SAR and QSAR in Environmental Research

Application of the linear interaction energy method for rational design of artemisinin analogues as haeme polymerisation inhibitors

M. Srivastava [a] , H. Singh [a] & P.K. Naik [a]

[a] Department of Biotechnology and Bioinformatics , Jaypee
University of Information Technology, Waknaghat , Solan 173215,
Himachal Pradesh, India
Published online: 04 Dec 2010.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Application of the linear interaction energy method for rational design of artemisinin analogues as haeme polymerisation inhibitors

M. Srivastava, H. Singh and P.K. Naik*

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173215, Himachal Pradesh, India

The anti-malarial activity of artemisinin-derived drugs appears to be mediated by an interaction of the drug's endoperoxide bridge with intra-parasitic haeme. The binding affinity of artemisinin analogues with haeme were computed using linear interaction energy with a surface generalised Born (LIE-SGB) continuum solvation model. Low levels of root mean square error (0.348 and 0.415 kcal/mol) as well as significant correlation coefficients ($r^2 = 0.868$ and 0.892) between the experimental and predicted free energy of binding (FEB) based on molecular dynamics and hybrid Monte Carlo sampling techniques establish the SGB-LIE method as an efficient tool for generating more potent inhibitors of haeme polymerisation inhibition.

Keywords: artemisinin; free energy of binding; binding affinity; docking; surface generalised Born continuum solvation model; linear interaction energy

1. Introduction

Malaria is one of the most widespread and prevalent endemic diseases; it threatens approximately 40% of the world's population in more than 107 countries. This disease is estimated to cause approximately 350–500 million clinical illnesses and up to 3 million deaths each year [1]. Most deaths are attributed to the parasite *Plasmodium falciparum*. The enzymes in the parasite digestive vacuole (cysteine- and aspartic-proteinases) break down haemoglobin into amino acids and haeme [2]. While all of the amino acid content is used to build parasite proteins, only a small portion of the haeme is incorporated into the parasite haemoproteins; the rest of the haeme is detoxified (polymerised) by parasite enzymes [3].

A number of drugs have been investigated for their use in the treatment of malaria. However, new strains of *Plasmodium falciparum* resistant to some of these drugs, e.g. chloroquine, quinine and mefloquine, are causing substantial deterioration in clinical treatment [4–8]. This has motivated the search for new anti-malarial drugs that are effective against this form of malaria, thus having a very high priority in anti-malarial drug design [9–11]. This led to Chinese researchers introducing a new compound, qinghaosu (or artemisinin, as it is known in the West), present in extracts of Qinghao or *Artemisia annua* L., which has been used in China for thousands of years [12]. It is a potent anti-malarial drug against the multi-drug resistant strains of *Plasmodium falciparum* [13,14].

---

*Corresponding author. Email: pknaik73@rediffmail.com
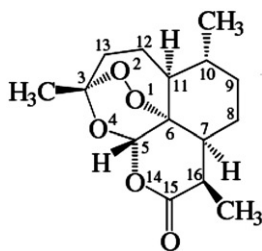
Figure 1. Sterochemistry and atomic numbering scheme of artemisinin.
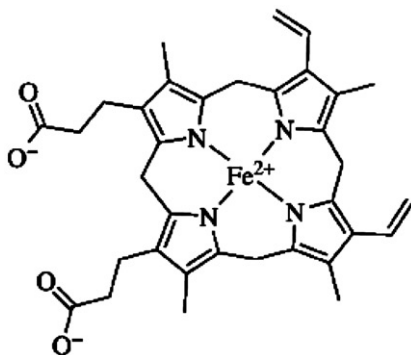
Figure 2. The structure of the haeme compound.

The structure of artemisinin was identified as an endoperoxide containing sesquiterpene lactone (Figure 1) and the presence of the 1,2,4-trioxane-ring system seems to be essential for its anti-malarial activity [15–19]. Studies on the mode of action of artemisinin and its derivatives have shown that free haeme (Figure 2) could be the molecule targeted by artemisinin in biological systems and that Fe^{+2} ions interact with the peroxide when artemisinin react with haeme [9,11,20–23]. An initial step in the action of artemisinin includes haeme-catalysed artemisinin activation into a very reactive radical, which binds to the parasite proteins or haeme [5,11,21–24] and haemozoin [4,21–25]. It has been proposed that haeme iron attacks the endoperoxide linkage of artemisinin either at the O1 [26] or O2 position [27]. In pathway A, haeme iron attacks the compound at the O2 position and produces a free radical at the O1 position. Later it rearranges to form the C4 free radical. In pathway B, haeme iron attacks the compound at the O1 position and produces a free radical at the O2 position. After that the C3–C4 bond is cleaved to give a carbon radical at C4. It has been suggested that the C4 free radical in both pathways is an important substance in anti-malarial activity [28].

The effectiveness of artemisinin and its derivatives as anti-malarial drugs for the treatment of multi-drug resistant *P. falciparum* has received considerable attention in recent years. More often than not the focus of these studies has been to demonstrate anti-malarial efficacy *in vitro* for new structural classes or modification of the natural product architecture. As a wide variety of molecular scaffolds are available for optimisation, this diversity presents a significant challenge in determining the essential features for activity. A rational approach for the discovery of a pharmaceutically acceptable, economically

viable, peroxide-based anti-malarial awaits the development of a global mechanism of action model for organic peroxides [29,30] and/or a predictive quantitative structure–activity relationship (QSAR). With the advent of parallel synthesis methods and technology, we might expect the number of anti-malarial artemisinin analogues to be tested to grow dramatically. Combinatorial methods could also be envisioned as a semi-rational approach to the above discovery strategy.

QSAR models using calculated molecular descriptors of ligand have been developed earlier for prediction of biological activity of artemisinin analogues [31,32]. However, these ligand-based models predict the biological activity with high prediction error (0.76) [31] and also use a smaller dataset (19 compounds) [32]. Another method of orchestrating these strategies is to make use of structure-based linear interaction energy (LIE) models for the rapid prediction and virtual prescreening of anti-malarial activity. The LIE approximation is a way of combining molecular mechanics calculations with experimental data to build a model scoring function for the evaluation of ligand–protein binding free energies. A LIE method for rational design of artemisinin analogues for inhibition of haeme polymerisation has not yet been determined.

The availability of the X-ray structure of haeme helps to facilitate the understanding the structure–activity relationships (SARs) for haeme polymerisation and enables molecular modelling techniques to be applied for designing novel and more potent inhibitors. In this study we have applied a structure-based LIE method implementing a surface generalised Born (SGB) [33] continuum model for solvation, SGB-LIE, to build a binding affinity model for estimating the free energy of binding for a diverse set of haeme inhibitors. The LIE method [34,35] has been applied to a number of protein–ligand systems with promising results [36–38] producing small errors of the order of 1 kcal/mol for free energy prediction [39]. The magnitude of free energy changes upon binding of inhibitors to haeme that directly correlates with the experimental potency of these inhibitors. Hence, fast and accurate estimation of binding free energies provides a means to screen the compound libraries for lead optimisation and rational design. This could bring about the development of new and more effective drugs.

2. Materials and methods

2.1 *LIE methodology*

The LIE method employs experimental data on binding free energy values for a set of ligands (referred to as the training set) to estimate the binding affinities for a set of novel compounds. The method is based on the linear response approximation (LRA), which dictates that the binding free energy of a protein–ligand system is a function of polar and non-polar energy components that scale linearly with the electrostatic and van der Waals (VDW) interactions between a ligand and its environment. The free energy of binding (FEB) for the complex is derived from considering only two states: (1) free ligand in the solvent and (2) ligand bound to the solvated protein. The conformational changes and entropic effects pertaining to unbound receptor are taken into account implicitly and only interactions between the ligand and either the protein or solvent are computed during molecular mechanics calculations. Among the various formulations of the LIE methodology developed in the past, the SGB-LIE method [40] has been shown to be an order of

magnitude faster than the methods based on explicit solvent with the same order of accuracy. In the LIE method,

$$\Delta G_{\text{bind}} = \alpha \langle \Delta U_{\text{ele}} \rangle + \beta \langle \Delta U_{\text{vdw}} \rangle + \gamma \langle \Delta_{\text{SASA}} \rangle \tag{1}$$

where $\langle \Delta U_{\text{ele}} \rangle$ and $\langle \Delta U_{\text{vdw}} \rangle$ denote the average change in the electrostatic and VDW interaction energy of the ligand in the free and bound states, respectively and $\langle \Delta_{\text{SASA}} \rangle$ is the change in the solvent-accessible surface area (SASA) of the ligand. The α, β and γ terms are adjustable parameters that need to be determined by fitting the experimental data on the training set compounds.

The SGB-LIE method also offers better accuracy in treating the long-range electrostatic interactions. However, the SGB-LIE method used in this studied is based on the original formulation proposed by Carlson and Jorgensen [40] and implemented in *Liaison* (Schrödinger, Inc. Portland, OR) using the OPLS-2005 force field. A novel feature of Liaison is that the simulation takes place in implicit (continuum) rather than explicit solvent—hence the name Liaison, for Linear Interaction Approximation in Implicit Solvation. The explicit-solvent version of the methodology was first suggested by Hansson and Åqvist [41], based on approximating the charging integral in the free-energy-perturbation formula with a mean-value approach, in which the integral is represented as half the sum of the values at the endpoints, namely the free and bound states of the ligand. The empirical relationship used by Liaison is

$$\Delta G_{\text{bind}} = \alpha(\langle U_{\text{ele}}^{\text{b}} \rangle - \langle U_{\text{ele}}^{\text{f}} \rangle) + \beta(\langle U_{\text{vdw}}^{\text{b}} \rangle - \langle U_{\text{vdw}}^{\text{f}} \rangle) + \gamma(\langle U_{\text{cav}}^{\text{b}} \rangle - \langle U_{\text{cav}}^{\text{f}} \rangle) \tag{2}$$

Here $\langle \rangle$ represents the ensemble average, b represents the bound form of the ligand, f represents the free form of the ligand, and α, β and γ are the coefficients; U_{ele}, U_{vdw} and U_{cav} are the electrostatic, VDW and cavity energy terms in the SGB continuum solvent model. The cavity energy term, U_{cav}, is proportional to the exposed surface area of the ligand. Thus, the difference $\langle U_{\text{cav}}^{\text{b}} \rangle - \langle U_{\text{cav}}^{\text{f}} \rangle$ measures the surface area lost by contact with the receptor.

The energy terms involved can be computed using energy minimisation, molecular dynamics or Monte Carlo calculations. In the SGB model of solvation, there is no explicit VDW or electrostatic interaction between the solute and solvent. The contribution for net free energy of solvation comes from two energy terms, namely, reaction field energy (U_{rxn}) and cavity energy (U_{cav}): $U_{\text{SGB}} = U_{\text{rxn}} + U_{\text{cav}}$. The cavity and reaction field energy terms implicitly take into account the VDW and the electrostatic interactions, respectively, between the ligand and solvent. The application of the SGB-LIE method for a given protein–ligand system essentially involves computing four energy components, i.e. the VDW and Coulombic energy between the ligand and protein and the reaction field and cavity energy between the ligand and continuum solvent. The total electrostatic energy in the SGB-LIE method is the sum of Coulombic and reaction field energy terms.

2.2 *Computational details*

Preparation of the receptor and ligands was done using the Schrödinger package from Schrödinger Inc [42]. All of the calculations for the SGB-LIE method were performed in the Liaison package from Schrödinger Inc [43]. The Liaison module performs LIE calculations in the OPLS force field with a residue-based cutoff of 15 Å. The OPLS force field was also used for charge assignment and all energy calculations.

2.3 *Receptor preparation*

The X-ray structure of haeme-pdb was taken from the Protein Data Bank (PDB ID: 1CTJ) and has been used as the initial structure in the preparation of the haeme receptor site. Haeme is a planar molecule with a strong positive charge on its central iron atom, which lies slightly above the porphyrin plane (Figure 2). The charge on the iron was assigned as +2 but the structure was kept the same. Hydrogen was added to the model automatically via the Maestro interface leaving no lone pair and using an explicit all-atom model. The multi-step Schrödinger's protein preparation tool (PPrep) was used for final preparation of receptor model. The complex structure was energy minimised using the OPLS-2005 force field and the conjugate gradient algorithm, keeping all atoms except hydrogen fixed. The minimisation was stopped either after 1000 steps or after the energy gradient converged below 0.01 KJ/mol. Complete geometry optimisation was carried out using LACVP** [44] for the iron atoms, followed by single-point calculations using LACVP** for the iron atom. An unrestricted density functional theory (DFT) was employed to model effectively the open shell orbital on the two iron atoms. The Jaguar suite of *ab initio* quantum chemical program [45] was used to carry out all quantum mechanics (QM) calculations.
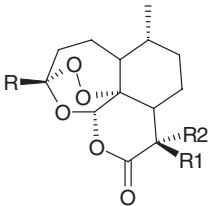
2.4 *Preparation of ligands*

An initial dataset of 158 artemisinin analogues were collected from published data [46–51] in which several different ring systems were represented. All of the analogues were either peroxides or trioxanes, which should act via similar mechanisms of action and were categorised into different classes (Table 1a–j). Each of these compounds had associated *in vitro* bioactivity values (IC_{50} values reported in ng/ml) against the drug-resistant malaria strain *P. falciparum* (W-2 clone). The log value of the relative activity (RA) of these compounds was used for analysis and was defined as

$$\log(RA) = \log[(\text{artemisinin } IC_{50}/\text{analogue } IC_{50})(\text{analogue MW}/\text{artemisin in MW})]$$

Molecular models of the artemisinin and its analogues (Table 1a–j) were built using the Builder feature in Maestro (Schrödinger package) and energy minimised in a vacuum using Impact. Each structure was assigned an appropriate bond order using the ligprep script shipped by Schrödinger and optimised initially by means of the OPLS-2005 force field using the default settings. Complete geometrical optimisation of these structures was carried out with the HF/3-21G method (in this work) using the Jaguar (Schrödinger Inc.). In order to check the reliability of the geometry obtained, we compared the structural parameters of the artemisinin 1,2,4-trioxane ring with theoretical [52] and experimental [53,54] values from the literature. All calculations reproduced most of the structural parameters of the artemisinin 1,2,4-trioxane ring seen in X-ray structures (Table 2). This applies especially to the bond length of the endoperoxide bridge, which seems to be responsible for the anti-malarial activity [15–19].
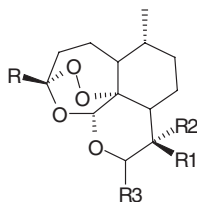
2.5 *Docking of the ligands*

All of the ligands were docked to the haeme receptor using Glide. After ensuring that protein and ligands are in the correct form for docking, the receptor-grid files were generated using a grid-receptor generation program, using VDW scaling of the

Table 1a. Artemisinin analogues with anti-malarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Compounds	R	R1	R2	Log (RA)	IC_{50} (ng/ml)
Training set					
1	CH_3	CH_3	H	1.00	0.040
2	C_4H_8Ph	H	H	0.45	0.194
3	CH_3	H	2-Z-Butenyl	−1.10	5.750
4	CH_3	H	H	0.79	0.065
5	CH_3	H	2-E-Butenyl	−0.60	1.818
6	CH_3	Allyl	H	−0.10	0.550
7	CH_3	C_4H_9	H	0.17	0.311
8	C_4H_8Ph	C_4H_9	H	−0.32	1.310
9	$CH_2CH_2CO_2Et$	C_4H_9	H	1.36	0.025
10	C_4H_9	C_4H_9	H	−0.48	1.568
11	CH_3	C_2H_5	H	1.40	0.017
12	CH_3	C_6H_{13}	H	0.86	0.069
13	CH_3	i-C_6H_{13}	H	−0.04	0.547
14	CH_3	i-C_5H_{11}	H	0.07	0.408
15	C_3H_6(p-Cl-Ph)	H	H	0.10	0.457
16	C_4H_9	H	H	−0.74	2.416
17	$CH_2CH_2CO_2Et$	H	H	0.37	0.214
18	CH_3	C_3H_6(p-Cl-Ph)	H	1.37	0.025
Test set					
19	CH_3	Br	CH_2Br	−1.64	27.244
20	CH_3	$=CH_2$	–	−0.89	3.083
21	CH_3	CH_2CH_3	–	−0.36	1.053
22	CH_3	$–CH_2CH_2–$	–	−0.94	3.632
23	CH_3	C_5H_{11}	H	1.02	0.046
24	CH_3	C_4H_8Ph	H	0.63	0.133
25	CH_3	C_2H_4Ph	H	0.12	0.400
26	CH_3	C_3H_7	H	1.13	0.033

receptor at 0.4. The default size was used for the bounding and enclosing boxes. The grid box was generated at the centroid of the haeme. The ligands were docked initially using the 'standard precision' method and further refined using the 'Extra precision' Glide algorithm. For the ligand docking stage, VDW scaling of the ligand was set at 0.5. Of the 50,000 poses that were sampled, 4000 were taken through minimisation (conjugate gradients 1000) and the 30 structures having the lowest energy conformations were further evaluated for the favourable Glide docking score. A single best conformation for each ligand was considered for further analysis.

Table 1b. 10-Substituted artemisinin derivatives with anti-malarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.
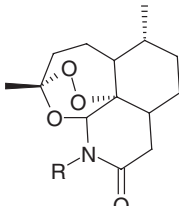

Compounds	R	R1	R2	R3	log (RA)	IC_{50} (ng/ml)
Training set
27	CH_3	CH_3	H	H	0.75	0.068
28	CH_3	CH_3	H	OH	0.55	0.114
29	CH_3	CH_3	H	OEt	0.34	0.202
30	CH_3	CH_3	H	OH	0.96	0.051
31	CH_3	H	Br	H	0.28	0.248
32	CH_3	CH_3	Br	NH-2-(1,3-thiazole)	0.66	0.134
33	CH_3	CH_3	Br	p-Cl-aniline	0.79	0.105
34	CH_3	CH_3	Br	aniline	0.18	0.397
35	CH_3	Br	CH_3	NH-2-pyridine	−0.09	0.768
36	CH_3	CH_3	Br	NH-2-pyridine	−0.77	3.667
37	CH_3	CH_3	H	α -OEt	0.32	0.212
38	CH_3	C_4H_9	H	H	1.32	0.021
39	CH_3	C_2H_5	H	H	0.67	0.086
40	CH_3	C_3H_7	H	OEt	−0.04	0.529
41	CH_3	H	H	OEt	0.43	0.157
42	CH_3	CH_3	H	C_3H_6OH	0.78	0.077
43	CH_3	CH_3	H	C_4H_9	0.06	0.400
44	CH_3	CH_3	H	OCH_2CO_2Et	0.52	0.158
45	CH_3	CH_3	H	$OC_2H_4CO_2Me$	0.10	0.433
46	CH_3	CH_3	H	$OC_3H_6CO_2Me$	−0.03	0.605
47	CH_3	CH_3	H	$OCH_2(4-PhCO_2Me)$	−0.07	0.720
48	CH_3	CH_3	H	(R)-$OCH_2CH(CH_3)CO_2Me$	1.79	0.009
49	CH_3	CH_3	H	(S)-$OCH_2CH(CH_3)CO_2Me$	2.25	0.003
50	CH_3	CH_3	H	(R)-$OCH(CH_3)CH_2CO_2Me$	0.87	0.073
51	CH_3	CH_3	H	(S)-$OCH(CH_3)CH_2CO_2Me$	1.70	0.011
52	$CH_2CH_2CO_2Et$	H	H	H	0.70	0.096
53	C_4H_9	H	H	H	0.75	0.075
Test set
54	C_4H_8Ph	H	H	H	0.58	0.139
55	CH_3	$-OCH_2-$	–	OOH	−0.62	1.857
56	CH_3	$-CH_2O-$	–	OOH	−0.57	1.655
57	CH_3	$=CH_2$	–	OOH	−0.99	4.131
58	CH_3	C_5H_{11}	H	H	0.16	0.318
59	CH_3	C_3H_6Ph	H	H	1.40	0.021
60	CH_3	C_3H_7	H	H	0.74	0.076
61	CH_3	CH_3	H	$OOt-C_4H_9$	0.92	0.061
62	–	CH_3	OH	α-OH	−0.89	3.303
63	–	CH_3	H	CH_2CHF_2	0.11	0.366
64	–	CH_3	OH	OCH_2CF_3	0.33	0.243
65	–	CH_3	OH	OEt	−0.44	1.281

Table 1c. 11-Aza-artemisinin derivatives with anti-malarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.
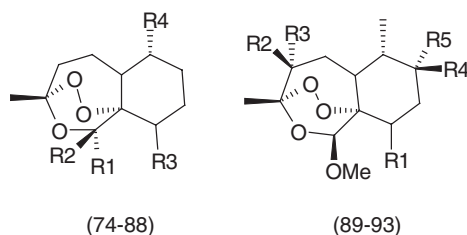

Compounds	R	log RA	IC_{50} (ng/ml)
Training set			
66	C_3H_6Ph	0.02	0.522
67	C_2H_4Ph	0.16	0.364
68	C_5H_{11}	−0.20	0.758
69	i-C_5H_{11}	−0.04	0.524
70	CH_2(p-Cl-Ph)	−0.16	0.802
Test set			
71	CH_2Ph	0.34	0.231
72	CH_2–(2-C_5H_4N)	1.46	0.018
73	Acetaldehyde	1.47	0.015

2.6 *LIE calculations*

The docked complex corresponding to each analogue was transported to the Liaison package for subsequent SGB-LIE calculations. Two sampling techniques, molecular dynamics (MD) and hybrid Monte Carlo (HMC), were used for LIE conformation space sampling in the present work. A conjugate gradient minimisation was performed first, starting from the initial docked structures and then a 15 ps MD equilibration is followed with temperature smoothly increasing from 0 to 310 K by velocity scaling and re-sampling. Finally, a 25 ps MD simulation was run for the SGB-LIE data collection. A residue-based cutoff of 12 Å was set for the non-bonding interactions. The non-bonded pair list was updated every 10 fs. The time integration step of 1.0 fs and sampling of LIE energies in every 10 steps was used. Similarly, the average LIE energies for the ligand were obtained using the OPLS-2005 force field. The average LIE energy terms were used for building the binding affinity model and free energy estimation for artemisinin analogues. The α, β and γ LIE fitting parameters were determined based on Gaussian elimination method using Matlab 6.5 as described by Thomas and Finny [55] and by fitting the experimental data on the training set compounds.

In order to explore the reliability of the proposed model the cross validation method was used. Prediction error sum of squares (PRESS) is a standard index to measure the accuracy of a modelling method based on the cross validation technique. The cross validation analysis performed by using the leave one out (LOO) method in which one compound is removed from the dataset and its activity predicted using the model derived

Table 1d. Artemisinin derivatives lacking the D-ring with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

(74–88)　　　　　(89–93)

Compounds	R1	R2	R3	R4		log (RA)	IC_{50} (ng/ml)
Training set							
74	–O_2CCH_2Ph	H	H	CH_3		−0.51	1.648
75	H	H	H	CH_3		−0.32	0.628
76	H	OCH_3	H	H		−0.31	0.660
77	OCH_2Ph	H	H	H		−0.09	0.530
78	OCH_3	H	$C_2H_4O_2CNEt$	H		−0.65	0.118
79	H	OCH_3	$C_2H_4OCH_3$	H		−0.39	0.996
80	H	OCH_3	$C_2H_4OCH_2Ph$	H		0.75	0.091
81	H	OCH_3	C_2H_4O-allyl	H		0.40	0.184
82	H	OCH_3	$C_2H_4O_2Ph$	H		−0.59	2.086
83	H	OCH_3	$C_2H_4O_2C(4\text{-}PhCO_2Me)$	H		0.27	0.343
84	H	OCH_3	$C_2H_4O_2C(4\text{-}PhCO_2H)$	H		−0.81	3.856
85	H	OCH_3	$C_2H_4O_2C(4\text{-}PhCONEt_2)$	H		0.230	0.398
86	H	OCH_3	$C_2H_4O_2C(4\text{-}PhCO_2C_2H_4NMe_2)$	H		−0.600	2.790
Test set							
87	H	OCH_3	$C_2H_4O_2CCH_2NCO_2\text{-}(t\text{-}C_4H_9)$	H		−0.04	0.670
88	H	OCH_3	$C_2H_4OCH_2(4\text{-N-Me-pyridine})$	H		−0.90	4.439
89	C_2H_4OH	H	CH_3	H	H	−1.80	26.849
90	C_2H_4OH	CH_3	H	H	H	0.23	0.251
91	C_2H_4OH	CH_3	CH_3	H	H	−1.80	28.102
92	$C_2H_4OCH_2Ph$	CH_3	CH_3	H	H	−1.80	36.157
93	$C_2H_4OCH_2(4\text{-py})$	–	–	–	–	0.14	0.373

from the rest of the data points. The cross-validated correlation coefficient (q^2) that resulted in the optimum number of components and lowest standard error of prediction were considered for further analysis and calculated using the following equations:

$$q^2 = 1 - \frac{\sum_y (y_{\text{predicted}} - y_{\text{observed}})^2}{\sum_y (y_{\text{observed}} - y_{\text{mean}})^2}$$

$$\text{PRESS} = \sum_y (y_{\text{predicted}} - y_{\text{observed}})^2$$

where $y_{\text{predicted}}$, y_{observed} and y_{mean} are the predicted actual and mean values of the inhibitory activities of the artemisinin analogues and PRESS is the sum of the predictive sum of squares.

Table 1e. Miscellaneous artemisinin derivatives with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Compounds	Structure	log (RA)	IC$_{50}$ (ng/ml)
Training set			
94		0.78	0.063
95		−4.00	6.339
96		0.23	0.259
97		−1.20	6.340
98		−3.30	684.899
Test set			
99		−0.96	3.622
100		−0.79	2.344
101		−0.64	1.573
102		−2.09	56.889

Table 1e. Continued.
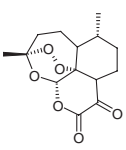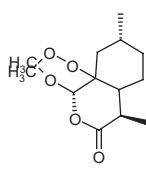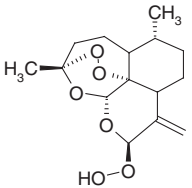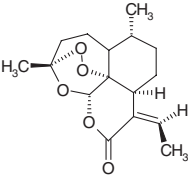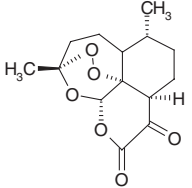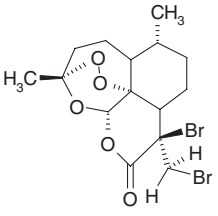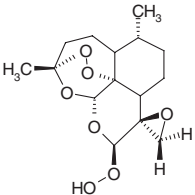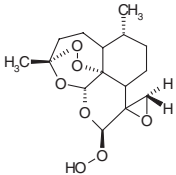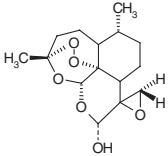
Compounds	Structure	log (RA)	IC$_{50}$ (ng/ml)
103		−2.49	123.612
104		−0.80	2.309

Table 1f. 9-Substituted artemisinin derivatives with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Compounds	Structure	log (RA)	IC$_{50}$ (ng/ml)
Training set			
105		−0.739	2.320
106		−0.197	0.657
107		−2.298	79.429
108		−1.487	19.143

(*continued*)

Table 1f. Continued.

Compounds	Structure	log (RA)	IC$_{50}$ (ng/ml)
Test set			
109		−0.460	1.286
110		−0.409	1.143
111		−0.361	0.971

3. Results and discussions

We have applied the SGB-LIE method to a training set of 101 artemisinin analogues to build a binding affinity model that was then used to compute the absolute FEB and RA for a test set of 57 analogues. The whole dataset of 158 analogues was randomly split into 64% of training set and 36% of test set. The training set used for building the binding affinity model comprised different classes of artemisinin analogues (Table 1a–j). Among these are endoperoxide artemisinin analogues, deoxy-artemisinin analogues, 10-substituted artemisinin derivatives, 3-substituted artemisinin derivatives, 9-substituted artemisinin derivatives, 11-Aza-artemisinin derivatives, artemisinin derivatives without D-ring and miscellaneous artemisinin derivatives. The experimental RA values for all of the compounds in the training set were calculated against the drug-resistant malarial strain *P. falciparum* (W-2 clone). The IC$_{50}$ value of these analogues was used for calculation of the absolute $\Delta G_{\text{binding}}$ energy (Table 3). With the wide range of difference in IC$_{50}$ values and the large diversity in the structures, the combined set of 101 ligands is ideal to be considered as a training set as the set does not suffer from bias due to the similarity of the structures. Also, the training set of 101 analogues has enough data points not to suffer from over-parameterisation by the SGB-LIE model. Training set compounds were docked into the haeme and the SGB-LIE calculations were performed using the Liaison module. The simulations were performed both for the ligand-free and ligand-bound state. The various interaction energy terms described in the methods were collected and are presented

Table 1g. Dihydroartemisinin derivatives with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.
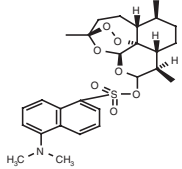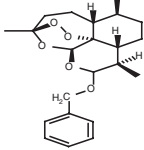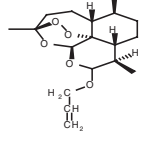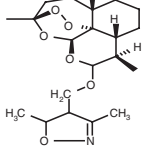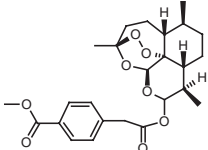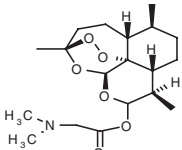
Compounds	R	log (RA)	IC_{50} (ng/ml)
Training set			
112	OR=H	0.487	0.123
113	(S)-$CH_2CH(CH_3)COOCH_3$	2.104	0.004
114	(S)-$CH(CH_3)CH_2COOCH_3$	0.599	0.137
115	(R)-$CH(CH_3)CH_2COOCH_3$	1.429	0.020
116	1-adamantylmethyl	0.007	0.603
117	(S)-$CH_2CH(CH_3)COOH$	−0.658	2.380
118	(S)-$CH(CH_3)CH_2COOH$	−0.608	2.123
119	(R)-$CH(CH_3)CH_2COOH$	−0.383	1.263
Test set			
120	OR==O	−0.269	0.743
121	$CH_2PhCOOH$	0.176	0.394
122	(R)-$CH_2CH(CH_3)COOCH_3$	1.524	0.016
123	(R)-$CH_2CH(CH_3)COOH$	−0.463	1.520

Table 1h. Tricyclic 1.2.4–trioxanes derivatives with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Compounds	Structure	log(RA)	IC_{50} (ng/ml)
Training set			
124		−0.475	1.886
125		0.995	0.057

(*continued*)

Table 1h. Continued.

Compounds	Structure	$\log(RA)$	IC_{50} (ng/ml)
126		−0.413	1.771
127		0.632	0.171
128		0.968	0.057
129		0.905	0.057
130		0.991	0.057
Test set 131		0.660	0.143
132		0.787	0.086

(*continued*)

Table 1h. Continued.

Compounds	Structure	$log(RA)$	IC_{50} (ng/ml)
133		0.717	0.057
134		0.434	0.229
135		0.129	0.314

Table 1i. N-Alkyl-11-aza-9-desmethylartemisinins derivatives with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.
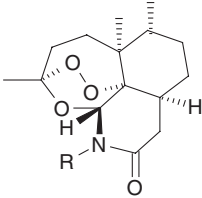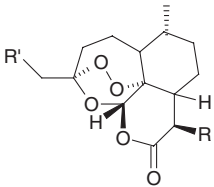
Compounds	R	$log(RA)$	IC_{50} (ng/ml)
Training set			
136	–	0.328	0.400
137	$C_5H_{11}(n)$	0.041	0.435
138	$C_5H_{11}(i)$	0.173	0.321
139	$(CH_3)_2NCH_2CH_2$	−0.432	1.300
140	$HO_2C(CH_2)_5$	−0.921	4.492
Test set			
141	$C_6H_5CH_2$	0.276	0.268
142	$p\text{-}ClC_6H_4CH_2$	0.045	0.500
143	$C_6H_5(CH_2)_2$	0.294	0.267
144	$C_6H_5(CH_2)_3$	0.312	0.266

Table 1j. 3C-substituted artemisinin derivatives with anti-malarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in this work.
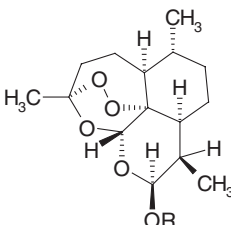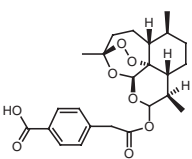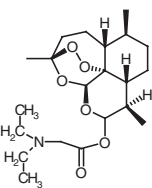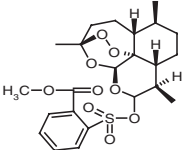

Compounds	R1	R	log (RA)	IC$_{50}$ (ng/ml)
Training set				
145	CH_3	H	0.049	0.357
146	CH_3CH_2	H	0.828	0.062
147	$CH_3(CH_2)_2$	H	−0.745	2.427
148	CH_3CH	H	−0.347	0.977
149	EtO_2CCH_2	H	0.365	0.216
150	$C_6H_5CH_2$	H	−2.000	50.780
151	p-$ClC_6H_4(CH_2)_2$	H	0.104	0.453
152	$C_6H_5(CH_2)_3$	H	0.449	0.195
153	CH_3	$CH_3(CH_2)_3$	0.410	0.187
Test set				
154	$CH_3(CH_2)_2$	$CH_3(CH_2)_3$	−0.481	1.573
155	$C_6H_5CH_2$	$CH_3(CH_2)_3$	−2.000	58.723
156	p-$ClC_6H_4(CH_2)_2$	$CH_3(CH_2)_3$	−0.276	1.239
157	$C_6H_5(CH_2)_3$	$CH_3(CH_2)_3$	−0.319	1.306
158	EtO_2CCH_2	$CH_3(CH_2)_3$	1.359	0.025

in Table 3. The largest contribution for the binding energy comes from the VDW interactions. This is because the artemisinin analogues used in the study are mostly lipophilic molecules that interact favourably with a hydrophobic binding site.

The energy values in Table 3 were used to fit Equation (1) to obtain distinct and unique solution for calculating the values of three unknown parameters, α, β and γ (see [55]). The values obtained for the three fitting parameters,

$$\Delta G_{bind} = \alpha \langle \Delta U_{ele} \rangle + \beta \langle \Delta U_{vdw} \rangle + \gamma \langle \Delta U_{cav} \rangle \tag{3}$$

α, β and γ are −0.0271, −0.0902 and −1.44 respectively. The large value of the cavity energy term signifies the fact that binding is largely driven by the ligand's ability to bury itself in the binding cavity, which is understandable given that most of the ligands are highly hydrophobic in nature. Even though the R value is low, VDW interactions contribute significantly to the FEB owing to the large magnitude of the VDW interaction term. In Table 3 the experimental free energy values obtained from the RTIC$_{50}$ and the predicted free energy values estimated using fitting parameters have been presented.

Figure 3 graphically shows the quality of fit between the SGB-LIE binding energy predictions and the experimental values. If a predicted binding energy agrees exactly with the experimental value, a data point (represented by diamonds) exactly in the diagonal line would be shown. To help visualise these data points, a lower and upper bound line are also

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

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

[a]Atoms are numbered according to Figure 1.
[b]From this work.
[c]Values from [52].
[d]Values from [53] (experimental estimated standard deviations in brackets).
[e]Values from [54] (experimental estimated standard deviations in brackets).

plotted in the figure, with 1.0 kcal/mol below or above the experimental values. From the figure most of the data points (100 out of 101) are within or very close to these two bound lines, which means most of them have either less than or about 1.0 kcal/mol error. The only data point that shows large deviation from the experimental value is ligand 98 which has 2.187 kcal/mol errors. The overall root mean square error (RMSE) between the experimental values and the values obtained by the fit was 0.328 kcal/mol, which is an indicator to the robustness of the fit. The correlation coefficient r^2 is 0.845 indicating a good correlation with experiment. The statistical significance of the SGB-LIE model is evaluated by the correlation coefficient (r^2), standard error (s), F-test value, significance level of the model (P), LOO cross-validation coefficient (q^2) and predictive error sum of squares (PRESS):

$$\Delta G = (-0.0271)\langle\Delta U_{ele}\rangle + (-0.0902)\langle\Delta U_{vdw}\rangle + (-1.44)\langle\Delta U_{cav}\rangle$$
$$n = 102, \ r^2 = 0.845, \ s = 0.465, \ F = 234.1, \ P = 0.0001, \ q^2 = 0.844, \ \text{PRESS} = 21.38$$

(4)

The SGB-LIE model developed in this study (Equation (4)) is statistically ($q^2 = 0.844$, $r^2 = 0.845$, $F = 234.1$) best fitted and consequently used for the prediction of anti-malarial activities (pIC_{50}) of the training and test sets of molecules as reported in Tables 3 and 4. The predicted activity calculated from the FEB is satisfactory with small deviation

Table 3. Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the training set inhibitors using the SGB-LIE method.

Ligand	$\langle U_{ele} \rangle^a$ kcal/mol	$\langle U_{vdw} \rangle^a$ kcal/mol	$\langle U_{cav} \rangle^a$ kcal/mol	$\Delta G_{bind,expt}^b$ kcal/mol	$\Delta G_{bind,LIE}^c$ kcal/mol	RA_{expt}^d	RA_{pred}^d
1	−2.516	−17.145	2.118	−1.906	−1.435	10.000	4.514
2	3.418	−16.108	1.619	−0.970	−0.971	2.818	2.821
3	−3.078	−18.372	1.323	1.036	−0.165	0.079	0.603
4	−3.853	−14.546	2.051	−1.620	−1.537	6.166	5.360
5	−1.981	−18.983	1.162	0.354	0.092	0.251	0.391
6	−1.958	−19.826	1.308	−0.354	−0.043	0.794	0.470
7	−2.338	−18.840	1.213	−0.692	0.016	1.479	0.447
8	−1.081	−22.970	1.939	0.160	−0.691	0.479	2.012
9	−2.327	−14.082	2.291	−2.175	−1.966	22.909	16.074
10	−5.208	−20.827	1.119	0.266	0.408	0.331	0.261
11	−0.814	−9.738	2.163	−2.423	−2.214	25.119	17.656
12	−2.665	−18.442	2.426	−1.584	−1.758	7.244	9.713
13	−0.697	−17.834	1.180	−0.357	−0.072	0.912	0.564
14	−4.569	−12.307	0.958	−0.531	−0.146	1.175	0.613
15	−1.426	−22.362	1.882	−0.463	−0.654	1.259	1.737
16	−2.462	−17.382	0.729	0.522	0.585	0.182	0.164
17	−2.614	−19.968	2.039	−0.912	−1.064	2.344	3.028
18	−2.953	−14.843	2.393	−2.175	−2.027	23.442	18.261
27	−3.078	−18.792	2.191	−1.595	−1.377	5.623	3.885
28	−3.834	−19.631	2.379	−1.288	−1.551	3.548	5.529
29	−2.707	−18.188	1.967	−0.946	−1.119	2.188	2.926
30	−2.849	−14.030	1.794	−1.766	−1.241	9.120	3.757
31	−3.244	−14.159	1.724	−0.826	−1.118	1.905	3.117
32	−3.131	−21.131	2.154	−1.191	−1.111	4.571	3.990
33	−1.162	−18.594	2.534	−1.333	−1.940	6.166	17.204
34	−2.097	−15.063	1.988	−0.546	−1.447	1.514	6.920
35	−3.044	−15.843	1.056	−0.156	−0.009	0.813	0.634
36	−2.157	−19.793	0.994	0.769	0.412	0.170	0.310
37	−1.920	−17.825	1.928	−0.919	−1.116	2.089	2.915
38	−0.706	−13.997	2.419	−2.286	−2.202	20.893	18.102
39	−3.972	−18.487	2.191	−1.456	−1.380	4.677	4.111
40	−3.144	−16.665	1.376	−0.377	−0.393	0.912	0.937
41	−1.522	−16.466	1.654	−1.096	−0.855	2.692	1.791
42	−1.155	−15.515	2.148	−1.520	−1.662	6.026	7.657
43	−2.053	−18.557	0.903	−0.542	0.430	1.148	0.222
44	−2.414	−17.032	2.148	−1.091	−1.491	3.311	6.512
45	−1.365	−19.248	1.247	−0.496	−0.023	1.259	0.566
46	−3.597	−18.110	0.966	−0.298	0.340	0.933	0.318
47	−1.575	−23.515	1.558	−0.195	−0.080	0.851	0.702
48	−1.994	−11.342	2.789	−2.822	−2.939	61.660	75.041
49	−1.754	−9.730	3.019	−3.449	−3.422	177.828	169.651
50	−2.485	−18.383	2.019	−1.546	−1.182	7.413	4.007
51	−2.661	−14.790	2.954	−2.678	−2.848	50.119	66.731
52	−2.601	−16.296	2.197	−1.386	−1.623	5.012	7.476
53	−2.497	−17.867	2.303	−1.536	−1.637	5.623	6.662
66	−0.334	−21.951	1.454	−0.385	−0.105	1.047	0.652
67	−6.696	−19.333	1.102	−0.598	0.338	1.445	0.297
68	−2.908	−18.273	1.455	−0.164	−0.368	0.631	0.890
69	−1.570	−16.242	1.024	−0.383	0.033	0.912	0.452
70	−1.949	−13.471	1.085	−0.131	−0.294	0.692	0.912
74	−0.811	−14.935	0.948	0.296	0.004	0.309	0.506
75	−2.003	−15.996	0.746	−0.275	0.423	0.479	0.147
76	−3.262	−13.216	0.874	−0.246	0.021	0.490	0.312
77	−0.448	−18.855	1.058	−0.375	0.190	0.813	0.313

Table 3. Continued.

Ligand	$\langle U_{ele}\rangle^{a}$ kcal/mol	$\langle U_{vdw}\rangle^{a}$ kcal/mol	$\langle U_{cav}\rangle^{a}$ kcal/mol	$\Delta G_{bind,expt}^{b}$ kcal/mol	$\Delta G_{bind,LIE}^{c}$ kcal/mol	RA_{expt}^{d}	RA_{pred}^{d}
78	−3.355	−15.598	1.739	−1.266	−1.006	4.467	2.878
79	−1.541	−13.995	0.408	−0.002	0.716	0.407	0.121
80	−0.327	−15.482	1.739	−1.417	−1.099	5.623	3.283
81	−2.100	−16.519	1.843	−1.002	−1.107	2.512	2.998
82	−3.419	−15.003	0.852	0.435	0.219	0.257	0.370
83	−2.704	−21.131	1.634	−0.634	−0.373	1.862	1.199
84	−1.926	−12.919	0.496	0.799	0.503	0.155	0.255
85	−2.775	−11.353	0.835	−0.545	−0.103	1.698	0.805
86	−2.526	−11.029	0.090	0.607	0.934	0.251	0.145
94	−3.020	−16.659	1.824	−1.641	−1.042	6.026	2.192
95	−0.808	−13.499	0.387	1.093	0.682	0.054	0.108
96	−2.531	−15.943	1.721	−0.800	−0.972	1.698	2.268
97	−1.705	−16.422	0.406	1.094	0.943	0.063	0.081
98	−3.483	−21.475	0.244	3.866	1.680	0.001	0.020
105	−8.768	−13.714	0.696	0.498	0.472	0.182	0.190
106	−5.534	−16.037	1.213	−0.249	−0.149	0.635	0.537
107	−6.589	−17.617	0.229	2.590	1.438	0.005	0.035
108	−7.804	−15.997	0.352	1.748	1.148	0.033	0.090
112	−3.988	−13.850	1.473	−1.242	−0.764	3.071	1.370
113	−1.586	−11.374	3.017	−3.229	−3.276	127.092	137.458
114	−1.020	−20.888	2.248	−1.176	−1.325	3.972	5.106
115	−2.372	−21.331	2.839	−2.308	−2.100	26.850	18.879
116	−3.801	−21.520	1.347	−0.300	0.104	1.016	0.514
117	−2.574	−15.816	0.757	0.513	0.407	0.220	0.263
118	−2.678	−16.410	0.819	0.446	0.373	0.247	0.279
119	−2.424	−15.593	1.151	0.138	−0.185	0.414	0.716
124	−2.894	−20.405	0.787	0.376	0.786	0.335	0.168
125	−2.022	−17.400	2.432	−1.695	−1.878	9.879	13.450
126	−1.210	−21.164	0.849	0.339	0.719	0.386	0.203
127	−1.761	−16.138	1.842	−1.044	−1.149	4.286	5.114
128	−2.492	−18.765	2.484	−1.695	−1.817	9.284	11.402
129	−1.684	−17.108	2.319	−1.695	−1.751	8.043	8.833
130	−3.207	−14.210	2.117	−1.695	−1.680	9.805	9.556
136	−0.640	−19.898	1.455	−0.543	−0.283	1.000	0.645
137	−3.500	−18.708	1.107	−0.493	0.188	1.100	0.348
138	−3.401	−13.899	1.004	−0.673	−0.100	1.490	0.566
139	−2.207	−16.793	0.772	0.155	0.463	0.370	0.220
140	−1.375	−19.543	0.788	0.890	0.665	0.120	0.175
145	−2.199	−10.135	0.534	−0.610	0.205	1.120	0.283
146	−3.199	−16.606	2.338	−1.643	−1.782	6.730	8.511
147	−2.036	−12.098	0.337	0.525	0.661	0.180	0.143
148	−2.081	−17.176	1.154	−0.014	−0.056	0.450	0.483
149	−2.937	−20.664	2.147	−0.906	−1.148	2.320	3.490
150	−1.725	−19.374	0.286	2.326	1.382	0.010	0.049
151	−1.522	−20.360	1.647	−0.468	−0.494	1.270	1.326
152	−1.084	−21.884	2.107	−0.968	−1.031	2.810	3.121
153	−1.788	−21.281	2.014	−0.994	−0.932	2.570	2.314

The data are collected from a 15 ps MD simulation after a 15 ps MD equilibration.

[a] U_{ele}, U_{vdw} and U_{cav} energy terms represent the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment.

[b] $\Delta G_{bind,expt}$ refers to free energy of binding with haeme and is computed using the relationship $\Delta G_{binding} \approx RT \ln(IC_{50,expt})$, where 298 K is used in the work for temperature T.

[c] $\Delta G_{bind,LIE}$ refer to the absolute free energy values obtained using SGB-LIE method.

[d] RA_{expt} and RA_{pred} refers to the experimental and predicted relative activity and is calculated as $RA = IC_{50}$ of artemisinin/IC_{50} of the analogue) × (MW of the analogue/MW of the artemisinin).
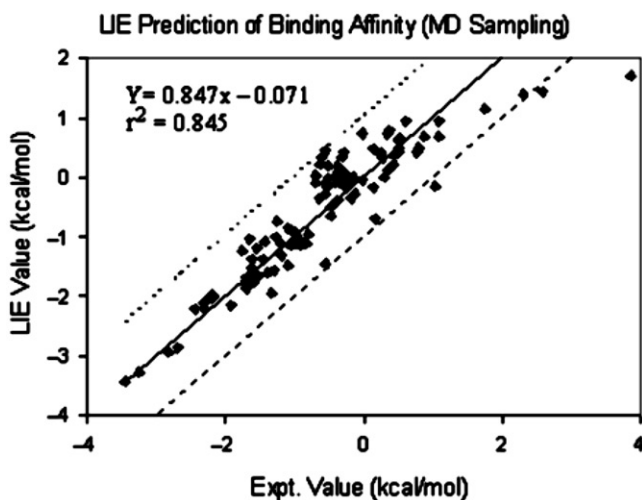
Figure 3. LIE binding energies for the training set from MD sampling. The overall RMS error is 0.328 kcal/mol for 101 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

compared with the experimental activity of the training and test sets of molecules. The calculated FEB represents the experimental activity well.

Satisfied with the robustness of the binding affinity model developed using the training set, we applied the LIE model to the artemisinin analogues comprising the test set. The test set includes 56 compounds categorised into different subgroups as mentioned in Table 1. The analogues comprising the test set were obtained from different sources [46–51]. As the experimental values of IC_{50} for these inhibitors are already available, this set of molecules provides an excellent dataset for testing the prediction power of SGB-LIE method for new ligands. Table 4 presents the free energy values estimated for the 56 test compounds. The free energy values were estimated based on optimised SGB-LIE parameters α, β and γ from the training set. The quality of fit between the SGB-LIE binding energy predictions and the experimental values is shown in Figure 4. We can see from the figure that most of the data points (55 out of 56) are within or very close to these two bound lines, which means most of them have either less than or about 1.0 kcal/mol error. The only data point that shows large deviation from the experimental value is ligand 102, which has 2.025 kcal/mol errors. The reason for the large error in 102 is not very clear: it could be due to the force field parameters used since this is the only case whose R group has been substituted with oxygen atoms. The root mean square (RMS) error between the experimental and predicted free energy values was 0.348 kcal/mol, which is comparable to the level of accuracy achieved by the most accurate method such as free energy perturbation. The squared correlation coefficient between experimental and SGB-LIE estimates for the free energy of the test set compounds is also significant ($r^2 = 0.868$). The predicted relative anti-malarial activity of artemisinin derivatives estimated using LIE free energy is also very close to experimental RA for the test set (Table 4).

To test how sensitive the LIE method is to the underlying sampling techniques, in other words, how good the sampling technique is in surfing the local conformation space, we also implemented LIE with the HMC sampling [56,57]. A Metropolis accept/reject

Table 4. Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the test set inhibitors using SGB-LIE method.

Ligand	$\langle U_{ele} \rangle^a$ kcal/mol	$\langle U_{vdw} \rangle^a$ kcal/mol	$\langle U_{cav} \rangle^a$ kcal/mol	$\Delta G_{bind,expt}^b$ kcal/mol	$\Delta G_{bind,LIE}^c$ kcal/mol	RA_{expt}^d	RA_{pred}^d
19	−3.281	−21.784	0.487	1.957	1.353	0.023	0.064
20	−2.613	−18.200	0.694	0.667	0.713	0.129	0.119
21	−2.740	−16.044	1.053	0.030	0.006	0.437	0.455
22	−3.833	−14.495	0.429	0.764	0.794	0.115	0.109
23	−2.919	−18.198	2.060	−1.826	−1.246	10.471	3.929
24	−3.139	−19.419	1.938	−1.195	−0.954	4.266	2.841
25	−0.555	−21.367	1.692	−0.542	−0.494	1.318	1.215
26	−2.136	−18.608	2.248	−2.027	−1.501	13.490	5.543
54	−3.294	−19.805	2.180	−1.168	−1.264	3.796	4.456
55	−1.837	−16.930	0.576	0.366	0.748	0.240	0.126
56	−1.815	−11.207	0.534	0.298	0.291	0.269	0.272
57	−2.520	−17.954	0.693	0.840	0.690	0.102	0.132
58	−1.891	−16.900	1.168	−0.678	−0.106	1.445	0.549
59	−2.806	−13.592	2.332	−2.287	−2.056	25.119	16.986
60	−1.133	−20.846	2.348	−1.523	−1.470	5.495	5.025
61	−2.241	−18.266	2.207	−1.659	−1.470	8.318	6.040
62	−2.837	−14.729	0.481	0.708	0.713	0.129	0.128
63	−3.090	−16.208	0.905	−0.596	0.243	1.288	0.313
64	−1.302	−17.980	1.877	−0.838	−1.046	2.138	3.035
65	−3.185	−16.301	0.936	0.147	0.209	0.363	0.327
71	−1.949	−13.565	1.463	−0.868	−0.830	2.192	2.055
72	−2.525	−9.035	1.969	−2.364	−1.952	28.840	14.354
73	−2.217	−14.830	2.495	−2.493	−2.195	29.512	17.842
87	−3.360	−21.074	1.579	−0.237	−0.282	0.912	0.983
88	−3.718	−13.614	0.490	0.883	0.623	0.126	0.195
89	−2.617	−21.764	0.342	1.948	1.542	0.016	0.032
90	−3.202	−14.273	1.356	−0.820	−0.578	1.698	1.130
91	−3.343	−20.539	0.350	1.975	1.439	0.016	0.039
92	−3.115	−20.933	0.287	2.124	1.559	0.016	0.041
93	−3.708	−18.669	1.060	−0.584	0.258	1.380	0.333
99	−2.202	−20.718	0.973	0.762	0.527	0.110	0.163
100	−1.739	−17.415	0.473	0.504	0.937	0.162	0.078
101	−2.296	−15.944	0.688	0.268	0.510	0.229	0.152
102	−3.180	−18.111	0.939	2.393	0.368	0.008	0.248
103	−3.304	−21.937	0.263	2.852	1.690	0.003	0.023
104	−2.504	−12.637	0.533	0.496	0.441	0.158	0.174
109	−6.627	−15.768	0.884	0.149	0.330	0.346	0.255
110	−7.339	−10.394	0.453	0.079	0.484	0.390	0.197
111	−6.277	−15.199	0.811	−0.017	0.374	0.435	0.225
120	−1.345	−12.865	0.572	−0.176	0.373	0.538	0.213
121	−1.874	−22.404	2.144	−0.551	−1.016	1.500	3.289
122	−2.973	−20.845	2.894	−2.438	−2.207	33.445	22.608
123	−2.698	−14.775	0.847	0.248	0.186	0.344	0.382
131	−3.370	−20.572	2.342	−1.152	−1.426	4.567	7.244
132	−3.772	−18.720	2.473	−1.455	−1.770	6.123	10.429
133	−3.675	−23.555	2.664	−1.695	−1.612	10.823	9.404
134	−2.567	−21.808	2.349	−0.874	−1.346	2.718	6.029
135	−2.910	−13.930	1.174	−0.685	−0.355	1.345	0.770
141	−3.131	−18.025	1.822	−0.780	−0.913	1.890	2.366

(*continued*)

Table 4. Continued.

Ligand	$\langle U_{ele} \rangle^a$ kcal/mol	$\langle U_{vdw} \rangle^a$ kcal/mol	$\langle U_{cav} \rangle^a$ kcal/mol	$\Delta G_{bind,expt}^b$ kcal/mol	$\Delta G_{bind,LIE}^c$ kcal/mol	RA_{expt}^d	RA_{pred}^d
142	−2.166	−19.803	1.773	−0.411	−0.708	1.110	1.833
143	−1.677	−22.034	2.181	−0.782	−1.108	1.970	3.416
144	−1.197	−18.325	1.843	−0.783	−0.969	2.050	2.802
154	−1.597	−21.077	1.347	0.268	0.005	0.330	0.515
155	−2.852	−21.381	0.704	2.412	1.626	0.010	0.038
156	−1.402	−19.480	1.361	0.127	−0.165	0.530	0.867
157	−1.654	−22.698	1.482	0.158	−0.043	0.480	0.674
158	−2.392	−18.213	2.581	−2.174	−2.009	22.850	17.291

The data are collected from a 15 ps MD simulation after a 15 ps MD equilibration.
[a]U_{ele}, U_{vdw} and U_{cav} energy terms represents the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment.
[b]$\Delta G_{bind,expt}$ refers to free energy of binding with haeme and is computed using the relationship $\Delta G_{binding} \approx RT \ln(IC_{50,expt})$, where 298 K is used in the work for temperature T.
[c]$\Delta G_{bind,LIE}$ refer to the absolute free energy values obtained using SGB-LIE method.
[d]RA_{expt} and RA_{pred} refers to the experimental and predicted relative activity and is calculated as $RA = IC_{50}$ of artemisinin/IC_{50} of the analogue) × (MW of the analogue/MW of the artemisinin).
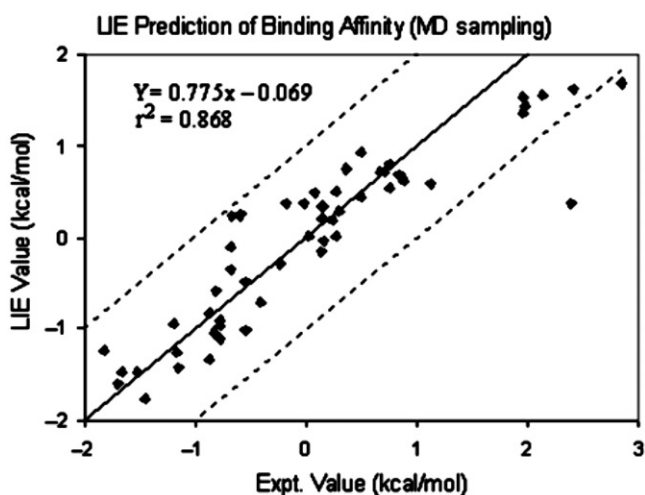

Figure 4. LIE binding energies for the test set from MD sampling. The overall RMS error is 0.348 kcal/mol for 57 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

criterion is checked every five steps of the HMC's underlying MD simulation. The time step used in the HMC's underlying MD is 3.0 fs with the RESPA algorithm [58,59]. Using the same three-parameter model, the LIE predictions are shown in Figure 5. Again, as we can see from the figure, most of the data points (99 out of 101) are within or very close to these two bound lines, which mean most of them have either less than or about 1.0 kcal/mol error. The data point 98 again has a large error, 2.992 kcal/mol but the data points 105 and 113 also show some significant error of 1.603 and 1.315 kcal/mol, respectively.
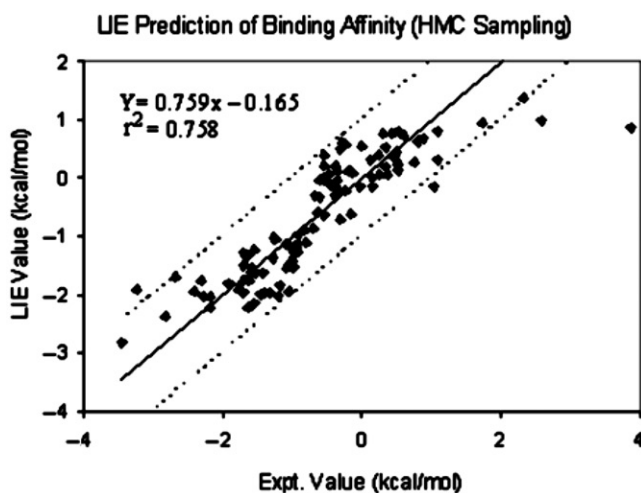
Figure 5. LIE binding energies for the training set from HMC sampling. The overall RMS error is 0.415 kcal/mol for 101 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

The reason for the large errors in these data points is not very clear. The overall RMS error is 0.415 kcal/mol and the correlation coefficient is 0.758, which are comparable to those from MD sampling. The new parameters are found to be $\alpha = -0.0078$, $\beta = -0.0735$ and $\gamma = -1.20$:

$$\Delta G = (-0.0078)\langle \Delta U_{ele}\rangle + (-0.0735)\langle \Delta U_{vdw}\rangle + (-1.20)\langle \Delta U_{cav}\rangle$$
$$n = 102,\ r^2 = 0.758,\ s = 0.588,\ F = 133.6,\ P = 0.0001,\ q^2 = 0.757,\ \text{PRESS} = 36.51$$

(5)

The SGB-LIE model developed in this study (Equation (5)) is statistically ($q^2 = 0.757$, $r^2 = 0.758$, $F = 133.6$) best fitted and consequently used for prediction of anti-malarial activity of the training and test sets of molecules. For the test set of 56 compounds the SGB-LIE model was able to predict their activity with an overall RMS error of 0.371 kcal/ mol. Figure 6 graphically shows the quality of fit between the SGB-LIE binding energy predictions and the experimental values of the test set. We can see from the figure that most of the data points (55 out of 56) are within or very close to the two bound lines. The only data point that shows large deviation from the experimental value is ligand 103, which has 1.677 kcal/mol errors. The squared correlation coefficient between the experimental and SGB-LIE estimates for the free energy of the test set compounds is also significant ($r^2 = 0.891$).

Except for the analogue 98 mentioned above, the LIE calculations agree with experiments quite well. A close look at the components in the LIE binding energy for each ligand reveals some important points. For example, the experiments showed that analogues of artemisinin substituted at C-3 were found to be less active than those at C-9. For increasing alkyl bulk at C-3 a drop in anti-malarial efficacy was noted (145, RA = 1.12; 147, RA = 0.2). Upon butyl substitution at C-9, the corresponding dual substituted analogues (3-alkyl, 9-butyl) showed a doubling of activity (153, RA = 2.6; 154, RA = 0.33). For the C-3 arylalkyl-substituted analogues alone, an increase in
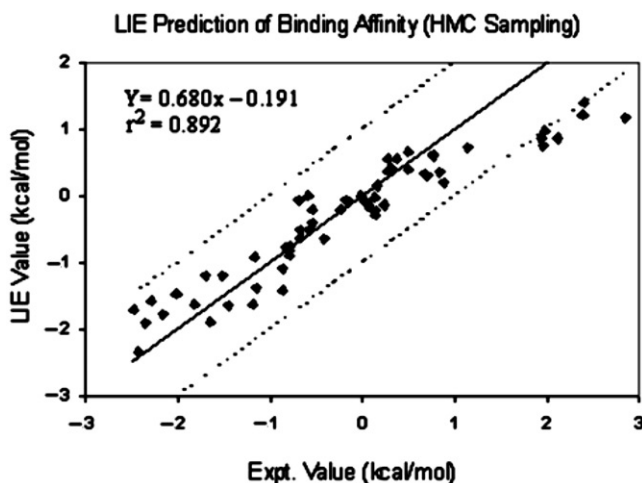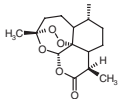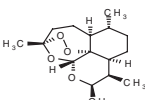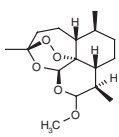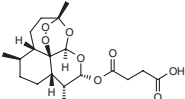
Figure 6. LIE binding energies for the test set from HMC sampling. The overall RMS error is 0.371 kcal/mol for 57 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

activity was observed with increasing chain length between the ring system and the aryl ring (two carbons for 150, RA = 0.01; three carbons for 151, RA = 1.3; four carbons for 152, RA = 2.8). Dual substituted arylalkyl analogues (3-arylalkyl, 9-butyl) were generally less active than 3-substituted arylalkyl analogues alone (e.g. 151, RA = 1.3; 156, RA = 0.53). This has also been confirmed in our SGB-LIE predictions. The reason behind this is the fact that the binding site or acceptor for artemisinin and analogues exists with limited dimensions at both C-9 and C-3, more tolerant of aryl and ester substitution on n-alkyl chains than of branches alkanes of any length. This is evident from the loss of cavity energy due to burial of solvent accessible surface area. Further, dual substitution at C-3 and C-9, explored for only 9-butyl analogues, was on the whole detrimental to activity with the exception to 158. The high potency of the dual substituted analogue 158 is due to the formation of hydrogen bond between haeme and 158 as suggested by Avery et al. [50]. In another class of analogues (tricyclic 1,2,4-trioxanes) it has been seen that a benzyl ether substituent is more potent as anti-malarial activity than a methyl ether substituent (135, RA = 1.34; 128, RA = 9.28). Our SGB-LIE model also revealed a similar conclusion. Apparently, one important factor contributing to very high anti-malarial potency is the presence of a lipophilic and bulky substituent capable perhaps of sterically 'protecting' the trioxane moiety from biological reducing agents, thereby making the trioxane a more selective oxidising agent [60]. The superior anti-malarial activity of the substituted esters over the substituted acid groups once again suggested that lipophilicity may play an important role in determining the anti-malarial activity. For example, the ester derivatives (113–115 and 122) possess superior *in vitro* activity to artemisinin in comparison to their corresponding acids (117–119 and 123). The probable reason behind this may be the fact that the ester derivatives being more lipophilic have strong VDW interaction (SGB-LIE model) than acid derivatives with haeme receptor. Also it is evident from the result that increased polarity and increased water solubility is associated with decreased

Table 5. Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the validation set inhibitors using the SGB-LIE method.

Name	Structure	$\langle U_{ele} \rangle$	$\langle U_{vdw} \rangle$	$\langle U_{cav} \rangle$	Expt IC$_{50}$ ng/ml	$\Delta G_{bind,expt}$ kcal/mol	$\Delta G_{bind,LIE}$ kcal/mol
Artemisinin		−2.516	−17.145	2.118	0.099	−1.371	−1.435
Dihydroartemisinin		−3.988	−13.85	1.473	0.282	−0.750	−0.764
Artemether		−2.910	−13.930	2.474	0.023	−2.239	−2.227
Artesunate		−2.147	−15.930	1.251	0.557	−0.347	−0.306

anti-malarial activity (e.g. analogues 105–111) [61] and resulted in low VDW interaction (SGB-LIE model).

To evaluate the accuracy of the SGB-LIE estimation for haeme polymerisation inhibition potencies, we have taken a separate dataset called the validation set consisting of four analogues of artemisinin (Table 5). Their experimental activity and chemical structures were obtained from the literature [62,63]. The experimental activity (IC$_{50}$ value) of these compounds has been obtained from an *in vitro* study of haeme polymerisation inhibition. For all of the compounds, SGB-LIE predictions produce exactly the same trend for haeme polymerisation inhibition even though the exact magnitudes of these values do not match very well to experimental values (Table 5). The difference in the exact magnitudes of estimated versus experimental FEB for compounds in the training set and a few ligands in the test set may be due to the limitations imposed by inadequate sampling and force field parameterisation. In addition, the calculation of absolute binding free energy from experimental IC$_{50}$ values for anti-malarial activity obtained from the *in vitro* cell line is only an approximation. Practically the IC$_{50}$ value of a drug molecule is dependent upon a number of factors including solubility, membrane permeability, p-glycoprotein activity against the compound, etc. However, the SGB-LIE model developed is able to predict the binding energy of the validation set quite accurately in comparison to the binding kinetics *in vitro*.

Overall, we found that the binding affinities for this binding set of artemisinin derivatives are largely coming from the VDW interaction between ligands and haeme receptor (i.e. needs a good geometric fit) and the net loss of the cavity energy, which is the

same as the burial of solvent accessible surface area. As the parameters α, β and γ are crucial to the LIE method, a natural question arises: how close are these parameters from one fit to another fit? If the LIE model really has the ability to predict binding affinities, one might expect that the parameters should be comparable to different fittings for the same binding set. Of course, we should not expect them to be identical due to the 'best possible fit' fitting procedure. As we have already seen from above experiment, the parameters are indeed comparable to the LIE fitting using either MD sampling or HMC sampling.

4. Conclusion

We have demonstrated that the SGB-LIE method can be applied to estimate the FEB with a high level of accuracy for a range of compounds with varying inhibition potencies. Despite the limitation imposed by the insufficient sampling inherent in the MD and HMC protocols, the methods have reproduced experimental data with reasonably small error for the majority of artemisinin analogues. A detailed study on the SARs for artemisinin analogues can throw light on the moieties and functional groups important in determining the inhibition potency. The close estimation of inhibition potencies of a wide range of compounds has established the LIE methodology as an efficient tool for screening novel compounds with very different structures. Compared with the empirical methods, such as scoring function approaches, the LIE method is more accurate owing to the semi-empirical approach adopted in which experimental data are used to build the binding affinity model. The SGB-LIE method seems promising when compared with the free energy perturbation or thermodynamic integration methods in achieving comparable accuracy at much faster speed even for structurally very different ligands.

References

[1] World Health Organization. *The World Health Report*, WHO, Geneva, 1999.
[2] A.V. Pandey, B.L. Tekwani, R.L. Singh, and V.S. Chauhan, *Artemisinin, an endoperoxide antimalarial disrupts the hemoglobin catabolism and hemin detoxification systems in malarial parasite*, J. Biol. Chem. 274 (1999), pp. 19383–19388.
[3] S. Kamchonwongpaisan, E. Samoff, and S.R. Meshnick, *Identification of hemoglobin degradation products in* Plasmodium falciparum, Mol. Biol. Parasitol. 86 (1997), pp. 179–186.
[4] P. Olliaro, *Mode of action and mechanisms of resistance for antimalarial drugs*, Pharmacol. Ther. 89 (2001), pp. 207–219.
[5] F.P. Mockenhaupt, *Mefloquine resistance in* Plasmodium falciparum, Parasitol. Today 11 (1995), pp. 248–253.
[6] R.G. Ridley, *Medical need, scientific opportunity and the drive for antimalarial drugs*, Nature 415 (2002), pp. 686–693.
[7] N.J. White, *Antimalarial drug resistance*, Clin. Invest. 113 (2004), pp. 1084–1092.
[8] R. Arav-Boger and T.A. Shapiro, *Molecular mechanisms of resistance in antimalarial chemotherapy: the unmet challenge*, Annu. Rev. Pharmacol. Toxicol. 45 (2005), pp. 565–585.
[9] F. Cheng, J. Shen, X. Luo, W. Zhu, J. Gu, R. Ji, H. Jiang, and K. Chen, *Molecular docking and 3-D-QSAR studies on the possible antimalarial mechanism of artemisinin analogues*, Bioorg. Med. Chem. 10 (2002), pp. 2883–2891.

[10] A.K. Bhattacharjee, M.G. Hartell, D.A. Nichols, R.P. Hicks, B. Stanton, J.E. van Hamont, and W.K. Milhous, *Structure–activity relationship study of antimalarial indolo[2,1-b]quina-zoline-6,12-diones (tryptanthrins): three dimensional pharmacophore modeling and identification of new antimalarial candidates*, Eur. J. Med. Chem. 39 (2004), pp. 59–67.

[11] C.W. Jefford, *Why artemisinin and certain synthetic peroxides are potent antimalarials. Implication for the mode of action*, Curr. Med. Chem. 8 (2001), pp. 1803–1826.

[12] R.K. Haynes and S.C. Vonwiller, *From Qinghao, marvelous herb of antiquity, to the antimalarial trioxane qinghaosu- and some remarkable new chemistry*, Acc. Chem. Res. 30 (1997), pp. 73–79.

[13] D.L. Klayman, *Qinghaosu (artemisinin): an antimalarial drug from China*, Science 228 (1985), pp. 1049–1055.

[14] X.D. Luo and C.C. Shen, *The chemistry, pharmacology, and clinical applications of qinghaosu (artemisinin) and its derivatives*, Med. Res. Rev. 7 (1987), pp. 29–57.

[15] G. Bernardinelli, C.W. Jefford, D. Maric, C. Thomson, and J. Weber, *Computational studies of the structures and properties of potential anti-malarial compounds based on the 1,2,4-trioxane ring structure: I. Artemisinin-like molecules*, Int. J. Quant. Chem.: Quant. Biol. Symp. 21 (1994), pp. 117–131.

[16] G.H. Posner, J.N. Cumming, P. Ploypradith, and C.H. Oh, *Evidence for Fe(IV)AO in the molecular mechanism of action of the trioxane antimalarial artemisinin*, J. Am. Chem. Soc. 117 (1995), pp. 5885–5886.

[17] G.H. Posner, D. Wang, J.N. Cumming, P. Ploypradith, C.H. Oh, A.N. French, A.L. Bodley, and T.A. Shapiro, *Further evidence supporting the importance of and the restrictions on a carbon-centered radical for high antimalarial activity of 1, 2, 4-trioxanes like artemisinin*, J. Med. Chem. 38 (1995), pp. 2273–2275.

[18] R.K. Haynes and S.C. Vonwiller, *The behaviour of qinghaosu (artemisinin) in the presence of hemin iron (II) and (III)*, Tetrahedron Lett. 37 (1996), pp. 253–256.

[19] M.A. Rafiee, N.L. Hadipour, and H. Naderi-manesh, *The role of charge distribution on the antimalarial activity of artemisinin analogues*, J. Chem. Inf. Model. 45 (2005), pp. 366–370.

[20] W.M. Wu, Y. Wu, Y.L. Wu, Z.J. Yao, C.M. Zhou, Y. Li, and F. Shan, *Unified mechanistic framework for the Fe(II)-induced cleavage of qinghaosu and derivatives/analogues. The first spin-trapping evidence for the previously postulated secondary C-4 radical*, J. Am. Chem. Soc. 120 (1998), pp. 3316–3325.

[21] S.R. Meshnick, *Artemisinin: mechanisms of action, resistance and toxicity*, Int. J. Parasitol. 32 (2002), pp. 1655–1660.

[22] R.K. Haynes and S. Krishna, *Artemisinins: activities and actions*, Microbes Infect. 6 (2004), pp. 1339–1346.

[23] R. Kannan, K. Kumar, D. Sahal, S. Kukreti, and V.S. Chauhan, *Reaction of artemisinin with haemoglobin: implications for antimalarial activity*, Biochem. J. 385 (2005), pp. 409–418.

[24] S. Kamchonwongpaisan, E. Samoff, and S.R. Meshnick, *Identification of hemoglobin degra-dation products in Plasmodium falciparum*, Mol. Biol. Parasitol. 86 (1997), pp. 179–186.

[25] Y.L. Hong, Y.Z. Yang, and S.R. Meshnick, *The interaction of artemisinin with malarial hemozoin*, Mol. Biochem. Parasitol. 63 (1994), pp. 121–128.

[26] C.W. Jefford, M.G.H. Vicente, Y. Jacquier, F. Favarger, J. Mareda, P. Millasson-Schmidt, G. Brunner, and U. Burger, *The deoxygenation and isomerization of artemisinin and artemether and their relevance to antimalarial action*, Helvetica Chim. Acta. 79 (1996), pp. 1475–1487.

[27] G.H. Posner, J.N. Cumming, P. Ploypradith, and C.H. Oh, *Evidence for Fe(IV)=O in the molecular mechanism of action of the trioxane antimalarial artemisinin*, J Am. Chem. Soc. 117 (1995), pp. 5885–5887.

[28] G.H. Posner, C.H. Oh, D. Wang, L. Gerena, W.K. Milhous, S.R. Meshnick, and W. Asawamahasadka, *Mechanism-based design, synthesis, and in vitro antimalarial testing of new 4-methylated trioxanes structurally related to artemisinin: the importance of a carbon-centered radical for antimalarial activity*, J. Med. Chem. 37 (1994), pp. 1256–1258.

[29] A. Robert and B. Meunier, *Is alkylation the main mechanism of action of the antimalarial drug artemisinin?*, Chem. Soc. Rev. 27 (1998), pp. 273–274.

[30] A.V. Pandey, B.L. Tekwani, R.L. Singh, and V.S. Chauhan, *Artemisinin, an endoperoxide antimalarial, disrupts the hemoglobin catabolism and hemin detoxification systems in malarial parasite*, J. Biol. Chem. 274 (1999), pp. 19383–19388.

[31] R. Guha and P.C. Jurs, *Development of QSAR models to predict and interpret the biological activity of artemisinin analogues*, J. Chem. Inf. Comput. Sci. 44 (2004), pp. 1440–1449.

[32] J.C. Pinheiro, R. Kiralj, and M.M.C. Ferreira, *Artemisinin derivatives with antimalarial activity against Plasmodium falciparum designed with the aid of quantum chemical and partial least squares methods*, QSAR Comb. Sci. 22 (2003), pp. 830–842.

[33] R.H. Zhou, R.A. Friesner, A. Ghosh, R.C. Rizzo, W.L. Jorgensen, and R.M. Levy, *New linear interaction method for binding affinity calculations using a continuum solvent model*, J. Phys. Chem. 105 (2001), pp. 10388–10397.

[34] J. Aqvist, C. Medina, and J.E. Samuelsson, *New method for predicting binding-affinity in computer-aided drug design*, Protein Eng. 7 (1994), pp. 385–391.

[35] J. Aqvist and J. Marelius, *The linear interaction energy method for predicting ligand binding free energies*, Combin. Chem. High Throughput Screening 4 (2001), pp. 613–626.

[36] Y. Tominaga and W.L. Jorgensen, *General model for estimation of the inhibition of protein kinases using Monte Carlo simulations*, J. Med. Chem. 47 (2004), pp. 2534–2549.

[37] H.K.S. Leiros, B.O. Brandsdal, O.A. Andersen, V. Os, I. Leiros, R. Helland, J. Otlewski, N.P. Willassen, and A.O. Smalas, *Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements*, Protein Sci. 13 (2004), pp. 1056–1070.

[38] D. Ostrovsky, M. Udier-Blagovic, and W.L. Jorgensen, *Analyses of activity for factor Xa inhibitors based on Monte Carlo simulations*, J. Med. Chem. 46 (2003), pp. 5691–5699.

[39] M.M.H. van Lipzig, N.P.E. Vermeulen, M. Wamelink, D. Geerke, A. Jongejan, T.M. ter Laak, and J.H.N. Meerman, *Molecular dynamics simulations with the estrogen receptor: prediction of ligand binding affinity and orientation by the linear interaction energy method*, Drug Metab. Rev. 35 (2003), p. 107.

[40] H.A. Carlson and W.L. Jorgensen, *An extended linear-response method for determining free-energies of hydration*, J. Phys. Chem. 99 (1995), pp. 10667–10673.

[41] T. Hansson and J. Åqvist, *Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations*, Protein Eng. 8 (1995), pp. 1137–1145.

[42] *FirstDiscovery2.7*, Schrödinger Inc.: Portland, OR, 2004.

[43] *Liaison*, version 4.0, Schrödinger, LLC, New York, NY, 2005.

[44] P.J. Hay and W.R. Wadt, *Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals*, J. Chem. Phys. 82 (1985), pp. 299–310.

[45] *Jaguar*, version 4.1: Schrödinger, Inc.: Portland, OR, 2000.

[46] J.R. Woolfrey, M.A. Avery, and A.M. Doweyko, *Comparison of 3D quantitative structure-activity relationship methods: Analysis of the in vitro antimalarial activity of 154 artemisinin analogues by hypothetical active-site lattice and comparative molecular field analysis*, J. Comp. Aided Mol. Design. 12 (1998), pp. 165–181.

[47] N. Acton, J.M. Karle, and R.E. Miller, *Synthesis and antimalarial activity of some 9-substituted artemisinin derivatives*, J. Med. Chem. 36 (1993), pp. 2552–2557.

[48] A.J. Lin, M. Lee, and D.L. Klayman, *Antimalarial activity of new water-soluble dihydroartemisinin derivatives*, J. Med. Chem. 32 (1989), pp. 1249–1252.

[49] G.H. Posner, C.H. Oh, L. Gerena, and W.K. Milhous., *Extraordinarily potent antimalarial compounds: new, structurally simple, easily synthesizes, tricyclic 1,2,4-trioxanes*, J. Med. Chem. 35 (1992), pp. 2459–2467.

[50] M.A. Avery, J.D. Bonk, S. Mehrotra, W.K. Chong, R. Miller, W. Milhous, D.K. Goins, S. Venkatesan, C. Wyandt, I. Khan, and B.A. Avery, *Structure–activity relationships of the antimalarial agent artemisinin*, J. Med. Chem. 38 (1995), pp. 5038–5044.

[51] M.A. Avery, S. Mehrotra, J.D. Bonk, J.A. Vroman, and D.K. Goins, *Structure–activity relationships of the antimalarial agent artemisinin*, J. Med. Chem. 39 (1996), pp. 2900–2906.

[52] J.C. Pinheiro, M.M.C. Ferreira, and O.A.S. Romero, *Antimalarial activity of dihydroartemisinin derivatives against* P. falciparum *resistant to mefloquine: a quantum chemical and multivariate study*, J. Mol. Struct. 572 (2001), pp. 35–44.

[53] I. Leban, L. Golic, and M. Japelj, *Crystal and molecular structure of qinghaosu: a redetermination*, Acta. Pharm. Jugosl. 38 (1988), pp. 71–77.

[54] J.N. Lisgarten, B.S. Potter, C. Bantuzeko, and R.A. Palmer, *Structure, absolute configuration, and conformation of the antimalarial compound, artemisinin*, J. Chem. Cryst. 28 (1998), pp. 539–543.

[55] G.B. Tomas and R.L. Finney, *Calculus and Analytic Geometry*, 9th ed., Addison-Wesley, Reading, MA, 2001.

[56] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B. (1987), pp. 195–216.

[57] R. Zhou and B.J. Berne, *Smart walking: a new method for Boltzmann sampling of protein conformations*, J. Chem. Phys. 107 (1997), pp. 9185–9196.

[58] M. Tuckerman, B.J. Berne, and G.J. Martyna, *Reversible multiple time scale molecular dynamics*, J. Chem. Phys. 97 (1992), pp. 1990–2001.

[59] R. Zhou and B.J. Berne, *A new molecular dynamics method combing the reference system propagator algorithm with a fast multipole method for simulating proteins*, J. Chem. Phys. 103 (1995), pp. 9444–9459.

[60] D.L. Klayman, *Quinghaosu (Artemisinin): An antimalarial drug from China*, Science 228 (1985), pp. 1049–1055.

[61] F. Oisebenoit-Vical, A. Robert, and B. Meunier, *Trioxaquines Are new antimalarial agents active on all erythrocytic forms, including gametocytes*, Antimicrob. Agents Chemo. 51 (2000), pp. 2836–2841.

[62] D.J. Creek, W.N. Charman, F.C.K. Chiu, R.J. Prankerd, Y. Dong, J.L. Vennerstrom, and S.A. Charman, *Relationship between antimalarial activity and heme alkylation for spiro- and dispiro-1,2,4-trioxolane antimalarials*, Antimicrob. Agents Chemo. 52 (2008), pp. 1291–1296.

[63] R.C. Rizzo, J. Tirado-Rives, and W.L. Jorgensen, *Estimation of binding affinities for HEPT and nevirapine analogs with reverse transcriptase via Monte Carlo simulations*, J. Med. Chem. 44 (2001), pp. 145–154.