Research Article

JOURNAL OF CHEMOMETRICS

Received: 5 June 2009, Revised: 31 August 2009, Accepted: 17 September 2009, Published online in Wiley InterScience: 2009

(www.interscience.wiley.com) DOI: 10.1002/cem.1261

Quantitative structure–activity relationship (QSAR) of artemisinin: the development of predictive *in vivo* antimalarial activity models

Mani Srivastava[a], Harvinder Singh[a] and Pradeep Kumar Naik[a]*

A quantitative structure–activity relationship (QSAR) analysis has been performed on a data set of 194 artemisinin analogs for antimalarial activity. Several types of descriptors including topological, spatial, thermodynamics, information content, lead likeness, and E-state indices have been used to derive a quantitative relationship between antimalarial activity and structural properties. A systematic approach of zero tests, missing value test, simple correlation test, multicollinearity test, and genetic algorithm method of variable selection was used to generate the model. Statistically significant model ($r^2 = 0.845$, $q^2_{cv} = 0.799$, *F*-test = 53.40) was obtained with the descriptors like molecular connectivity indexes, E-state index, length-to-breadth ratio of compounds, MLog P, HOMO, electron density, Balabans topological index, and strain energy of the molecules. The robustness of the QSAR models was characterized by the values of the internal leave one out cross-validated regression coefficient (q^2_{cv}) for the training set and determination coefficient in prediction, q^2_{test} for the test set. The value of $q^2_{test} = 0.876$ for the test set; revealed good external predictability of the QSAR model. Also, for an external data set (validation set) of four artemisinin analogs, the QSAR model was able to predict the antimalarial activity very well in comparison to experimental values. The model was also tested successfully for external validation criteria. The QSAR model developed in this study should aid further design of novel potent artemisinin derivatives. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: artemisinin; quantitative-structure–activity relationship; antimalarial activity

1. INTRODUCTION

Artemisinin (*Qinghaosu*), a sesquiterpene endoperoxide isolated from *Artemisia annua*, is a remarkable life saving antimalarial compound, effective against drug-resistant *Plasmodium falciparum* and cerebral malaria [1–4]. Artemisinin and its derivatives have many advantages: quick reduction of fevers, fast clearing parasites in blood (90% of malaria patients recovered within 48 h), and no significant side effects. As a consequence, they are of special interest for severe malaria. The first decline in the number of parasites is also beneficial for combination therapies. Prompted by the clinical successes of artemisinin, significant efforts have been made to identify new analogs that have a similar mechanism of action yet are superior in activity. Subsequent research led to derivatives [5,6] of artemisinin such as artemether, arteether, and artesunate. A consistent number of structural modifications have been introduced in the original structure of artemisinin in order to overcome the solubility as well as the neurotoxic problem associated with its utilization as an antimalarial drug. The artemisinin family of molecules has been extensively studied to elucidate its mechanism of action as an antimalarial drug and to develop more potent and selective antimalarial agents [3–6]. An essential feature of artemisinin (and analogs) activity is hypothesized to be the presence of a peroxide bridge (Figure 1), which forms a bond with a high valence non-heme iron molecule, leading to the generation of free radicals [4,5].

A number of QSAR studies have also been reported for the prescreening of prospective artemisinin analogs for antimalarial activity. A number of these studies [7–9] have used comparative molecular field analysis (CoMFA) [10,11] as a tool to model the activity of artemisinin analogs in terms of active site binding. Although comparative molecular field analyses (CoMFA) are statistically excellent and offer good predictive performance, they are inherently limited to the need to align with the database molecules correctly within 3D space. The determination of the "active" conformation that each compound will retain is a critical issue due to the unavailability of an X-ray structure. We should have some knowledge or hypothesis regarding active conformations of the molecules under study as a prerequisite for structural alignment. Hence, the developed models based on CoMFA may not suit drug design, because of a false conformational hypothesis. However, we were motivated to explore possible alternatives that would use alignment free descriptors derived from 2D or 3D molecular topology and thus alleviate frequent ambiguity of structural alignment typical of 3D QSAR methods. A QSAR equation is a mathematical equation that

* *Correspondence to: P. K. Naik, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Distt. Solan 173215, Himachal Pradesh, India.*
E-mail: pknaik73@rediffmail.com

a *M. Srivastava, H. Singh, P. K. Naik*
Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173215, Himachal Pradesh, India

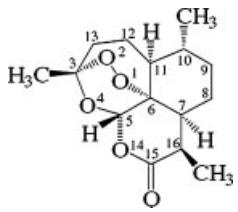
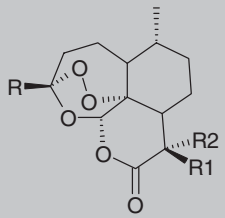


Figure 1. Stereochemistry and atomic numbering scheme of artemisinin.

correlates the biological activity to a wide variety of physical or chemical parameters [12,13]. There are many examples available in the literature in which QSAR models have been used successfully for the screening of compounds for biological activity [14–17].

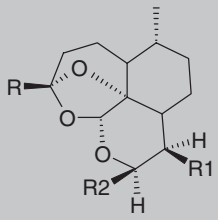
In this QSAR study, we have applied E-state, electronic, structural, topological quantum mechanics, and physicochemical based descriptors which can be calculated without structural alignments, for the development of QSAR equation correlating *in vivo* antimalarial activity. The model developed in the present study is the first of its kind for the antimalarial activity prediction of artemisinin congeners because of its high statistical quality. Further, the behavior of the QSAR model is examined by a variety of statistical parameters and the contributions of various descriptors are analyzed. The methodology used in the present study is in line with that used by Deswal and Roy [18] for the development of thrombin inhibitors.

Table Ia. Artemisinin analogs with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work



Compounds	R	R1	R2	Log RA	pIC_{50} (ng/ml)
1	CH_3	CH_3	H	1.00	1.398
2	C_4H_8Ph	H	H	0.45	0.712
3	CH_3	H	2-Z-Butenyl	−1.10	−0.760
4	CH_3	H	H	0.79	1.188
5	CH_3	H	CH_3	−0.17	0.228
6	CH_3	H	2-E-Butenyl	−0.60	−0.260
7	CH_3	Allyl	H	−0.10	0.260
8	CH_3	C_4H_9	H	0.17	0.674
9	C_4H_8Ph	C_4H_9	H	−0.32	−0.117
10	C_3H_6(P-Cl-Ph)	C_4H_9	H	−0.28	−0.097
11	C_4H_9	C_4H_9	H	−0.48	−0.195
12	CH_3	C_2H_5	H	1.40	1.777
13	CH_3	C_6H_{13}	H	0.86	1.162
14	CH_3	i-C_4H_9	H	−0.55	−0.212
15	CH_3	i-C_6H_{13}	H	−0.04	0.262
16	CH_3	i-C_3H_7	H	−0.04	0.317
17	CH_3	i-C_5H_{11}		0.07	0.389
18	C_3H_6(*p*-Cl-Ph)	H	H	0.10	0.340
19	C_4H_9	H	H	−0.74	−0.383
20	$CH_2CH_2CO_2Et$	H	H	0.37	0.669
21	C_2H_5	H	H	0.05	0.448
22	i-C_4H_9	H	H	−0.35	0.007
23	CH_3	Br	CH_2Br	−1.64	−1.435
24	CH_3	$=CH_2$		−0.89	−0.489
25	CH_3	CH_2CH_3	$R_1=R_2$	−0.36	−0.022
26	CH_3	C_5H_{11}	H	1.02	1.339
27	CH_3	C_4H_8Ph	H	0.63	0.876
28	CH_3	C_2H_4Ph	H	0.12	0.398
29	CH_3	C_3H_6Ph	H	0.78	1.042
30	CH_3	C_3H_7	H	1.13	1.487

Table Ib. Deoxy-artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work



Compounds	R	R1	R2	MW	Log RA	pIC_{50} (ng/ml)
31	CH_3	CH_3	OEt	296	−4	−3.623
32	CH_3	CH_3	OH	268	−4	−3.580
33	CH_3	C_4H_8Ph	—	370	−4	−3.720
34	CH_3	C_3H_7	—	280	−4	−3.599
35	CH_3	C_6H_{13}	—	322	−4	−3.660
36	CH_3	C_4H_9	H	294	−4	−3.620
37	CH_3	i-C_5H_{11}	—	324	−4	−3.662
38	$CH_2CH_2CO_2Et$	H	H	328	−4	−3.668
39	C_2H_4Ph	H	—	252	−4	−3.553
40	CH_2 CH_3	H	—	252	−4	−3.553
41	i-C_4H_9	H	—	280	−4	−3.599
42	i-C_4H_9	H	H	280	−4	−3.599
43	CH_3	C_2H_4Ph	—	342	−4	−3.686
44	CH_3	C_3H_6Ph	—	356	−4	−3.703
45	CH_3	CH_3	—	266	−4	−3.577

2. MATERIALS AND METHODS

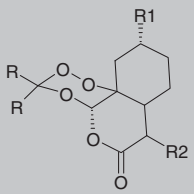
2.1. Data set

An initial dataset of 194 artemisinin analogs was collected from the published data [19–24] in which several different ring systems were represented. All the analogs were either peroxides or trioxanes, which should act via similar mechanisms of action and were categorized into different classes (Tables Ia–Im). These molecules were rationally designed as functional mimics of natural artemisinin with the goal of simplifying the chemical synthesis and improving the antimalarial activity. Structural modifications are mainly introduced at varying radicals at positions R, R_1, and R_2 in artemisinin scaffold. Each of these compounds had associated *in vitro* bioactivity values (IC_{50} values reported in ng/ml) against the drug-resistant malaria strain *P. falciparum* (W-2 clone). The log value of the relative activity (RA) of these compounds was used for analysis and was defined as

$$\text{Log(RA)} = \log[(\text{artemisinin } IC_{50}/\text{analog } IC_{50})(\text{analog MW}/\text{artemisinin MW})]$$

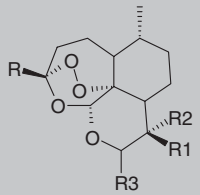
Molecular models of artemisinin and its analogs (Tables Ia–Im) were built using the builder feature in Maestro (Schrodinger package) and energy minimized in a vacuum using Impact. Each structure was assigned an appropriate bond order using ligprep script shipped by Schrödinger and optimized initially by means of the OPLS 2005 force field using default setting. Complete geometrical optimization of these structures was carried out with the HF/3-21G method (in this work) using the Jaguar (Schrodinger Inc.) [25]. In order to check the reliability of the geometry obtained, we compared the structural parameters of the artemisinin 1,2,4-trioxane ring with theoretical [26] and experimental [27,28] values from the literature. All calculations reproduced most of the structural parameters of the artemisinin 1,2,4-trioxane ring seen in X-ray structures (Table II). This applies especially to the bond length of the endoperoxide bridge which

Table Ic. Seco-artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work



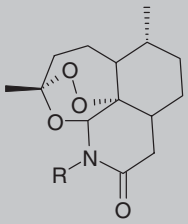
Compounds	R	R1	R2	Log RA	pIC_{50} (ng/ml)
46	CH_3	H	H	−2.37	−1.906
47	C_2H_5	H	H	−1.13	−0.713
48	CH_3	CH_3	CH_3	−0.60	−0.183
49	—	—	—	−0.15	0.245
50	CH_3	H	CH_3	−0.86	−0.420
51	—	—	$-(CH_2)_4-$	−0.26	0.097

Table Id. 10-Substituted artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work



Compounds	R	R1	R2	R3	Log RA	pIC_{50} (ng/ml)
52	CH_3	CH_3	H	H	0.75	1.170
53	CH_3	CH_3	H	OH	0.55	0.945
54	CH_3	CH_3	H	OEt	0.34	0.694
55	CH_3	CH_3	H	OH	0.96	1.295
56	CH_3	CH_3	H	OEt	−1.08	−0.740
57	CH_3	H	Br	H	0.28	0.606
58	CH_3	CH_3	Br	NH-2-(1,3-thiazole)	0.66	0.874
59	CH_3	CH_3	Br	*p*-Cl-aniline	0.79	0.977
60	CH_3	CH_3	Br	aniline	0.18	0.401
61	CH_3	Br	CH_3	NH-2-pyridine	−0.09	0.115
62	CH_3	CH_3	Br	NH-2-pyridine	−0.77	−0.564
63	CH_3	CH_3	H	OMe	0.28	0.654
64	CH_3	CH_3	H	α-OEt	0.32	0.674
65	CH_3	C_4H_9	H	H	1.32	1.677
66	CH_3	C_2H_5	H	H	0.67	1.068
67	CH_3	C_3H_7	H	OEt	−0.04	0.277
68	CH_3	H	H	OEt	0.43	0.804
69	CH_3	C_2H_5	H	OEt	0.50	0.835
70	CH_3	CH_3	H	C_3H_6OH	0.78	1.115
71	CH_3	CH_3	H	C_4H_9	0.06	0.398
72	CH_3	CH_3	H	OCH_2CO_2Et	0.52	0.800
73	CH_3	CH_3	H	$OC_2H_4CO_2Me$	0.10	0.364
74	CH_3	CH_3	H	$OC_3H_6CO_2Me$	−0.03	0.218
75	CH_3	CH_3	H	OCH_2(4-PhCO_2Me)	−0.07	0.143
76	CH_3	CH_3	H	(R)-$OCH_2CH(CH_3)CO_2Me$	1.79	2.070
77	CH_3	CH_3	H	(R)-$OCH(CH_3)CH_2CO_2Me$	0.87	1.134
78	CH_3	CH_3	H	(S)-$OCH(CH_3)CH_2CO_2Me$	1.70	1.964
79	$CH_2CH_2CO_2Et$	H	H	H	0.70	1.017
80	C_3H_6(*p*-Cl-Ph)	H	H	H	−0.55	−0.295
81	C_2H_5	H	H	H	−1.00	−0.580
82	C_3H_7	H	H	H	0.84	1.238
83	CH_3	$-OCH_2-$		OOH	−0.62	−0.269
84	CH_3	$-CH_2O-$		OOH	−0.57	−0.219
85	CH_3	$=CH_2$		OOH	−0.99	−0.616
86	—	CH_3	OH	α-OH	−0.89	−0.519
87	CH_3	C_5H_{11}	H	H	0.16	0.498
88	CH_3	C_3H_7	H	H	0.74	1.117
89	—	CH_3	H	CH_2CF_2	0.11	0.437
90	—	CH_3	OH	CH_2CF_3	0.33	0.615
91	—	CH_3	OH	OEt	−0.44	−0.108
92	—	OH	CH_3	OEt	−1.13	−0.798
93	CH_3	CH_3	H	OOt-C_4H_9	0.92	1.217

Table Ie. 11-Aza-artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work



Compounds	R	Log RA	pIC_{50} (ng/ml)
94	C_3H_6Ph	0.02	0.283
95	C_2H_4Ph	0.16	0.439
96	C_5H_{11}	−0.20	0.121
97	i-C_5H_{11}	−0.04	0.281
98	CH_2(*p*-Cl-Ph)	−0.16	0.096
99	i-C_4H_9	0.02	0.359
100	CH_2Ph	0.34	0.636
101	CH_3	0.70	1.099
102	C_3H_7	0.05	0.408
103	2-Thiophene	0.17	0.458
104	2-Furan	0.11	0.418

seems to be responsible for the antimalarial activity [29–33]. These molecules were randomly divided into 156 molecules in training set and 38 molecules in test set.

2.2. Descriptor calculation

E-state indices [34], M log P [35], Superpendentic index [36], structural [37], symmetrical, topological, lead likeness [38], electronic Wang–Ford atomic charge [23] and extended Huckel partial charge [18,39,40], bulk, moments, orbital energies, molecular connectivity indexes [41], gravitational indexes [42], hydrophobicity [13], steric [14,15] and thermodynamic factors [16], and topological descriptors were calculated using ADME Model Builder software package (version 4.5). The Superpendentic index is computed from the pendent matrix. These descriptors help differentiate the molecules mostly according to their size, degree of branching, flexibility, and overall shape. Some of the descriptors included in the study are listed and described in Table III.

2.3. Regression analysis

The total number of descriptors calculated initially was 372. A systematic search in the order of missing value test, zero test, correlation coefficient, multi-colinearity, and genetic algorithm was performed to determine significant descriptors using ADME Model Builder (version 4.5) software package (Fujitsu Inc.). Any parameter which is not calculated (missing value) for any number of the compounds in the data set is rejected in the first step. Some of the descriptors were rejected because they contained a value of zero for all the compounds and have been removed (zero tests). In order to minimize the effect of colinearity and to avoid redundancy, the correlation matrix was developed with a cut-off value of 0.6 the variables that show exact linear dependences between subsets of the variables and multi-colinearity (high multiple correlations between subsets of the variables) were physically removed from the analysis. From descriptors thus remained, the set of descriptors that would give the statistically best QSAR models was selected from the large pool using a genetic function approach implemented in ADME model Builder (version 4.5) software package (Fujitsu Inc.). The genetic algorithm starts with the creation of a population of randomly generated parameter sets. The usage probability of a given parameter from the active set is 0.5 in any of the initial population sets. The sets are then compared according to their objective functions. The form of objective function favors sets that have r^2 as high as possible, while minimizing the number of parameters used as descriptors. The higher the score the higher the probability of a given set will be used for the creation of the next generation of sets. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The parameters set used for the genetic algorithm includes: mutation 0.1, crossover 0.9, population 300, number of generations 1000, R^2 floor limit 50%, and the objective function was R^2/N_par. The form of the objective function favors sets that have the R^2 as high as possible, while minimizing the number of parameters used as descriptors. The higher the score, the higher the probability that a given set will be used for the creation of the next generation of sets. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The algorithm runs until the desired number of generations is reached. Equations were developed between the observed activity and the descriptors. The best equation was taken based on the statistical parameters such as regression coefficient (r^2), adjusted regression coefficient (r^2_{adj}), regression coefficient cross validation (q^2_{cv}), and *F*-test values.

2.4. Validation test

The predictive capability of the QSAR equation is determined using the leave-one-out cross validation method. The cross validation regression coefficient (q^2_{cv}) was calculated by the following equation:

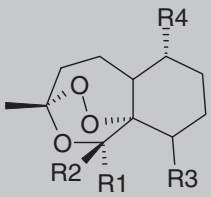
$$q^2_{cv} = 1 - \frac{\text{PRESS}}{\text{TOTAL}} = 1 - \frac{\sum_{i=1}^{n} (Y_{exp} - Y_{pred})^2}{\sum_{i=1}^{n} (Y_{exp} - \overline{Y})^2}$$

Where, Y_{pred}, Y_{exp}, and Y are the predicted, experimental, and mean values of experimental activity, respectively of the training set compounds. Also, the accuracy of the prediction of the QSAR equation was validated by *F*-value, r^2, and r^2_{adj}. A large *F* indicates that the model fit is not a chance occurrence. It has been shown that a high value of statistical characteristics need not be the proof of a highly predictive model [43,44]. Hence, in order to evaluate the predictive ability of our QSAR model, we used the method described by Golbraikh and Tropsha [43] and Roy and Roy [44]. The determination coefficient in prediction (q^2_{test}) was calculated using the following equation [44]:

$$q^2_{test} = 1 - \frac{\sum (Y_{pred_{text}} - Y_{Test})^2}{\sum (Y_{Test} - \overline{Y})^2}$$

where $Y_{pred_{test}}$ and Y_{Test} are the predicted values based on the QSAR equation (model response) and experimental activity values, respectively, of the external test set compounds. *Y* is the

Table If. Artemisinin derivatives lacking the D-ring with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work



Compounds	R1	R2	R3	R4	Log RA	pIC_{50} (ng/ml)
105	$-O_2CCH_2Ph$	H	H	CH_3	−0.51	−0.217
106	H	H	H	CH_3	−0.32	0.202
107	H	OCH_3	H	H	−0.31	0.180
108	OCH_3	H	H	H	−1.04	−0.550
109	H	H	H	—	−0.41	0.007
110	OCH_2Ph	H	H	H	−0.09	0.275
111	C_2H_4OH	H	CH_3	—	−1.80	−1.429
112	C_2H_4OH	CH_3	H	—	0.23	0.601
113	C_2H_4OH	CH3	CH3	—	−1.80	−1.449
114	$C_2H_4OCH_2Ph$	CH3	CH3	—	−1.80	−1.558
115	OCH_3	H	$C_2H_4O_2CNEt_2$	H	0.65	0.929
116	OCH_3	H	$C_2H_4O_2CNPh_2$	—	0.65	0.829
117	H	OCH_3	$C_2H_4OCH_3$	H	−0.39	0.002
118	H	OCH_3	$C_2H_4OCH_2Ph$	H	0.75	1.039
119	H	OCH_3	C_2H_4O-allyl	H	0.40	0.735
120	H	OCH_3	$C_2H_4O_2Ph$	H	−0.59	−0.319
121	H	OCH_3	$C_2H_4O_2C$(4-$PhCO_2Me$)	H	0.27	0.465
122	H	OCH_3	$C_2H_4O_2C$(4-$PhCO_2H$)	H	−0.81	−0.586
123	H	OCH_3	$C_2H_4O_2C$(4-$PhCONEt_2$)	—	0.23	0.400
124	H	OCH_3	$C_2H_4O_2C$(4-$PhCO_2C_2H_4NMe_2$)	—	−0.60	−0.446
125	H	OCH_3	$C_2H_4O_2CCH_2NCO_2$-(t-C_4H_9)	H	−0.04	0.174
126	OCH_3	—	$C_2H_4OCH_2$(4-F-Ph)	—	0.38	0.648
127	OCH_3	—	$C_2H_4OCH_2$(4-Py)	—	0.14	0.428
128	H	OCH_3	$C_2H_4OCH_2$(4-N-Me-pyridine)	H	−0.90	−0.647

mean activity value of the training set compounds. Further evaluation of the predictive ability of the QSAR model for the external test set compounds was done by determining the value of rm^2 by the following equation [44]:

$$rm^2 = r^2(1 - |\sqrt{r^2 - r_0^2}|)$$

where r^2 is the squared Pearson correlation coefficient for regression calculated using $Y = a + bx$; "a" is referred to as the y-intercept, "b" is the slope value of regression line, and r^2_0 is the squared correlation coefficient for regression without using y-intercept and the regression equation was $Y = bx$. Both r^2 and r^2_0 between experimental and predicted values for the external test set compounds were calculated using the regression of analysis Toolpak option of Excel. The values of k *and* k', slopes of the regression line of the predicted activity versus actual activity and *vice versa*, were calculated using the following equations [45]:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2} \quad k' = \frac{\sum y_i \tilde{y}_i}{\sum y_i^2}$$

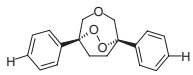
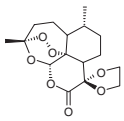
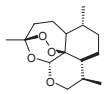
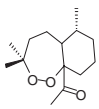
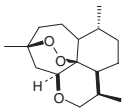
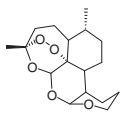
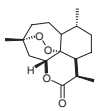
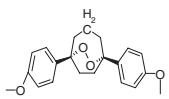
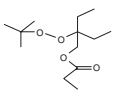
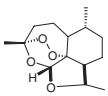
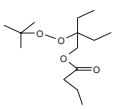
where cy_i and y_i are the predicted and actual activities, respectively.

To further check the inter-correlation of descriptors variance inflation factor (VIF) analysis was performed. The VIF value is calculated from $1/1 - r^2$, where r^2 is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. If the VIF value is larger than 10, information of the descriptor could be hidden by correlation of descriptors [45,46].

3. RESULTS AND DISCUSSION

The 194 active compounds considered as potential of W-2 strain of *P. falciparum* inhibition were randomly divided into a training set of 156 compounds and a test set of 38 compounds. The experimental IC_{50} values against the W-2 strain of *P. falciparum* for these compounds are available from *in vitro* analysis. With the wide range of difference between the IC_{50} values and the large diversity in the structures, the combined data sets of 156 molecules and 38 molecules are ideal to be considered as training and test set, as both the sets do not suffer from bias, due to the similarity of the structures. The various molecular descriptors (372 in total) as described in Table III were calculated initially. By

Table Ig. Miscellaneous artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)	Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
129		0.78	1.203	136		−2.26	−1.862
130		−2.09	−1.755	137		−0.24	0.180
131		−1.27	−0.802	138		−2.59	−2.167
132		0.23	0.587	139		−0.96	−0.559
133		−0.67	−0.353	140		−0.79	−0.370
134		−4.00	−3.543	141		−0.64	−0.197
135		−4.00	−3.567	142		−0.353	0.090

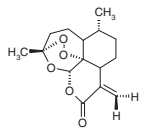
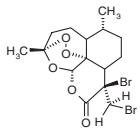
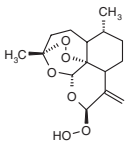
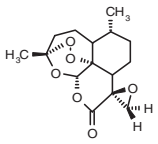
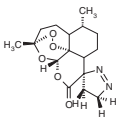
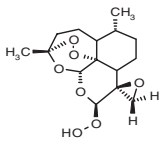
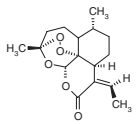
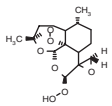
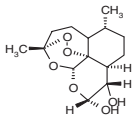
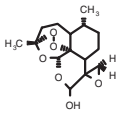
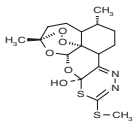
applying missing value test, zero test, correlation test with a cutoff value of 0.6, and multicollinearity test with a cutoff value of 0.9 we have discarded the most likely parameters that resulted in 117 parameters. Further additional parameters were discarded by applying genetic algorithm and finally 13 parameters were selected for the development of the QSAR equation. At first step, the QSAR equation was developed using only one parameter (V7CH), which showed significant correlation with the biological activity in comparison to the remaining 12 parameters. Taking a brute force approach, we increased the number of parameters in the QSAR equation developed at the first step one by one and evaluated the effect of addition of new term on the statistical quality of the model. As the correlation coefficient, r^2 can be easily increased by the number of terms in the QSAR equation; we took the cross-validation correlation coefficient, q^2_{cv}, as the limiting factor for a number of descriptors to be used in the final model. It was observed that the q^2_{cv} value increased until the number of descriptors in the equation reached up to 13 as shown in Table IV. With further addition of parameters to equation 13 (in Table IV), there was a decrease in the q^2_{cv} value of the model. Therefore, the number of descriptors was restricted to 13 in the final QSAR model. The best significant relationship between the molecular descriptors and antimalarial activity has been deduced to

$$\begin{aligned} pIC_{50} = {} & -1.34 - 11.3V7CH + 0.161EMAX1 + 0.352LOGP \\ & - 0.668GEOM3 + 0.0002STRA6 + 0.004STRA4 \\ & + 0.054STRA2 - 0.875L/B2 - 9.92FVMN \\ & - 3.25HOMO - 2.53BOMX + 0.475MOLC9 \\ & - 9.89V6C \end{aligned} \quad (1)$$

$$(N = 154; r^2 = 0.777; s = 0.662; \text{PRESS} = 79.105;$$
$$r^2_{adj} = 0.756; q^2_{cv} = 0.713; F - \text{test} = 37.53)$$

It was found that the compound numbers **31**–**45** were outliers with prediction error in between 2.233 and 3.150. The reason for those compounds being found as outliers could probably be their very low activity. The antimalarial action of artemisinin appears to be mediated by the generation of free radicals from the

Table Ih. 9-Substituted artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)	Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
143		−0.729	−0.328	149		−1.487	−1.282
144		−0.739	−0.365	150		−1.926	−1.549
145		−2.447	−2.106	151		−0.460	−0.109
146		−0.198	0.182	152		−0.409	−0.058
147		−0.717	−0.325	153		−0.361	0.013
148		−2.469	−2.207				

endoperoxide bridge of the drug [30,47]. All these compounds belong to deoxyartemisinin derivatives (Table Ib) that lack the endoperoxide moiety. This group of compounds has single oxygen instead of the peroxide bridge. The interaction between heme and these analogs is mediated by three nonperoxide oxygens leading to inactivity or low activity [48,49]. The quality of the above QSAR model has been improved further by removing these compounds and is as follows:

$$\begin{aligned} pIC_{50} = {} & 4.04 - 9.76V7CH + 0.069EMAX1 + 0.117LOGP \\ & - 0.581GEOM3 - 0.00014STRA6 + 0.0035STRA4 \\ & + 0.039STRA2 - 1.09L/B2 - 4.82FVMN \\ & - 1.02HOMO - 0.285BOMX - 1.36MOLC9 \\ & - 5.24V6C \end{aligned} \quad (2)$$

$$(N = 141; r^2 = 0.845; s = 0.343; \text{PRESS} = 19.386;$$
$$r^2_{adj} = 0.830; q^2_{cv} = 0.799; F - \text{test} = 53.40)$$

where N is the number of compounds in the training set, r^2 is the squared correlation coefficient, s is the estimated standard deviation about the regression line, r^2_{adj} is the square of adjusted correlation coefficient for degree of freedom, F-test is the measure of variance which compares two models differing by one or more variables to see if the more complex model is more reliable than the less complex one, the model is supposed to be good if the F-test is above a threshold value, and q^2_{cv} is the square of the correlation coefficient of the cross validation. The QSAR model developed in this study is statistically ($r^2 = 0.845$, $q^2_{cv} = 0.799$, F-test = 53.40) best fitted and consequently used for the prediction of antimalarial activity (pIC_{50}) of training and test sets of molecules as reported in Tables V and VI. The quality of the prediction models for the training compounds before and after the removal of outliers has been shown in Figures 2 and 3. The r^2 and q^2_{cv} values of 0.845 and 0.799, respectively, of the model corroborates with the criteria for a QSAR model to be highly predictive [27]. The standard error of estimate for the model was 0.343, which is an indicator of the robustness of the fit and suggested that the predicted pIC_{50} based on Equation (2) is reliable.

Table Ii. Dihydroartemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

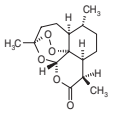
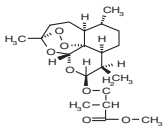
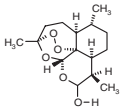
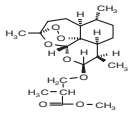
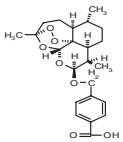
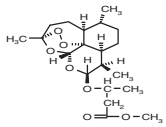
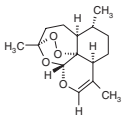
Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)	Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
154		−0.269	0.129	158		1.524	1.788
155		0.310	0.705	159		2.104	2.368
156		0.176	0.404	160		0.599	0.863
157		0.487	0.911				

Table Ij. Tricyclic 1,2,4-trioxanes derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

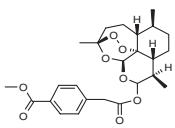
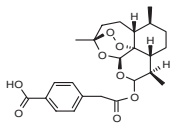
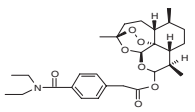
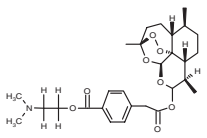
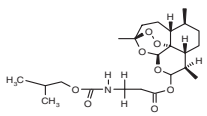
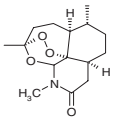
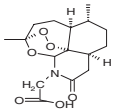
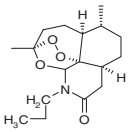
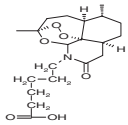
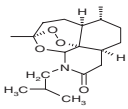
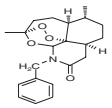
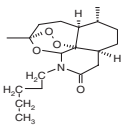
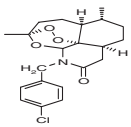
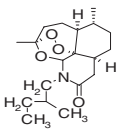
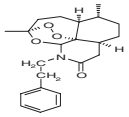
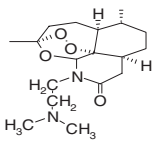
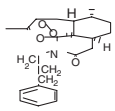
Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
161		0.660	0.845
162		−0.475	−0.275
163		0.551	0.699
164		0.205	0.340
165		0.312	0.503

Table Ik. *N*-Alkyl-11-aza-9-desmethylartemisinins derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)	Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
166		0.328	0.728	172		−1.222	−0.886
167		−0.125	0.233	173		−0.921	−0.652
168		0.161	0.500	174		0.276	0.572
169		0.041	0.362	175		0.045	0.301
170		0.173	0.494	176		0.294	0.573
171		−0.432	−0.114	177		0.312	0.574

V7CH is the seventh order chain molecular connectivity indexes. However, V6C measures sixth order cluster molecular connectivity indexes. These descriptors contain information about the size and the degree of branching in a molecule [41]. EMAX1 is the maximum atomic E-state index for each atom of each compound in the data set which is a measure of the reactivity of each atom, analogous to the concept of free valence. It provides information regarding intermolecular interactions [50]. L/B2 is the length-to-breadth ratio of compounds calculated by rotating the molecule in the *Z*-axis in increments of *N* degrees. MLog P calculate Octanol/Water Partition coefficient of the molecule based on the algorithm by [51]. It is the most popular and traditional. It explains one of the principal characteristics of any preparation, the lipophilicity. The higher its value, the more probable the transfer of the preparation from the aqueous medium into the biological membrane. This property is critical for medicinal preparations that are administered orally and must be absorbed through the GI tract. Log *P* value less than 0.5 will be absorbed appropriately. HOMO is the highest occupied molecular orbital energy (calculated using single point MOPAC(AMI)-based semiempirical quantum mechanical methods); this descriptor considers only interactions of valence π electrons for adjacent atoms. The descriptor BOMX is a measure of the electron density between adjacent atoms, representative to the strength of that bond. FVMN is a measure of the available bonding capacity left in an atom. It is considered to be a measure of the likelihood that the specified atom will be the site of a radical attack. The descriptors such as STRA2, STRA4, STRA6 comprise each of the strain energy terms like bond, torsional, and total energy terms of the molecules used in the molecular mechanics force field. The descriptor MOLC9 includes Balabans topological index J. It measures the degree of branching in structures [41]. GEOM3 is

Table II. 3C-substituted artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

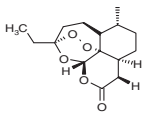
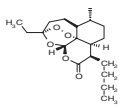
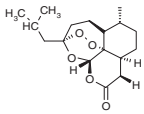
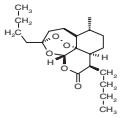
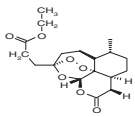
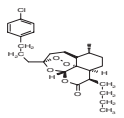
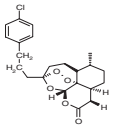
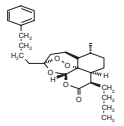
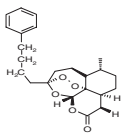
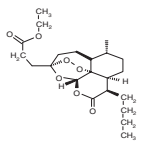
Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)	Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
178		0.049	0.447	183		0.410	0.729
179		−0.347	0.010	184		−0.481	−0.197
180		0.365	0.665	185		−0.276	−0.093
181		0.104	0.343	186		−0.319	−0.116
182		0.449	0.710	187		1.359	1.594

Table Im. Various derivatives of artemisinin and artemether with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work

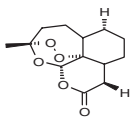
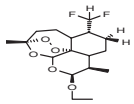
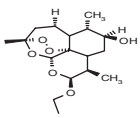
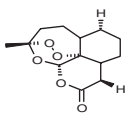
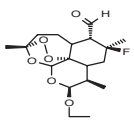
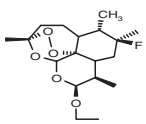
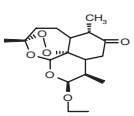
Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)	Compound no.	Analog structure	Log RA	pIC_{50} (ng/ml)
188		0.437	0.083	192		2.570	0.717
189		2.188	0.672	193		0.016	−1.347
190		1.622	0.504	194		−0.120	0.192
191		1.445	0.495				

Table II. Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees)

Parameters[a]	Theoretical			Experimental[d]	Experimental[e]
	3-21G[b]	3-21G**[c]	6-31G[c]		
O1-O2	1.463	1.462	1.447	1.475(4)	1.469(2)
O2-C3	1.441	1.440	1.435	1.417(4)	1.416(3)
C3-O4	1.436	1.436	1.435	1.448(4)	1.445(2)
O4-C5	1.407	1.408	1.403	1.388(4)	1.379(2)
C5-C6	1.529	1.530	1.533	1.528(5)	1.523(2)
C6-O1	1.478	1.477	1.469	1.450(4)	1.461(2)
O1-O2-C3	106.9	107.070	108.800	107.600(2)	108.100(1)
O2-C3-O4	107.0	107.310	106.760	107.200(2)	106.600(2)
C3-O4-C5	115.6	115.700	117.300	113.500(3)	114.200(2)
O4-C5-C6	112.0	112.030	112.280	114.700(2)	114.500(2)
C5-C6-O1	111.1	111.589	110.910	111.100(2)	110.700(2)
C6-O1-O2	111.2	111.286	113.240	111.500(2)	111.200(2)
O1-O2-C3-O4	−74.9	−74.680	−71.840	−75.500(3)	−75.500(2)
O2-C3-O4-C5	31.8	32.150	33.390	36.300(4)	36.000(2)
C3-O4-C5-C6	29.4	28.400	25.320	24.800(4)	25.300(2)
O4-C5-C6-O1	−51.8	−50.769	−49.410	−50.800(4)	−51.300(2)
C5-C6-O1-O2	10.1	9.792	12.510	12.300(3)	12.700(2)
C6-O1-O2-C3	50.8	50.522	46.700	47.700	47.800(2)

[a] Atoms are numbered according to Figure 1.
[b] This work.
[c] Values from Reference [26].
[d] Values from Reference [27] (experimental estimated standard deviations in brackets).
[e] Values from Reference [28] (experimental estimated standard deviations in brackets.

Table III. List of descriptors used in the study

Type	Descriptors
E-state indices	Electro-topological-state indices.
Electronic	Partial positive surface area, partial negative surface area, relative positive charge, relative negative charge, relative positive charged surface area, relative negative charged surface area, weighted positive charged partial surface area, weighted negative charged partial surface area, fractional negative charged partial surface area, fractional positive charged partial surface area, Huckel molecular orbital indices, highest occupied molecular orbital, lowest unoccupied molecular orbital, free valence value, nucleophilic superdelocalizability, free radical superdelocalizability, heat of formation, dipole moments, energy of the highest occupied orbital, energy of the lowest unoccupied orbital, electronegativity, hardness.
Information content	Information of atomic composition index, superpendentivity index
Spatial	Radius of gyration, Jurs descriptors, shadow indices, area, density, length-to-breadth ratios.
Structural	Topological symmetry, geometrical symmetry, combined symmetry, conformational flexibility indices, molecular distance edge descriptors, moment of inertia indices, geometric moment indices, number of single bonds, number of aromatic bonds.
Thermodynamic	Average energy, bond strain energy, angle strain energy, non-bonded strain energy, torsional strain energy, total strain energy of molecule.
Leadlikeness	LogP (Meylan, Howard), LogS, LogP(Moriguchi, Hirono).
Topological	Wiener index, Kier and Hall molecular connectivity indices, path count and length descriptors, topological polar surface area (TPSA), Balban indices.

Table IV. Statistical assessment of QSAR equations with varying number of descriptors

No. of descriptors	QSAR equation	r^2	Press	q^2
1	$pIC_{50} = 1.78 - 6.00$ V7CH	0.186	261.73	0.181
2	$pIC_{50} = 0.637 - 5.61$ V7CH $+ 0.095$ EMAX1	0.223	252.28	0.212
3	$pIC_{50} = 0.073 - 5.75$ V7CH $+ 0.084$ EMAX1 $+ 0.143$ LOGP	0.237	251.25	0.221
4	$pIC_{50} = 0.140 - 5.77$ V7CH $+ 0.085$ EMAX1 $+ 0.149$ LOGP $- 0.116$ GEOM3	0.237	253.10	0.217
5	$pIC_{50} = -0.489 - 6.09$ V7CH $+ 0.087$ EMAX1 $+ 0.192$ LOGP $- 0.471$ GEOM3 $+ 0.003$ STRA6	0.304	233.91	0.281
6	$pIC_{50} = -0.626 - 6.78$ V7CH $+ 0.074$ EMAX1 $+ 0.215$ LOGP $- 0.461$ GEOM3 $+ 0.002$ STRA6 $+ 0.004$ STRA4	0.319	243.36	0.292
7	$pIC_{50} = -1.04 - 7.67$ V7CH $+ 0.072$ EMAX1 $+ 0.216$ LOGP $- 0.710$ GEOM3 $+ 0.001$ STRA6 $+ 0.005$ STRA4 $+ 0.052$ STRA2	0.424	198.28	0.397
8	$pIC_{50} = 0.134 - 8.29$ V7CH $+ 0.061$ EMAX1 $+ 0.353$ LOGP $- 0.716$ GEOM3 $+ 0.001$ STRA6 $+ 0.006$ STRA4 $+ 0.046$ STRA2 $- 1.15$ L/B2	0.467	182.97	0.438
9	$pIC_{50} = -0.736 - 7.75$ V7CH $+ 0.057$ EMAX1 $+ 0.309$ LOGP $- 0.696$ GEOM3 $+ 0.001$ STRA6 $+ 0.005$ STRA4 $+ 0.039$ STRA2 $- 0.993$ L/B2 $- 4.95$ FVMN	0.516	150.18	0.486
10	$pIC_{50} = -1.45 - 6.95$ V7CH $+ 0.058$ EMAX1 $+ 0.348$ LOGP $- 0.748$ GEOM3 $+ 0.001$ STRA6 $+ 0.004$ STRA4 $+ 0.045$ STRA2 $- 1.05$ L/B2 $- 4.97$ FVMN $- 1.38$ HOMO	0.534	164.87	0.501
11	$pIC_{50} = 0.51 - 6.98$ V7CH $+ 0.055$ EMAX1 $+ 0.336$ LOGP $- 0.701$ GEOM3 $+ 0.001$ STRA6 $+ 0.005$ STRA4 $+ 0.046$ STRA2 $- 1.04$ L/B2 $- 5.90$ FVMN $- 1.38$ HOMO $- 2.67$ BOMX	0.542	163.93	0.507
12	$pIC_{50} = 3.88 - 8.40$ V7CH $+ 0.064$ EMAX1 $+ 0.115$ LOGP $- 0.673$ GEOM3 $+ 0.001$ STRA6 $+ 0.002$ STRA4 $+ 0.041$ STRA2 $- 1.08$ L/B2 $- 4.43$ FVMN $- 1.51$ HOMO $- 0.73$ BOMX $- 1.45$ MOLC9	0.606	140.13	0.573
13	$pIC_{50} = 3.34 - 8.98$ V7CH $+ 0.088$ EMAX1 $+ 0.172$ LOGP $- 0.461$ GEMO3 $+ 0.0001$ STRA6 $+ 0.004$ STRA4 $+ 0.038$ STRA2 $- 1.04$ L/B2 $- 4.78$ FVMN $- 0.816$ HOMO $- 1.18$ BOMX $- 1.12$ MOLC9 $- 4.89$ V6C	0.782	78.403	0.738

Table V. Observed and predicted activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) of training set of artemisinin derivatives

Compound No.	W-2 clone inhibition (pIC_{50})			Compound No.	W-2 clone inhibition (pIC_{50})		
	Observed	Predicted	Residual		Observed	Predicted	Residual
187	1.594	0.899	0.695	111	−1.429	−1.336	0.093
52	1.170	0.474	0.696	112	0.601	0.103	0.498
2	0.712	0.118	0.594	114	−1.558	−1.026	0.532
56	−0.740	0.137	0.877	115	0.929	0.687	0.242
188	0.083	0.354	0.271	116	0.829	1.133	0.304
1	1.398	1.016	0.382	31	−3.623	−3.346	0.723
4	1.188	0.518	0.670	32	−3.580	−3.079	0.299
5	0.228	0.492	0.264	41	−3.599	−2.759	0.840
6	−0.260	0.302	0.562	35	−3.660	−2.693	0.967
189	0.672	1.070	0.398	36	−3.620	−2.868	0.752
7	0.260	1.070	0.810	37	−3.662	−3.075	0.587
60	0.401	0.227	0.174	39	−3.553	−2.651	0.902
61	0.115	−0.372	0.487	40	−3.553	−2.771	0.782
63	0.654	0.446	0.208	42	−3.599	−3.407	0.308
64	0.674	0.450	0.224	43	−3.686	−3.220	0.134
8	0.508	0.295	0.213	118	1.039	0.848	0.191
9	−0.117	0.562	0.679	119	0.735	0.456	0.279
11	−0.195	0.495	0.690	120	−0.319	0.139	0.458
66	1.068	1.079	0.011	121	0.465	0.139	0.326
67	0.277	0.245	0.032	122	−0.586	0.139	0.725

(*Continues*)

Table V. (Continued)

Compound No.	W-2 clone inhibition (pIC_{50})			Compound No.	W-2 clone inhibition (pIC_{50})		
	Observed	Predicted	Residual		Observed	Predicted	Residual
68	0.804	0.379	0.425	123	0.400	−0.395	0.795
69	0.835	0.231	0.604	124	−0.446	−0.094	0.352
13	1.162	1.192	0.030	126	0.648	1.213	0.565
14	−0.212	−0.304	0.092	128	−0.647	−0.065	0.582
15	0.262	−0.270	0.532	134	−3.543	−3.060	0.483
16	0.317	0.325	0.008	135	−3.567	−3.197	0.370
17	0.389	−0.056	0.445	136	−1.862	−1.740	0.122
94	0.283	−0.121	0.404	138	−2.167	−1.693	0.474
95	0.439	0.174	0.265	139	−0.559	−0.526	0.033
96	0.121	−0.024	0.145	140	−0.370	−0.292	0.078
97	0.281	−0.243	0.524	141	−0.197	−1.011	0.814
98	0.096	−0.082	0.178	104	0.418	0.373	0.045
70	1.115	0.103	1.012	143	−0.328	0.654	0.982
129	1.203	0.023	1.180	145	−2.106	−1.313	0.793
72	0.800	0.794	0.006	146	0.182	0.253	0.071
73	0.364	0.966	0.602	147	−0.325	0.653	0.978
75	0.143	0.823	0.680	149	−1.282	−1.054	0.228
76	2.070	1.220	0.850	151	−0.109	−0.348	0.239
77	1.134	0.359	0.775	152	−0.058	−0.498	0.440
78	1.964	1.677	0.287	153	0.013	−0.295	0.308
79	1.017	0.211	0.806	154	0.129	0.320	0.191
80	−0.295	−0.380	0.085	155	0.705	0.827	0.122
18	0.340	−0.266	0.606	156	0.404	0.834	0.430
20	0.669	0.192	0.477	157	0.911	0.195	0.716
21	0.448	0.182	0.266	158	1.788	1.102	0.686
82	1.238	1.054	0.184	159	2.368	1.276	1.092
100	0.636	0.407	0.229	160	0.863	0.940	0.077
102	0.408	0.429	0.021	162	−0.275	0.664	0.939
23	−1.435	−1.080	0.355	165	0.503	0.785	0.282
83	−0.269	−1.067	0.798	167	0.233	−0.331	0.564
25	−0.022	0.098	0.120	168	0.500	−0.226	0.726
87	0.498	0.879	0.381	169	0.362	−0.617	0.979
26	1.339	0.284	1.055	170	0.494	−0.227	0.721
27	0.876	0.358	0.518	171	−0.114	−0.612	0.498
28	0.398	0.187	0.211	172	−0.886	−0.041	0.845
29	1.042	0.242	0.800	173	−0.652	−0.284	0.368
88	1.117	0.709	0.408	174	0.572	−0.134	0.706
30	1.487	0.620	0.867	177	0.574	−0.245	0.819
194	0.192	0.427	0.235	176	−0.385	−0.875	0.490
89	0.437	0.080	0.357	179	0.010	0.107	0.097
92	−0.108	0.358	0.466	180	0.665	0.317	0.348
191	0.504	0.895	0.391	181	0.343	−0.132	0.475
192	0.495	1.142	0.647	182	0.710	0.316	0.394
193	0.717	0.938	0.221	183	0.729	−0.054	0.783
93	0.497	0.659	0.162	184	−0.197	0.284	0.481
46	1.217	0.390	0.827	185	−0.093	0.475	0.568
47	−0.713	−0.056	0.657	186	−0.116	0.833	0.949
48	−0.183	−0.226	0.043	12	1.777	0.785	0.992
49	0.245	−0.163	0.408	38	−3.668	−2.900	0.768
50	−0.420	−0.302	0.118	33	−3.720	−2.759	0.961
105	−0.217	−0.021	0.196	44	−3.703	−2.850	0.853
132	0.587	0.660	0.073	190	−0.798	−0.861	0.063
106	0.202	−0.001	0.203	148	−2.207	0.019	2.226

(*Continues*)

Table V. (Continued)

Compound No.	W-2 clone inhibition (pIC_{50})			Compound No.	W-2 clone inhibition (pIC_{50})		
	Observed	Predicted	Residual		Observed	Predicted	Residual
107	0.180	0.513	0.333	150	−1.549	0.532	2.081
51	0.097	−0.226	0.323	130	−1.755	0.242	1.997
108	−0.550	−0.542	0.008	85	−0.616	1.152	1.768
109	0.007	0.395	0.388	3	−0.760	0.942	1.702
110	0.275	0.411	0.136	65	1.677	−0.223	1.900

Table VI. Observed and predicted activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) of test set of artemisinin derivatives

Compound No.	W-2 clone inhibition (pIC_{50})			Compound No.	W-2 clone inhibition (pIC_{50})		
	Observed	Predicted	Residual		Observed	Predicted	Residual
53	0.945	0.685	0.260	133	−0.353	−0.330	0.023
54	0.694	0.484	0.210	34	−3.599	−2.850	0.749
55	1.295	1.304	0.009	117	0.002	0.013	0.011
57	0.606	0.361	0.245	125	0.174	0.552	0.378
58	0.874	0.211	0.663	127	0.428	0.403	0.025
59	0.977	0.797	0.180	137	0.180	0.958	0.778
62	−0.564	−0.537	0.027	142	0.090	0.536	0.446
10	−0.097	−0.181	0.084	103	0.458	0.199	0.259
99	0.359	0.299	0.060	144	−0.365	−0.978	0.613
71	0.398	0.944	0.546	161	0.845	0.215	0.630
74	0.218	0.045	0.173	163	0.699	0.146	0.553
19	−0.383	−0.281	0.102	164	0.340	0.010	0.330
81	−0.580	−0.256	0.324	166	0.728	0.250	0.478
22	0.007	0.217	0.210	175	0.301	0.313	0.012
84	−0.219	−0.622	0.403	178	0.447	0.880	0.433
24	−0.489	−0.365	0.124	101	1.099	1.199	0.100
86	−0.519	−0.124	0.395	113	−1.449	−1.820	0.371
90	0.615	0.320	0.295	45	−3.577	−3.190	0.387
131	−0.802	−0.499	0.303	91	−0.415	−0.694	0.279

$pIC_{50} = -\log_{10} IC_{50}$.

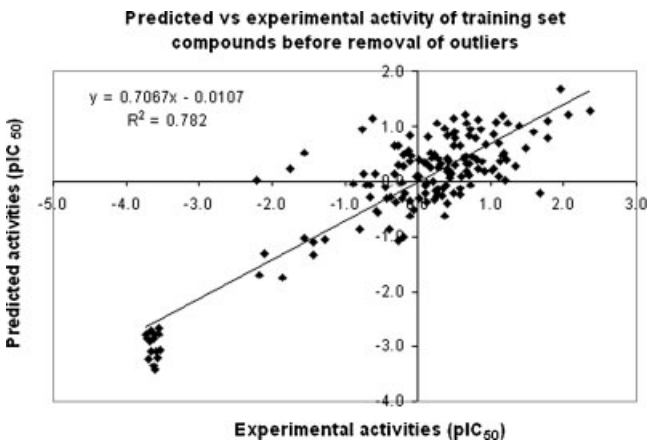


Figure 2. Relationship between predicted and experimental activities as per Equation (1) before removal of outliers.

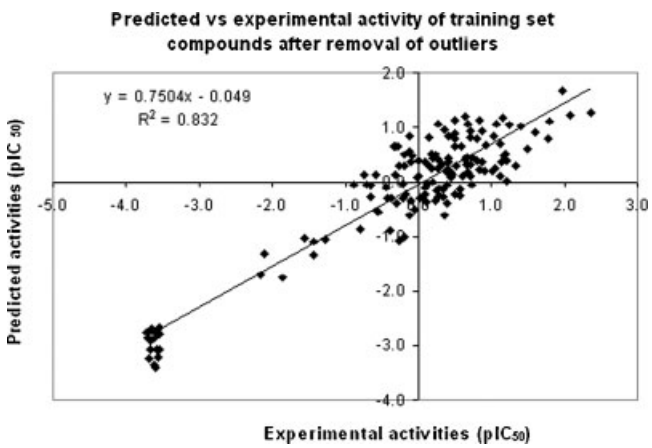


Figure 3. Relationship between predicted and experimental activities as per Equation (2) after removal of outliers (deoxy-artemisinin derivatives).

Table VII. Correlation matrix of the descriptors used in the QSAR model

	V7CH	EMAX1	Log P	GEOM3	STRA6	STRA4	STRA2	L/B2	FVMN	HOMO	BOMX	MOLC9	V6C
V7CH	1.00												
EMAX1	−0.14	1.00											
Log P	0.05	0.17	1.00										
GEOM3	−0.06	0.15	0.20	1.00									
STRA6	0.06	−0.01	−0.07	0.28	1.00								
STRA4	0.29	0.09	−0.11	0.14	0.63	1.00							
STRA2	0.15	−0.01	0.04	0.22	0.25	0.12	1.00						
L/B2	−0.12	0.03	0.43	0.07	0.01	0.06	−0.16	1.00					
FVMN	0.06	−0.10	−0.09	−0.01	0.13	−0.01	−0.15	0.08	1.00				
HOMO	0.37	−0.02	0.21	−0.09	−0.26	−0.14	0.21	0.32	−0.08	1.00			
BOMX	0.01	−0.01	−0.03	0.07	−0.08	0.07	0.14	−0.06	−0.44	0.05	1.00		
MOLC9	−0.41	−0.04	−0.50	−0.13	−0.27	−0.35	−0.21	−0.18	0.12	−0.21	0.11	1.00	
V6C	0.22	−0.07	0.03	0.11	−0.09	0.03	−0.02	−0.13	−0.10	0.17	0.12	−0.05	1.00

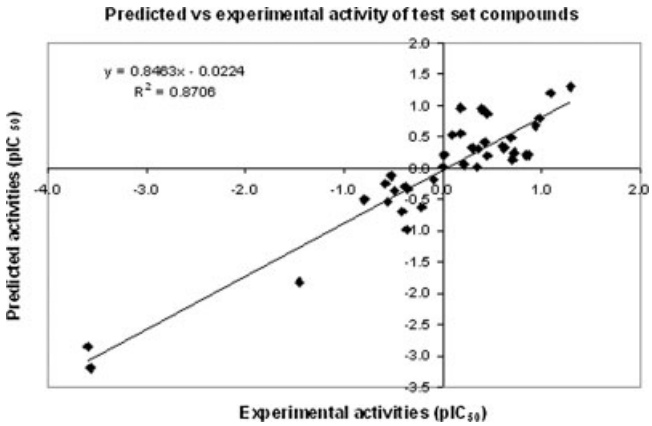


Figure 4. Relationship between predicted and experimental activities as per Equation (2).

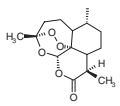
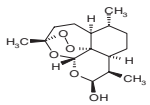
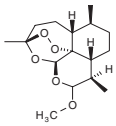
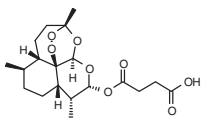
the mass weighted thickness descriptor. The calculation involves diagonalization of the covariance matrix formed from the (*x*, *y*, *z*) coordinates of the atoms, translated to the center of mass of the structure. The contribution of each atom is weighted by its mass.

The inter-correlation of the descriptors used in the final model (Equation 2) was very low (below 0.6) which is in conformity to the study that for a statistically significant model, it is necessary that the descriptors involved in the equation should not be inter-correlated with each other [18]. The correlation matrix for the used descriptors is shown in Table VII. To further check the inter-correlation of descriptors VIF analysis was performed. In this model, the VIF values of these descriptors are 1.70 (V7CH), 1.152 (EMAX1), 2.079 (Log P), 1.271 (GEOM3), 2.252 (STRA6), 2.331 (ATRA4), 1.344 (STRA2), 1.398 (L/B2), 1.441 (FVMN), 1.479 (HOMO), 1.402 (BOMX), 2.257 (MOLC9), and 1.155 (V6C). Based on the VIF analysis it has been found that the descriptors used in the final model have very low inter-correlation.

Satisfied with the robustness of the QSAR model developed using training set, we have applied the QSAR model to an external data set of artemisinin analogs comprising the test set. As the experimental values (pIC_{50}) for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the QSAR model for new ligands. Table VI presents the predicted pIC_{50} values of the test set based on Equation (2). The overall root mean square error (RMSE) between the experimental and predicted pIC_{50} value was 0.325 which revealed good predictability. The squared correlation coefficient between experimental and predicted pIC_{50} values for the test set is also significant ($r^2 = 0.871$). Figure 4 shows the quality of the fit. The estimated correlation coefficients between experimental and predicted pIC_{50} values with intercept (r^2) and without intercept (r_0^2) are 0.871 and 0.862, respectively. The value of $[(r^2 - r_0^2)/r^2] = (0.871 - 0.862)/0.871 = 0.010$, which is less than 0.1 (stipulated value) [45]. Also, the values of k and k' were 1.028 and 0.847, which are well within the specified range of 0.85 and 1.15 [45]. The values of $q^2_{\text{test}} = 0.876$ and $rm^2 = 0.788$ were found to be in the acceptable range [46], thereby indicating the good external predictability of the QSAR model.

To evaluate the accuracy of the QSAR model for antimalarial activities, we have taken a separate data set called validation set consisting of four analogs of artemisinin (Table VIII). Their experimental activity and chemical structures were obtained from the literature [52]. The experimental activity (IC_{50} value) of these compounds was obtained from *in vitro* study in parasitized whole blood (human) against drug-resistant strains of *P. falciparum* (W-2 clone) [53,54]. The W-2 clone is chloroquinone resistant. For all the compounds, QSAR predictions produce exactly the same trend for antimalarial activity, even though the exact magnitudes of these values do not match very well to experimental values (Table VIII). Coupled with the good predictive ability of the QSAR model developed in this study, we believe that this model would perform well as rapid screening tools to uncover new and more potent antimalarial drugs based on artemisinin derivatizations.

Table VIII. Observed and predicted activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) of validation set of artemisinin derivatives

Compound name	W-2 Clone line inhibition (pIC_{50})			
	Structure	Observed	Predicted	Residual
Artemisinin		1.004	0.228	0.776
Dihydroartemisinin		0.694	0.480	0.213
Artemether		1.638	0.606	1.03
Artesunuate		0.259	1.119	0.861

4. CONCLUSION

The QSAR analysis of a series of artemisinin derivatives enabled consistent models of structure–activity relationships to be obtained for several descriptors. The models that had the best predictive ability contained topological, theromodynamic, electronic, E-state indices, and physicochemical descriptors. In this study, we used a more systematic way of variable selection in order of missing value test → zero test → simple correlation test → multicollinearity test → genetic algorithm to obtain QSAR models for 194 artemisinin derivatives. Using a combination of topological, electro-topological state indices, electronic, and thermodynamic descriptors of chemical structures, we have built several robust QSAR models with high values of q^2 (for training sets) and predictive r^2 (for test set). The high predictive ability of the models allows virtual screening of chemical databases or virtual libraries determined by either synthetic feasibility or commercial availability of starting materials to prioritize the synthesis of most promising candidates. Therefore, these models should facilitate the rational design of novel derivatives, guide the design of focused libraries based on the artemisinin skeleton, and facilitate the search for related structures with similar biological activity from large databases.

REFERENCES

1. Haynes RK, Vonwille SC. From Qinghao, marvelous herb of antiquity, to the antimalarial trioxane Qinghaosu—and some remarkable new chemistry. *Acc. Chem. Res.* 1997; **30**: 73–79.
2. Klayman DL. Qinghaosu (artemisinin): an antimalarial drug from *China*. *Science* 1985; **228**: 1049–1055.
3. Kamchonwongpaisan S, Meshnick SR. The mode of action of the antimalarial artemisinin and its derivatives. *Gen. Pharmac.* 1996; **27**: 587–592.
4. Posner GH, Cumming JN, Ploypradith P, Oh CH. Evidence for Fe(IV)O in the molecular mechanismof action of the trioxane antimalarial artemisinin. *J. Am. Chem. Soc.* 1995; **117**: 5885–5886.
5. Posner GH, Park SB, Gonzalez L, Wang D, Cumming JN, Klinedinst D, Shapiro TA, Bachi MD. Evidence for the importance of high valent FeO and of a diketone in the molecular mechanism of antimalarial trioxane analogs of artemisinin. *J. Am. Chem. Soc.* 1996; **118**: 3537–3538.
6. Robert A, Meunier B. Is alkylation the main mechanism of action of the antimalarial drug atemisinin*? Chem. Soc. Rev.* 1998; **27**: 273–274.
7. Avery MA, Alvin-Gaston M, Rodrigues CR, Barreiro EJ, Cohen FE, Sabnis YA, Woolfrey JR. Structure activity relationships of the antimalarial agent artimisinin.The development of predictive *in vitro* potency models using CoMFA and HQSAR methodologies. *J. Med. Chem.* 2002; **45**: 292–303.
8. Tommuphean S, Kokpol S, Parasuk V, Wolschann P, Winger RH, Liedl KR, Rode BM. Comparative molecular field analysis of artemisinin derivatives: ab intio versus semiempirical optimized structures. *J. Comput. Aided Mol. Des.* 1998; **12**: 397–409.
9. Avery MA, Gao F, Chong WK, Mehrotra S, Milhous WK. Structure–activity relationships of the antimalarial agent artemisinin. 1. Synthesis and comparative molecular field analysis of C-9 analogs of artemisinin and 10-deoxoartemisinin. *J. Med. Chem.* 1993; **36**: 4264–4275.
10. Cramer RD III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 1988; **110**: 5959–5967.
11. Cramer RD III, Patterson DE, Bunce JD, Bunce JD, Frank IE. Cross-validation, bootstrapping and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct. Act. Relat. Pharmacol. Chem. Soc.* 1998; **7**: 18–25.
12. Hansch C, Kurup A, Garg R, Gao H. Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. *Chem. Rev.* 2001; **101**: 619–672.
13. Livingstone DJ. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* 2000; **40**: 195–209.
14. Shi LM, Fan Y, Myers TG, Paul JN. Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR

study of anticancer ellipticine analogs. *J. Chem. Inf. Comput. Sci.* 1998; **38**: 189–199.
15. Oloff S, Mailman RB, Trospha A. Application of validated QSAR models of D_1 dopaminergic antagonists for database mining. *J. Med. Chem.* 2005; **48**: 7322–7332.
16. Meneses-Marcel A, Marrero-Ponce Y, Machado-Tugores Y, Monterro-Torres A, Pereira DM, Escario JA, Nogel-Ruiz JJ, Ochao C, Aran VJ, Martinez-Fernadez AR, Garcia Sanchez RN. A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results. *Bioorg. Med. Chem. Lett.* 2005; **15**: 3838–3843.
17. Santana L, Uriarte H, Gonzalez-Diaz H, Zagotto R, Soto-Otero E, Mendez-Alvarez J. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J. Med. Chem.* 2006; **49**: 1149–1156.
18. Deswal S, Roy N. Quantitative structure activity relationship studies of aryl heterocycle-based thrombin inhibitors. *J. Med. Chem.* 2006; **41**(11): 1339–1346.
19. Woolfrey JR, Avery MA, Doweyko AM. Comparison of 3D quantitative structure–activity relationship methods: analysis of the in vitro antimalarial activity of 154 artemisinin analogs by hypothetical active-site lattice and comparative molecular field analysis. *J. Comp. Aided Mol. Design.* 1998; **12**: 165–181.
20. Acton N, Karle JM, Miller RE. Synthesis and antimalarial activity of some 9-substituted artemisinin derivatives. *J. Med. Chem.* 1993; **36**: 2552–2557.
21. Lin AJ, Lee M, Klayman DL. Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives. *J. Med. Chem.* 1989; **32**: 1249–1252.
22. Posner GH, Oh CH, Gerena L, Milhous WK. Extraordinarily potent antimalarial compounds: new, structurally simple, easily synthesizes, tricyclic 1,2,4-trioxanes. *J. Med. Chem.* 1992; **35**: 2459–2467.
23. Avery MA, Bonk JD, Mehrotra S, Chong WK, Miller R, Milhous W, Goins DK, Venkatesan S, Wyandt C, Khan I, Avery BA. Structure–activity relationships of the antimalarial agent artemisinin. *J. Med. Chem.* 1995; **38**: 5038–5044.
24. Avery MA, Mehrotra S, Bonk JD, Vroman JA, Goins DK. Structure–activity relationships of the antimalarial agent artemisinin. *J. Med. Chem.* 1996; **39**: 2900–2906.
25. Jaguar, version 4.1: *Schrodinger*. 2000; Inc.: Portland, OR.
26. Pinheiro JC, Ferreira MMC, Romero OAS. Antimalarial activity of dihydroartemisinin derivatives against *P. falciparum* resistant to mefloquine: a quantum chemical and multivariate study. *J. Mol. Struct.* 2001; **572**: 35–44.
27. Leban I, Golic L, Japelj M. Crystal and molecular structure of qinghaosu: a redetermination. *Acta. Pharm. Jugosl.* 1988; **38**: 71–77.
28. Lisgarten JN, Potter BS, Bantuzeko C, Palmer RA. Structure, absolute configuration, and conformation of the antimalarial compound, artemisinin. *J. Chem. Cryst.* 1998; **28**: 539–543.
29. Bernardinelli G, Jefford CW, Maric D, Thomson C, Weber J. Computational studies of the structures and properties of potential antimalarial compounds based on the 1,2,4-trioxane ring structure:I. Artemisinin-like molecules. *Int. J. Quant. Chem.: Quant. Biol. Symp.* 1994; **21**: 117–131.
30. Posner GH, Cumming JN, Ploypradith P, Oh CH. Evidence for Fe(IV)AO in the molecular mechanism of action of the trioxane antimalarial artemisinin. *J. Am. Chem. Soc.* 1995; **117**: 5885–5886.
31. Posner GH, Wang D, Cumming JN, Ploypradith P, Oh CH, French AN, Bodley AL, Shapiro TA. Further evidence supporting the importance of and the restrictions on a carbon-centered radical for high antimalarial activity of 1, 2, 4-trioxanes like artemisinin. *J. Med. Chem.* 1995; **38**: 2273–2275.
32. Haynes RK, Vonwiller SC. The behaviour of qinghaosu (artemisinin) in the presence of hemin iron (II) and (III). *Tetrahedron. Lett.* 1996; **37**: 253–256.
33. Rafiee MA, Hadipour NL, Naderi-manesh H. The role of charge distribution on the antimalarial activity of artemisinin analogs. *J. Chem. Inf. Model.* 2005; **45**: 366–370.
34. Gregorio deC, Kier LB, Hall LH. QSAR modeling with the electrotopological state indices: corticosteroids. *J. Comput. Aid. Mol. Des.* 1998; **12**: 557–561.
35. Meylan WM, Howard PH. Atom/fragment contribution method for estimating octanolwater partition coefficients. *J. Pharm. Sci.* 1995; **84**(1): 83–92.
36. Gupta S, Singh M, Madan AK. Superpendentic index: a novel topological descriptor for prediction of biological activity. *J. Chem. Inf. Comput. Sci.* 1999; **39**: 272–277.
37. Liu SS, Cao CZ, Li ZL. Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance edge (MDE) vector λ. *J. Chem. Inf. Comput. Sci.* 1998; **38**: 387–394.
38. Lipinski A, Lombardo F, Dominy B, Feeney P. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* 2001; **46**(1–3): 3–26.
39. Eliopoulos GM, Moellering RC Jr, Antimicrobial combinations: In *Antibiotics in Laboratory Medicine*, V Lorian (ed.). (4th edn), Williams Wilkins, Baltimore, 1996; 330–396.
40. Brenwald NP, Gill MJ, Wise R. Prevalence of a putative efflux mechanism among fluoroquinolone-resistant clinical isolates of *Streptococcus pneumoniae*. *Antimicrob. Agent. Chemother.* 1998; **42**: 2032–2035.
41. Kier LB, Hall LH. *Molecular Connectivity in Chemistry and Drug Research*. Academic Press: New York, 1976.
42. Katritzky AR, Mu Lan, Lobanov VS, Karelson M. Correlation of boiling points with molecular structure. I A training set of 298 diverse organics and a test set of 9 simple inorganics. *J. Phys. Chem.* 1996; **100**: 10400–10407.
43. Golbraikh A, Tropsha A. Beware of q^2. *J. Mol. Graph. Model.* 2002; **20**(4): 269–276.
44. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* 2008; **27**: 302–313.
45. Jaiswal M, Khadikar PV, Scozzafava A, Supuran CT. Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorg. Med. Chem. Lett.* 2004; **14**: 3283–3290.
46. Shapiro S, Guggenheim B. Inhibition of oral bacteria by phenolic compounds. Part 1 QSAR analysis using molecular connectivity. *Quant. Struct. Act. Relat.* 1998; **17**: 327–337.
47. Jefford CW, Vicente MGH, Jacquier Y, Favarger F, Mareda J, Millasson-Schmidt P, Brunner G, Burger U. The deoxygenation and isomerization of artemisinin and artemether and their relevance to antimalarial action. *Helvetica Chim Acta* 1996; **79**: 1475–1487.
48. Meshnick SR, Tsang TW, Lin FB, Pan HZ, Chang CN, Kuypers F, Chiu D, Lubin B. Activated oxygen mediates the antimalarial activity of qinghaosu. *Prog. Clin. Biol. Res.* 1989; **313**: 95–104.
49. Shukla KL, Gund TM, Meshnick SR. Molecular modeling studies of the artemisinin (qinghaosu)-hemin interaction: docking between the antimalarial agent and its putative receptor. *J. Mol. Graph.* 1995; **13**: 215–222.
50. Kier LB, Hall LH. The E-state as an extended free valence. *J. Chem. Inf. Comput. Sci.* 1997; **37**: 548–552.
51. Moriguchi I, Hirono S, Liu Q, Nakagome I, Matsushita Y. Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* 1992; **40**(1): 127–130.
52. Darren JC, William NC, Francis CKC, Richard JP, Yuxiang D, Jonathan LV, Susan AC. Relationship between antimalarial activity and heme alkylation for spiro- and dispiro-1,2,4-trioxolane antimalarials. *Antimicrob. Agent. Chemother.* 2008; **52**: 1291–1296.
53. Desjardins RE, Canfield CJ, Haynes DE, Chulay JD. Quantitative assessment of antimalarial activity in vitro by a semiautomatyed microdilution technique. *Antimicrob. Agent. Chemother.* 1979; **16**: 710–718.
54. Milhous WK, Weatherley NF, Bowdre JH, Desjardins RF. In vitro activities of and mechanisms of resistance to antifolate antimalarial drug. *Antimicrob. Agent. Chemother.* 1985; **27**: 525–530.

