filter-4 estimates the external consistency of the equation in terms of cross-validated R^2 or Q^2 with leave-one-out (LOO) cross-validation as a default option (default acceptable limits are $0.3 \leq Q^2 \leq 1.0$). In the first iteration the data files corresponding to each individual descriptor class will be evaluated separately for their ability to form a baseline model and accordingly they will be classified as the primary contributors (category I) and the residual descriptor classes (leftover group). The second iteration is meant for identifying the collective information content of the leftover descriptor classes' vis-à-vis the activity under study. In this, the leftover descriptor classes of the first iteration will be merged and recycled for their 'collective' influence in evolving the baseline models. Accordingly, the residual descriptor classes will be classified as the 'collective' contributors (category II) and the leftover of second iteration. At the end of second iteration, if no descriptor class is selected under the categories II and I, the process will be terminated to redefine the filters' threshold in CP-MLR for new baseline models to facilitate the capture of the descriptor class. However, if no descriptor class is selected under the category I alone, then the leftovers, if any, of the second iteration will be excluded from the study by treating them as noncontributing descriptor classes and the process will be continued with the second stage of the flowchart.

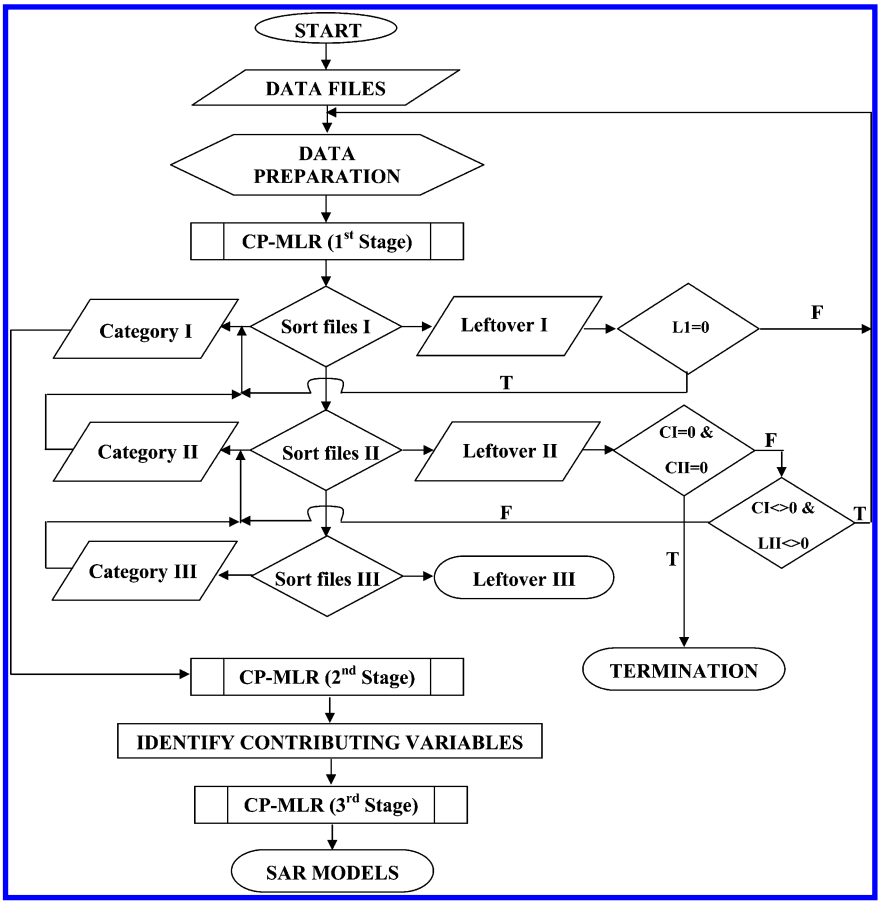

Figure 2. Procedure of the model(s) development strategy used in the study. It is embedded with Combinatorial Protocol in Multiple Linear Regression (CP-MLR) and shows the progress of selection of descriptors classes into categories I, II, and III (CI, CII, CIII) and leftovers I, II, and III (LI, LII, LIII) (first stage), individual descriptors (second stage), and final structure−activity models (third stage). In this 'T' stands for 'true' and 'F' stands for 'false'. In each stage the CP-MLR has been used for distinct function namely categorizing descriptors classes, sieving contributing descriptors from the identified descriptor categories, and finally identifying higher models and descriptors involved therein.

Otherwise, the leftovers of the second iteration will be carried forward to the next generation iteration to examine their possibility of making a 'secondary' contribution in association with the primary descriptor classes (category I). Thus, the third iteration will classify the corresponding descriptor classes as the 'secondary' contributors (category III) and the noncontributing descriptor classes, which will be excluded from the study from this point onward. On identifying the contributing descriptor classes in the form of categories I, II, and/or III, the process will continue in the second stage with the collated descriptor classes to create all possible baseline models that could possibly emerge from them. These models give out the individual contributing descriptors across the categories. The identified individual descriptors will be recycled in the last stage for higher models and comprehensive diagnosis of the phenomenon. All the models identified in the last stage have been further put to a randomization test[24,29] by repeated randomization of the activity to discover the chance correlations, if any, associated with them. For this every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted to express the percent chance correlation of the model under scrutiny. This flowchart (Figure 2) has been implemented with the descriptor class listed in Table 3 to model the antimalarial activity of 4-(3′,5′-disubstituted anilino)quinoline analogues (Tables 1 and 2). For this, the default filter thresholds and one- and two-descriptor combinations have been considered to explore the baseline models. The model development procedure has been further verified by creating divergent training and test sets from the compounds of the study.

RESULTS AND DISCUSSION

An early QSAR study of tebuquine (compounds related to amodiaquine) and other similar analogues suggested a correlation of their antimalarial activity with the decreasing size and electron donation property of the 4-anilino substituents of these molecules.[1] In a subsequent investigation of the antimalarial activity of 7-substituted-4-aminoquinolines, De and co-workers have observed that an iodo or a bromo group as the choice of substituents for the 7-position of the quinoline moiety.[4] This deviation from the more common choice of 7-chloro was attributed to the differences in the source of malarial strains/parasites. They also suggested a separation of either 2−3 or 10−12 carbon linkers between the two nitrogens of the diaminoalkane of 4-aminoquinoline for the optimum activity.[4] However, O'Neill and co-workers study had put emphasis on cellular accumulation of the drug for the activity. Their investigations have suggested no obvious relationship between the internitrogen separation and the antimalarial activity in 4-aminoquinolines.[3] All these studies collectively imply that the antimalarial activity of these compounds is due to the composite influence of all

the observed phenomena. Keeping these in view, we attempted to analyze the structure−activity relations of the 4-(3′,5′-disubstituted anilino)quinolines (Tables 1 and 2) to highlight the characteristics of 4-anilino substituents for the antimalarial activity. These compounds (Tables 1 and 2) have dissimilar substituent groups R_1 and R_2 at the meta (3′- and 5′-) positions of the 4-anilino moiety (Figure 1). In view of this dissimilarity in R_1 and R_2 groups, it is considered appropriate to analyze them separately to understand the structural and functional requirements of respective domains. It may be mentioned here that the compounds 9, 19, 20, and 21 of Table 1 are associated with uncertain activity values (active at doses >1 μM or <6.00 in terms of $-\log IC_{50}$).[15] Hence, they are not included in the data set for the model development. However, as the order of the activity of these compounds is known, they have been used as one part of verification of the predictions. Furthermore, for the compounds of Table 2, our preliminary attempts with this divergent descriptor base have shown that the calculated (also the predicted) activity of compound 37 deviates to a very large extent from its experimental value in a wide cross-section of different models. Interestingly, compound 37 is the only one in this set wherein the two nitrogens of the substituent group (Table 2, substituent variation under 'X') are separated by a three carbon linker. The huge loss of activity in compound 37 when compared to its lower homologue (compound 36) clearly indicates the significance of the spacing between the outermost nitrogen and its preceding one in the expression of the activity. However, in case of molecular descriptors, in normal course the difference in their successive values will be limited to the increment (or decrement) of the structural fragment. Under these circumstances compound 37 has been excluded from the data set of the compounds of Table 2. This makes the total number of compounds under Table 2 as 42. With these data sets, the first stage of the strategy (Figure 2) has identified eight descriptor classes under different categories (Table 3) as contributing ones for modeling the antimalarial activity of 4-(3′,5′-disubstituted anilino)quinoline analogues (Tables 1 and 2). Among these, the descriptors from Constitutional (CONST) and Topological (TOPO) classes have been found to be the primary contributors (category I) for modeling the antimalarial activity of the compounds (Tables 1 and 2). In addition to these two classes, the Atom Centered Fragments (ACF) class has also become the part of the primary contributors for modeling the activity of compounds listed in Table 2. All other descriptor classes, except for the Empirical (EMP), have shown significance either as 'collective' contributors (category II) or as the 'secondary' contributors (category III) in explaining the activity (Table 3). From the collated descriptor classes of categories I to III, the second stage of the flowchart (Figure 2) has identified 104 and 113 descriptors across the classes for modeling the activity of the compounds of Tables 1 and 2, respectively. At this level, some of these descriptor-combinations resulted in two variable models with correlation coefficients of the order of 0.83 (Table 1) and 0.86 (Table 2). In the last stage of the strategy (Figure 2) these correlation coefficients have been used as pointers for searching the higher dimension models among the 104 (correspond to Table 1) and 113 (correspond to Table 2) descriptors. Since filter-1 in CP-MLR controls the association of variables, setting a higher value for this filter facilitates the free mixing of descriptors. Accordingly for three-parameter models of the compounds of Table 1, the filters 1 and 3 of CP-MLR have been redefined as 0.79 and 0.83, respectively. This has resulted in 84 three-parameter models with a maximum correlation coefficient of 0.89. Similarly, for the four-parameter models of these compounds, the filter-3 has been redefined as 0.89. This has resulted in 70 four-parameter models with a maximum correlation coefficient of 0.943. From the 104 descriptors identified for the compounds of Table 1, these three- and four-parameter models have jointly shared 70 descriptors among themselves. All these 70 descriptors are shown in Table 4 along with their brief meaning, average regression coefficients, and total incidence corresponding to three- and four-parameter models which will serve as a measure of their estimate across the models. A more elaboration of all descriptors identified in this study are available in ref 19 and cross-references provided therein. Of these 70 descriptors 52 belong to category I and 18 belong to category II. To maintain brevity, the complete regression equations have been shown for selected models with three- and four-parameters only (eqs 1 and 2).

$$-\log IC_{50} = 12.463 - 0.284(0.091)\mathrm{AMW} - 4.327(1.175)\mathrm{BIC4} - 1.451(0.466)\mathrm{GGI4}$$

$$n = 18,\ r = 0.876,\ Q^2 = 0.689,\ s = 0.274,\ F = 15.34 \quad (1)$$

$$-\log IC_{50} = 7.584 + 0.012(0.004)\mathrm{ZM2V} + 0.775(0.134)\mathrm{CIC4} - 4.424(0.830)\mathrm{SEige} - 0.061(0.013)\mathrm{MPC09}$$

$$n = 18,\ r = 0.943,\ Q^2 = 0.814,\ s = 0.197,\ F = 25.92 \quad (2)$$

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 the 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. Also, in the randomization study (100 simulations per model) none of the identified models has shown any chance correlation. The above models have been further validated through four test sets, each containing four compounds out of the 18 active ones listed in Table 1 (Table 5). Of the four test sets, three were constructed from the cluster analysis of the 18 compounds using the bit-packed version of the MACCS fingerprints (FP BIT MACCS),[30] the total descriptors of the study and the activity. The cluster analysis of the MACCS fingerprints has been carried out in MOE using the standard procedure described therein.[30] As these compounds are closely related, for the MACCS fingerprints, an 85% similarity level has been used to cluster them into different groups. The remaining two cluster patterns of these compounds have been generated in the SYSTAT[31] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the respective descriptors or the activity as the case may be. The selection of the test set from the cluster tree has been done in such a way to keep the test compounds at a

Table 4. Descriptors Identified from the Third Stage of the Flowchart (Figure 2) for Modeling the Antimalarial Activity of 7-Chloro-4-(3′,5′-disubstituted anilino)quinolines (Table 1) along with the Average Regression Coefficients and the Total Incidence

descrip[a]	av reg coef (total incidence)[b]	descrip[a]	av reg coef (total incidence)[b]	descrip[a]	av reg coef (total incidence)[b]
		CONST			
AMW	−0.266(8)	Ss	0.055(2)	nBO	0.071(2)
		TOPO			
ZM1	0.016(1)	IC1	1.782(5)	SEigv	0.569(5)
ZMIV	0.014(6)	SIC2	−4.862(9)	SEige	−3.280(86)
ZM2	0.015(1)	BIC2	−4.958(20)	VRD2	0.357(2)
ZM2V	0.012(13)	SIC3	−3.301(3)	VRZ2	0.362(2)
Ram	−0.247(3)	CIC3	0.858(12)	VRm2	0.362(2)
RDSUM	0.021(1)	BIC3	−4.248(8)	VRv2	0.355(2)
Har	0.046(2)	IC4	0.373(1)	VRe2	0.362(2)
RhyDp	0.072(1)	SIC4	−4.081(7)	VRp2	0.354(2)
Wap	−0.0002(1)	CIC4	0.864(19)	NGS	0.056(1)
DELS	−0.064(2)	BIC4	−4.110(46)	MPC05	−0.033(1)
X3	0.250(2)	IC5	0.388(1)	MPC06	−0.026(4)
X4	0.353(3)	SIC5	−4.288(3)	MPC07	−0.027(8)
X5A	9.842(2)	CIC5	0.899(9)	MPC08	−0.040(21)
X1sol	0.131(1)	BIC5	4.488(16)	MPC09	−0.046(64)
X3sol	0.222(1)	SEigZ	−0.794(2)	TPC	−0.002(1)
X4sol	0.299(1)	SEigm	−0.784(2)	T(N..O)	−0.052(40)
X5sol	−0.338(2)				
		MWC			
MWC01	0.071(2)	MWC04	0.084(1)	SRW02	0.036(2)
MWC02	0.031(1)	MWC08	−3.411(11)	SRW04	0.010(1)
MWC03	0.045(1)	TWC	0.012(1)		
		GVZ			
GGI2	−0.750(9)	GGI4	−1.320(8)	GGI6	−2.441(1)
		2DAUTO			
MATS1m	−11.024(6)	GATS8v	−0.510(5)	GATS8p	−0.595(3)
GATS2v	1.735(10)	GATS2p	1.411(3)		
		ACF			
H-047	−0.074(3)	H-052	0.075(3)		

[a] Descriptors identified are from the three- and four-parameter models emerged from CP-MLR with filter-1 as 0.79, filter-2 as 2.0, filter-4 as $0.3 \leq Q^2 \leq 1.0$, filter-3 as 0.83 (3-descriptor model), and 0.89 (4-descriptor model); number of compounds in the study are 18; CONST: AMW average molecular weight, Ss sum of Kier-Hall electrotopological states, nBO number of non-H bonds; TOPO: ZMn and ZMnV for n Zagreb index and n Zagreb index by valence vertex degrees; Ram ramification index; RDSUM, reciprocal distance Wiener-type index; Har, Harary H index; RhyDp, reciprocal hyperdistance-path index; Wap, all-path Wiener index; DELS, molecular electrotopological variation; X*k*, X*k*A, X*k*Sol, chi-index of order *k*, A is average, Sol is salvation; IC*k*, BIC*k*, SIC*k*, CIC*k*, information content index of order *k*, in this B is bond, S is structural, C is complementary, SEig*w*, eigenvalue sum from *w* weighted distance matrix; VR*w*2, average Randic-type eigenvector-based index from *w* weighted distance matrix; NGS, Gordon-Scantlebury index; MPC*k*, molecular path count of order *k*; TPC, total path count; T(N..O), sum of topological distances between N..O; MWC: MWC*k*, molecular walk counts of order *k;* TWC, total walk count; SRW*k*, self-returning walk count of order *k*; GVZ: GGI*k*, topological charge index of order *k*; 2DAUTO: MATS1m, Moran autocorrelation − lag 1/weighted by atomic masses; GATS*kw*, Geary autocorrelation − lag *k*/weighted by property *w*; ACF: H-047, H attached to C1(sp3)/C0(sp2); H-052, H attached to C0(sp3) with 1X attached to next C (weighing term *w* may be D − distance matrix, Z − matrix, m − mass, v − van der Waals volume, e − electronegativity, p − polarizability); also see ref 19. [b] The average regression coefficient of the descriptor corresponding to all models and the total number of its incidence; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

maximum possible distance from each other. The fourth test set of the compounds corresponds to the random selection procedure. Additionally, all four compounds with uncertain activity (compounds 9, 19, 20, and 21) have also been added

Table 5. Predicted Residual Activity of Different Test Sets (4 Compounds Each) of the Compounds of Table 1 and the Resultant Predictive R^2 Values

	residual[a]							
	MACCS[b]		Des Clus[c]		Act Clus[d]		Rand[e]	
compd no.	M1[f]	M2[g]	D1[f]	D2[g]	A1[f]	A2[g]	Ra1[f]	Ra2[g]
2			0.35	0.00				
3							−0.34	−0.24
4					−0.16	0.07		
5	−0.25	−0.04						
6					−0.09	0.29		
7							0.53	0.39
9[h]	6.71	6.12	6.69	6.09	6.68	6.05	6.64	6.01
10			−0.32	−0.07				
12	−0.10	−0.16					0.01	−0.06
13	−0.07	−0.27						
14					−0.02	0.21		
16							0.12	0.29
17			0.11	−0.1	0.07	−0.01		
19[h]	5.84	5.98	5.92	5.89	5.72	5.85	5.66	5.82
20[h]	6.63	5.87	6.62	5.79	6.54	5.75	6.47	5.71
21[h]	6.72	7.67	6.74	7.68	6.67	7.76	6.61	7.69
22	−0.22	−0.15	−0.22	−0.05				
pred R^2	0.80	0.81	0.58	0.97	0.97	0.91	0.59	0.71

[a] The difference of observed $-\log IC_{50}$ and predicted $-\log IC_{50}$; the training models have 14 compounds each. [b] Test set from the cluster analysis of the MACCS fingerprint codes of the compounds. [c] Test set from the cluster analysis of full descriptor data set of the compounds. [d] Test set from the cluster analysis of the activity of the compounds. [e] Test set from random selection of the compounds. [f] M, D, A, and Ra respectively represent MACCS, Descriptor Cluster, Activity Cluster, and Random selections; the training set equations are formed with the descriptors of eq 1. [g] The training set equations are formed with the descriptors of eq 2. [h] Predicted activity against the reported activity of <6.00.

to each one of the test sets. With this, these test sets represent different cross-sections of compounds. The predictions of the test sets have been done with the models developed using the 14 compounds remaining in the training sets (all training set equations corresponding to Table 1 as well as Table 2 have been given in the Supporting Information). The residuals of the predictions and the corresponding predictive R^2 value of each test set have been shown in Table 5. The predictions corresponding to all the test sets compounds are within the reasonable limits of their actual values (Table 5). Also, all the models from the training sets have predicted the three compounds out of the four (compounds 9, 19, 20, and 21; Table 1) with uncertain activity values as low active ones (Table 5). The exception is with compound 21 where the part of the varying group is a phenyl moiety (Table 1). Especially, for these compounds the predictions from the four-parameter equations corresponding to eq 2 are more satisfactory compared to those from the three-parameter equations corresponding to eq 1 (Tables 1 and 5). The overall satisfactory performance of all the test sets gave us the confidence about the procedure adopted in the study and the variables selected therein. Now in terms of physical meaning of the descriptors, in eq 1, AMW (average molecular weight), BIC4 (bond information content of order 4), and GGI4 (Galvez topological charge index of order 4) respectively belong to CONST, TOPO, and GVZ classes. And all the descriptors of eq 2 are from the TOPO class. The negative regression coefficient of AMW sets restrictions on the R_1 group of these compounds (Table 1). The descriptors BCI4 (eq 1) and CIC4 (complementary information content of order 4) (eq 2) are closely related. While BIC4 represents

Table 6. Descriptors Identified from the Third Stage of the Flowchart (Figure 2) for Modeling the Antimalarial Activity of 7-Chloro-4-(3′,5′-disubstituted anilino)quinolines (Table 2) along with the Average Regression Coefficients and the Total Incidence

descriptor[a]	av reg coef (total incidence)[b]	descriptor[a]	av reg coef (total incidence)[b]
	CONST		
nDB	−0.682(68)		
	TOPO		
HNar	−0.614(3)	X3Av	2.187(10)
X2A	1.684(12)	PHI	−0.153(68)
X3A	2.133(13)	SIC1	1.864(4)
X2Av	2.115(15)		
	BCUT		
BEHm5	0.326(7)	BEHe5	0.414(9)
BELm5	0.645(10)	BEHp5	0.358(13)
BEHv5	0.336(10)	BELp5	0.645(9)
BELv5	0.625(10)		
	GVZ		
JGI1	0.833(1)		
	2DAUTO		
MATS4m	0.682(2)	MATS3e	−0.652(3)
MATS5m	0.660(3)	GATS3v	0.305(1)
	ACF		
C-002	0.069(1)	H-051	0.239(28)

[a] Descriptors identified are from the three-, four-, and five-parameter models emerged from CP-MLR with filter-1 as 0.79, filter-2 as 2.0, filter-4 as $0.3 \leq Q^2 \leq 1.0$, filter-3 as 0.86 (3-descriptor model), 0.87 (4-descriptor model), and 0.89 (5-descriptor model); number of compounds in the study are 42; CONST: nDB, double bond in the form of carbonyls; TOPO: HNar, Narumi simple topological index; X*k*A, X*k*Av, average chi-index of order *k*, v is valence; PHI, Kier flexibility index; SIC1, structural information content of 1-order; BCUT: BEH*wk*, Burden highest eigenvalues of rank *k* and atomic property *w*; BEL*wk*, Burden lowes eigenvalues of rank *k* and atomic property *w*; GVZ: JGI1, mean topological charge index of order1; 2DAUTO: MATS*kw*, Moran autocorrelation − lag *k*/weighted by property *w*; GATS*kw*, Geary autocorrelation − lag *k*/weighted by property *w*; ACF: C-002, count of R−CH2−R′; H-051, count of H on alpha C to C−X; (weighing term *w* may bem − mass, v − van der Waals volume, e − electronegativity, p − polarizability) also see ref 19. [b] See footnote b of Table 4.

the measures of number of bonds and their multiplicity in the chosen structural fragment, CIC4 measures the deviation of the information content from its maximum value. Interestingly, the Galvez topological charge index of order 4 (GGI4) which corresponds to the fourth eigenvalue of the corrected adjacency matrix of the R_1 group has shown its presence in the model equations. ZM2V (second Zagreb index by valence vertex degrees) in eq 2 represents the sum over all bonds of the product of the valence vertex degree of the two atoms incident to the considered bond. This and other Zagreb indices collectively indicate the preference for aromatic ring substituted with F or NO_2 groups and nitrogen containing aromatic and heterocyclic systems for better activity. The other descriptors of interest in eq 2 are MPC09 (molecular path count of order 9) and SEige (eigenvalue sum from electronegativity weighted distance matrix). The coefficients of these descriptors suggest that a smaller number of ninth-order path counts and smaller eigenvalue sums would lead to better activity. All these descriptors of eqs 1 and 2 are associated with high frequency and have shown significance in many models (Table 4). T(N..O) is another descriptor which took part in as many as 40 models. This represents the sum of topological distances between N and O in the chosen structural fragment. The negative regression coefficient of this descriptor suggests that a closer topological path between N and O in the structural frame would be better for the activity.

Similarly, for exploring the higher dimensional models from the 113 descriptors (identified at the end of second stage of the flowchart i.e., Figure 2) of compounds listed in Table 2 the filter-1 of CP-MLR has been kept as 0.79 and the filter-3 has been successively incremented to the highest correlation coefficient of preceding descriptor level. Since the number of compounds (Table 2) is more here, the higher dimensional models have been explored up to five parameters. In this, the successive filter-3 values for three-, four-, and five-descriptor models have been kept, respectively, as 0.86, 0.87 and 0.89. This has yielded 1 three-descriptor model, 4 four-descriptor models, and 29 five-descriptor models. All these models collectively shared 22 descriptors among themselves (Table 6). Of these 22 descriptors, 10 belong to category I, 11 belong to category II, and 1 belongs to category III. The following are the selected structure−activity models of these compounds (eqs 3−5).

$$-\mathrm{LogIC}_{50} = 7.430 - 0.626(0.103)\mathrm{nDB} + 0.373(0.078)\mathrm{BEHe5} - 0.105(0.034)\mathrm{PHI}$$

$$n = 42,\ r = 0.873,\ Q^2 = 0.705,\ s = 0.241,\ F = 40.60 \quad (3)$$

$$-\mathrm{LogIC}_{50} = 7.368 - 0.589(0.098)\mathrm{nDB} + 2.487(0.623)\mathrm{X3A} - 0.142(0.036)\mathrm{PHI} + 0.617(0.145)\mathrm{BELv5}$$

$$n = 42,\ r = 0.894,\ Q^2 = 0.744,\ s = 0.224,\ F = 36.91 \quad (4)$$

$$-\mathrm{LogIC}_{50} = 7.519 - 0.780(0.118)\mathrm{nDB} + 2.282(0.587)\mathrm{X3A} - 0.202(0.041)\mathrm{PHI} + 0.698(0.139)\mathrm{BELv5} + 0.219(0.086)\mathrm{H\text{-}051}$$

$$n = 42,\ r = 0.911,\ Q^2 = 0.764,\ s = 0.209,\ F = 35.24 \quad (5)$$

For these compounds also four test sets, originating respectively from the MACCS fingerprints, the descriptors of the study, the activity, and the random selection, have been constructed using the same approach as explained in the case of the compounds of Table 1. In this case, for each test set 12 compounds have been selected leaving the remaining 30 compounds in the training set. All the training set models with the descriptors of eqs 3−5 have been generated for the evaluation of the test sets. Table 7 shows the residuals of the test sets compounds' predictions from the corresponding training sets and the associated predictive R^2 values. All the predictions of the test sets compounds are within the satisfactory limits of their actual values and hence offer confidence to the models. Here, the two descriptors, nDB and PHI, are common to all the models of the compounds of Table 2. Of these, nDB is from the CONST class and PHI is from the TOPO class. Here, nDB accounted for the double bond in the form of the carbonyl function in the R_2 group of these compounds. This alone has correlated with activity to the tune of 0.787 ($r = 0.787$). It suggests

Table 7. Predicted Residual Activity of Different Test Sets (12 Compounds Each) of the Compounds of Table 2 and the Resultant Predictive R^2 Values

	residual[a]											
	MACCS[b]			Des Clus[c]			Act Clus[d]			Rand[e]		
compd no.	M3[f]	M4[g]	M5[h]	D3[f]	D4[g]	D5[h]	A3[f]	A4[g]	A5[h]	Ra3[f]	Ra4[g]	Ra5[h]
7				−0.15	0.18	−0.03				−0.08	−0.28	−0.14
24										−0.32	−0.42	−0.23
26	0.36	0.35	0.27				0.24	0.24	0.23			
27							−0.41	−0.37	−0.31			
28	0.38	0.30	0.13							0.32	0.18	0.11
30	−0.09	−0.11	−0.23	−0.19	−0.22	−0.32						
31	−0.22	−0.27	−0.26	−0.35	−0.41	−0.37	−0.34	−0.38	−0.32	−0.26	−0.31	−0.27
32				−0.24	−0.32	−0.34						
33							−0.35	−0.29	−0.15			
34	0.12	0.22	0.35									
36	−0.06	−0.17	−0.07				−0.03	−0.16	−0.07	−0.05	−0.13	−0.07
39				−0.10	0.02	0.03	−0.05	0.06	0.09			
40										0.11	0.13	0.15
42				0.14	−0.22	−0.13	0.21	−0.12	−0.04			
44	0.01	−0.04	−0.08	−0.02	−0.07	−0.13				−0.01	−0.03	−0.06
45							0.23	0.13	0.10			
48	0.16	0.14	0.10									
49	−0.12	−0.12	−0.11				−0.10	−0.10	−0.09	−0.13	−0.10	−0.10
50							0.09	0.17	0.15			
52				−0.25	−0.13	−0.10						
53				−0.22	−0.15	−0.13						
54	−0.19	−0.17	−0.15									
55	−0.28	−0.25	−0.24	−0.30	−0.24	−0.22	−0.27	−0.24	−0.23	−0.29	−0.25	−0.23
58										−0.07	−0.03	−0.05
59				−0.31	−0.32	−0.38						
61										−0.11	0.03	0.02
62							0.32	0.40	0.42			
63	0.21	0.26	0.26	0.19	0.25	0.25						
64										0.12	0.21	0.20
pred R^2	0.72	0.71	0.74	0.78	0.76	0.76	0.76	0.77	0.83	0.81	0.76	0.87

[a−e] The training models have 30 compounds each; see footnotes *a*−*e* of Table 5. [f] M, D, A, and Ra respectively represent MACCS, Descriptor Cluster, Activity Cluster, and Random selections; the training set equations are formed with the descriptors of eq 3. [g] The training set equations are formed with the descriptors of eq 4. [h] The training set equations are formed with the descriptors of eq 5.

that the carbonyl function in the R_2 group has a negative influence on activity. The descriptor PHI (Kier flexibility index) is derived from the Kier alpha-modified shape indices. It suggests that unsaturation, loops and branches in R_2 group would lead to better activity. Also, the second- and third-order chi (χ) descriptors in terms of average connectivity (X2A and X3A) and average valence connectivity (X2Av and X3Av) have participated in several models. These descriptors also indicate that branching in R_2 would be better for the activity. The other TOPO class descriptors that have taken part in models are HNar and SIC1. Here, HNar is a Narumi harmonic topological index related to molecular branching and represents the number of non-hydrogen atoms divided by the reciprocal vertex degree sum and SIC1 is the structural information content (neighborhood symmetry of order 1). H-051 in eq 5 is an ACF class descriptor, and it represents the hydrogens attached to the singly bonded carbons alpha to functions such as C−X where C and X are connected by double or triple bonds. The positive regression coefficient of this descriptor indicates the necessity of such unsubstituted alpha carbons for better activity. Another interesting ACF class descriptor that took part in the models is C-002. This accounts for the number of carbons with two hydrogens and the remaining two valences satisfied by other functional groups, that is R−CH_2−R′ where R and R′ may be the same or different. This descriptor has entered in only one model with a positive regression coefficient. Several of the fifth highest (H) as well as the lowest (L) eigenvalues of the Burden matrix corresponding to atomic masses (m), van der Waals volumes (v), Sanderson electronegativities (e), and polarizabilities (p) (BEHm5, BEHv5, BEHe5, BEHp5, BELm5, BELv5, BELp5) have taken part in the model formation. These BCUT (category II) descriptors indicate the role of the corresponding property measures of the R_2 group in the activity. The 2DAUTO class descriptors MATS*kw* and GATS*kw* have also participated in modeling the activity of these compounds. In this, MATS is Moran Autocorrelation of Topological Structure and GATS is Geary Autocorrelation of Topological Structure. In these descriptors *k* indicates the path length (*lag*) in the graph, and the *w* corresponds to atomic property used to weight the graph. Here graph lengths 3, 4, and 5 in association with atomic properties namely masses (m), van der Waals volumes (v), and Sanderson electronegativities (e) have shown significance in correlating the activity. The regression coefficients of these descriptors indicate that the higher path lengths rich in mass, volume, and electronic content would be favorable for the activity. The GVZ class descriptor JGI1 represents the mean topological charge index of path of length one. It is the only category III descriptor that took part in a model.

CONCLUDING REMARKS

The two series of compounds considered in the study shared common descriptor classes in modeling the activity.

However, the individual descriptors participated in the models of each set of compounds were different from one another. This clearly suggests that the R_1 and R_2 groups of these 4-(3′,5′-disubstituted anilino)quinolines (Figure 1) map different domains in the activity space. For the models involving the R_1 group (Table 1) MPC09 (molecular path count of order 9), BIC4 (Bond Information Content of order 4), SEige (eigenvalue sum from electronegativity weighted distance matrix), and T(N..O) (sum of topological distances between N..O) are some of the highly participated descriptors. Models involving these descriptors in association with others suggest that a compact structural framework having aromatic, heterocyclic ring(s) substituted with closely spaced F, NO_2, and O functional groups would be favorable for the activity. The two descriptors, nDB (double bond in the form of carbonyls) and PHI (Kier flexibility index), are common to all the models corresponding to the R_2 group of the compounds (Table 2). Here, nDB has accounted for the double bond in the C═O of the carboxyl's carbonyl portion of the R_2 group of these compounds. The regression coefficient associated with this descriptor suggests that C═O in R_2 has a negative influence on activity. As the esters move the system toward a neutral environment, this gives a clear indication in favor of the electron rich centers for better antimalarial activity, also an observation drawn in several of the previous reports.[3,5−7] The descriptor PHI suggests that unsaturation, loops and branches in R_2 group would lead to better activity. Apart from these, the neighborhood of functional groups should be free from substitutions for better activity. Since compound 37 (with nitrogens separated by a propylene linker) is excluded from the study, the models developed here may not hold well for this type of compounds. Another exception may be compound 21 (a phenyl in place of NRR′ of R_1) type analogues. All the test sets (from the MACCS fingerprints, the descriptors of the study, the activity, and the random selection procedures) have been predicted well by their corresponding training set equations. The models obtained and the participating descriptors suggest that the substituent groups of the 4-anilino moiety of the 4-(3′,5′-disubstituted anilino)quinolines hold scope for further modification. The flowchart (Figure 2) used in this study has efficiently treated the data set and identified the information rich descriptors corresponding to the activity.

Supporting Information Available: The complete data set of independent descriptors and training set regression equations. This material is available free of charge via the Internet at htpp://pubs.acs.org

REFERENCES AND NOTES

(1) Werbel, L. M.; Cook, P. D.; Elslager, E. F.; Hung, J. H.; Johnson, J. L.; Kesten, S. J.; McNamara, D. J.; Ortwine, D. F.; Worth, D. F. Synthesis, Antimalarial Activity, and Quantitative Structure−Activity Relationships of Tebuquine and a Series of Related 5-[(7-Chloro-4-quinolinyl)amino]-3-[(alky1amino)methyl][1, l′-biphenyl]-2-ols and $N^{(\omega)}$-Oxides. *J. Med. Chem.* **1986**, *29*, 924−939 and previous reports of the series.

(2) Koh, H. L.; Go, M. L.; Ngiam, T. L.; Mak, J. W. Conformational and Structural Features Determining in vitro Antimalarial Activity in some Indolo[3,2-*c*]quinolines, Anilinoquinolines and Tetrahydroindolo[3,2-*d*]benzazepines. *Eur. J. Med. Chem.* **1994**, *29*, 107−113.

(3) O'Neill, P. M.; Willock, D. J.; Hawley, S. R.; Bray, P. G.; Storr, R. C.; Ward, S. A.; Park, B. K. Synthesis, Antimalarial Activity, and Molecular Modeling of Tebuquine Analogues. *J. Med. Chem.* **1997**, *40*, 437−448.

(4) De, D.; Krogstad, F. M.; Byers, L. D.; Krogstad, D. J. Structure−Activity Relationships for Antiplasmodial Activity among 7-Substituted 4-Aminoquinolines. *J. Med. Chem.* **1998**, *41*, 4918−4926.

(5) Hawley, S. R.; Bray, P. G.; Mungthin, M.; Atkinson, J. D.; O'Neill, P. M.; Ward, S. A. Relationship between Antimalarial Drug Activity, Accumulation, and Inhibition of Heme Polymerization in *Plasmodium falciparum* In Vitro. *Antimicrob. Agents Chemother.* **1998**, *42*, 682−686.

(6) O'Neill, P. M.; Bray, P. G.; Hawley, S. R.; Ward, S. A.; Park, B. K. Aminoquinolines-Past, Present, and Future: A Chemical Perspective. *Pharmacol. Ther.* **1998**, *77*, 29−58.

(7) Egan, T. J. Structure−Function Relationships in Chloroquine and Related 4-Aminoquinoline Antimalarials. *Mini Rev. Med. Chem.* **2001**, *1*, 113−123.

(8) Rieckmann, K. Determination of the Drug Sensitivity of *Plasmodium falciparum*. *JAMA J. Am. Med. Assoc.* **1971**, *217*, 573−578.

(9) Chou, A. C.; Fitch, C. D. Control of Heme Polymerase by Chloroquine and other Quinoline Derivatives. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 422−427.

(10) Slater, A. F. G.; Chloroquine: Mechanism of Drug Action and Resistance in *Plasmodium falciparum*. *Pharmacol. Ther.* **1993**, *57*, 203−235.

(11) Hatton, C. S. R.; Bunch, C.; Peto, T. E. A.; Pasvol, G.; Russell, S. J.; Singer, C. R. J.; Edwards, G.; Winstanley, P. Frequency of Severe Neutropenia associated with Amodiaquine Prophylaxis against Malaria. *Lancet* **1986**, 411−414.

(12) Neftel, K. A.; Woodtly, W.; Schmid, M.; Frick, P. G.; Fehr, J. Amodiaquine Induces Agranulocytosis and Liver Damage. *Br. Med. J.* **1986**, *292*, 721−723.

(13) Maggs, J. L.; Kitteringham, N. R.; Park, B. K. Drug−Protein Conjugates. Mechanism of Formation of Protein Arylating Intermediates from Amodiaquine in Man. *Biochem. Pharmacol.* **1988**, *37*, 303−311.

(14) Harrison, A. C.; Kitteringham, N. R.; Clarke, J. B.; Park, B. K. Mechanism of Bioactivation and Antigene Formation of Amodiaquine in the Rat. *Biochem. Pharmacol.* **1992**, *43*, 1421−1430.

(15) Delarue, S.; Girault, S.; Maes, L.; Debreu-Fontaine, M.; Labaeid, M.; Grellier, P.; Sergheraert, C. Synthesis and in Vitro and in Vivo Antimalarial Activity of New 4-Aminoquinolines. *J. Med. Chem.* **2001**, *44*, 2827−2833.

(16) Basak, S. C.; Harriss, D. K.; Magnuson, V. R. POLLY, University of Minnesota: Duluth, MN, 1988.

(17) Molconn-Z, ver. 2.07, eduSoft Lc, a Virginia Corporation, Ashland, VA 23005 U.S.A. www.edusoft-lc.com.

(18) (a) Katritzky, A. R.; Lobnov, V.; Karelson, M. *CODESSA* (*Comprehensive Descriptors for Structural and Statistical Analysis*); University of Florida: Gainesville, FL, 1994. (b) Katritzky, A. R.; Perumal, S.; Petrukhin, R.; Kleinpeter, E. CODESSA-Based Theoretical QSPR Model for Hydantoin HPLC-RT Lipophilicities. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 569−574.

(19) DRAGON software version 3.0-2003. By Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Milano, Italy. http://disat.unimib.it/chm/Dragon.htm.

(20) Gonzalez, M. P.; Helguera, A. M. TOPS-MODE verces DRAGON Descriptors to Predict Permeability Coefficients through Low-Density Polymer. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 665−672.

(21) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure Activity Relationship and Quantitative Structure Property Relationship. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.

(22) Kubinyi, H. Variable Selection in QSAR Studies. I. An Evolutionary Algorithm. *Quant. Struct.-Act. Relat.* **1994**, *13*, 285−294.

(23) Prabhakar, Y. S. 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.

(24) Prabhakar, Y. S.; Solomon, V. R.; Rawal, R. K.; Gupta, M. K.; Katti, S. B. 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. Sci.* **2004**, *23*, 234−244.

(25) Prabhakar, Y. S. 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.

(26) Gupta, M. K.; Sagar, R.; Shaw, A. K.; Prabhakar, Y. S. CP-MLR Directed QSAR Studies on the Antimycobacterial Activity of Functionalised Alkenols − Topological Descriptors in Modeling the Activity. *Bioorg. Med. Chem.* **2005**, *13*, 343−351.

(27) ChemDraw Ultra 6.0 and ChemBats3D Ultra, Cambridge Soft Corporation, Cambridge, U.S.A.

(28) The complete data set of independent and dependent descriptors will be provided on request.

(29) So, S. S.; Karplus, M. Three-Dimensional Quantitative Structure−Activity Relationship from Molecular Similarity Matrices and Genetic Neural Networks. 1. Method and Validations. *J. Med. Chem.* **1997**, *40*, 4347−4359.
(30) (a) MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1255 University Street, Suite 1600, Montreal, Quebec, Canada H3B 3X3. (b) Brown, R. D.; Martin, Y. C. 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.
(31) SYSTAT, Version 7.0: SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611.

CI0501140