Available online at www.sciencedirect.com

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry

Bioorganic & Medicinal Chemistry 14 (2006) 2333–2357

Antimalarial activity: A QSAR modeling using CODESSA PRO software

Alan R. Katritzky,[a,*] Oleksandr V. Kulshyn,[a] Iva Stoyanova-Slavova,[a] Dimitar A. Dobchev,[a,c] Minati Kuanar,[a] Dan C. Fara[a] and Mati Karelson[b]

[a] *Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, FL 32611, USA*
[b] *Institute of Chemistry, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia*
[c] *Department of Chemistry, University of Tartu, 2 Jakobi street, Tartu EE2400, Estonia*

Received 21 August 2005; revised 4 November 2005; accepted 8 November 2005
Available online 19 January 2006

Abstract—A quantitative structure–activity relationship (QSAR) modeling of the antimalarial activity of two diverse sets of compounds for each of two strains D6 and NF54 of *Plasmodium falciparum* is presented. The molecular structural features of compounds are presented by molecular descriptors (geometrical, topological, quantum mechanical, and electronic) calculated using the CODESSA PRO software. Satisfactory multilinear regression models were obtained for data sets of the D6 and NF54 strains, with $R^2 = 0.84$ and 0.89, respectively. The models were also satisfactorily validated internally. The descriptors involved in these equations were related to the mechanism of antimalarial protection.

1. Introduction

Malaria, well known as an infectious lethal disease since ancient times, remains a major cause of death. Spread by *soporoza* of the genus *plasmodium*, it is characterized clinically by periodic fever, anemia, and enlargement of the liver and spleen. Hundreds of millions of new clinical cases arise annually with a high percentage of fatalities, especially among children[1], in the tropical and subtropical countries of Asia, Africa, and South America.

Only a limited number of drugs can now prevent and cure malaria: the most common being artemisine. There are numerous studies on the artemisine activity and synthesis.[2–4] A few groups of potentially antimalarial drugs are used as chemotherapeutics. These include: (i) quinoline derivatives, e.g., primaquine, chloroquine, and mefloquine[5], and/or (ii) simple sulfonamides, e.g., sulfadoxine[6,7], pyrimidine derivatives; pyrimethamine.[8]

The clones of *Plasmodium falciparum* used most often for in vitro testing of the antimalarial activity are : (i) Sierra Leone (D6), Thailand (Thai), and NF54 clones, all of which are mefloquine resistant and chloroquine sensitive, and (ii) Indochina (W2), and Colombia (FcB1) clones, which are chloroquine resistant, but mefloquine sensitive. With a wide range of molecular structures and their complementary activities, it has been assumed that the most important criteria for a systematic study of 3D quantitative structure–activity relationships (QSAR) have been satisfied. Perhaps, as a consequence, few QSAR studies on antimalarial activity have been reported during the last 15 years.

Grigorov et al.[1] correlated the antimalarial activity of a series of synthetic 1,2,4-trioxanes with molecular structure by the pharmacophore search method and the CATALYST package.[9] Either IC_{90} values against *P. falciparum* W2 (Indochina) and D6 (Sierra Leone) clones in vitro or ED_{90} values against *P. herghei* N in vivo for 28 organic compounds were measured. The authors divided the full data into 2 subsets as follows: (i) a training set (21 datapoints), and (ii) a test set (7 datapoints), and analyzed the assessment of 3D-QSAR for both in vitro and in vivo antimalarial activities. It was shown that (i) the in vitro and in vivo hypotheses share the same three features and are very nearly superimposable because the training set comprises similar, conformationally mobile *cis*-fused bicyclic structures, (ii) two hydrophobic sites and a hydrogen

Keywords: QSAR; Antimalarial activity; CODESSA-PRO; Multilinear regression.

* Corresponding author. Tel.: +1 352 392 554; fax: +1 352 392 9199; e-mail: katritzky@chem.ufl.edu

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.11.015

acceptor site located on the drug seem to be essential for antimalarial activity, and (iii) the peroxide bond of the trioxane lies close to the central ferrous atom of heme thereby fulfilling a criterion for parasiticidal action.

The automated molecular dockings of 30 artemisinin derivatives to heme, performed by Tonmunphean et al.[10], revealed that biological activity was significantly related to binding energy ($r = -0.93$) and less significantly to the O1–Fe distance ($r = -0.55$). The authors used relative activities (ratios of the activity of a compound to that of artemisinin) to reduce the inconsistencies from individual experimental environments. These activities were measured in vitro against the D6 and the W2 clones of *P. falciparum*. Based on the docking results, and to investigate the predictivity of the binding energy for antimalarial activity, QSAR models were constructed for both activities using the multiple linear regression method. The predicted biological activities of five additional artemisinin derivatives were in good agreement with the corresponding experimental values. By docking, Tonmunphean et al.[10] also observed that the artemisinin compounds approach heme by pointing O1 at the endoperoxide linkage toward the iron center, a mechanism controlled by steric hindrance.

Gironés et al.[11,12] studied the application, within a quantum similarity framework, of the kinetic energy-based quantum similarity measures in the evaluation of the antimalarial activity. The authors used two molecular sets composed of artemisinin derivatives, in which the 50% inhibition of synthesis and reduction of hidrofolate (IC_{50}) in different *P. falciparum* clone were analyzed. Satisfactory correlations were obtained for all antimalarial activities in all the studied molecular sets. Four-parameter QSAR models that relate IC_{50} (NF54 clone, $r = 0.754$) and $\log IC_{50}$ (D6 clone, $r = 0.767$, and W2 clone, $r = 0.821$) with the principal components (PCs) were proposed by the authors to be used for the prediction of antimalarial activity. The authors observed that the PCs accounting for the maximal variance are not always those best related to the activity.

Roy et al.[13] used the Wang–Ford charges of the non-hydrogen common atoms, obtained from molecular electrostatic potential surface of the energy minimized geometries, to model the antimalarial activity against *P. falciparum* for a series of 20 antimalarial cyclic peroxy ketals. The proposed QSAR model is a three-parameter regression equation ($r^2 = 0.791$) that correlates $\log(1/IC_{50})$ with the Wang–Ford charges of the (i) two oxygen atoms of the peroxide bridge, and (ii) the methoxy carbon of the common fragment, and an additional parameter I, which denotes the presence or absence of a seven-membered alicyclic ring attached to the peroxy bridge-containing ring. The authors stated that the difference in charges between the peroxy oxygens (i) contributes positively to the activity and (ii) possibly facilitates the bond breakage of the peroxy bridge by the heme-ion within the parasite. They also concluded that the activity increase with the increase in negative charge of the methoxy carbon of the common fragment of the molecule is related to a possible secondary electronic interaction with the positively charged side chains of the histidine-rich protein of *P. falciparum*.

Quantitative structure–activity relationship (QSAR) methodology is an essential tool in medicinal chemistry.[14] In recent years, methodology for a general QSAR approach has been developed and encoded in the CODESSA PRO software package.[15] CODESSA PRO enables the calculation of numerous quantitative descriptors solely on the basis of molecular structural information. Research using CODESSA PRO has successfully correlated and predicted many diverse physico-chemical properties. Recently, the QSAR treatments of (i) in vitro minimum inhibitory concentration (MIC) of 3-aryloxazolidin-2-one antibacterials required to inhibit growth of *Staphylococcus aureus*[16] and (ii) study on milk to plasma concentration ratio[17] have been reported.

In the present work, the aim is to develop QSAR models and to explain the antimalarial activity of various organic compounds against *P. falciparum* various clones (D6, NF54, etc.) in vitro using theoretical molecular descriptors.

2. Results

2.1. Data set

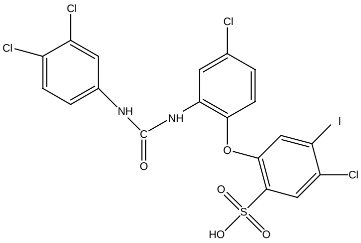
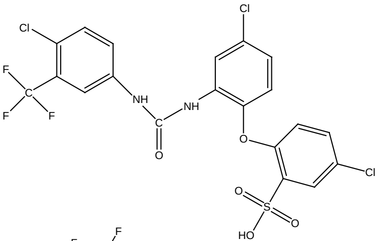
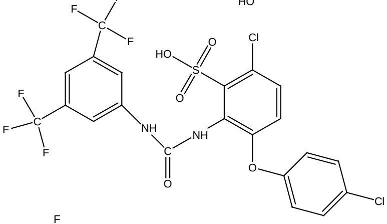
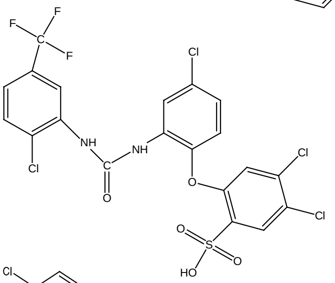
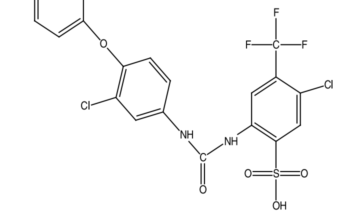
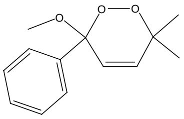
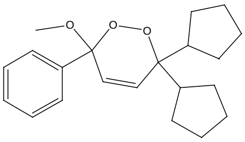
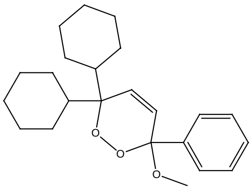
Two data sets of $\log IC_{50}$ were taken for the QSAR studies. The data set based on the D6 strain of *P. falciparum* consisting of 57 organic compounds was collected from four literature sources.[12,18–20] The second data set based on the NF54 strains of *P. falciparum* consisting of 69 organic compounds was collected from another four sources.[4,11,13,21] These data sets are shown in Tables 1 and 2 together with their log IC_{50} (nM) values.

2.2. Methodology

The structures of the compounds were drawn using ISIS/Draw as implemented in the ISIS 2.4 package and pre-optimized using Molecular Mechanics Force Fields (MM+).[22] The molecular geometries were refined using AM1 Hamiltonian (Austin Method 1) calculations together with eigenvector following a geometry optimization procedure available in the quantum chemical program MOPAC 7.05 and implemented in the CODESSA PRO package.[15] The gradient norm criterion 0.01 kcal/Å was applied in the geometry optimization for all structures.

The CODESSA PRO package was used to calculate up to 961 different molecular descriptors, derived solely from the molecular structure and divided into the following classes: (i) constitutional, (ii) geometrical, (iii) topological, (iv) electrostatic, (v) quantum chemical, and (vi) thermodynamic. These descriptors are based on the molecular geometry, LCAO MO wave, and ther-

Table 1. Experimental and predicted log IC_{50} according to the multilinear QSAR model in Table 3 of 57 compounds for D6 strain

Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
1		1.73	1.85	18
2		1.58	1.80	18
3		1.08	1.02	18
4		1.37	1.61	18
5		3.04	3.30	12
6		2.28	2.36	12
7		2.45	2.08	12
8		2.34	1.81	12

(*continued on next page*)

Table 1 (*continued*)

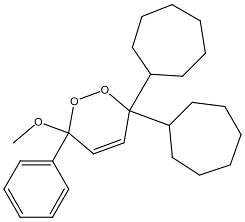
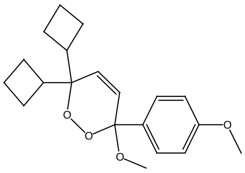
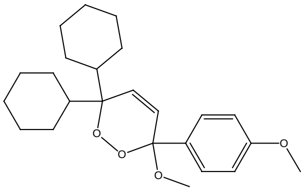
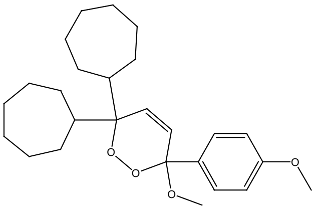
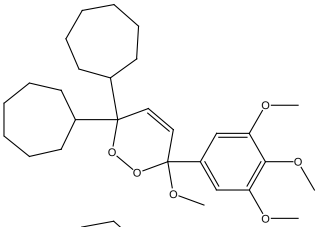
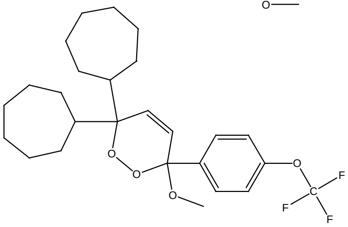
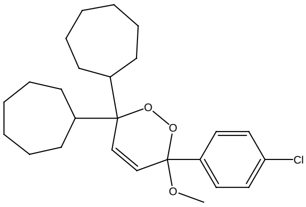
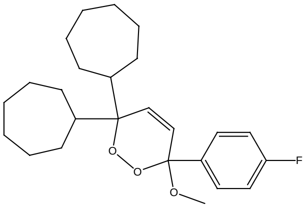
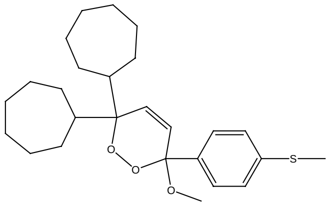
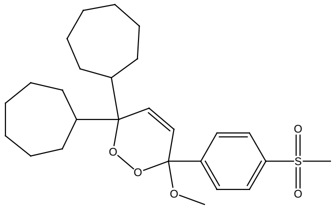
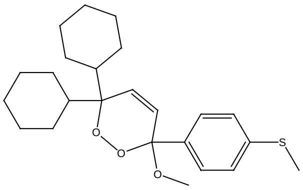
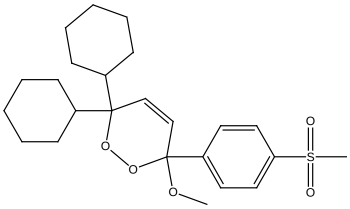
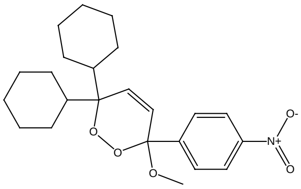
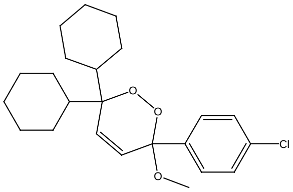
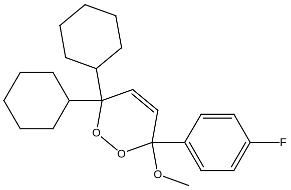
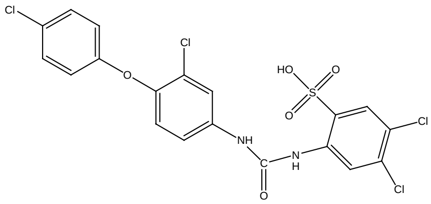
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
9		2.20	2.94	12
10		2.26	2.31	12
11		2.32	2.03	12
12		2.08	2.04	12
13		1.79	1.78	12
14		1.76	1.69	12
15		1.93	1.90	12
16		1.89	2.01	12

Table 1 (*continued*)

Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
17		1.49	1.94	12
18		2.20	2.28	12
19		1.75	2.26	12
20		1.66	2.24	12
21		2.00	1.97	12
22		2.30	2.19	12
23		1.68	1.54	18
24		1.70	1.56	18

(*continued on next page*)

Table 1 (*continued*)

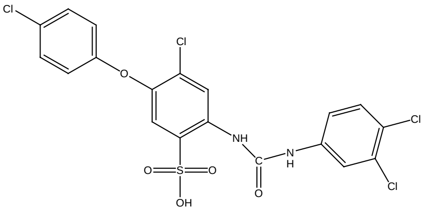
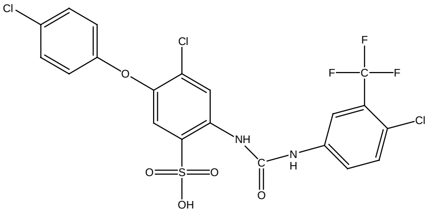
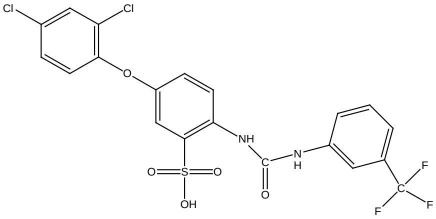
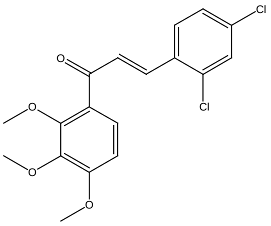
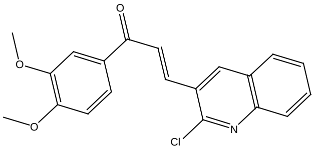
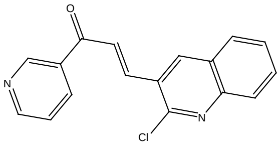
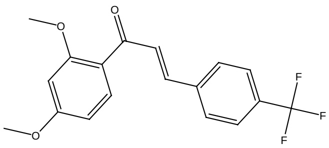
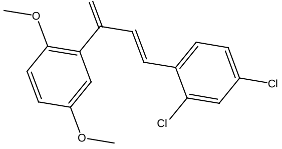
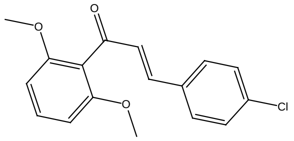
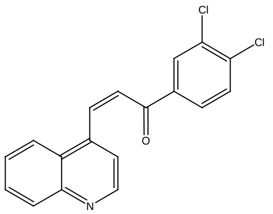
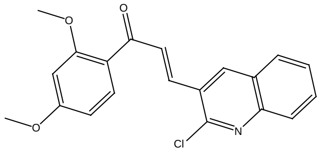
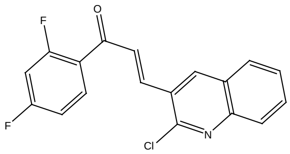
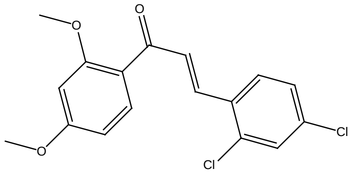
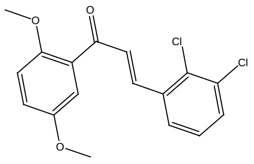
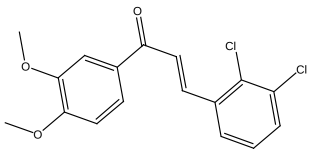
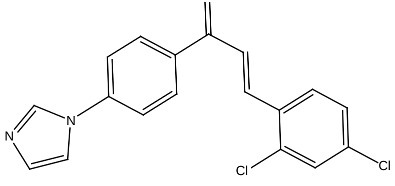
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
25		1.62	1.73	18
26		1.58	2.26	18
27		1.53	1.16	18
28		2.28	2.79	19
29		2.63	3.26	19
30		2.86	2.65	19
31		3.40	3.30	19
32		2.83	3.11	19

Table 1 (*continued*)

Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
33		2.96	3.11	19
34		3.09	2.75	19
35		3.23	3.36	19
36		3.11	2.67	19
37		2.99	3.07	19
38		2.89	3.04	19
39		3.04	2.99	19
40		3.29	3.11	19

(*continued on next page*)

Table 1 (*continued*)

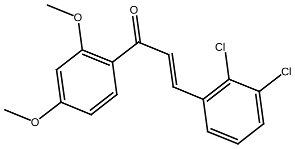
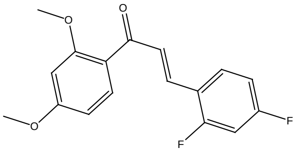
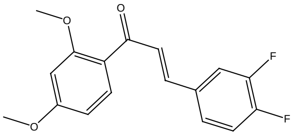
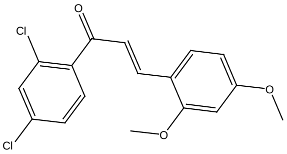
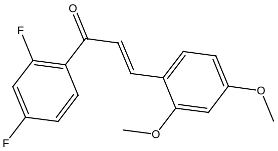
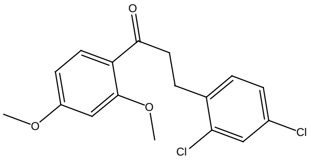
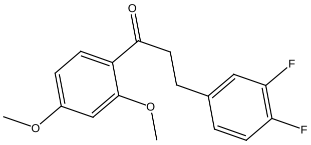
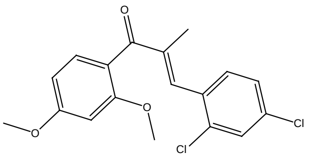
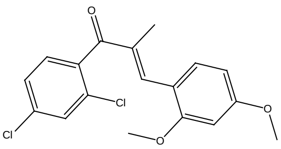
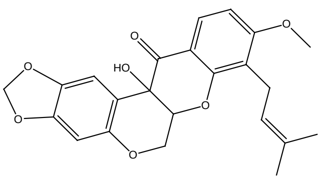
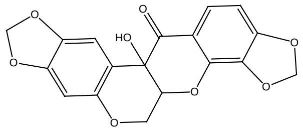
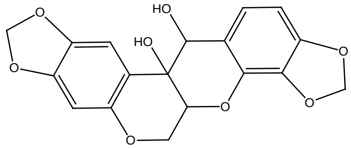
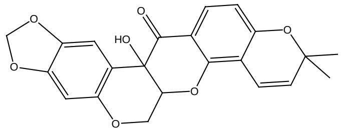
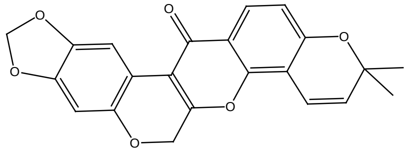
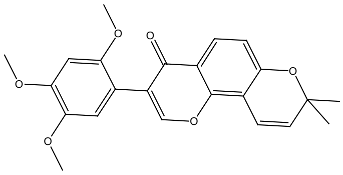
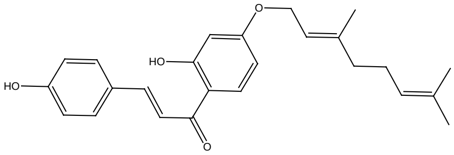
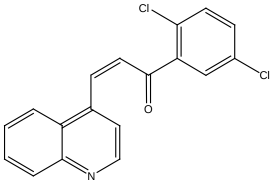
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
41		2.95	3.10	19
42		3.11	3.67	19
43		3.29	3.58	19
44		3.32	3.11	19
45		3.53	3.71	19
46		4.28	3.13	19
47		4.18	3.73	19
48		4.20	3.24	19
49		3.80	3.30	19
50		4.15	4.91	20

Table 1 (*continued*)

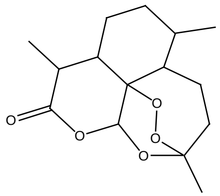
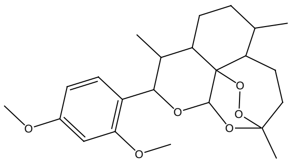
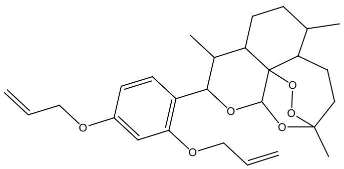
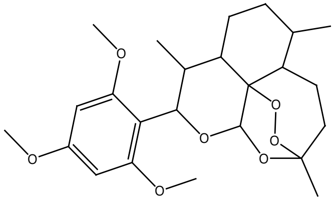
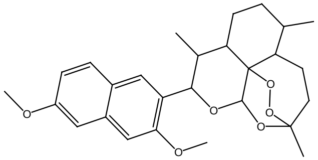
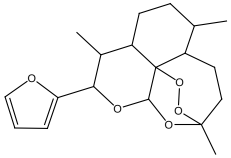
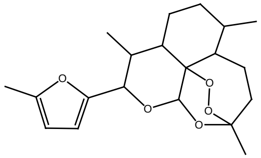
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
51		4.85	4.70	20
52		4.78	4.96	20
53		5.00	4.88	20
54		4.29	4.25	20
55		4.59	3.69	20
54		4.44	3.88	20
57		2.11	2.82	19

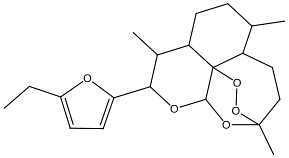
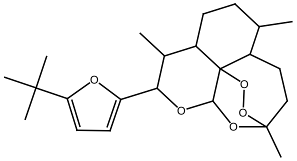
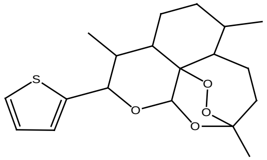
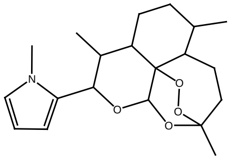
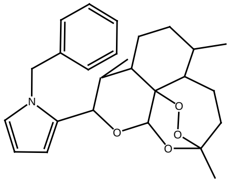
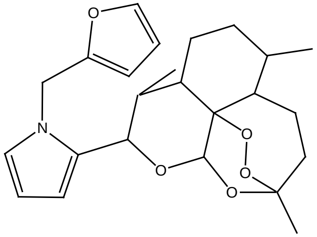
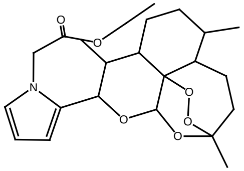
modynamic functions calculated by using the MOPAC program package.[23]

The best multilinear regression (BMLR) procedure[24] was used to find the best correlation models from selected non-collinear descriptors. The BMLR selects the best two-parameter regression equation, the best three-parameter regression equation, etc., based on the highest R^2 value in the stepwise regression procedure. During the BMLR procedure the descriptor scales are normalized, centered automatically, and the final result is given in natural scales. This result has the best representation of the property in the given descriptor pool.

A major decision in developing successive QSAR models is when to stop adding descriptors to the model during the stepwise regression procedure. The lack of an adequate control leads to over-correlated equations, which contain an excess of descriptors and are difficult to interpret in terms of interaction mechanisms. In this

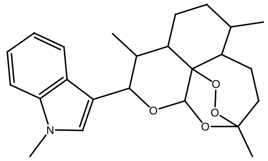
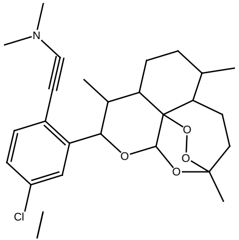
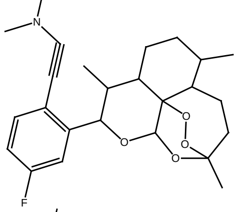
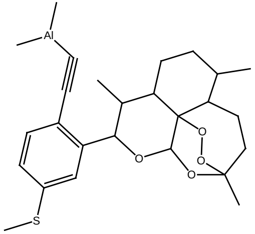
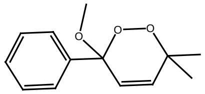
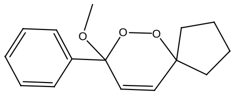
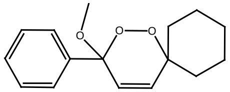
Table 2. Experimental and predicted $\log IC_{50}$ according to the multilinear QSAR model in Table 4 of 69 compounds for NF54 strain

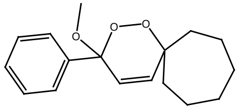
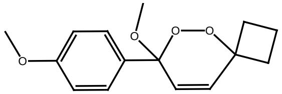
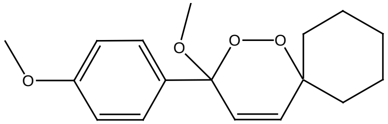
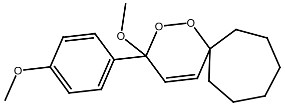
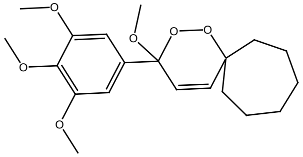
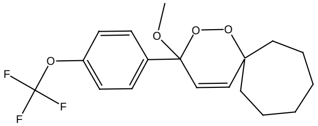
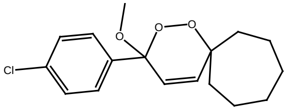
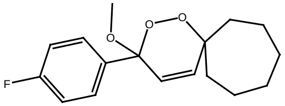
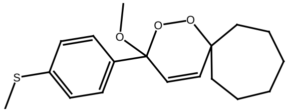
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
1		1.00	0.83	11
2		0.62	0.89	11
3		0.82	0.83	11
4		0.89	0.86	11
5		0.95	0.83	11
6		0.15	0.93	11
7		0.72	0.92	11

8		0.93	0.91	11
9		1.00	0.88	11
10		0.71	1.16	11
11		0.66	1.15	11
12		1.20	1.03	11
13		0.97	1.05	11
14		0.96	1.05	11

(*continued on next page*)

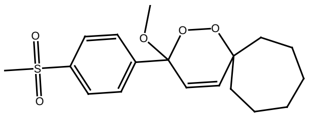
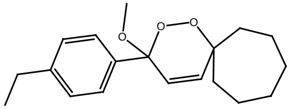
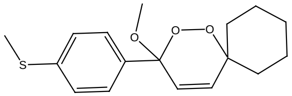
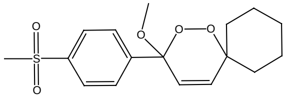
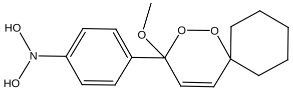
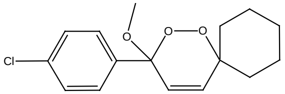
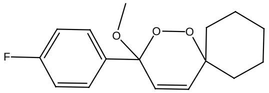
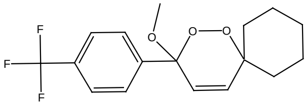
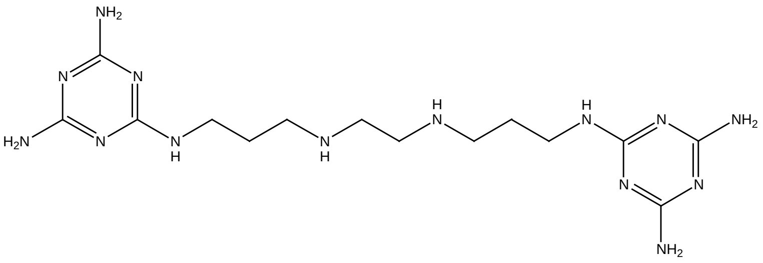
Table 2 (*continued*)

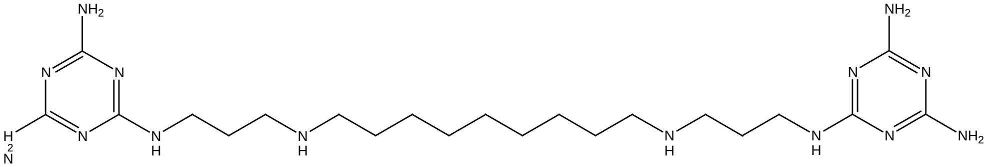
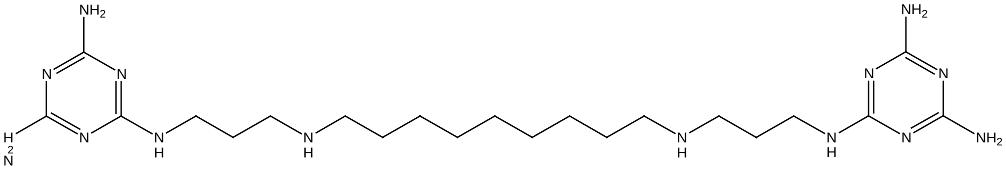
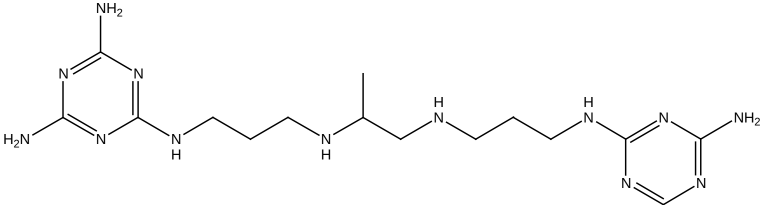
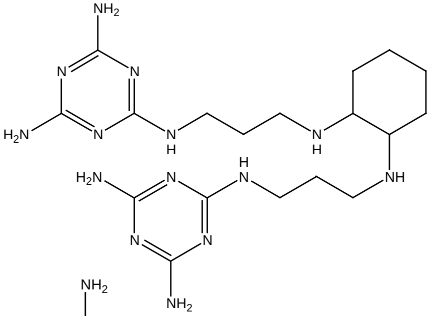
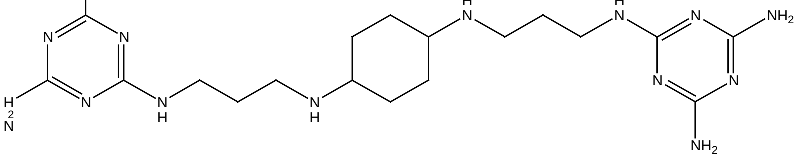
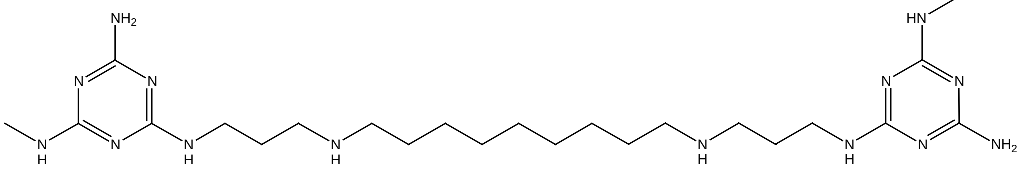
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
15		0.60	1.07	11
16		1.04	0.87	11
17		0.92	0.87	11
18		0.92	1.02	11
19		3.04	2.13	13
20		2.28	2.11	13
21		2.45	2.11	13

22		2.34	2.10	13
23		2.20	2.10	13
24		2.26	2.09	13
25		2.32	2.08	13
26		2.08	2.03	13
27		1.79	2.04	13
28		1.76	2.09	13
29		1.93	2.09	13
30		1.89	2.42	13

(*continued on next page*)

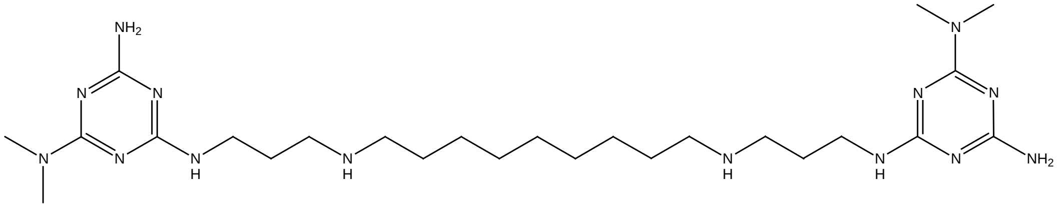
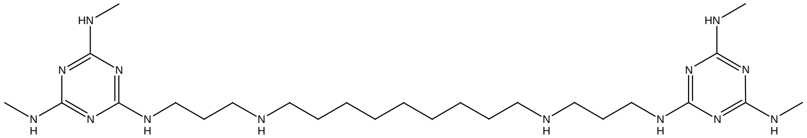
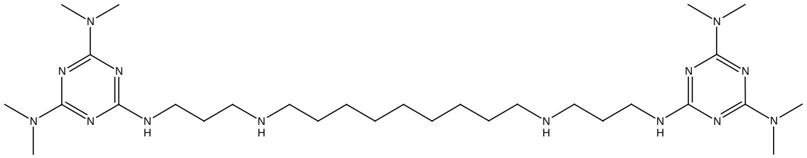
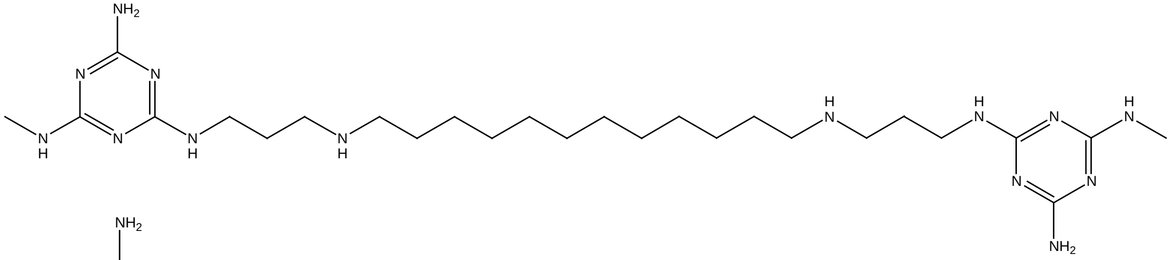
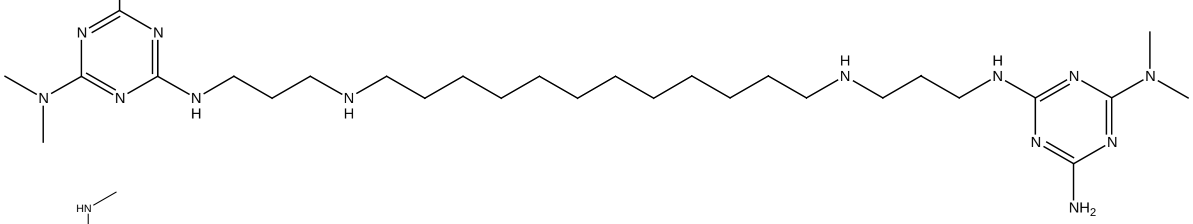
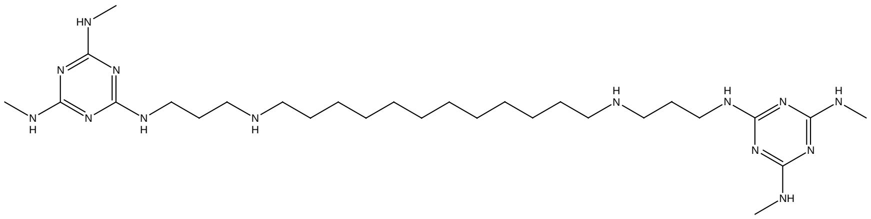
Table 2 (*continued*)

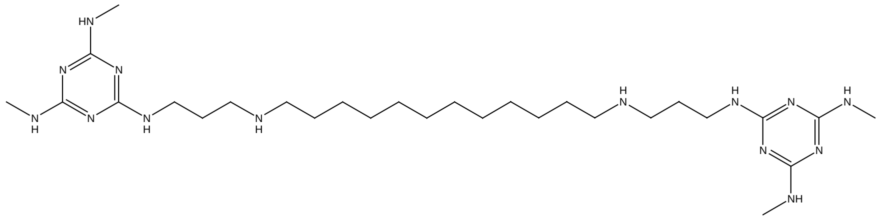
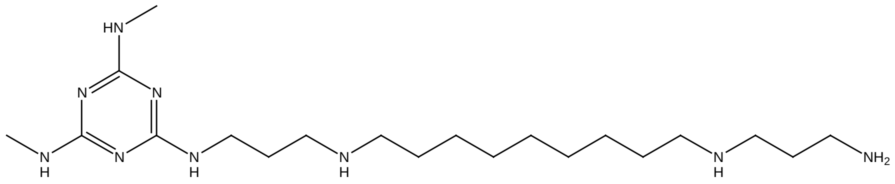
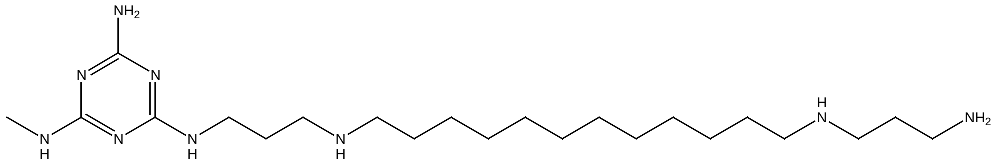
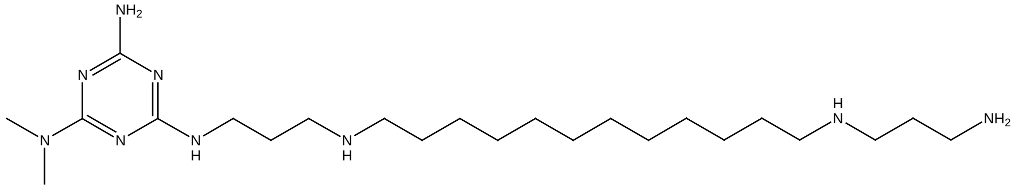
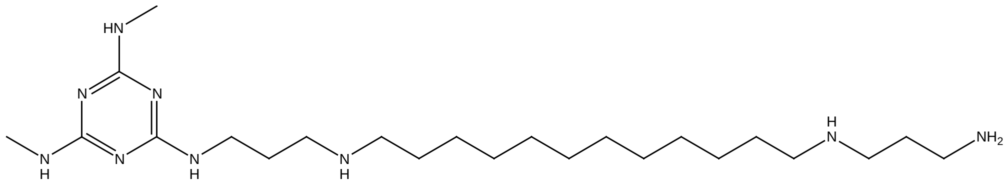
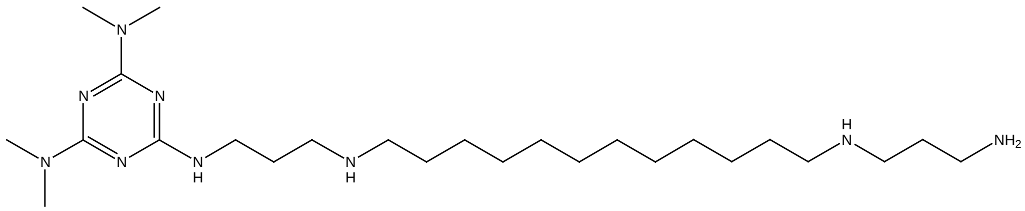
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
31		1.49	2.06	13
32		2.26	2.08	13
33		2.20	2.46	13
34		1.75	2.07	13
35		1.66	2.08	13
36		2.00	2.10	13
37		2.30	2.09	13
38		2.15	2.07	13
39		4.25	4.67	21

40		3.64	3.67	21
41		2.79	3.23	21
42		4.26	4.48	21
43		4.23	4.09	21
44		4.23	4.15	21
45		3.60	3.21	21

(*continued on next page*)

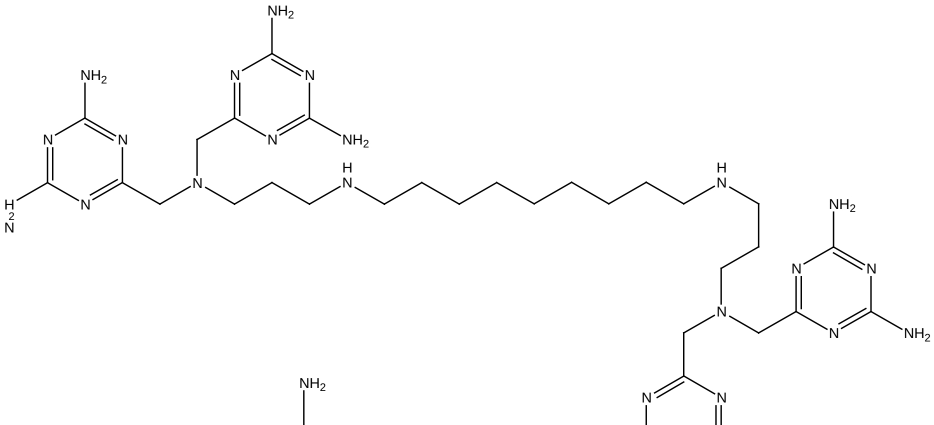
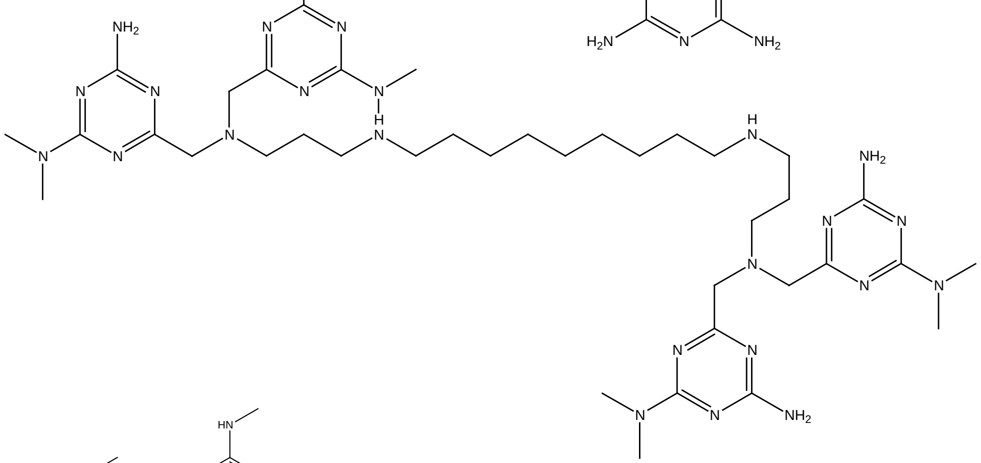
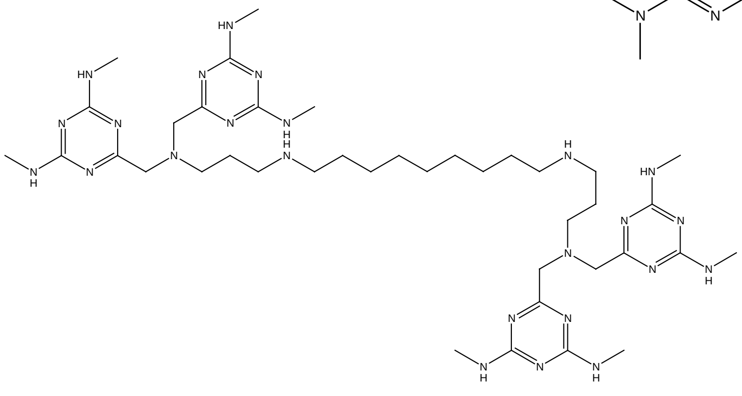
Table 2 (*continued*)

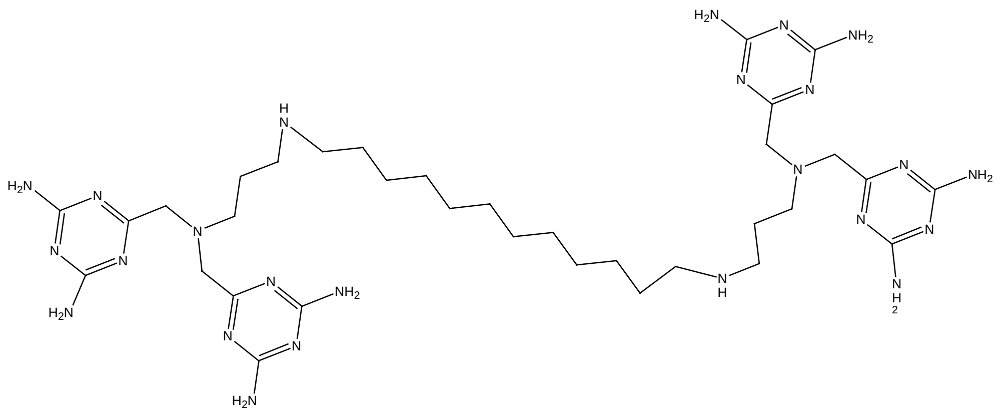
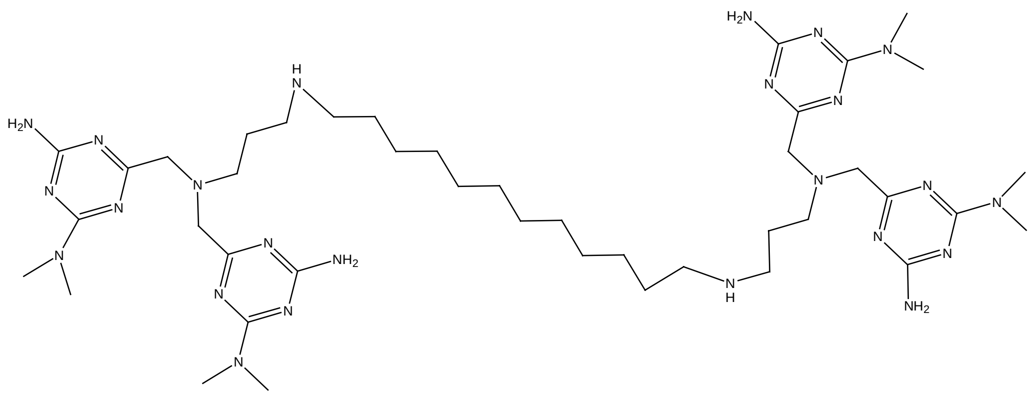
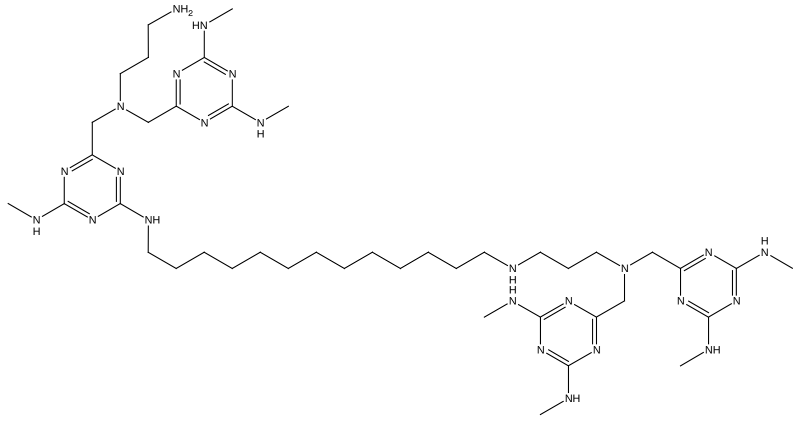
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
46		3.29	3.07	21
47		3.17	2.98	21
48		2.76	2.60	21
49		2.91	2.86	21
50		2.69	2.65	21
51		2.67	2.67	21

52		2.24	2.67	21
53		2.76	2.82	21
54		3.31	2.69	21
55		2.79	2.57	21
56		3.24	2.57	21
57		2.92	2.42	21

(*continued on next page*)

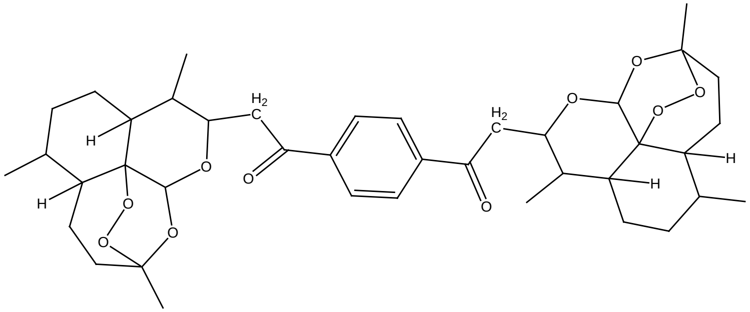
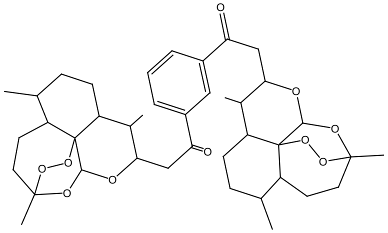
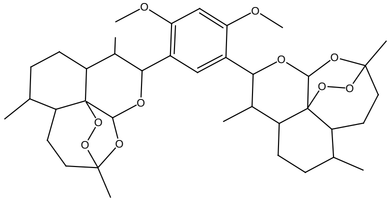
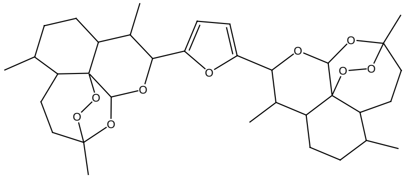
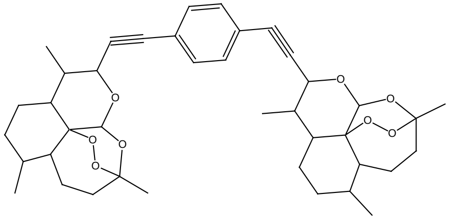
Table 2 (*continued*)

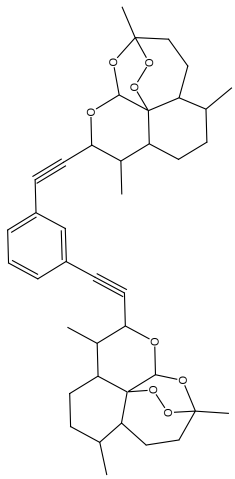
Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
58		3.88	3.68	21
59		2.21	2.32	21
60		1.86	2.16	21

61		3.19	3.18	21
62		1.84	1.95	21
63		1.65	1.87	21

(*continued on next page*)

Table 2 (*continued*)

Number	Structure	Experimental log IC_{50}	Predicted log IC_{50}	Reference
64		0.28	0.57	4
65		0.28	0.55	4
66		0.11	0.52	4
67		0.51	0.61	4
68		1.48	0.38	4

69		1.56	0.41	4

work, attempts were made to develop QSAR equations based on an optimum small number of descriptors with a significantly high quality of the regression equations.

The QSAR models obtained were validated automatically, (i) by the leave-one-out method (ii) by internal correlation whereby one-third of the compounds is predicted with the model fitted with two thirds of the compounds (see description below).

2.3. QSAR models for D6 and NF54 strains

By using the best multilinear regression method equations for both the strains were constructed with up to six descriptors. A simple rule ("breaking point" rule) was used to decide the optimum number of descriptors by considering the improvement of the R^2 by addition of a further descriptor to the model. If the difference between the models with n and $n + 1$ descriptors is improved by a value of less than 0.04, then the optimum model is taken to have n descriptors. The selection of the optimum number of descriptors is shown in Figure 1. In addition, the Fisher criterion was also monitored for a significant improvement in the correlation coefficient value with respect to the number of descriptors. The final QSAR models selected for the two strains (D6 and NF54) are shown in Tables 3 and 4, respectively.

In Tables 3 and 4, X, ΔX, t test, R^2, R^2_{cv}, and s^2 are the regression coefficients, standard errors of the regression coefficients, t significance, coefficient of determination, and squared cross-validated coefficient (leave-one-out method), respectively. In the development of these models, we added descriptors only when, the intercollinearity among the descriptors is <0.50. The predicted log IC_{50} values for the compounds are listed in Tables 1 and 2. The linear plot of experimental and predicted log IC_{50} according to the QSAR model presented in Table 3 is given in Figure 2. The experimental and predicted values according to Table 4 are plotted in Figure 3.

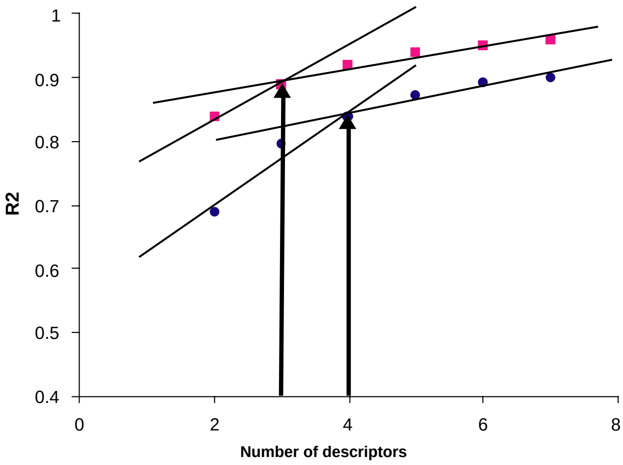


Figure 1. Number of descriptors versus R^2 of the models (circles-D6, squares-NF54).

Table 3. The multilinear QSAR model obtained for 57 organic compounds for strain D6 ($R^2 = 0.84$, $F = 68.15$, and $s = 0.43$)

Number	X	$\pm\Delta X$	t test	R^2	R^2_{cv}	S^2	Descriptor
0	−0.257	0.526	−0.488				Intercept
1	−0.639	0.058	−11.02	0.014	0.040	1.06	Average atom weight, d1
2	9.232	0.898	10.28	0.609	0.568	0.427	Average bonding information content (order 2), d2
3	24.39	3.298	7.397	0.768	0.739	0.258	HA-dependent HDCA-2/SQRT (TMSA) (Zefirov PC), d3
4	24.971	5.184	4.817	0.840	0.809	0.182	Relative number of benzene rings, d4

Table 4. The multilinear QSAR model obtained for 69 organic compounds for strain NF54 ($R^2 = 0.89$, $F = 176.1$, and $s = 0.37$)

Number	X	$\pm\Delta X$	t test	R^2	R^2_{cv}	S^2	Descriptor
0	3.679	0.219	16.76				Intercept
1	43.87	2.060	21.28	0.555	0.528	0.543	HA dependent HDSA-2/TMSA(Zefirov PC), d6
2	−44.95	4.472	−10.05	0.778	0.758	0.274	Maximum partial charge (Zefirov) for atoms for atom H, d5
3	−6.0E−5	7.0E−6	−8.150	0.890	0.878	0.138	Wiener index, d7

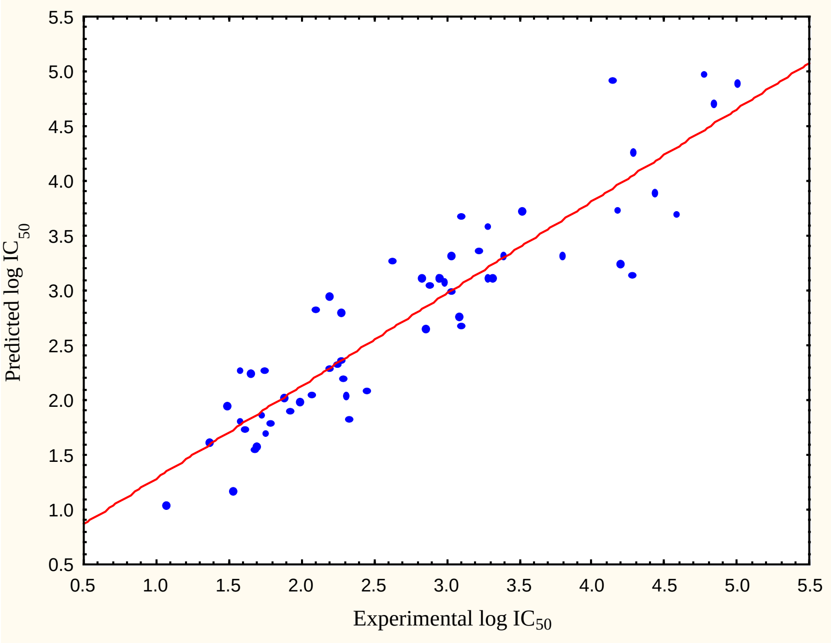


Figure 2. The experimental and predicted log IC$_{50}$ for D6 strain according to the model in Table 3 ($R^2 = 0.840$).

3. Discussion

3.1. Discussion on the QSAR models for the prediction of antimalarial activity

The predicted ranges of the biological activity according to QSAR models developed are in good agreement with the experimental data. The predicted range for the model in Table 3 is 1.017–4.956 compared with the experimental range of 1.080–5.00. There are three outliers from this model, compounds 46, 48, and 55. These compounds have somewhat underestimated values of descriptor d3. For the second NF54 strain model (Table 4), the predicted range is −0.108–4.622, that compares favorably with the experimental range of 0.114–4.258. In this model there are two outliers (see Table 2, compounds **6** and **11**).The reason for these outliers is similar to that for the previous three outliers, i.e., the descriptor d6, is somewhat overestimated.

The most significant descriptor in Table 3 according to the t test is the *Average atom weight, d1*. The most active compounds in Table 1 seemed to possess higher values of this descriptor as compared with the values of the less active compounds. Most of the active compounds include heavier atoms (as S, Cl) in their structure. The second significant descriptor (d2) involved in the equation is *Average bonding information content (order 2)$^{\wedge 2}$(IC),* defined as in Eq. 1

$$^{\wedge 2}(\mathrm{IC}) = -n\sum_{i}^{k}\frac{n_i}{n}\log_2\frac{n_i}{n}, \qquad (1)$$

where n_i is the number of atoms in the ith class, n is the total number of atoms in the molecule, and k denotes

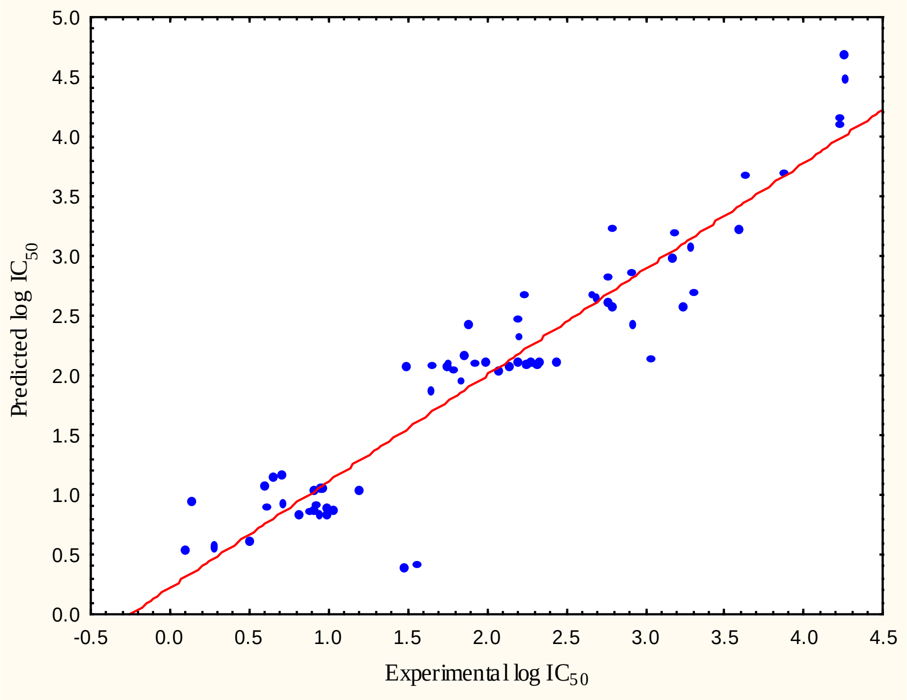


Figure 3. Experimental and predicted log IC_{50} for NF54 strain according to the model in Table 4 ($R^2 = 0.890$).

the number of atomic layers in the coordination sphere around a given atom.[25] The diffusion of molecules in and between different media is another essential factor that probably influences the activity of compounds. In general, it depends both on the shape and the symmetry of the drug molecules. The topological descriptor, d2, describes the connectivity and branching in a molecule and relates to molecular shape and symmetry.[26] In addition, d2 possesses larger values for the less active compounds compared with the active ones.

The third descriptor in the QSAR model is *HA-dependent HDCA-2/SQRT(TMSA), d3*

$$\mathrm{d3} = \sum_{\mathrm{A}} q_{\mathrm{A}} \frac{\sqrt{S_{\mathrm{A}}}}{\sqrt{S^2}_{\mathrm{tot}}}, \tag{2}$$

where the summation is carried out over all possible hydrogen-bonding acceptor atoms A q_A is the partial charge on H-bonding acceptor atoms, selected by a threshold charge, and S_A is the solvent-accessible surface area of H-bonding acceptor atoms, selected also by the threshold charge. This descriptor reflects possible hydrogen bonding donor and acceptor interactions between the molecules that seem to be important for the processes behind the antimalarial activity.[27] The last descriptor in the model is the *Relative number of benzene rings, d4*. All compounds for D6 (and NF54) strain include benzene rings. It was not surprising that d4 appear in the model. The antiparasitic efficacy depends largely on the number and nature of rings,[28] whether aromatic or not.

The most significant descriptor in the model of Table 4 according to the *t* test is *HA-dependent HDSA-2/TMSA (Zefirov PC), d6*. This descriptor is similar to *HA-dependent HDCA-2* (see Eq. 2); both related to the surface area and the charge distribution of the molecule. The second significant descriptor for activity (log IC_{50}) of the NF54 strain is the *Maximum partial charge (Zefirov) for atoms for atom H, d5* (see Table 4). Atomic partial charges are referred to as static chemical reactivity. According to this view, the calculated σ- and π-electron densities on a particular atom characterize the possible direction of the chemical transformation and are thus considered as reactivity indices.[26]

The final descriptor is *Wiener index (W) d7*: (Eq. 3), where d_{ij} corresponds to the number of bonds in the shortest path connecting the pair of atoms *i* and *j*, and N_{SA} the number of non-hydrogen atoms in the molecule.

$$W = 0.5 \sum_{(i,j)}^{N_{\mathrm{SA}}} d_{ij}. \tag{3}$$

The above descriptor d7 characterizes the “compactness” of a molecule, being larger for extended chains and smaller for branched compounds. In our case, most of the active compounds have smaller values of this descriptor. Previous studies showed that descriptor d7 correlated well with antimalarial activity.[29]

3.2. Cross-validation

To validate the models, the parent data points of the both sets were divided according to the experimental values into three subsets (A, B, and C) as follows: the 1st, 4th, 7th, etc., data points comprise th first subset (A), the 2nd, 5th, 8th, etc., the second subset (B), and

Table 5. Internal validation of the QSAR models

Training set	N	R^2 (Fit)	R^2_{cv} (Fit)	S^2 (Fit)	Test set	N	R^2 (Pred)
Validation for the model in Table 3							
A + B	38	0.827	0.802	0.231	C	19	0.785
A + C	38	0.832	0.797	0.216	B	19	0.798
B + C	38	0.807	0.783	0.240	A	19	0.807
Average		0.822	0.794	0.229			0.796
Validation for the model in Table 4							
A + B	46	0.891	0.869	0.328	C	23	0.866
A + C	46	0.887	0.863	0.537	B	23	0.857
B + C	46	0.881	0.857	0.443	A	23	0.846
Average		0.886		0.863			0.856

the 3rd, 6th, 9th, etc., the third subset (C). Three training sets were prepared as combinations of two subsets (A and B), (A and C), and (B and C), respectively. For each training set, the correlation equation was derived with the descriptors of Tables 3 and 4. The equation obtained was then used to predict log IC_{50} values for the compounds from the appropriate test set. The efficiency of QSAR models to predict log IC_{50} value was also estimated using the internal cross-validation. The correlation coefficients and standard deviations of linear correlations between experimental and predicted for test sets of log IC_{50} values were also calculated and the results are shown in Table 5. The average values of R^2 (Fit) and R^2 (Pred) are close which suggests considerable stability and consequently satisfactory predictivity for both models.

4. Conclusions

A quantitative structure–activity relationship (QSAR) study was applied to two diverse sets of potentially active compounds against the D6 and NF54 strains of malaria. For each set, statistically significant models were obtained using the BMLR method encoded in CODESSA PRO software. These models may be considered as mathematical equations for the prediction of antimalarial activities of the compounds structurally similar to those used in this study. The internal consistency of the models was verified using the R^2_{cv} and leave-1/3-out cross-validation. However, the ultimate test of these models is their ability to predict activities for newly reported compounds.

The mechanism of antimalarial activity was discussed by analyzing the physico-chemical meaning of the descriptors involved in the QSAR models. These descriptors could be related to the shape (volume, mass) and branching of the molecule as well as to the charge-related interactions. The present work, based on 2D QSAR and the models developed, provides for the first step toward the prediction of novel active compounds.

Acknowledgment

The Estonian Science Foundation Grant No. 4548 is acknowledged for the partial support of this work.

References and notes

1. Grigorov, M.; Weber, J.; Tronchet, J. M. J.; Jefford, C. W.; Milhous, W. K.; Maric, D. *J. Chem. Inf. Sci.* **1997**, *37*, 124.
2. O'Dowd, H.; Ploypradith, P.; Xie, S.; Shapiro, T.; Posner, G. *Tetrahedron* **1999**, *55*, 3625.
3. Cheng, F.; Shen, J.; Luo, X.; Zhu, W.; Jiande, G.; Ji, R.; Jiang, H.; Chen, K. *Bioorg. Med. Chem.* **2002**, *10*, 2883.
4. Posner, G.; Ploypradith, P.; Parker, M.; O'Dowd, H.; Woo, S.-H.; Northrop, J.; Krasavin, M.; Dolan, P.; Kensler, T.; Xie, S.; Shapiro, T. *J. Med. Chem.* **1999**, *42*, 4275.
5. Fitch, Coy. D. *Life Sci.* **2004**, *74*, 1957.
6. Posner, G.; Maxwell, J.; O'Dowd, H.; Krasavin, M.; Xie, S.; Shapiro, T. *Bioorg. Med. Chem.* **2000**, *8*, 1361.
7. Ryckebusch, A.; Deperez-Poulain, R.; Debreu-Fontaine, M.-A.; Vandaele, R.; Mouray, E.; Grellier, P.; Sergheraert, C. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2595.
8. Agarwal, Anu.; Srivastava, K.; Puri, S.; Chauhan, P. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1881.
9. CATALYST Version 2.0 software; Molecular simulations Inc.: Burlington, MA, 1993.
10. Tonmunphean, S.; Parasuk, V.; Kokpol, S. *Quant. Struct.-Act. Relat.* **2000**, *19*, 475.
11. Girones, X.; Gallegos, A.; Carbo-Dorca, R. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1400.
12. Girones, X.; Gallegos, A.; Carbo-Dorca, R. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1053.
13. Roy, K.; Sengupta, C. *Quant. Struct.-Act. Relat.* **2001**, *20*, 319.
14. Katritzky, A. R.; Fara, D. C.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1333.
15. *CODESSA PRO Software*, University of Florida, 2002.
16. Katritzky, A. R.; Fara, D. C.; Karelson, M. *Bioorg. Med. Chem.* **2004**, *12*, 3027.
17. Katritzky, A.; Dobchev, D.; Hur, E.; Fara, D.; Karelson, M. *Bioorg. Med. Chem.* **2005**, *13*, 1623.
18. Jiang, S.; Prigge, S.; Wei, L.; Gao, Y.-E.; Hudson, T.; Gerena, L.; Dame, J.; Kyle, D. *Antimicrob. Agents Chemother.* **2001**, *45*, 2577.
19. Li, R.; Kenyon, G.; Cohen, F.; Chen, X.; Gong, B.; Dominigues, J.; Davidson, E.; Kurzban, G.; Miller, R.; Nuzum, E.; Rosenthal, P.; McKerrow, J. *J. Med. Chem.* **1995**, *38*, 5031.
20. Yenesew, A.; Derese, S.; Midiwo, J.; Oketch, H.; Lisgarten, J.; Palmer, R.; Heydenreich, M.; Peter, M.; Akala, H.; Wangui, J.; Liyala, P.; Waters, N. *Phytochemistry* **2003**, *64*, 773.
21. Klenke, B.; Barret, M.; Brun, R.; Gilbert, I. *J. Antimocrob. Chemother.* **2003**, *52*, 290.

22. Hyperchem, v. 7.5; Hypercube Inc.; Gainesville, FL, 2003.
23. Stewart, J. J. P. *MOPAC 6.0*; QCPE No 455, 1989.
24. Katritzky, A.; Mu, L.; Lobanov, V. *J. Phys. Chem.* **1996**, *100*, 10400.
25. Basak, S. C.; Harriss, D. K.; Magnuson, V. R. *J. Pharm. Sci.* **1984**, *73*, 429.
26. Karelson, M. *Molecular Descriptors in QSAR/mQSPR*; Wiley Interscience: New York, 2000.
27. Santos, E.; Carneiro, J.; Ferreira, V. *Bioorg. Med. Chem.* **2004**, *12*, 87.
28. Copp, B.; Kayser, O.; Brun, R.; Kiderlen, A. *Planta Med.* **2003**, *69*, 527.
29. Agrawal, V.; Sharma, R.; Khadikar, P. *Bioorg. Med. Chem.* **2002**, *10*, 1361.