



Structural requirements of 3-carboxyl-4(1H)-quinolones as potential antimalarials from 2D and 3D QSAR analysis



Jiazhong Li^{a,*}, Shuyan Li^b, Chongliang Bai^a, Huanxiang Liu^a, Paola Gramatica^c

^a School of Pharmacy, Lanzhou University, Donggang West Road 199, 730000 Lanzhou, China

^b Department of Chemistry, Lanzhou University, Tianshui South Road 222, 730000 Lanzhou, China

^c QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Theoretical and Applied Sciences, University of Insubria, via Dunant 3, 21100 Varese, Italy

ARTICLE INFO

Article history:

Received 3 April 2013

Received in revised form 8 July 2013

Accepted 9 July 2013

Available online 19 July 2013

Keywords:

4(1H)-quinolones

Antimalarial

MLR

CoMFA

CoMSIA

ABSTRACT

Malaria is a fatal tropical and subtropical disease caused by the protozoal species *Plasmodium*. Many commonly available antimalarial drugs and therapies are becoming ineffective because of the emergence of multidrug resistant *Plasmodium falciparum*, which drives the need for the development of new antimalarial drugs. Recently, a series of 3-carboxyl-4(1H)-quinolone analogs, derived from the famous compound endochin, were reported as promising candidates for orally efficacious antimalarials. In this study, to analyze the structure–activity relationships (SAR) of these quinolones and investigate the structural requirements for antimalarial activity, the 2D multiple linear regressions (MLR) method and 3D comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) methods are employed to evolve different QSAR models. All these models give satisfactory results with highly accurate fitting and strong external predictive abilities for chemicals not used in model development. Furthermore, the contour maps from 3D models can provide an intuitive understanding of the key structure features responsible for the antimalarial activities. In conclusion, we summarize the detailed position-specific structural requirements of these derivatives accordingly. All these results are helpful for the rational design of new compounds with higher antimalarial bioactivities.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Malaria was originated more than 30 million years ago [1] and remains one of the most significant world-wide public health problems, especially in tropical and subtropical regions. In 2010, an estimated 3.3 billion people were at risk of contracting malaria, and there were 655,000 malaria deaths. Approximately 86% of malaria deaths globally were of children under the age of 5 [2]. In 2011, malaria was considered endemic in a total of 106 countries [2].

Malaria is caused by five species of parasites of the genus *Plasmodium* that affect humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium oval* and *Plasmodium knowlesi*. Malaria due to *P. falciparum* is the most deadly variety, as it is responsible for the majority of malaria deaths. The primary drugs for treatment of malaria have been the quinolones chloroquine (CQ), quinine (QN) and mefloquine; the antifolate combination of pyrimethamine and sulfadoxine; and others [3]. However, during the past 3 decades, *P. falciparum* has developed resistance to every commonly available antimalarial drug [4], including

chloroquine-resistant (CQ-R), multi-drug-resistant (MDR) and others [5–7]. More recently, artemisinin-based combination therapies (ACTs) have been recommended by the World Health Organization (WHO) as the first-line treatment in malaria endemic countries [8]. However, suspected resistance to artemisinins has been identified in 4 countries in the Greater Mekong subregion, and this resistance has been reported in a growing number of countries [2]. History has shown that once resistance to a new drug emerges, it spreads rapidly and causes the organism to become cross-resistant to the whole class of drugs [4,9].

To meet this challenge, much effort has been invested in modifying existing drugs using structural changes to counteract the induced resistance. There is an urgent need to develop new antimalarial drugs with new chemotypes. In 2011, the WHO published the Global Plan for Artemisinin Resistance Containment (GPARC), recommending five key activities for successful management of artemisinin resistance, one of which is invest in research related to artemisinin resistance [2]. Many research groups throughout the world have been dedicated to this problem for years and have reported various structures with potential antimalarial efficiency. Among these structures, 4(1H)-quinolone endochin is well-known. Endochin (Fig. 1a), discovered during the 1940s as a causal prophylactic and potent erythrocytic stage agent in avian malaria models

* Corresponding author. Tel.: +86 13659427727; fax: +86 9318915686.

E-mail address: lijiazhong@lzu.edu.cn (J. Li).

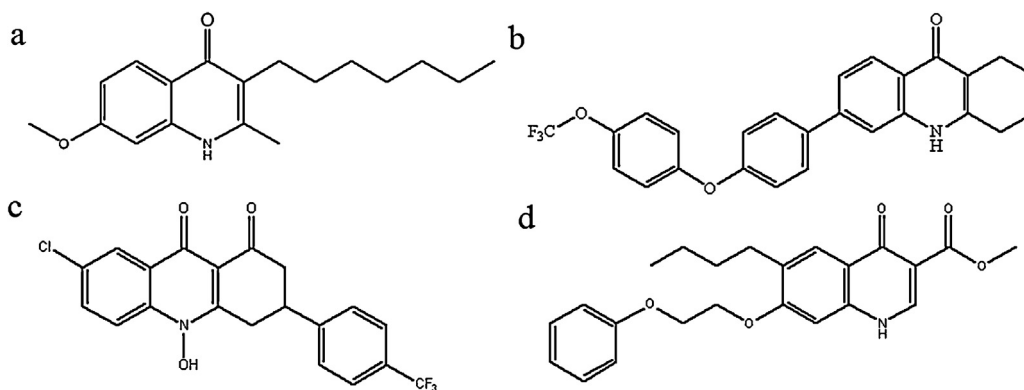


Fig. 1. The structures of endochin and several ELQs.

[10], is actually not effective against mammalian malaria [11]. During the past several decades, structural modifications to endochin have been widely investigated to hunt for potential antimalarial agents for humans, and various endochin-like quinolones (ELQs) have been reported, such as THA (Fig. 1b) [12], floxacrine (Fig. 1c) [13–15] and ICI56780 (Fig. 1d) [16]. Among the modifications, it is important to optimize the 3-position of endochin, and an alkyl side chain here is sometimes regarded as a critical feature in antimalarial potency [3,17]. Since the identification of ICI56780, the 3-carboxyl substituents also show their potential as antimalarial agents [16]. Most recently, Zhang et al. reported a series of 3-carboxyl endochin-like quinolones with several unique substituents, such as a carboxyl ester at the 3-position avoiding the presence of a long side fat chain and a *meta* substituted aromatic ring at the 2-position [18]. Their research led to the development of a series of potential efficacious oral antimalarials.

To better understand the structural requirements of the 3-carboxyl analogs for antimalarial activities, herein we present our quantitative structure–activity relationship (QSAR) analysis of the series of 3-carboxyl-4(1H)-quinolones, reported by Zhang et al., by using 2D multiple linear regressions method and 3D CoMFA [19] and CoMSIA [20] methods. Combining the results of this study with reported conclusions, we will summarize the detailed position-specific SAR of these derivatives to aid the design of new chemical entities with potential antimalarial activities.

2. Experiments and methods

2.1. Dataset

In this study, the 3-carboxyl-4(1H)-quinolone analogs were selected from Ref. [18]. Their antimalarial activities EC_{50} (μ M) against the clinically relevant multidrug resistant malarial strain TM90-C2B, which is resistant to chloroquine, mefloquine, pyrimethamine and atovaquone, are converted to the corresponding pEC_{50} ($-\log EC_{50}$) and used as dependent variable. The molecular structures and corresponding biological activities are listed in Table 1. The whole dataset is randomly divided into a training set containing 38 molecules, which is used to develop QSAR models, and a prediction set containing 11 molecules (marked by “a” in Table 1), never used in model development which is used *a posteriori* to evaluate the predictive ability of the built models for new chemicals.

2.2. Structural conformations generation

The 3D structures of these quinolones are sketched in SYBYL 6.9 program [21]. Partial atomic charges are calculated using the Gasteiger–Hückel method, and energy minimization is performed

using the Tripos force field with a convergence criterion of $0.01 \text{ kcal mol}^{-1}$. Then, the conformational search is conducted using the multisearch routine to obtain the lowest energy conformation for each molecule.

2.3. MLR model

2.3.1. Descriptors generation

In this work, the molecular descriptors are generated in three different ways. First, the molecular conformations with lowest energies are exported to the DRAGON 5.5 program [22] to calculate 2914 descriptors including zero-, one-, two-, three-dimensional descriptors, charge descriptors and molecular properties. Then, the Maestro package [23] is used to generate 50 descriptors related to ADME properties using QikProp [24], including properties such as skin permeability and octanol/water partition coefficients.

In addition, the MOE program [25] is also employed to calculate 327 different 2D and 3D descriptors. The 2D descriptors from MOE are defined to be numerical properties that can be calculated from the connection table representation of a molecule, including physical properties and pharmacophore feature descriptors. There are two types of 3D molecular descriptors: those that depend on internal coordinates only, defined as ‘i3D’, and those that depend on absolute orientation, defined as ‘x3D’. There are 133 ‘i3D’ and 10 ‘x3D’ descriptors in MOE, such as potential energy descriptors, MOPAC descriptors and conformation dependent charge descriptors.

The descriptors from different programs can represent a molecule from different aspects, but some of them may describe the same meanings with the same or similar values. To minimize duplication, all 3291 descriptors are combined together and the constant or near-constant descriptors are deleted. If the pairwise correlation of two descriptors is higher than 0.97, the one showing higher pair correlation with other descriptors is excluded. Thus, 861 descriptors undergo the subsequent variable selection process.

2.3.2. Model development and validation

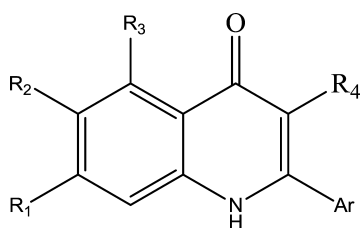
After the molecular descriptors calculation and pre-reduction, a genetic algorithm (GA) is used to select the descriptors relevant to the molecular antimalarial activities [26,27]. The GA first generates an initial population (chromosomes), which is actually a set of linear models. The Friedman’s ‘lack-of-fit’ (LOF) function is used as the fitness function to evaluate the fitness of these chromosomes:

$$\text{LOF} = \left\{ \frac{\text{SSE}}{(1 - (c + dp/n))} \right\}^2 \quad (1)$$

where SSE is the sum of squares of errors, c is the number of basis function, d is the smoothness factor (default 0.5), p is the number of features in the model, and n is the number of samples for model

Table 1

The studied 4(1H)-quinolone analogs and corresponding experimental and calculated antimalarial activities.



No.	R ₁	R ₂	R ₃	R ₄	Ar	Activity			
						Exp.	MLR	CoMFA	CoMSIA
1	—OMe	—H	—H	—COOEt		6.04	6.38	6.10	5.99
2	—OMe	—H	—H	—COOEt		6.74	6.34	6.45	6.84
3	—OMe	—H	—H	—COOEt		6.68	6.45	6.57	6.95
4	—OMe	—H	—H	—COOEt		9.00	8.48	9.03	8.86
5 ^a	—OMe	—H	—H	—COOEt		5.11	5.36	5.43	5.17
6	—OMe	—H	—H	—COOEt		5.74	6.48	5.90	5.97
7	—OMe	—H	—H	—COOEt		5.20	5.07	5.17	5.08
8 ^a	—OMe	—H	—H	—COOEt		5.64	5.87	6.06	4.99
9	—OMe	—H	—H	—COOEt		7.70	7.87	7.58	7.60
10	—OMe	—H	—H	—COOEt		5.06	5.47	4.87	4.89
11	—OMe	—H	—H	—COOEt		8.00	7.59	8.04	8.19
12	—OMe	—H	—H	—COOEt		6.54	6.80	6.39	6.46
13 ^a	—OiPr	—H	—H	—COOEt		5.80	5.24	5.68	5.83

Table 1 (Continued)

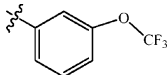
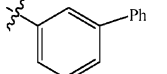
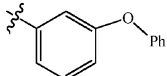
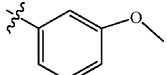
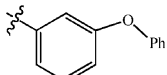
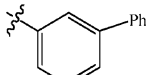
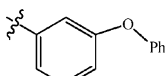
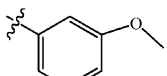
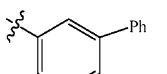
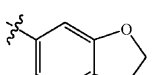
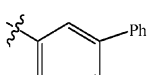
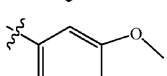
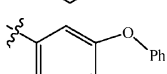
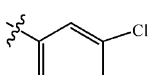
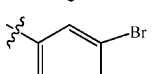
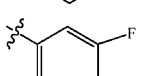
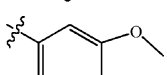
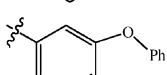
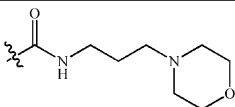
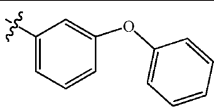
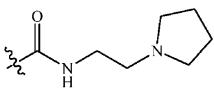
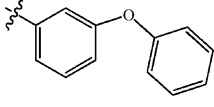
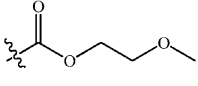
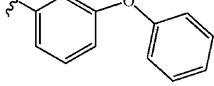
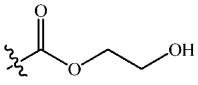
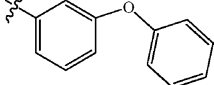
No.	R ₁	R ₂	R ₃	R ₄	Ar	Activity			
						Exp.	MLR	CoMFA	CoMSIA
14	—OCF ₃	—H	—H	—COOEt		5.24	5.23	5.30	5.28
15	—OH	—H	—H	—COOEt		8.00	8.11	7.85	8.09
16 ^a	—OH	—H	—H	—COOEt		6.66	6.52	6.51	6.79
17	—SMe	—H	—H	—COOEt		5.29	5.10	5.50	5.37
18	—F	—H	—F	—COOEt		5.92	5.93	5.97	6.10
19	—OMe	—OMe	—H	—COOEt		7.52	7.35	7.80	7.59
20	—OMe	—OMe	—H	—COOEt		6.33	6.79	6.22	6.42
21	—Cl	—OMe	—H	—COOEt		5.82	5.85	5.99	5.88
22 ^a	—Cl	—Cl	—H	—COOEt		7.70	8.64	7.45	7.86
23	—OMe	—Cl	—H	—COOEt		7.30	7.08	7.40	7.32
24 ^a	—OMe	—Cl	—H	—COOEt		8.00	8.25	7.56	8.27
25	—OMe	—Cl	—H	—COOEt		6.92	7.08	6.73	7.09
26	—OMe	—Cl	—H	—COOEt		8.00	8.18	8.01	7.82
27	—OMe	—Cl	—H	—COOEt		6.96	6.81	6.81	6.92
28 ^a	—OMe	—Cl	—H	—COOEt		6.92	7.27	6.92	7.17
29	—OMe	—Cl	—H	—COOEt		7.52	6.85	7.61	7.71
30	—OMe	—F	—H	—COOEt		6.77	7.22	6.60	6.56
31	—OMe	—F	—H	—COOEt		8.00	8.15	7.91	7.72

Table 1 (Continued)

No.	R ₁	R ₂	R ₃	R ₄	Ar	Activity			
						Exp.	MLR	CoMFA	CoMSIA
32 ^a	—OMe	—F	—H	—COOEt		7.00	6.95	6.40	6.91
33	—OMe	—F	—H	—COOEt		5.72	6.03	6.02	5.95
34	—OMe	—F	—H	—COOEt		6.70	5.91	6.57	6.41
35	—OMe	—F	—H	—COOEt		6.82	7.28	6.89	6.63
36	—OMe	—F	—H	—COOEt		7.30	6.85	7.10	7.04
37 ^a	—OMe	—F	—H	—COOEt		8.00	8.24	8.20	8.62
38	—OMe	—F	—H	—COOEt		8.00	7.79	8.04	8.10
39	—OMe	—F	—H	—COOEt		7.10	6.72	7.22	7.14
40	—OMe	—H	—H			6.96	6.96	6.95	6.94
41	—OMe	—H	—H			7.10	7.43	7.17	7.10
42	—OMe	—H	—H			6.35	6.37	6.35	6.22
43 ^a	—OMe	—F	—H			7.40	7.18	6.65	7.42
44	—OMe	—F	—H			7.70	7.69	7.64	7.76
45 ^a	—OMe	—F	—H			6.74	6.72	7.16	6.76

Table 1 (Continued)

No.	R ₁	R ₂	R ₃	R ₄	Ar	Activity			
						Exp.	MLR	CoMFA	CoMSIA
46	—OMe	—F	—H			7.00	6.96	7.21	6.89
47	—OMe	—F	—H			5.74	6.14	5.82	5.89
48	—OMe	—F	—H			7.30	7.46	7.34	7.20
49	—OMe	—F	—H			6.96	6.79	6.97	7.07

^a Compounds contained in the prediction set.

construction. Then, the operators of selection, crossover and mutation are used to produce the next generations. The evolution steps continue until the stopping conditions are satisfied. The important parameters that control the GA performance in the current research are listed as follows: population size (200), maximum generations (5000) and mutation probability (0.1).

The final model is built by using MLR method based on the selected descriptors, called GA-MLR, implemented in the software QSARINS [28,29]. This software, free for academia and research groups, is based on the GA-MLR method and implements various tools for a rigorous external validation of the models, based on different validation criteria, as well for the check of model applicability domain. The linear model is vigorously validated by both internal techniques (determination coefficient R^2 , leave-one-out (LOO) cross-validation Q_{LOO}^2 , leave-many-out (LMO) Q_{LMO}^2 , bootstrapping Q_{boot}^2 and root mean squared error (RMSE)) and external criteria (four different ways to calculate the correlation coefficient for prediction set samples Q_{F1}^2 [30], Q_{F2}^2 [31], Q_{F3}^2 [32], and CCC [33,34,35]). LMO and bootstrapping are repeated 5000 times with 35% of the objects left out randomly from the training set each time, and then the mean values of Q_{LMO}^2 and Q_{boot}^2 are reported. Additionally, the Y scrambling technique, which consists of shuffling the response data, is performed with iterations of 5000 to exclude the possibility of chance correlation in the proposed model.

The applicability domain (AD), a theoretical region defined by the descriptors used in modeling, is evaluated by leverage analysis [36,37]. The leverage (hat) is calculated by $h_i = x_i(X^T X)^{-1} x_i^T$ ($i = 1, \dots, m$), where x_i is the descriptor row-vector of the query compound i , m is the number of query compounds and X is the $n \times p$ matrix of the training set (p is the number of model descriptors and n is the number of training set samples). The limit of the model domain is quantitatively defined by the leverage cut-off h^* , set as $3(p+1)/n$. A leverage greater than h^* for the training set means that the sample is highly influential in determining the model, while for the test set (X outlier), it means that the prediction is the substantial extrapolation of the model and could be unreliable. Meanwhile, a compound with a standardized residual greater than 3σ (3 standard deviation units) is recognized as a Y outlier. In QSARINS software [28,29] useful plots to visualize the results of internal validation procedures and of AD study are available.

2.4. CoMFA and CoMSIA models

The 3D CoMFA and CoMSIA models are developed in SYBYL 6.9 program [21]. To construct 3D QSAR models, the molecules should be aligned together based on a certain rule. Because the molecules in our investigation have a common substructure (atoms present as balls in Fig. 2), the alignment, automatically executed in SYBYL program, is made based on this common substructure using “database alignment”. The most active compound 4 is used as the template for alignment.

The alignment of the molecules in the training set is put into a lattice that extends 4 Å units beyond the aligned molecules in all directions. Then, by using the sp^3 carbon atom with a +1 charge as a probe, the steric (Lennard–Jones potentials) and electrostatic (Coulomb potentials) field energies between the probe and molecules at each lattice intersection are calculated with a grid step of 2 Å, which are automatically generated by the CoMFA-STD method with cut off value of 30 kcal mol^{−1}.

CoMSIA [20] analysis has the advantage in exploring more similarity fields: steric (S), electrostatic (E), hydrophobic (H), H-bond donor (D) and acceptor (A) fields. All these fields are calculated according to the method developed by Klebe et al. [38], with the same lattice box, probe atom and grid as used for CoMFA calculations. The only adjustable parameter for CoMSIA field calculation is

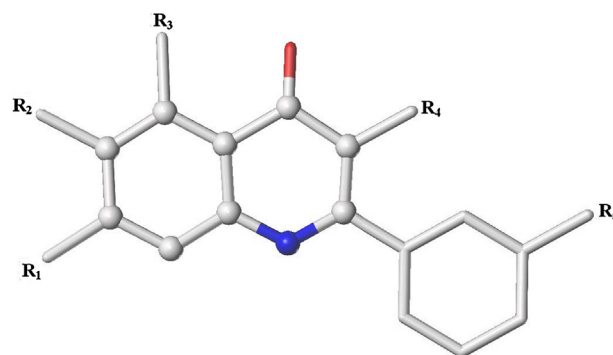


Fig. 2. Structure of the template (common atoms used for alignment are represented as balls).

Table 2

The selected descriptors used to build the linear model and corresponding meanings.

Descriptor	Meaning	Descriptor type	Stand coeff	P value
EEig12d	Eigenvalue 12 from edge adj. matrix weighted by dipole moments	Edge adjacency indices	1.426	0.000
HATS5p	Leverage-weighted autocorrelation of lag 5/weighted by atomic polarizabilities	GETAWAY descriptors	0.880	0.000
CIC3	Complementary information content (neighborhood symmetry of 3-order)	Information indices	0.659	0.000
BLI	Kier benzene-likeness index	2D Topological descriptors	−0.293	0.002
F10[O—O]	Frequency of O—O at topological distance 1	2D frequency fingerprints	−0.324	0.0001
E_strain	E minus energy of local minimum	MOE i3D descriptors	−0.455	0.000
B08[O—F]	Presence/absence of O—F at topological distance 8	2D binary fingerprints	−0.653	0.000
BEHv7	Highest eigenvalue n . 7 of Burden matrix/weighted by atomic van der Waals volumes	Burden eigenvalue descriptors	−1.112	0.000

the attenuation factor α , which will further affect the final model. The attenuation factor normally has an optimal value between 0.2 and 0.4; the default is 0.3 [39]. Large values result in a strong attenuation of the distance-dependent effects of molecular similarity.

Then, the partial least squares (PLS) regression analysis is used to derive 3D-QSAR models, both for CoMFA and CoMSIA. The leave-one-out cross-validation procedure in PLS is performed to obtain the optimal number of principal components. The final models, by noncross-validated PLS, are developed with the optimum number of components. Then, the correlation coefficient and the root mean squared error are subsequently calculated for the final models. The bootstrapping technique is employed for 100 runs to internally validate these models, and the values of Q^2_{boot} are reported. The same prediction set as MLR model is used to externally validate the predictive powers of the CoMFA and CoMSIA models.

3. Results and discussion

3.1. MLR model

The application of the GA provided a group of linear models containing different sets of molecular descriptors with different performances. A practical, useful QSAR model should have a high cross-validated performance (robust), a similarly high external predictive ability (predictive), and the least difference between internal and external predictive ability (stable and generalizable). On the basis of the above principles, the best linear model was built containing 8 descriptors, which are listed in Table 2. The intercorrelations between the selected descriptors are presented in Table 3. It can be seen that most of the variables used here are independent, except that the correlation coefficient between BEHV7 and EEig12d is a little high. This result is also acceptable, mainly because the resulting model demonstrates its ability in prediction for external compounds. When we deleted one of these two descriptors, the model performance decreased substantially. Furthermore, Table 2 shows that the P values of these two descriptors are very small, which means that the influence of the two descriptors is statistically significant; therefore, we kept both of them in the final model.

The linear model equation and the corresponding statistical items were described as follows:

$$\text{pEC}_{50} = 5.13\text{EEig12d} + 67.83\text{HATS5p} + 3.26\text{CIC3} - 6.13\text{BLI} \\ - 0.64\text{F10[O—O]} - 0.12\text{E_strain} - 2.69\text{B08[O—F]} \\ - 16.08\text{BEHV7} + 45.28$$

$$n_{\text{training}} = 38, \quad n_{\text{prediction}} = 11, \quad R^2 = 0.865, \quad Q^2_{\text{LOO}} = 0.769,$$

$$Q^2_{\text{LMO}} = 0.799, \quad Q^2_{\text{boot}} = 0.739, \quad Q^2_{\text{F1}} = 0.825, \quad Q^2_{\text{F2}} = 0.825$$

$$Q^2_{\text{F3}} = 0.826, \quad \text{CCC} = 0.928, \quad \text{RMSE}_{\text{train}} = 0.343,$$

$$\text{RMSE}_{\text{Prediction}} = 0.404, \quad F = 23.313$$

These parameters demonstrate that the model has good fitting ability with a high R^2 value and good stability with high Q^2 values from both LOO and LMO cross-validation techniques. The resulted correlation coefficients (R^2 and Q^2_{LOO}) from 5000 LMO procedures, together with those from the best linear model vs. K_{XY} (the inter-correlation among descriptors and response), were plotted in Fig. 3. In this figure, the majority of parameter values of LMO are around the model parameters, which indicates that the model is robust and stable. Fig. 4 shows the distribution of 5000 R^2 and Q^2_{LOO} from Y scrambling procedures and those from the best linear model. In this figure, it is evident that the correlation coefficients of the final model are much higher than those after endpoint scrambling because the relationship between the structure and response is broken. All these results indicate that the best linear model is not obtained by chance and that the relationship between the structures of 3-carboxyl-4(1H)-quinolone analogs and corresponding antimalarial activities does exist.

The predictive ability of a QSAR model for new chemicals is most important for its practical usage [36,37]. After the model was built, it was used to evaluate the prediction set samples, that is to predict new chemicals not used in model development and the predictions were listed in Table 1. Then, the Williams plot, standardized residuals vs. leverage values shown in Fig. 5, was used to illustrate

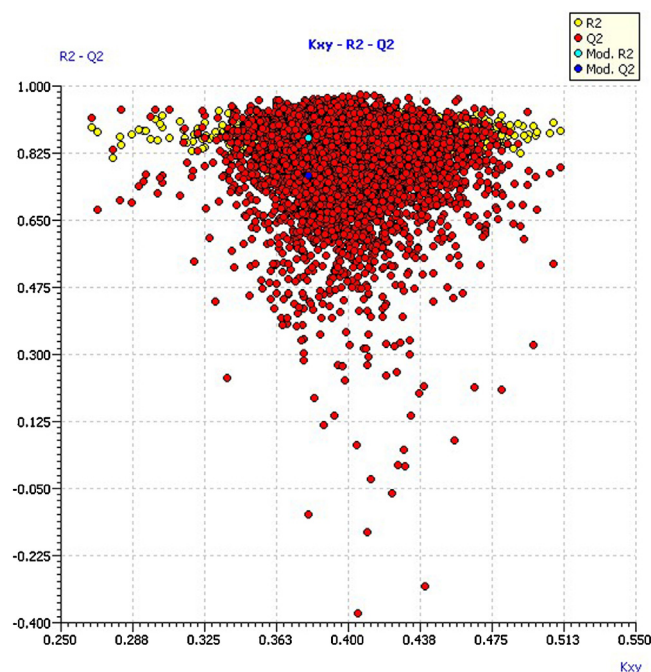
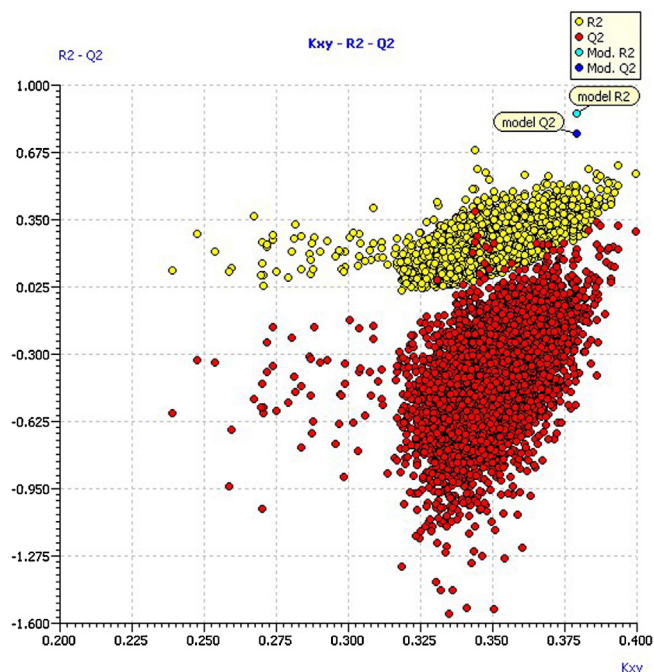
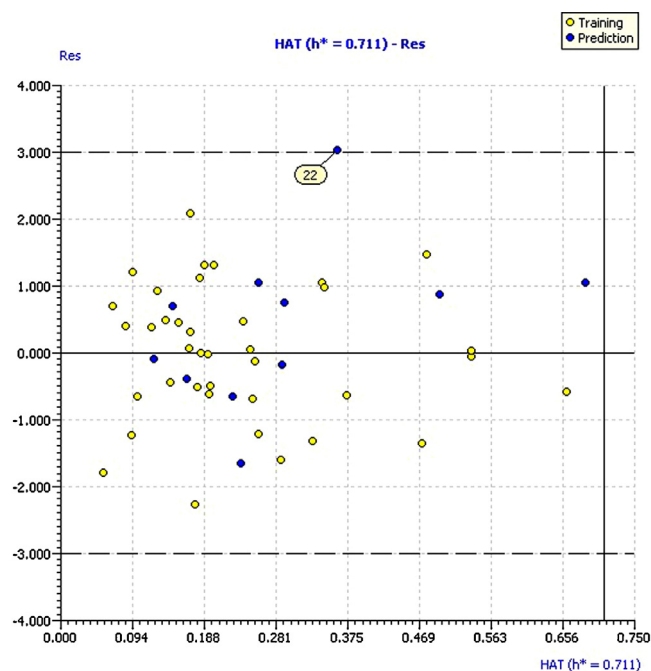


Fig. 3. The LMO scatter plot (plot of K_{XY} vs. the R^2 and Q^2_{LOO} from LMO procedure and the linear model).

Table 3

The intercorrelation coefficients of the selected descriptors in the linear model.

	EEig12d	HATS5p	CIC3	BLI	F10[O—O]	E_strain	B08[O—F]	BEHv7
EEig12d	1							
HATS5p	−0.17	1						
CIC3	0.55	−0.28	1					
BLI	−0.10	0.31	−0.13	1				
F10[O—O]	0.18	0.09	−0.07	0.05	1			
E_strain	0.55	0.00	0.26	0.02	0.12	1		
B08[O—F]	0.03	−0.03	0.04	−0.34	−0.09	−0.12	1	
BEHv7	0.84	0.01	0.68	0.07	0.08	0.52	−0.07	1

**Fig. 4.** Y-Scramble plot (plot of K_{XY} vs. the R^2 and Q_{LOO}^2 from Y-scrambling procedure and the linear model).**Fig. 5.** Williams plot of the linear model. The dashed lines are the cut-off 3σ and the warning value of h^* (0.711).

the prediction and express the applicability domain of the model. From Fig. 5, it can be seen that all the molecules are located in the applicability domain of the model with leverage values lower than the warning h^* of 0.711. The prediction samples are all well-predicted, except that the standard residue of compound 22 is a little higher, similar to the cut-off of 3σ . Here the values of Q_{F1}^2 , Q_{F2}^2 and Q_{F3}^2 for the prediction set are quite similar and all very high. The parameter CCC is higher than 0.85, as suggested in reference for predictive models [33–35]. The scatter plot of the experimental vs. the calculated antimalarial activities of these quinolones is shown in Fig. 6. In this figure, all the samples are close to the line $y = x$, which means that the predicted values are similar to the corresponding experimental values.

Among the selected descriptors, seven are from the DRAGON program. EEig12d, eigenvalue 12 from an edge-adjacency matrix weighted by dipole moments, belongs to the edge adjacency index. This index takes into account the molecular polarity. The molecular antimalarial activity increases with the increase of this descriptor. Generally, molecules with lower descriptor values are least active compounds, such as compounds 5 and 6. This descriptor is the most significantly related to the antimalarial activity. HATS5p is a GETWAY descriptor, representing the leverage-weighted autocorrelation of lag 5/weighted by atomic polarizabilities. This descriptor is favorable to the molecular antimalarial activities. Compounds with large descriptor values generally have high activities, for example compounds 4 and 38. CIC3 is an information index

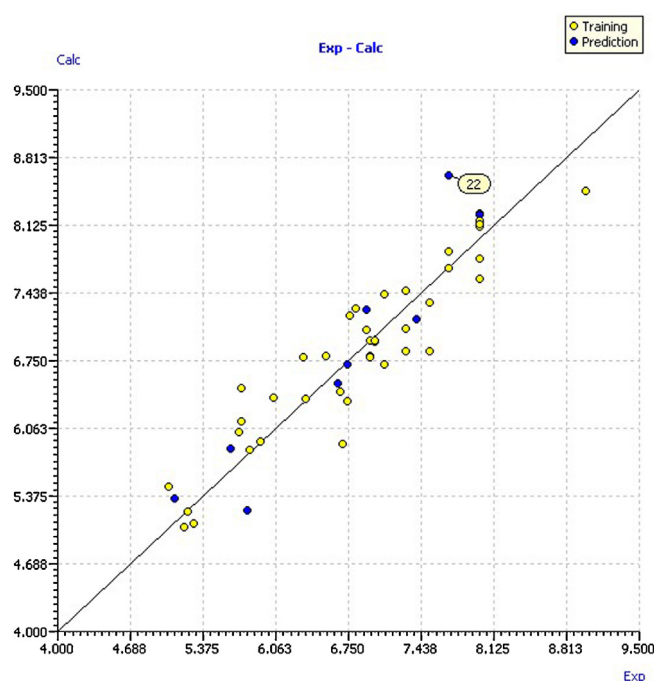
**Fig. 6.** The scatter plot of experimental vs. calculated antimalarial activities by the linear model.

Table 4
Summary of the CoMFA and CoMSIA analysis Parameters.

Parameter	CoMFA	CoMSIA
PLS statistics		
R^2 (squared correlation coefficient)	0.977	0.973
ONC (the optimal number of components)	6	6
Q_{LOO}^2 (Squared LOO cross-validated correlation coefficient)	0.650	0.777
Q_{Boot}^2 (bootstrapping analysis for 100 runs)	0.985	0.977
Q_{F1}^2	0.818	0.894
Q_{F2}^2	0.818	0.894
Q_{F3}^2	0.819	0.894
CCC	0.897	0.958
RMSE _{Train}	0.140	0.153
RMSE _{Prediction}	0.411	0.314
F	224.353	187.070
Field distribution (%)		
Steric	67.2	20.0
Electrostatic	32.8	34.8
Hydrophobic	–	45.2

describing the complementary information content (neighborhood symmetry of 3-order). The complementary information content (CIC) descriptor measures the deviation of the information content from its maximum value. That value corresponds with the vertex partition into equivalence classes containing one element each: $CIC = \log_2 nAT-IC$, where nAT is the total number of atoms and IC represents a measure of structural complexity per vertex. BLI, Kier benzene-likeness index, a 2D topological descriptor, indicates the Kier benzene-likeness index. BLI is calculated by dividing the first-order valence connectivity index $X1V$ by the number of non-H bonds (nBO) of the molecule and then normalizing on the benzene molecule. It was proposed to measure the molecule aromaticity. BLI is inversely related to the antimalarial activity. For example, compounds 7 and 17 have largest descriptor values of 1.037 and 1.004, but the activities of these two compounds are as lower as 5.20 and 5.29. F10[O–O], a frequency fingerprint descriptor, measures the frequency of O–O at topological distance 10. B08[O–F] is a 2D binary fingerprint, indicating the presence/absence of O–F at topological distance 8. The negative coefficients of these two fingerprint descriptors suggest that the presence of O–F and O–O with certain distances will decrease the antimalarial activities of the 3-carboxyl-4(1H)-quinolones. BEHV7, a burden eigenvalue descriptor, describes the highest eigenvalue n_7 of Burden matrix/weighted by atomic van der Waals volumes. The increase of the descriptor value is disadvantageous to the molecular antimalarial activity. E.strain is an i3D descriptor from MOE program, indicating the current energy E minus the value of the energy at a near local minimum. During minimization, the current chirality is preserved and charges are left undisturbed. This descriptor is also inversely related to the molecular bioactivity.

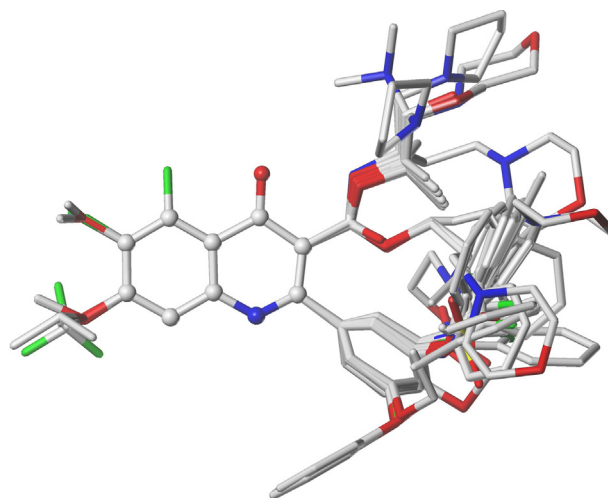
3.2. CoMFA model

To construct 3D models, all the compounds were aligned with the common atoms presented as balls in Fig. 2 by using the most active compound 4 as template. The resulting alignment is shown in Fig. 7. Because of the rigidity of the two-fused ring, the alignment is very ideal. Based on this alignment, 3D QSAR analyses CoMFA and CoMSIA were developed.

The CoMFA PLS analysis results are listed in Table 4. The model gives a R^2 of 0.977 and a LOO cross-validated Q_{LOO}^2 of 0.650 based on six components. The bootstrapping analysis was performed for 100 runs to further validate the model, and a correlation coefficient Q_{Boot}^2 as high as 0.985 was obtained. All these parameters suggest that the CoMFA model is robust and stable and a good internal consistency exists within the underlying data set.

The predicted results for the prediction set from the CoMFA model were listed in Table 1, and then the four different ways to calculate the predictivity performances for the prediction set samples: Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2 and CCC were all employed for the first time in a CoMFA analysis. The values of these four parameters for the CoMFA model reveal a similar performance with the linear model; both are very predictive. It can be seen clearly from the scatter plot of the experimental vs. the predicted antimalarial activities (Fig. 8) that the calculated pEC_{50} of the prediction set samples are in good agreement with the experimental data. The compound 22 from prediction set is much nearer to the line $y=x$ than in MLR model. The corresponding field contributions of steric and electrostatic are 67.2% and 32.8%.

To intuitively understand the relationships between molecular structure and antimalarial activities of these derivatives, the final partial least square analysis was further used to generate 3D contour maps using the field type StDEV*coeff. The contour maps can provide a more detailed understanding of the key structural features required for the antimalarial activity. The CoMFA steric and electrostatic contour maps were presented in Fig. 9a and b, respectively. For the steric field, the green regions indicate that bulky groups can increase molecular bioactivity, while yellow regions indicate that bulky groups will decrease bioactivity. In Fig. 9a, there are primarily two green regions and three yellow regions. The largest green contour appears at position R_5 , where a bulky group is preferred. For example, the activity of compound 31 with PhO– is higher than compound 30 with MeO– and 32 with CF_3O –. A

**Fig. 7.** The alignment of molecules using the lowest-energy conformations.

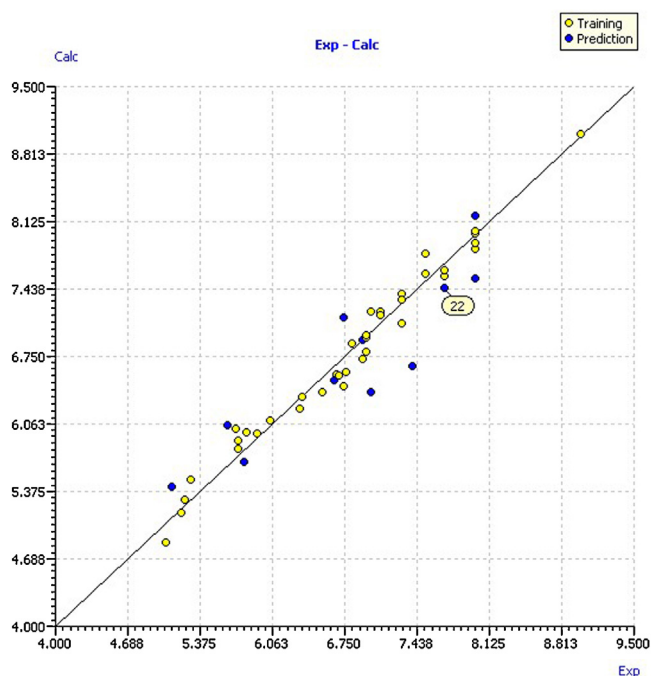


Fig. 8. The scatter plot of experimental vs. calculated antimalarial activities by the 3D CoMFA model.

similar example is the comparison of compounds 25 and 26. However, this substructure cannot be too large; otherwise, it will push the R_4 group up, which is unfavorable to the molecular activity because of the yellow region B in Fig. 9a. Another smaller green region appears near the R_2 position. It is reported that a halogen is preferred here [4]. From Fig. 9a, we can further conclude that

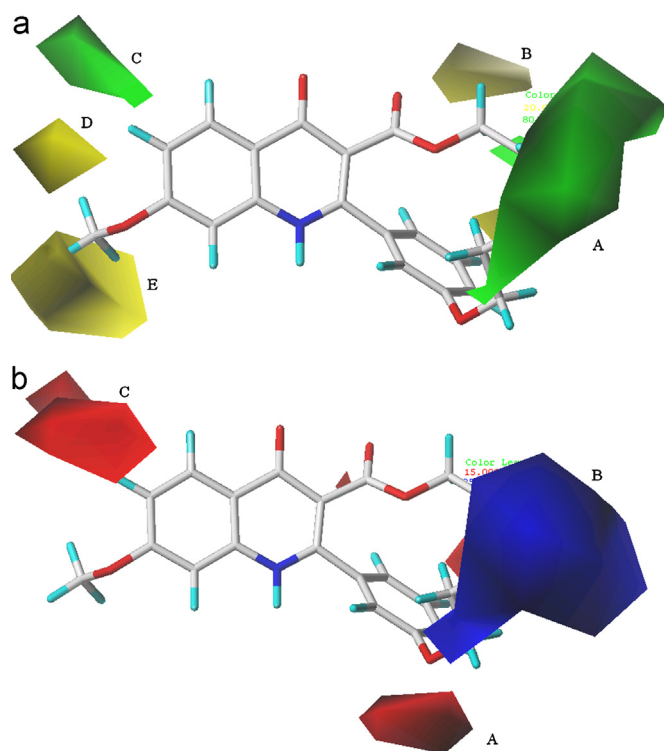


Fig. 9. The CoMFA STDEV*COEFF contour maps based on molecule 4. (a) The steric field. (b) The electrostatic field. (For interpretation of the references to color in text, the reader is referred to the web version of this article.)

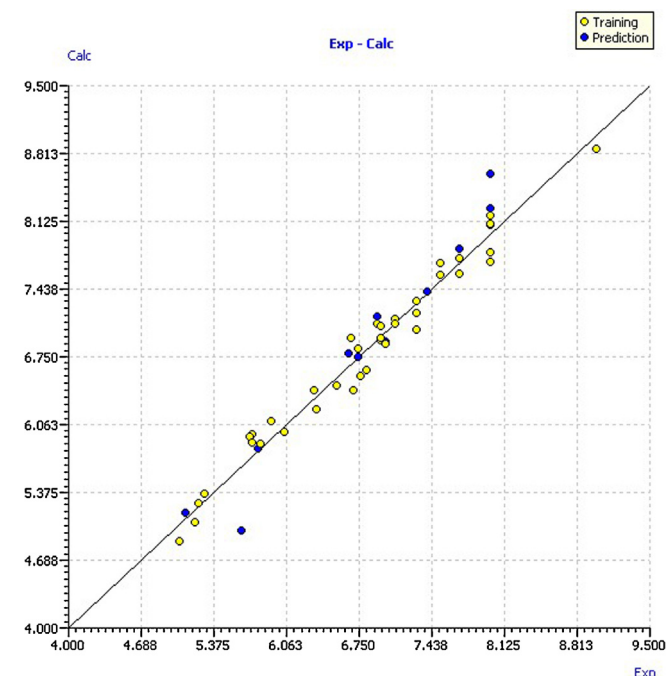


Fig. 10. The scatter plot of experimental vs. calculated antimalarial activities by the 3D CoMSIA model.

a halogen with larger volume will result in higher activity than smaller halogen and hydrogen, such as compounds 40 and 43, 41 and 44. The last two yellow regions are surrounding the R_1 position, demonstrating that a bulky group is not expected here. Actually, it has been proved that a small hydrophobic group at this position benefits the antimalarial potency, MeO— preferred [18].

For the electrostatic field, red contours indicate regions where electronegative or electron-donating groups will increase bioactivity, and blue contours represent regions where electropositive or electron-withdrawing groups will increase bioactivity. There are primarily two red regions and one blue region. The first red contour A in the CoMFA electrostatic contour map of Fig. 9b appears near the meta-substitution of the aromatic ring, indicating an electronegative group such as —O— is favorable to the molecular activity, whereas an electro-withdrawing group here generally decreases the activity. This point can be proved by the fact that compounds 5–10 have lower activities than others. Additionally, if an aromatic ring is connected to form a mode such as Ph—O—Ph at R_5 position, the density of the electron cloud near oxygen will be enhanced and will be impaired around Ph, which is the reason that a blue region B appears nearby the oxygen in this figure. Of course, phenyl is a special structure and can interact with the target through various mechanisms, but at least this could be a reason that a quinolone with Ph—O—Ph mode generally has high antimalarial potency [18]. Another red region C is located near the R_2 position, which means that an electronegative group here can improve the activity, which could be proven by the fact that a halogen atom at the R_2 position is superior to hydrogen from the perspective of electrostatic fields.

3.3. CoMSIA model

On the basis of steric and electrostatic fields used in CoMFA, we combined different fields in CoMSIA to evolve the best model. The results were summarized in Table 5. It is obvious that by adding hydrophobic field to S and E, the combination of SEH increases the Q^2_{LOO} value from 0.650 (CoMFA) to 0.771, whereas the addition of D and/or A will decrease, even dramatically in some cases, the total performance. Therefore, the model with steric, electrostatic and

Table 5
Different field combinations in CoMSIA analysis keeping S and E fields in all cases.

Field	ONC	Q^2_{LOO}	R^2	SEE ^a	F	Fraction
SEH	6	0.771	0.976	0.158	209.36	0.201:0.337:0.462
SED	2	0.013	0.621	0.591	28.71	0.306:0.527:0.167
SEA	3	0.092	0.671	0.559	23.08	0.218:0.475:0.307
SEHD	6	0.686	0.968	0.184	154.28	0.190:0.340:0.425:0.045
SEHA	6	0.701	0.971	0.175	170.26	0.172:0.261:0.424:0.143
SEDA	3	0.088	0.732	0.505	30.91	0.192:0.404:0.139:0.265
SEHDA	6	0.604	0.956	0.214	112.52	0.159:0.250:0.377:0.064:0.150

^a Standard error of estimate.

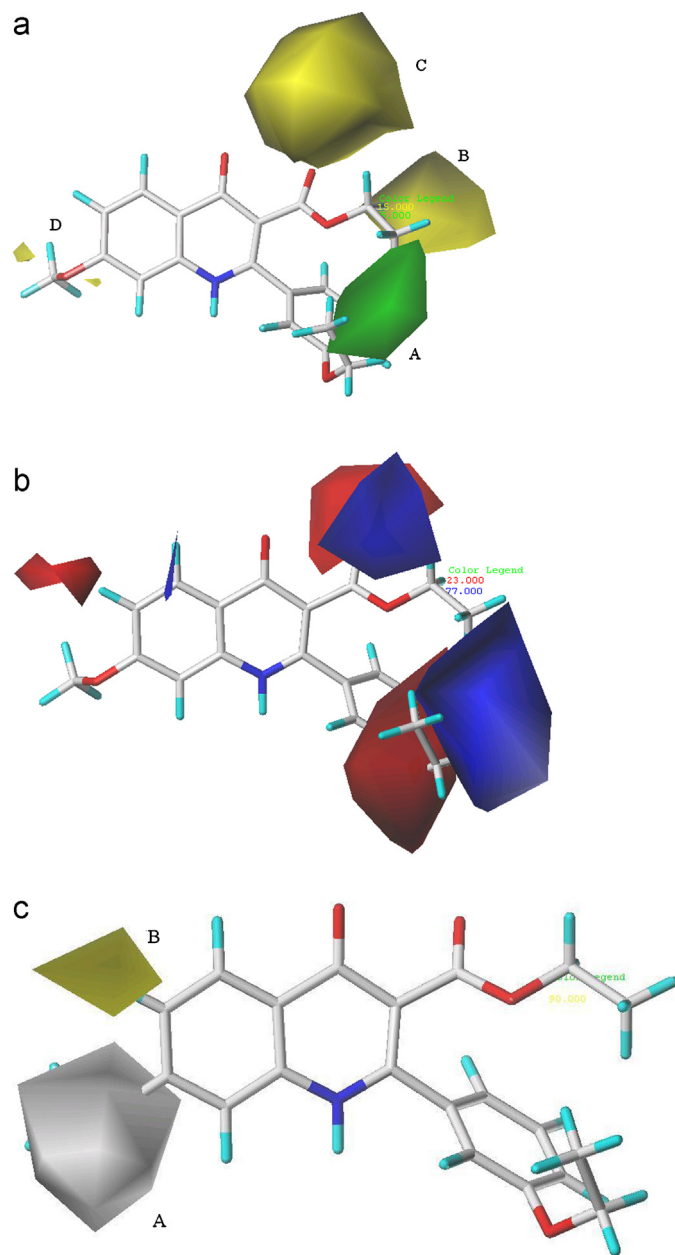


Fig. 11. The CoMSIA STDEV*COEFF contour maps based on molecule 4. (a) The steric field. (b) The electrostatic field. (c) The hydrophobic field.

hydrophobic fields is chosen for further analysis. Furthermore, the value of the attenuation factor α was optimized. We explored different values near 0.3 (between 0.2 and 0.4) with a step of 0.02. As a result, 0.26 was selected as an optimal value, which resulted in a LOO cross-validated Q^2 as high as 0.777.

By using the SEH fields and the optimal attenuation factor α , the final CoMSIA model was built based on six principal components. The corresponding statistical parameters are listed in Table 4. The CoMSIA model has similar fitting ability and bootstrapping performance as the CoMFA model. Comparatively, the CoMSIA model has higher Q^2_{LOO} value and its predictive ability, validated on the same prediction set samples, is better because all four of the different forms used to calculate correlation coefficient for the prediction set give higher values than CoMFA model. The scatter plot of the experimental activities vs. calculated activities from CoMSIA model for the training set and prediction set were depicted in Fig. 10. In this figure, all the compounds are located near the diagonal line. Overall, the CoMSIA model performs better than the CoMFA model.

The contour maps of the SEH CoMSIA model are shown in Fig. 11. For the steric and electrostatic contour maps, the CoMSIA model shows similar results as CoMFA model, especially on the R_4 and R_5 positions. On the left side there are two small regions for steric field, similar to CoMFA, but much smaller, while the electrostatic contours are quite similar. The CoMSIA hydrophobic contour map is shown in Fig. 11c. In this figure, a large white region overlaps the methoxy group at R_1 position, which indicates that the hydrophobic property on this position could affect the molecular bioactivity, and the higher the hydrophobic property is, the lower the bioactivity is. This point is consistent with the conclusion from the reference that a small hydrophobic group at 7-position (R_1) benefits the antimalarial potency of 4(1H)-quinolone analogs [18]. Another contour is represented as yellow near the R_2 position. The yellow color indicates that a hydrophobic substituent is favorable for activity, so a hydrophobic halogen at this position is anticipated, which exhibited the best activities, especially F and Cl, as stated in Ref. [4].

4. Conclusion

In this study, the structure–activity relationships of a series of 3-carboxyl-4(1H)-quinolone analogs with corresponding antimalarial activities were investigated by using 2D MLR and 3D CoMFA and CoMSIA methods. Three robust, stable and predictive QSAR models were established. The CoMFA and CoMSIA models were further used to generate the 3D contour maps to provide a more detailed understanding of the key structural features responsible for the antimalarial activity. Combining the results from the current study and the conclusions from other papers, we summarized the position-specific structural requirements to improve the antimalarial activities of the series of 3-carboxyl-4(1H)-quinolones:

R_1 : a small hydrophobic electro-donating group benefits the antimalarial activity;

R_2 : a hydrophobic and electronegative atom is expected, especially halogen, which will greatly improve the activity, and the larger the atom volume is, the higher the activity is;

R_1 – R_3 : dimethoxy or dihalogen substituted quinolones have reduced or even lost their activity;

R₄: important to the solubility of the series of quinolone analogs, an electro-withdrawing substructure is anticipated. The substructure should not be too large, and an ethyl ester group is favored;
R₅: a bulky (not too large, maybe moderate like a substituted benzene) group connected with the benzene ring by an electro-donating atom will greatly improve their activities.

All these results can provide helpful information to aid the rational design of new potent antimalarial quinolones with higher activity.

Acknowledgments

This research is supported by the Fundamental Research Funds for the Central Universities (Izujby-2012-88) and Gansu Provincial Natural Science Foundation (1208RJYA038).

References

- [1] G. Poinar Jr., *Plasmodium dominicana* n. sp. (Plasmodiidae: Haemospororida) from tertiary Dominican amber, *Systematic Parasitology* 61 (2005) 47–52.
- [2] WHO, World Malaria Report 2011, 2011 <http://www.who.int/malaria/world-malaria-report-2011/en/index.html>
- [3] R.W. Winter, J.X. Kelly, M.J. Smilkstein, R. Dodean, D. Hinrichs, M.K. Riscoe, Antimalarial quinolones: synthesis, potency, and mechanistic studies, *Experimental Parasitology* 118 (2008) 487–497.
- [4] R.M. Cross, A. Monastyrskyi, T.S. Mutka, J.N. Burrows, D.E. Kyle, R. Manetsch, Antimalarial quinolones: synthesis, potency, and mechanistic studies, *Journal of Medicinal Chemistry* 53 (2010) 7076–7094.
- [5] T.E. Wellems, C.V. Plowe, Chloroquine-resistant malaria, *Journal of Infectious Diseases* 184 (2001) 770–776.
- [6] A.B.S. Sidhu, D. Verdier-Pinard, D.A. Fidock, Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcrt mutations, *Science* 298 (2002) 210–213.
- [7] J.E. Hyde, Drug-resistant malaria, *Trends in Parasitology* 21 (2005) 494–498.
- [8] N.J. White, P. Olliaro, Artemisinin and derivatives in the treatment of uncomplicated malaria, *Medicina Tropical* 58 (1998) 54–56.
- [9] P.K. Ojha, K. Roy, Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection, *Chemometrics and Intelligent Laboratory Systems* 109 (2011) 146–161.
- [10] W. Salzer, H. Timmler, H. Andersag, A new type of compounds active against avian malaria, *Chemische Berichte* 81 (1948) 12–19.
- [11] A.C. Casey, 4(1H)-quinolones. 2. Antimalarial effect of some 2-methyl-3-(1-alkenyl)- or -3-alkyl-4(1H)-quinolones, *Journal of Medicinal Chemistry* 17 (1974) 255–256.
- [12] R.M. Cross, J.R. Maignan, T.S. Mutka, L. Luong, J. Sargent, D.E. Kyle, R. Manetsch, Optimization of 1,2,3,4-tetrahydroacridin-9(10H)-ones as antimalarials utilizing structure–activity and structure–property relationships, *Journal of Medicinal Chemistry* 54 (2011) 4399–4426.
- [13] W. Duerckheimer, W. Raether, H.G. Seliger, Tetrahydroacridone Chemotherapeutic Agent. 73-23374742337474, 19730724 (1975).
- [14] W. Raether, E. Fink, Anti-malarial activity of floxacrine (Hoe991). 1. Studies on blood schizontocidal action of floxacrine against plasmodium-berghei, plasmodium-vinckei and plasmodium-cynomolgi, *Annals of Tropical Medicine and Parasitology* 73 (1979) 505–526.
- [15] L.H. Schmidt, Antimalarial properties of floxacrine, a dihydroacridinedione derivative, *Antimicrobial Agents and Chemotherapy* 16 (1979) 475–485.
- [16] J.F. Ryley, W. Peters, Antimalarial activity of some quinolone esters, *Annals of Tropical Medicine and Parasitology* 64 (1970) 209–222.
- [17] R. Winter, J.X. Kelly, M.J. Smilkstein, D. Hinrichs, D.R. Koop, M.K. Riscoe, Optimization of endochin-like quinolones for antimalarial activity, *Experimental Parasitology* 127 (2011) 545–551.
- [18] Y. Zhang, J.A. Clark, M.C. Connelly, F. Zhu, J. Min, W.A. Guiguemde, A. Pradhan, L. Iyer, A. Furimsky, J. Gow, T. Parman, F.E. Mazouni, M.A. Phillips, D.E. Kyle, J. Mirsalis, R.K. Guy, Lead optimization of 3-carboxyl-4(1H)-quinolones to deliver orally bioavailable antimalarials, *Journal of Medicinal Chemistry* 55 (2012) 4205–4219.
- [19] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society* 110 (1988) 5959–5967.
- [20] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *Journal of Medicinal Chemistry* 37 (1994) 4130–4146.
- [21] SYBYL, Version 6.9, Tripos Inc., St. Louis, MO.
- [22] DRAGON for Windows (Software for molecular Descriptor Calculation), Version 5.5 – Talete srl. <http://www.talete.mi.it>, 2007.
- [23] Maestro, version 8.5 Schrödinger, LLC, New York, NY, 2008.
- [24] QikProp, Version 1.6, Schrödinger, New York, NY, 2001.
- [25] Molecular Operating Environment (MOE), 2009.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2011.
- [26] R. Leardi, R. Boggia, M. Terrile, Genetic algorithms as a strategy for feature selection, *Journal of Chemometrics* 6 (1992) 267–281.
- [27] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships, *Journal of Chemical Information and Computer Science* 34 (1994) 854–866.
- [28] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS, Software for QSAR MLR Model Development and Validation, University of Insubria, Varese, Italy, 2012 <http://www.qsar.it>
- [29] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS. A new software for the development, analysis and validation of QSAR MLR models, *Journal of Computational Chemistry* (2013) (in press).
- [30] <http://www.oecd.org/dataoecd/33/37/37849783.pdf> (accessed 22.01.12).
- [31] G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kühne, External validation prediction employing the predictive squared correlation coefficient test set activity mean vs. training set activity mean, *Journal of Chemical Information and Modeling* 48 (2008) 2140–2145.
- [32] V. Consonni, D. Ballabio, R. Todeschini, Comments on the definition of the Q² parameter for QSAR validation, *Journal of Chemical Information and Modeling* 49 (2009) 1669–1678.
- [33] L.I. Lin, A concordance correlation coefficient to evaluate reproducibility, *Biometrics* 45 (1989) 255–268.
- [34] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, *Journal of Chemical Information and Modeling* 51 (2011) 2320–2335.
- [35] N. Chirico, P. Gramatica, Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection, *Journal of Chemical Information and Modeling* 52 (2012) 2044–2058.
- [36] P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR & Combinatorial Science* 26 (2007) 694–701.
- [37] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, *QSAR & Combinatorial Science* 22 (2003) 69–77.
- [38] G. Klebe, U. Abraham, Comparative molecular similarity index analysis (CoMSIA) to study hydrogen bonding properties and to score combinatorial libraries, *Journal of Computer-Aided Molecular Design* 13 (1991) 1–10.
- [39] M. Böhm, J. Stürzebecher, G. Klebe, Three-dimensional quantitative structure–activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa, *Journal of Medicinal Chemistry* 42 (1999) 458–477.