Exploration of Rate-Limiting Conformational State for 5-[(7-Chloro-4-quinolinyl)amino]-3-[(alkylamino)methyl][1,1′-biphenyl]-2-ols and N^{ω}-Oxides (Tebuquine Analogues) for Antimalarial Activity Using Molecular Shape Analysis and Molecular Field Analysis Studies

Pooja Sharma,[†] Sandeep Chhabra,[†] Nitin Rai,[‡] and Nanda Ghoshal*,[†]

Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology (CSIR), Kolkata 700032, India, and Department of Pharmaceutics, Institute of Technology, Banaras Hindu University, Varanasi 221005, India

Received December 29, 2006

Tebuquine is a 4-aminoquinoline that shows significantly more potency as an antimalarial than amodiaquine and chloroquine both in vitro and in vivo. To explore the conformation in the rate-limiting step and to elucidate pharmacophoric properties of tebuquine-related analogues, molecular shape analysis (MSA) along with molecular field analysis (MFA) methods were applied on a series of 5-[(7-chloro-4-quinolinyl)amino]-3-[(alkylamino)methyl][1,1′-biphenyl]-2-ol analogues and their N^{ω}-oxides possessing antimalarial activity. The study was performed using 45 compounds in which 37 molecules were taken as a training set for the derivation of the 3D quantitative structure−activity relationship models and eight molecules were kept as a test set to evaluate the predictive ability of the derived models. Both methods were analyzed in terms of their predictive abilities and produced comparable results with good conventional and cross-validated r^2 values (0.908 and 0.886, respectively, for the MFA model and 0.846 and 0.812, respectively, for the MSA model). In external data set prediction, the MSA model scored much better than MFA. Steric, electrostatic, and hydrogen-bond donor/acceptor fields of molecules were found to be relevant descriptors for structure−activity relationships. The inclusion of polar solvent-accessible charged surface area and spatial descriptors in the MSA model generation resulted in a model with significant predictive ability for the test set molecules. This indicates the importance of the orientation of conformationally favored molecules inside the receptor site and solvation of the charged surfaces of the molecule by a polar solvent for the activity of the molecule. The results provided the appropriate tools for predicting the affinity of related compounds using a ligand-based approach, and for guiding the design and synthesis of novel and more potent antimalarial agents.

INTRODUCTION

Despite years of continuous effort, malaria is still one of the most deadly diseases affecting third-world countries, claiming more than 1 million lives annually, most of whom are children.[1] Today, approximately 40% of the world's population, mostly living in the world's poorest countries, is at risk for malaria. This disease is found throughout the tropical and subtropical regions of the world and causes more than 300 million acute illnesses and at least 1 million deaths annually.[2] The disease is caused by *sporozoa* of the genus *Plasmodium.* The parasite is transmitted through the bite of the female *Anopheles* mosquito. The entry points for the parasite are the phagocytic Kupfer cells of the liver. Although Kupfer cells normally digest and demolish anything foreign, the malaria parasite is able to alter this response, allowing it to spread to other liver cells and multiply.[3,4] Chloroquine, mefloquine, and other frontline drugs (8-aminoquinolines, quinolinemethanols, sulfonamides, and dihydrofolate reductase inhibitors) for the treatment and prevention of malaria are becoming increasingly ineffective.[5] The situation is rapidly worsening mainly due to the nonavailability of effective drugs and the development of drug resistance in areas where malaria is frequently transmitted.[6] This problem has necessitated a search for new chemical entities or modification of existing ones for enhanced antimalarial activity.

In light of previous studies, the mechanism of action of 4-aminoquinolines such as tebuquine is believed to involve complexation with heme in the food vacuole of the parasite.[7−10] Molecular modeling studies on the interaction of 4-aminoquinolines with the proposed cellular receptor heme revealed favorable interaction energies for chloroquine, amodiaquine, and tebuquine analogues. Tebuquine, the most potent antimalarial in the above series, had the most favorable interaction energy calculated in both the in vacuo and solvent-based simulation studies.[11] The lack of structural information about the binding mode of tebuquine and its analogs with its proposed biological target greatly hampers structure-based drug design. Therefore, we present a receptor-independent 3D quantitative structure−activity relationship (QSAR) study with partial use of 4D QSAR to explain the rate-limiting or probable conformation and the binding mode of 5-[(7-chloro-4-quinolinyl)amino]-3-[(alkylamino)methyl][1,1′-biphenyl]-2-ol (Figure 1) analogues and their N^{ω}-oxides.[12] The study

* Corresponding author phone: +91-33-2473-3491, ext. 254; fax: +91-33-2473-0284/5197; e-mail: nghoshal@iicb.res.in.
[†] Indian Institute of Chemical Biology (CSIR).
[‡] Banaras Hindu University.

10.1021/ci600570r CCC: $37.00
Published on Web 04/11/2007

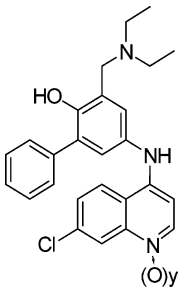

Figure 1. Structure of 5-[(7-chloro-4-quinolinyl)amino] and 3-[(alkylamino)methyl] [1,1′-biphenyl]-2-ol Y = 0, des-N-oxide; Y = 1, N-oxide

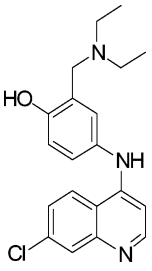

Figure 2. Amodiaquine.

has been carried out to evaluate the efficacy of receptor-independent methods to reach the results[11] obtained by the receptor-dependent simulation method. The approaches of molecular field analysis and molecular shape analysis were carried out in order to provide insight into the key structural features for the compounds, under study, obtained by the substitution of the phenyl ring at the 3′ position of amodiaquine (Figure 2). Molecular field analysis (MFA) methods attempt to postulate and represent the essential features of a receptor site from the aligned common features of the molecules in 3D space, whereas molecular shape analysis (MSA) provides information regarding the conformational variation of the molecules (partial use of 4D QSAR) with biological activity.[13] MFA is a 3D-QSAR approach that computes the steric, electrostatic, and hydrogen-bond donor/acceptor interactions of a given series of molecules, using probes within a regularly spaced grid. Appropriate statistical techniques are then applied to the above results to obtain a QSAR equation. This equation outlines the key features of the model, which explain variability in the biological activity, providing inputs for guiding and optimizing drug design efforts. MSA explores and analyzes multiple conformational states of each molecule under study and identifies the most accurate pose for bioactivity by optimizing the QSAR model. All these conformers are compared to the shape reference compound on the basis of shape descriptors and activity. If the QSAR equation shows good predictive ability for the test set molecules, then the conformers used for model generation will represent theoretical solutions of the conformations achieved during receptor binding. This can provide valuable information regarding the conformations present in the rate-limiting step of biological action.

MATERIALS AND METHODS

Biological Data. A data set of 45 molecules consisting of a series of 5-[(7-chloro-4-quinolinyl)amino]-3-[(alkylamino)methyl][1,1′-biphenyl]-2-ol analogues and their N^{ω}-oxides, synthesized by Werbel et al., was used for this study[12] (Table 1). The biological activities which were used as dependent variables in the MFA and MSA analyses were expressed as pED_{30} ($-\log$ molar ED_{30}), where ED_{30} represents the dose which increases the mean survival time of mice, infected by *Plasmodium berghei*, by 30 days.

The molecules in the series have two variable regions. Variation in the **A** region takes place due to different substitutions to the phenyl ring; whereas the **B** region varies with the presence or absence of N-oxide (Figure 3). The molecules were divided into training and test sets, so that both represent a wide range of activities. The structures of the training and test set molecules are given in Tables 1 and 2 respectively.

Molecular modeling software, Cerius2 (version 4.10) installed on an SGI Fuel workstation with the IRIX 6.5 operating system, was used for this study. The 3D structure of the molecules was modeled by using the Build/3d-sketcher module and energy minimized with the UNIVERSAL1.02 force field[14] using SMART MINIMIZATION and STANDARD CONVERGENCE with the preferences set to default. The charges were calculated at the beginning, and the charge method used was Gasteiger. To obtain the lowest-energy conformation for each molecule, energy-optimized structures were subjected to dynamics simulation using annealing dynamics with preferences of a constant number of atoms, volume, and energy (NVE); the number of annealing cycles equal to five; an initial temperature of 300 K; a midcycle temperature of 600 K; a temperature increment of 50 K; steps of dynamics per increment equal to 100; and a dynamics time step of 0.001 ps. The above combination of preferences resulted in 6000 steps of dynamics. This was followed by an energy-minimization procedure. A total of 8−10 runs were performed for each compound where the maximum number of iterations was 500. The lowest-energy structures thus obtained were taken as final models.

MFA is a method for quantifying the interaction energy between a probe and a set of aligned molecules.[15] A probe is placed at a random location and then moved about a target molecule within a defined 3D grid. At each defined point in the grid, an energy calculation is performed, measuring the interaction energy between the probe and the target molecule. Atoms in the target molecule are fixed, so that intramolecular energy in the target is ignored. When a complete probe map is calculated for each molecule in the target set, energy values for each point in the grid can be reported in columns and added to the study table. For a set of structures for which energy fields are generated, some or all of the grid data points can be used as descriptors in generating QSAR models and analyzing structure−activity relationships. Alignment was performed using the common substructure (CSS) method, which starts with defining a core model substructure to find a match in all of the molecules under study. It is composed of core atoms and substitution sites. Core atoms are those atoms in the core model that exactly match a substructure in all aligned models, and substitution sites represent sites where the matched alignment models differ. The 2-aminomethyl-4-aminophenol moiety (Figure 4) containing 10 atoms was used as the core model for alignment. The molecules of the training set were finally aligned using the following settings: align strategy, consensus; align method, RMS atoms; and align type, rigid. The consensus RMS value was found to be 0.1330.

Table 1. Structures and Experimental Activities of Compounds and Their Predicted Activities (G/PLS Pred) by MFA and MSA, for the Training Set

					MFA		MSA	
compd.	substituent (X)	N-oxide, (y)[a]	observed ED_{30} (mg/kg)	observed pED_{30} (molar)	predicted	residual	predicted	residual
1	H	0	10.4	1.62	1.080	0.540	1.617	0.003
3	2-Cl	0	15.3	1.48	1.444	0.036	1.107	0.373
4	2-Cl	1	27.4	1.24	0.748	0.492	1.291	−0.051
5	2-OCH_3	0	74.0	0.80	0.383	0.417	0.321	0.479
6	2-OCH_3	1	170	0.45	0.581	−0.131	−0.125	0.575
7	2,3-$(CHCH)_2$	0	4200	−0.94	−0.770	−0.170	−1.384	0.444
8	2,3-$(CHCH)_2$	1	14400	−1.46	−1.233	−0.227	−1.113	−0.347
9	2,5-$(OCH_3)_2$	0	9160	−1.27	-0.906	−0.364	−0.657	−0.613
11	3-CF_3	0	16.3	1.48	1.485	−0.005	1.830	−0.350
12	3-CF_3	1	113	0.66	0.860	−0.200	0.956	−0.296
13	3-Cl	0	17.3	1.43	1.125	0.305	1.201	0.229
14	3-Cl	1	27.4	1.24	0.790	0.450	1.126	0.114
15	3,4-$(OCH_3)_2$	0	927	-0.28	−0.084	−0.196	−0.501	0.221
17	3,4-Cl_2	0	22.7	1.34	1.351	−0.011	1.697	−0.357
19	4-CF_3	0	9.8	1.71	2.335	−0.625	1.726	−0.016
20	4-CF_3	1	7.4	1.84	1.597	0.243	1.347	0.493
21	4-CH_3	0	1530	-0.54	-0.551	0.011	−0.342	−0.198
22	4-CH_3	1	22100	-1.68	-1.432	−0.248	−0.602	−1.078
23	4-Cl	0	6.2	1.87	1.608	0.262	1.606	0.264
24	4-Cl	1	9.2	1.72	1.220	0.500	1.329	0.391
25	4-F	1	19.7	1.37	2.074	−0.704	0.845	0.525
26	4-OCH_3	0	112	0.61	0.662	−0.052	0.556	0.054
27	4-OCH_3	1	423	0.05	−0.211	0.261	−0.253	0.303
28	4-OH	0	320	0.15	−0.015	0.165	0.390	−0.240
29	2-CF_3	0	34.7	1.16	1.183	−0.023	1.525	−0.365
31	2-F	0	26.8	1.22	1.408	−0.188	1.229	−0.009
32	2-F	1	48.4	0.98	1.052	−0.072	1.232	−0.252
33	2,3-F_2	0	8.7	1.73	1.851	−0.121	1.555	0.175
34	2,3,4,5,6-F_5	0	7.4	1.85	2.199	−0.349	1.783	0.067
35	2,6-F_2	0	23.7	1.29	1.314	−0.024	1.904	−0.614
36	3-F	0	12.4	1.56	1.240	0.320	1.188	0.372
37	3-F	1	33.7	1.14	1.006	0.134	1.473	−0.333
40	4-SCH_3	1	158	0.50	0.353	0.147	0.136	0.364
41	2-aza	0	160	0.42	0.454	−0.034	0.022	0.398
42	2-aza	1	550	-0.09	0.085	−0.175	0.270	−0.360
43	3-aza	0	31	1.15	1.294	−0.144	1.248	−0.098
45	4-aza	1	330	0.14	0.357	−0.217	0.391	−0.251

[a] Y = 0, des-N-oxide; Y = 1, N-oxide.

MFA was performed using the QSAR module of Cerius2. A rectangular field, of dimensions 18 × 18 × 16 Å^3 was generated using the probes H^+, OH^-, and CH_3. A grid spacing of 2 Å was used, and fields at 900 points were generated. The energy cutoff was kept at −30 to +30 kcal. Only the top 10% of the total descriptors, which explained the maximum variance, were used in the generation of the QSAR. A regression analysis was performed using the G/PLS method that combines the best features of genetic function approximation (GFA)[16−19] and partial least-squares (PLS). GFA is an evolutionary algorithm that uses a combination of Friedman's multivariate adaptive regression splines and Holland's genetic algorithm to evolve a population of equations. Each generation has PLS applied to it to obtain equations that best fit the training set data. The linear PLS model finds "new variables" (latent variables) which are linear combinations of the original variables.[13] The population size for random equations was kept at 100, and 25 000 crossovers were performed to obtain the equations. The mutation probabilities were kept at system defaults. Smoothness (*d*) was kept at 1.00. Initial equation length values between 5 and 15 are typical for a G/PLS run, which is same as the final equation length.[13,15,19,20] The number of terms, number of components, and scaling were varied to obtain different equations. The final equation was obtained with 12 terms, three components, and no scaling PLS analysis. Fields of molecules are represented using grids in MFA, and each energy probe associated with an MFA grid point can serve as input for the calculation of a QSAR. These energies were added to the study table to form new column heads according to the probe type.

Molecular shape analysis deals with the quantitative characterization, representation, and manipulation of molecular shape in the construction of a QSAR with the incorporation of spatial molecular similarity data (described by Hopfinger and Burke).[21−24] The outcome of the MSA process is an optimized QSAR that can be used for activity estimation and ligand evaluation. The overall aim of MSA is to identify the biologically relevant conformation without knowledge of the receptor geometry and explain in a quantitative manner the activity of a series of analogues. There are seven operations involved in MSA. The final selection of the preferences for each operation is based upon optimizing the QSAR in terms of statistical significance, that is, the set of choices available for each operation are employed to generate trial QSARs. The QSAR, which corresponds to the best correlation between observed activities and computed molecular descriptors, defines the

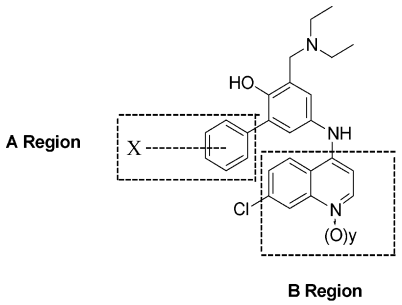

Figure **3a**

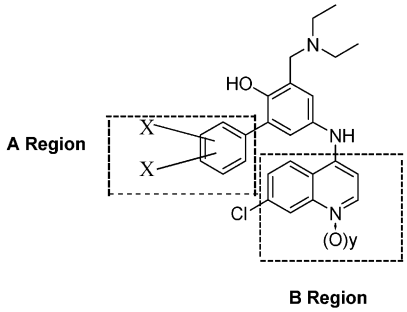

Figure **3b**

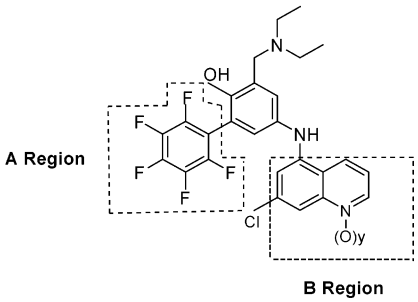

Figure **3c**

Figure 3. Variable regions in the data set molecules used in the present study. In Table 1, compounds 1−6, 11−14, 19−32, and 36−45 are corresponding to part a; compounds 7−10, 15−18, 33, and 35 are corresponding to part b, and compound 34 is corresponding to part c.

Table 2. Structures and Experimental Activities of Compounds and Their Predicted Activities (G/PLS Pred) by MFA and MSA, for the Test Set Molecules

					MFA		MSA	
compd.	substituent (X)	N-oxide (y)[a]	observed ED_{30} (mg/kg)	observed pED_{30} (molar)	predicted	residual	predicted	residual
2	H	1	26.3	1.23	0.236	0.994	1.316	−0.086
10	2,5-$(OCH_3)_2$	1	2450	−0.68	−0.496	−0.184	−0.455	−0.225
16	3,4-$(OCH_3)_2$	1	204	0.40	0.256	0.144	−0.159	0.559
18	3,4-Cl_2	1	36.6	1.15	1.261	−0.111	1.231	−0.081
30	2-CF_3	1	56.1	0.96	1.380	−0.420	1.230	−0.270
38	4-F	0	8.8	1.71	2.213	−0.503	1.424	0.286
39	4-SCH_3	0	151	0.50	1.078	−0.578	−0.152	0.652
44	4-aza	0	89	0.52	0.492	0.028	0.700	−0.180

[a] Y = 0, des-N-oxide; Y = 1, N-oxide.

specific requirements for each MSA operation.[13] The seven steps are as follows:

1. Conformational Analysis. The first operation in MSA is the conformational analysis of the analogues under study. A total of 50 conformers for each previously energy-minimized molecule were generated. The conformers were generated with the "use optimal search method" option of the shape analysis (MSA) module of Cerius2. The open force field was used to estimate the conformer energy.[25] A relative energy cutoff of 5 kcal/mol was used. Only those conformers which were located at the energy level within 5 kcal/mol of the energy minima were retained.

2. Hypothesizing an Active Conformer. The goal of this step is to select a structure that is present in the rate-limiting step for activity in a biological reaction. A number of conformers of different molecules were explored as being the active conformer. This was done by manually selecting the conformer. QSAR equations were generated with this as the active conformer and tested for statistical significance. A number of conformers of compounds **1**, **19**, **20**, **23**, **24**, **33**, and **34** (compounds having high activity) were tried as the active conformer for successive model generations until a statistically significant equation was obtained. The final QSAR equation was obtained with the minimum-energy conformer of **1** as the active conformer.

3. Selection of Shape Reference Compound (SRC). SRC is the molecule that is used when shape descriptors are calculated. MSA compares all other molecules to the SRC and provides information about each comparison. A number of analogues in the data set were evaluated for SRC selection (in the previous section). As mentioned earlier, selection of the SRC is ultimately dictated by maximizing the statistical significance of the resulting QSAR equation. A hypothesized active conformation (default option in MSA module) of the most active molecule, selected in the previous step, was chosen as the SRC for different QSAR trials.

4. Molecular Superposition. Each study molecule was aligned to the SRC using the maximum common subgraph (MCSG) search to calculate the shape descriptors. This method looks at molecules as points and lines and uses the techniques of graph theory to identify patterns. It finds the largest subset of atoms in the shape-reference compound that

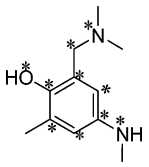

Figure 4. Common structure labeled with asterisks used as the core model in the CSS alignment of lowest-energy conformers.

is shared by all the structures in the study table and uses this subset for alignment.

5. Measurement of Molecular Shape Commonality. After alignment, various shape descriptors, based on relative shape similarity with the SRC, were calculated for each study molecule. The shape descriptors used in QSAR eq 2 are listed in Table 3.

6. Other Molecular Descriptors. In addition to the shape descriptors, other molecular descriptors were also considered while generating QSAR eq 2, shown in Table 3.

7. Construction of a Trial QSAR. Subsequent to the calculation of descriptors for all the conformers of each molecule, conformers were compared to the SRC, and one conformer for each study molecule was selected for regression analysis. Trial QSARs were generated using the G/PLS method described above. The final QSAR equation, eq 2, was obtained with the lowest-energy conformer of compound **1** as the active conformer and SRC. The equation had six components and seven terms and was obtained with scaled PLS analysis.

Model Validation. A number of QSAR equations were obtained, and each equation was evaluated for statistical significance by cross validation[26] using leave-one-out and leave-five-out methods and finally subjected to randomization to obtain the best QSAR model. A decrease in the value of r^2 indicates a good model as equations are generated with scrambled activity data during randomization.[26] Randomization was performed at the 99% confidence level for MFA and MSA models, using the same preferences that were earlier used to obtain the respective QSAR. A high value of r_{CV}^2 alone is not a sufficient criterion for a model to be predictive.[27,28] In light of several recent findings, the emphasis has been given, for the validation, on external test set compounds that were not used in model generation.[29−31] This external predictivity of the model is determined by calculating the predictive r^2 (r_{pred}^2) value, defined as

$$r_{pred}^2 = (SD - PRESS)/SD$$

where SD is the sum of the squared deviations between the actual activities of the compounds in the test set and the mean activity of the training set compounds and PRESS (predicted residual sum of squares) is the sum of the squared deviations between predicted and actual activities for every compound in the test set.

For G/PLS equations, r^2 and least-square error (LSE) were taken as statistical measures along with other validation parameters, namely, cross-validation r^2, PRESS, standard deviation of error of prediction (SDEP), bootstrap r^2 (BSr^2), standard error of estimate (s), and variance ratio (F) at specified degrees of freedom (d.f.).

Table 3. Molecular Descriptors Used in MSA (QSAR eq 2)

no.	descriptor	type	description
1	Vm	spatial	molecular volume
2	NCOSV	MSA(shape)	noncommon overlap steric volume
3	S-xz	spatial (shadow)	area of molecular shadow in the XZ plane
4	S-yz, Frac	spatial (shadow)	area of molecular shadow in the YZ plane over area of enclosing rectangle
5	FPSA-1	spatial (jurs)	fractional charged partial positive surface area
6	FPSA-2	spatial (jurs)	fractional total charge weighted positive surface area

Table 4. Statistical Indexes of MFA and MSA 3D-QSAR Models Using the G/PLS Method with Validation Tests

parameter	trials	MFA	MSA
r^2 (conventional)		0.908	0.846
BS r^2 (±SD)		0.889 (±0.004)	0.829 (±0.013)
F test		345.83	192.82
(d. f.)		1.35	1.35
s		0.064	0.0839
intercept		0.0004	0.0005
LSE		0.085	0.143
PRESS		11.502	9.030
SDEP		0.300	0.388
cv r^2 (standard)		0.886	0.812
cv r^2(leave-one-out)	10	0.812	0.805
cv r^2 (leave-five-out)	10	0.876	0.762
r_{pred}^2 with 8 test compounds		0.500	0.728
r_{pred}^2 with 7 test compounds		0.757	0.832
Randomization Test			
randomization r^2 (99% confidence level)	99	0.445	0.403

[a] CV r^2 is the cross-validated r^2, r_{pred}^2 is the test set predictive r^2, F test is the variance ratio at a specified d.f., s is standard error of estimate, LSE is the least-square error, BS r^2 is the bootstrap r^2, and SDEP is the standard deviation of error of prediction. Randomization r^2 is the lowest r^2 achieved during randomization.

RESULTS AND DISCUSSION

The MFA model was generated using CSS alignment of the energy-optimized structures. G/PLS analysis was carried out for 37 training set compounds, and the best model obtained is eq 1. The predicted activities from eq 1, for both training set and test set molecules, are listed in Tables 1 and 2, respectively, along with their observed activities. The statistical details are given in Table 4. The best QSAR eq 1 generated with MFA is

$$\begin{aligned}\text{Activity} = {} & 1.513\,53 - 0.054\,616 \times \text{“H}^+\text{/278”} - \\ & 0.040\,94 \times \text{“HO}-\text{/485”} + 0.024\,635 \times \text{“CH}_3\text{/323”} - \\ & 0.019\,62 \times \text{“HO}-\text{/683”} + 0.010\,081 \times \text{“H}^+\text{/ 483”} + \\ & 0.026\,587 \times \text{“H}^+\text{/ 589”} - 0.019\,003 \times \text{“CH}_3\text{/379”} - \\ & 0.035\,785 \times \text{“HO}-\text{/288”} - 0.026\,152 \times <\text{“H}^+\text{/138} - \\ & 5.187\,69\text{”}> - 0.014\,066 \times <\text{“H}^+\text{/216} + 27.1654\text{”}> - \\ & 0.031\,057 \times <\text{“H}^+\text{/279} + 9.23823\text{”}> \quad (1)\end{aligned}$$

with N*obs* = 37, $r^2 = 0.908$, standard cross-validated r^2 (r_{CV}^2) = 0.886, boot strap r_{BS}^2 (± SD) = 0.889 (±0.004), and LSE = 0.085 and where “H^+/ 278”, “HO−/ 485”, and so on are the probes and the numberings corresponding to their spatial positions, shown in Figure 4. The terms in angle brackets <...> are spline terms which contribute to the equation only

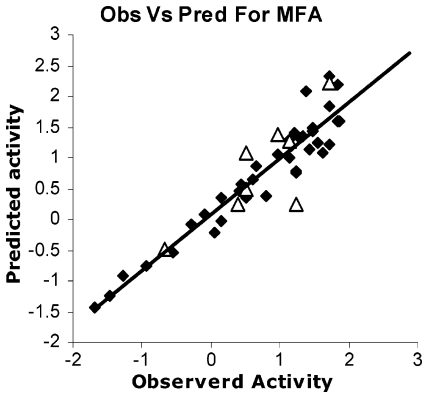

Figure 5. Predicted activities versus the experimental activities (pED_{30}) of data set compounds in MFA: (■) compounds of the training set; (Δ) compounds of the test set.

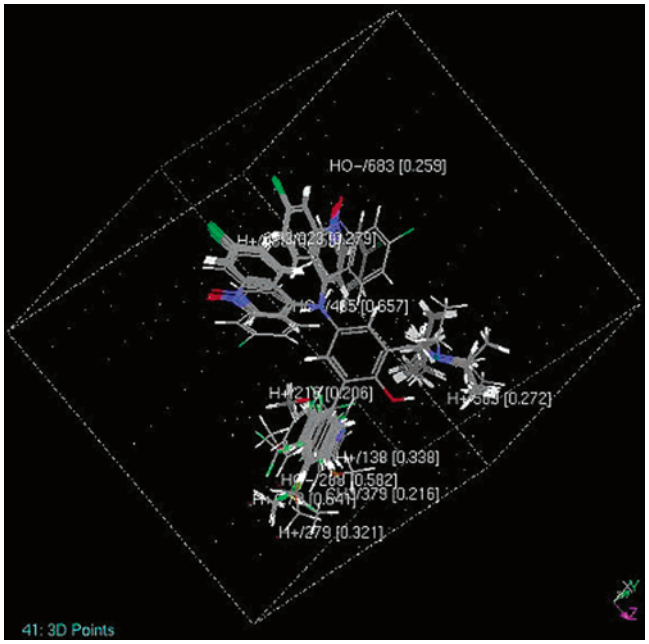

Figure 6. Aligned training set showing the field descriptors in 3D space, used in eq 1: Oxygen, red; carbon, gray; hydrogen, white; nitrogen, blue; chlorine, green; fluorine, light green.

Figure 7. Superimposition of training set molecules using "MCSG" method in the most probable bioactive orientations obtained by the best MSA QSAR model. The atoms of molecules are depicted as cylinder models with no hydrogen shown.

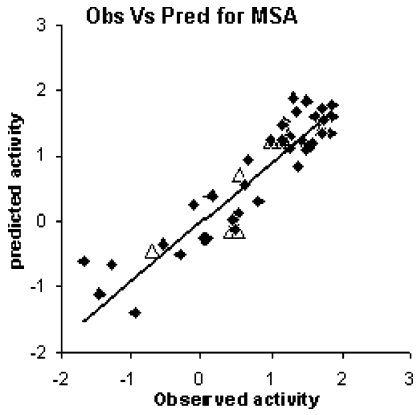

Figure 8. Predicted activities versus the experimental activities (pED_{30}) of the data set compounds using MSA: (■) compounds of the training set; (Δ) compounds of the test set.

when their value is greater than zero. N*obs* means the number of compounds studied; r^2 is the square of the correlation coefficient, used to describe the goodness of fit of the data to the QSAR model. BSr^2 is the average squared correlation coefficient calculated during the validation procedure. A high bootstrapped R^2 value (0.889) and a small value of SD (0.004) imply a high degree of confidence in the model. To determine the model's reliability and significance, both full cross-validation procedures[13] and randomization were performed. Standard cross-validation in G/PLS encompasses only the optimization of regression coefficients; it does not encompass optimization of the choice of descriptors. That is, the regression model is validated only for the specific subset of descriptors obtained from G/PLS. In contrast, full cross-validation encompasses the entire algorithm, including both the choice of descriptors and the optimization of regression coefficients.[33] Here, full cross-validation was performed by leave-one-out and leave-five-out methods. All the analyses reveal comparable cross-validated r^2 values for this model (Table 4). When subjected to a randomization test, eq 1 exhibited a confidence level of 99%, as the value of r^2 for all the 99 random trials was less than the conventional r^2. The randomization was done by repeatedly permuting the dependent variable set (i.e., the observed activity data). The value of r^2 decreased to 0.445 upon randomization. The results obtained suggest that there is a good internal consistency in the underlying data set. Although a good randomization result assures of no chance correlations, it does not guarantee a good predictive model; hence, the emphasis has been given to the validation on an external test set.[26,31] The MFA model has an r_{pred}^2 of 0.500 for eight test compounds, which increased to 0.757 with the exclusion of compound **2** from the test set. Compound **2** was found to be a true X relation outlier[34] in both training and test sets, and hence, the removal of this outlier was considered. The higher value of r_{pred}^2 shows that the model is able to predict the activity of molecules outside the training set and can be used for the prediction of new analogues.

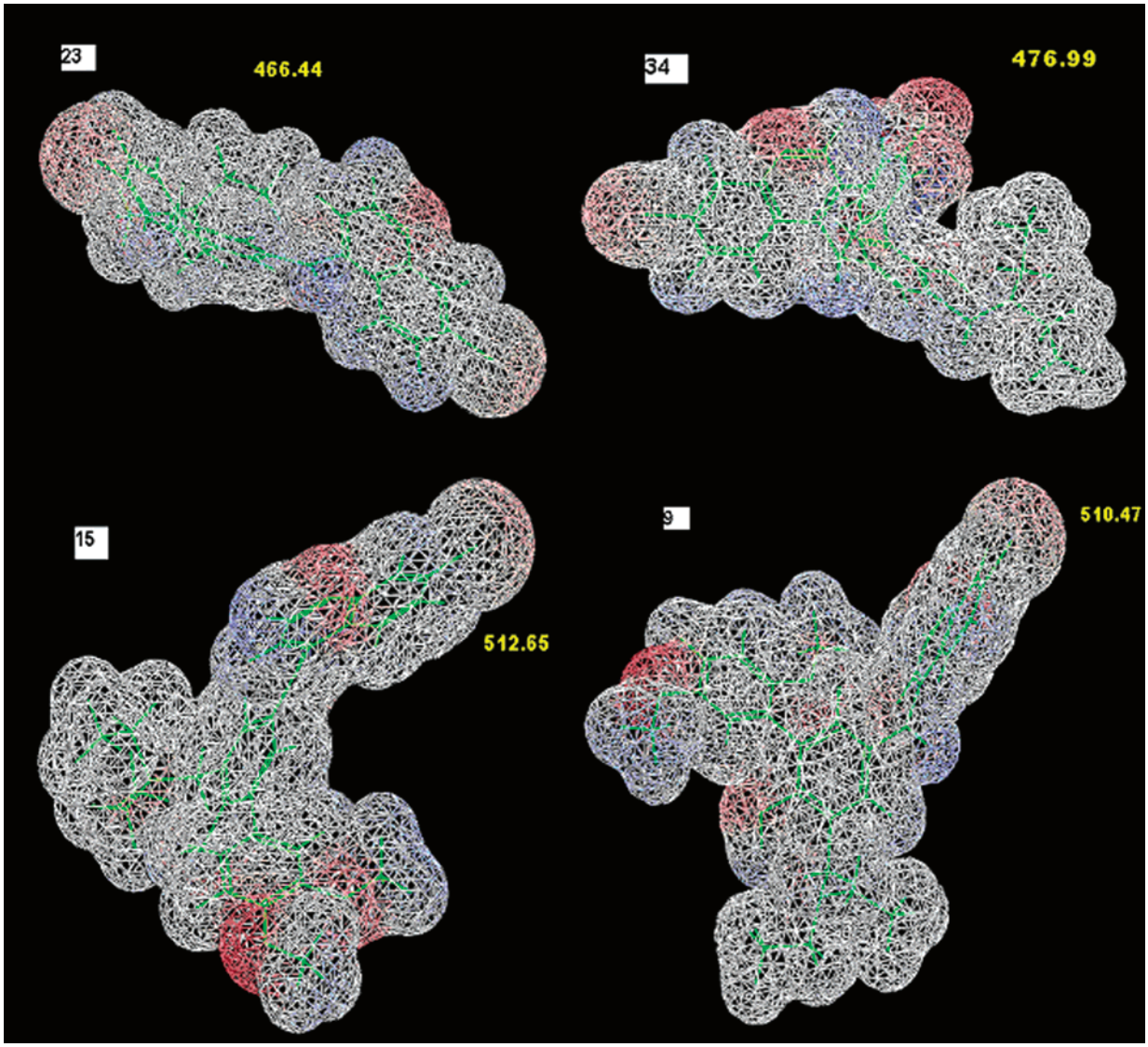

Figure 9. van der Waals surfaces of more active (23 and 34) and less active molecules (15 and 9) of the series.

The scatter plot of observed versus predicted activities, for both training set and test set molecules, obtained from eq 1 is shown in Figure 5.

Equation 1 has 11 molecular field descriptors. The field descriptors "H^+/278", "HO−/483", and "CH3/323" are the energy between the molecule and the proton, hydroxyl, and methyl probes, respectively. The descriptors "H^+/278", "HO−/288", "CH3/379", "H^+/279", "H^+/138", and "H^+/216" lie in the vicinity of the **A** region, whereas "HO−/485" and "HO−/683" lie near the **B** region. "H^+/589", "H^+/483", and "CH3/323" are not close to any variable part of the molecule (Figure 6).

The descriptors "H^+/278", "HO−/288", and "HO−/485" explain 64.1%, 58.2%, and 65.7% of the variance in activity, respectively. Hence, interactions at these points are most important for activity. The negative coefficients of "H^+/278", "H^+/279", "H^+/138", and "H^+/216", present in the vicinity of the **A** region, suggest that the molecules with higher interaction energies with these probes will have a lower activity. Thus, substituents which reduce the negative charge at the center of region **A** will increase the activity of the molecule by reducing its interaction energy with the positively charged H^+ probe. This explains the relatively higher activity of molecules **3**, **4**, **11**, **13**, **14**, **17**−**20**, **23**−**25**, and **29**−**38**, which have electron-withdrawing groups, when compared to molecules **5**, **6**, **9**, **10**, **15**, **16**, **26**−**28**, **39**, and **40**, which have electron-donating substitution. The presence of "CH_3/379" with a negative coefficient explains the lower activities of molecules **5**−**10**, **15**, **16**, **26**, **27**, **39**, and **40**, which have bulky steric substituents when compared to **1** and **2**. The negative coefficient of "HO−/288" suggests that a hydrogen-bond donor/acceptor will reduce the activity of the molecule as indicated by the lower activity of **28**. These results clearly indicate that the steric and electronic properties of specific moieties (**A** region in Figure 3) in the molecules affect the antimalarial potency of these compounds as suggested by Werbel et al.[12] in their 2D QSAR study.

The presence of "HO−/485" and "HO−/683" near the **B** region, with negative coefficients, suggests that the presence of a hydrogen-bond donor/acceptor in this region of the molecule will reduce its activity. This is in accordance with the fact that N-oxides of all but three molecules are less active than their des-N-oxide counterparts. However, few compounds in MFA alignment were oriented in different

directions, not suggesting a common binding mode. This may be due to a limitation of the MFA method as it uses only lowest-energy conformations of all the data set molecules in QSAR generation. This part has been discussed in detail under MSA.

MFA describes a self-consistent field around the molecules that can explain activity.[35,36] The MFA-generated model in this study was statistically significant and was used to correctly predict the activities of training and test set molecules (Tables 1, 2, and 4), indicating this model's usefulness to design potent antimalarials.

MSA. In an attempt to find spatial molecular similarity data along with some important parameters for these inhibitors, a MSA−QSAR study was performed for the same training set of 37 molecules. A view of aligned molecules is shown in Figure 7.

Each study molecule was aligned to the shape reference compound using the MCSG search. This places all the molecules in the same frame of reference, allowing for the calculation of shape descriptors. In addition to the shape descriptors, other molecular descriptors were also considered while generating the QSAR equation (as detailed in Materials and Methods section). The predicted activities, obtained from the best 3D QSAR model generated (eq 2), for both training set and test set molecules, are listed in Tables 1 and 2, respectively, along with their experimental activities. The scatter plot of actual versus predicted activities, for both training set and test set molecules, obtained from eq 2 are shown in Figure 8.

The QSAR equation generated by MSA was

$$\begin{aligned}\text{Activity} = 12.7079 - 0.02425 \times \text{“Vm”} - 0.053528 \times \\ <\text{“Shadow- XZ”} - 89.8083> + 2.13265 \times < \\ \text{“Jurs- FPSA- 2”} - 3.612137> + 18.8636 \times < \\ \text{“Shadow-YZ- Frac”} - 0.673308> - 22.7222 \times < \\ 0.418424 - \text{“Jurs- FNSA-1”}> + 0.019019 \times < \\ 102.51 - \text{“NCOSV”}> \quad (2)\end{aligned}$$

with $N = 37$, $r^2 = 0.846$, standard cross validated r^2 (r_{CV}^2) $= 0.812$, Boot strap r^2 ($\pm$SD) $= 0.829$ (± 0.013), and LSE $= 0.143$.

Statistical details for this MSA model (eq 2) are given in Table 4. It exhibited a 99% confidence level on randomization, and the value of r^2 decreased to 0.403. The predicted activities (pED_{30}) with eq 2 are in good concurrence with the experimental data (within a statistically tolerable error range) with a correlation coefficient of $r_{pred}^2 = 0.728$ for eight test set compounds. The comparable values of conventional r^2, r_{CV}^2, and predictive r^2 (Table 4) indicate that the model can be used for prediction of the activities of a wide range of molecules with considerable accuracy.

Equation 2 consists of one shape descriptor and five spatial descriptors. An analysis of the correlation matrix showed that no significant correlation among the independent variables exists. MSA involves a conformational analysis of the molecules under study and provides input on the relationship of conformational variations with biological activity.

Molecular volume (Vm) is a spatial descriptor that defines the volume of the molecule inside the contact surface. It is a function of conformation and is related to binding and transport. The negative coefficient of Vm indicates that a smaller molecular volume is required for better binding with the receptor. Among des-N-oxide analogues, the most active molecule (**23**) has a volume of 412.42 Å^3, whereas the least active molecule (**9**) has a volume of 450.32 Å.^{3} In order to measure the steric hindrance for these molecules in their MSA conformations, calculation of van der Waals surfaces were carried out using a 1.4 Å probe (Figure 9). The values obtained, in the range 451−523 Å^2, gave further evidence of the fact that the bulkiness and higher volume are detrimental to activity.

A similar relationship is also evident with N-oxide analogues. The Jurs fractional positive surface area-2 (Jurs FPSA-2),[37] present as spline terms in eq 2, increases the biological activity with an increase in its value, while the negative coefficient of Jurs fractional negative surface area-1 (Jurs FNSA- 1) indicates that the higher value of the relative negative charge is detrimental to the activity. This indicates the dependence of activity on the polar solvent-accessible charged surface area. Thus, solvation of the partially charged surfaces of the molecule by a polar solvent is essential for biological activity.

An increase in the area of shadow of the molecule over the *XZ* plane (Sxz) reduces the biological activity as indicated by the negative coefficient for shadow-XZ. This suggests that the orientation of the molecule inside the receptor site should be such that the large planar regions of the molecule are perpendicular to the XZ plane, that is, their projection over the XZ plane is minimum, while a positive coefficient for the Shadow-YZ fraction shows that an increase in fraction of the area of molecular shadow in the YZ plane over the enclosing rectangle (Syz, f) increases the activity.[38] Non-common overlap steric volume (NCOSV) as the spline term has no considerable effect on activity as the constant value (102.51) of this term is less than the descriptor value for almost the entire data set, resulting in no additive effect of this term on activity.[13]

Explanation of Binding Mode by MSA-Generated Conformation. The approach is effective for the analysis of data sets where activity information is available but the structure of the receptor site is unknown. The MSA-generated probable[23] bioactive conformation (by eq 2) of the most active molecule (**23**) has been discussed here, in an attempt to explain heme−drug interaction. This conformation has been shown in Figure 10, wherein the N-quinoline and N-diethyl distance is 7.857 Å. This interatomic distance is within the range 7.52−10.21 Å, required for the heme−drug complex as suggested by O'Neill et.al,[11] and represents theoretical solutions for the conformations achieved during receptor (heme) binding. In this conformation, an intramolecular H bond was observed between N-diethyl and the hydrogen atom of the hydroxyl function, whereas this bond was absent in the lowest-energy conformation of this molecule. It can be assumed that this intramolecular bond is responsible for maintaining a distance between N-quinoline and N-diethyl within the range 7.52−10.21 Å, thus proving the fact that the conformation used by MSA for developing the best QSAR model was probably the "bioactive conformation".

It was observed in the MSA alignment (Figure 7) that all of the compounds were in an almost similar orientation in terms of the direction of pharmacophoric groups, thus confirming the common binding mode for all molecules,

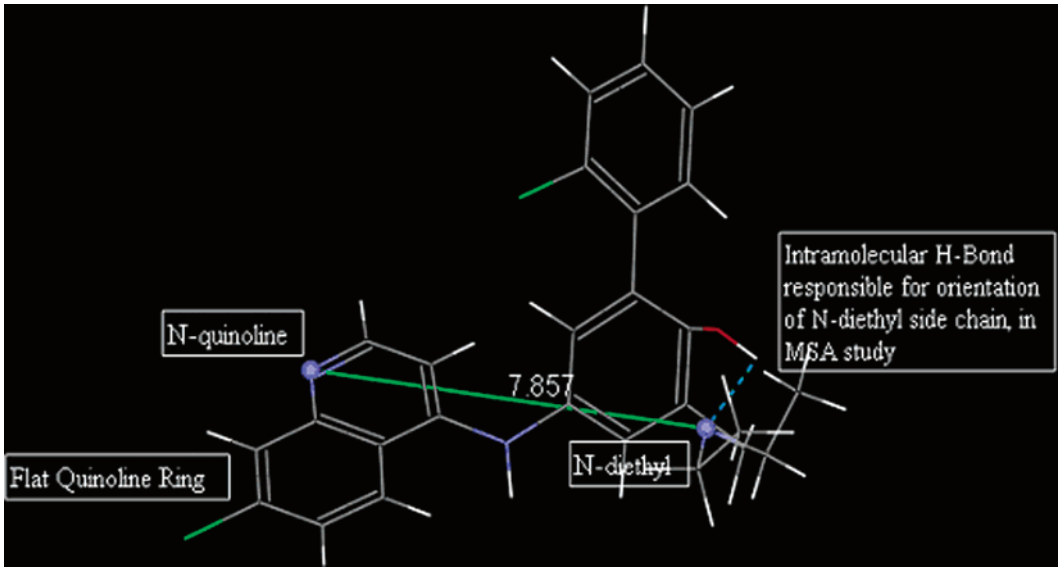

Figure 10. MSA-generated conformation of molecule 23 explaining essential binding constraints.

whereas MFA, which uses lowest-energy conformations for alignment, was not able to superimpose all the molecules in a similar orientation in terms of the direction of pharmacophoric groups (Figure 6). This is due to the fact that MSA explores and analyzes multiple conformational states of each molecule under study and identifies the most accurate pose for bioactivity. All these conformers are compared to the shape reference compound on the basis of shape descriptors and activity. As the MSA equation shows good predictive ability for the test set molecules, the conformers used for model generation could be assumed to represent theoretical solutions of the conformations achieved during receptor binding, which is the rate-limiting step of biological action. This could explain the reason for the better performance of MSA over MFA as the former attempts to select the conformation that is present in the rate-limiting step of biological action.

CONCLUSION

Two 3D-QSAR methods, MFA and MSA, were applied to rationalize the antimalarial activity of a set of 45 molecules of a series of 5-[(7-chloro-4-quinolinyl)amino]-3-[(alkylamino)- and methyl][1,1′-biphenyl]-2-ol analogues and their N^{ω}-oxides in order to evaluate their efficacy as a receptor-independent QSAR approach. Both models showed a high correlation and predictive ability. A high bootstrapped r^2 value and small standard deviations indicate that a similar relationship exists in all compounds. The MFA results show a good correlation between steric and electrostatic fields and antimalarial activity. The MFA model was able to explain the SAR of the molecules. Results obtained with MSA reveal the role of spatial properties and conformation in antimalarial activity. It suggests that a smaller molecular volume is better suited for activity. The MSA model has shown better predictive capability than the MFA model and was found to be more robust. This can be explained by the fact that MSA takes into consideration an array of conformers, whereas only the minimum-energy conformer was used for MFA. Both MFA and MSA models show that the presence of steric groups reduces the activity of the molecules. From this study, it has also been proved that MSA is capable of reaching the bioactive conformation and can be used as an efficient tool for 3D QSAR studies where information regarding receptor structure is not available.

ACKNOWLEDGMENT

The authors are thankful to the Council for Scientific and Industrial Research (CSIR), New Delhi, India, for financial assistance. N.R. is indebted to the Director of the Indian Institute of Chemical Biology (CSIR) for providing this research opportunity.

REFERENCES AND NOTES

(1) Wahlgren, M.; Bejarano, M. T. A Blueprint of 'Bad Air'. *Nature* **1999**, *400*, 506−507.
(2) *Roll Back Malaria*; WHO: Geneva, Switzerland. http://www.rbm-.who.int (accessed Mar 2007).
(3) Frevert, U.; Pradel, G. Malaria Sporozoites Actively Enter and Pass through Rat Kupffer Cells prior to Hepatocyte Invasion. *Hepatology* **2001**, *33*, 1154−65.
(4) Barnwell, J. W. Hepatic Kupffer Cells: The Portal that Permits Infection of Hepatocytes by Malarial Sporozoites. *Hepatology* **2001**, *33*, 1331−1333.
(5) Vroman, J. A.; Gaston, M. A.; Avery, M. A. Current Progress in the Chemistry, Medicinal Chemistry and Drug Design of Artemisinin Based Antimalarials. *Curr. Pharm. Des.* **1999**, *5*, 101−138.
(6) White, N. J. Why Is It that Antimalarial Drug Treatments Do Not Always Work? *Ann. Trop. Med. Parasitol.* **1998**, *92*, 449−58.
(7) Dorn, A.; Vippagunta, S. R.; Matile, H.; Jaquet, C.; Vennerstrom, J. L.; Ridle, R. G. An Assessment of Drug−Haematin Binding as a Mechanism for Inhibition of Haematin Polymerisation by Quinoline Antimalarials. *Biochem. Pharmacol.* **1998**, *55*, 727−736.
(8) Vippagunta, S. R.; Dorn, A.; Matile, H.; Bhattacharjee, A. K.; Karle, J. M.; Ellis, W. Y.; Ridle, R. G.; Vennerstrom, J. L. Structural Specificity of Chloroquine−Hematin Binding Related to Inhibition of Hematin Polymerization and Parasite Growth. *J. Med. Chem.* **1999**, *42*, 4630−4639.
(9) 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.
(10) 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.
(11) 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.
(12) Werbel, L. M.; Cook, P. D.; Elslager, E. F.; Hung, J. H.; Johnson, J. L.; Keston, 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-[(alkylamino)methyl][1,1′-biphenyl]-2-ols and N Omega-Oxides. *J. Med. Chem.* **1986**, *29*, 924−939.
(13) *MSI Cerius2*, version 4.10; Molecular Simulations, Accelrys Inc.: San Diego, CA. http:// www.msi.com (accessed Mar 2007).
(14) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., III. UFF, A Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc*. **1992**, *114*, 10024−10035.
(15) Hirashima, A.; Eiraku, T.; Kuwano, E.; Eto, M. Three-Dimensional Molecular Field Analyses of Agonists for Tyramine Receptor which Inhibit Sex−Pheromone Production in *Plodia interpunctella. Internet Electron. J. Mol. Des.* **2003**, *2*, 511−526.
(16) Friedman, J. H. *Multivariate Adaptive Regression Splines*; Technical Report 102, Stanford University, Department of Statistics: Stanford, CA, 1988.
(17) Holland, J. In *Adaptation in Artificial and Natural Systems*; University of Michigan Press: Ann Arbor, MI, 1975.
(18) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure−Activity Relationships and Quantitative Structure−Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.
(19) Ghoshal, N.; Mukherjee, P. 3-D-QSAR of N-Substituted 4-Amino-3, 3-dialkyl-2(3H)-furanone GABA Receptor Modulators Using Molecular Field Analysis and Receptor Surface Modelling Study. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 103−109.
(20) (a) Sharma, P.; Ghoshal, N. Exploration of a Binding Mode of Benzothiazol-2-yl Acetonitrile Pyrimidine Core Based Derivatives as Potent c-Jun N-Terminal Kinase-3 Inhibitors and 3D-QSAR Analyses. *J. Chem. Inf. Model.* **2006**, *46*, 1763−74. (b) Bhattacharya, P.; Leonard, J. T.; Roy, K. Exploring 3D-QSAR of Thiazole and Thiadiazole Derivatives as Potent and Selective Human Adenosine A3 Receptor Antagonists. *J. Mol. Model.* **2005**, *11*, 516 − 524.
(21) Hopfinger, A. J.; Tokarski, J. S. In *Practical Applications of Computer-Aided Design*; Charifson, P. S., Ed.; Marcel Dekker: New York, 1997; pp 105−164.
(22) Burke, B. J.; Hopfinger, A. J. 1-(Substituted-benzyl) imidazole-2(3H)-thione Inhibitors of Dopamine Beta-hydroxylase. *J. Med. Chem*. **1990**, *33*, 274−281.
(23) Burke, B. J.; Hopfinger, A. J. In *Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Ed.; John Wiley and Sons: New York, 1990; pp 11−73.
(24) Hopfinger, A. J.; Burke, B. J. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Ed.; John Wiley and Sons: New York, 1990; p 173.
(25) Accelrys. http://www.accelrys.com/doc/life/cerius410L/forcefield_engines/SimToolsTOC.doc.html (accessed Mar 2007).
(26) Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; Waterbeemd, H. V. D. Ed.; VCH Publishers, Inc.: New York, 1995; pp 309−318.
(27) Golbraikh, A.; Tropsha, A. Beware of q^2. *J. Mol. Graphics Modell.* **2002**, *20*, 269−276.
(28) Golbraikh, A.; Shen, M.; Xiao, Z; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational Selection of Training and Test Sets for the Development of Validated QSAR Models. *J. Comput.-Aided Mol. Des*. **2003**, *17* (2−4), 241−53.
(29) Wold, S.; Martens, H.; Wold, H. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. In Proc. Conf. Matrix Pencils, Pite Havasbad, Sweden, March 22−24, 1982; Springer-Verlag: New York, 1983.
(30) Ruhe, A.; Kågstro, B. M. In *Lecture Notes in Mathematics 973*; Springer-Verlag: Heidelberg, Germany, 1982; pp 286−293.
(31) Wold, S. Cross Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **1978**, *20*, 397−405.
(32) *Statistica 6.0*; Stat Soft Inc.: Tulsa, OK.
(33) Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. Quantitative Structure−Antitumor Activity Relationships of Camptothecin Analogues: Cluster Analysis and Genetic Algorithm-Based Studies. *J. Med. Chem.* **2001**, *44*, 3254−3263.
(34) Furusjo, E.; Svenson, A.; Rahmberg, M.; Andersson, M. The Importance of Outlier Detection and Training Set Selection for Reliable Environmental QSAR Predictions. *Chemosphere* **2006**, *63*, 99−108.
(35) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem*. **1995**, *38*, 2080−2090.
(36) Hirashima, A.; Eiraku, T.; Kuwano, E.; Taniguchi, E.; Eto, M. Three-Dimensional Pharmacophore Hypotheses of Octopamine Receptor Responsible for the Inhibition of Sex-Pheromone Production in *Plodia interpunctella*. *Internet Electron. J. Mol. Des*. **2002**, *1*, 37−51.
(37) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer Assisted Quantitative Structure Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2323−2329.
(38) Stouch, T. R.; Jurs, P. C. A Simple Method for the Representation, Quantification, and Comparison of the Volumes and Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 4−12.

CI600570R