



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

John R. Woolfrey^{a,b}, Mitchell A. Avery^{a,b,c,*} & Arthur M. Doweyko^{d,*}

^aDepartment of Medicinal Chemistry, ^bNational Center for the Development of Natural Products, School of Pharmacy, and ^cDepartment of Chemistry, University of Mississippi, University, MS 38677, U.S.A., ^dBristol-Myers Squibb, Pharmaceutical Research Institute, Princeton, NJ 08543, U.S.A.

Received 21 January 1997; Accepted 10 September 1997

Key words: antimalarial, artemisinin, comparative molecular field analysis (CoMFA), hypothetical active-site lattice (HASL), quantitative structure–activity relationships (QSARs)

Summary

Two three-dimensional quantitative structure–activity relationship (3D-QSAR) methods, comparative molecular field analysis (CoMFA) and hypothetical active site lattice (HASL), were compared with respect to the analysis of a training set of 154 artemisinin analogues. Five models were created, including a complete HASL and two trimmed versions, as well as two CoMFA models (leave-one-out standard CoMFA and the guided-region selection protocol). Similar r^2 and q^2 values were obtained by each method, although some striking differences existed between CoMFA contour maps and the HASL output. Each of the four predictive models exhibited a similar ability to predict the activity of a test set of 23 artemisinin analogues, although some differences were noted as to which compounds were described well by either model.

Introduction

In the pursuit of predictive pharmacophores in drug design, molecular modeling techniques have proven to be useful approaches for the generation of leads. Two such three-dimensional quantitative structure–activity relationship (3D-QSAR)-oriented computer programs, comparative molecular field analysis (CoMFA) and hypothetical active-site lattice (HASL), were designed to correlate biological activity to information inherent in the 3D structure of a compound. These methods can be used when drug receptor information is limited or unavailable, supplying QSAR information for the determination of active compound attributes. Each has shown repeated success in the development of predictive biological models [1–4].

The methods are similar in that both rely on an aligned set of compounds, typically energetically minimized, with a range of associated biological ac-

tivities. In addition, both methods work best when the molecules in the dataset contain relatively rigid features amenable to superpositioning with a broad variety of side chains or other small structural modifications. Both methods evaluate the developed model by mathematical methods, usually including a least-squares or related fitting routine to provide numerical information for use in grading the model.

The CoMFA method [2] is based on the placement of the template molecule in a lattice, and the interaction of this molecule with a ‘probe’ atom of designated size and charge, by default an sp^3 carbon atom of +1 charge, placed systematically at points, typically 2 Å apart, throughout this lattice. The electrostatic and steric fields of the molecule interact with those of the probe atom and the resulting field values are placed in a CoMFA table. Other compounds in the dataset are aligned to this template molecule, and the interactions of the molecular fields with the probe atom are assessed. These values are then compared

* To whom correspondence should be addressed.

a

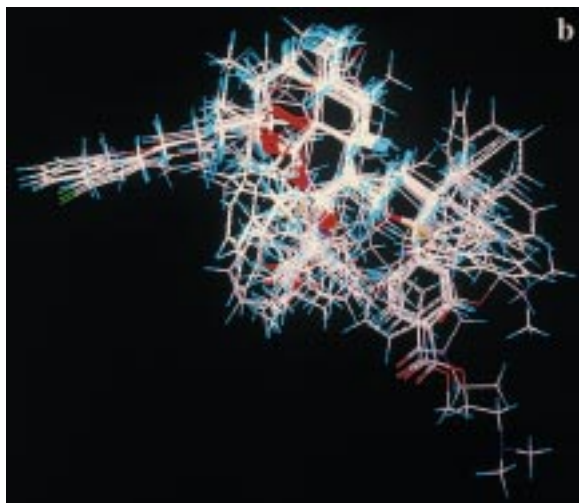
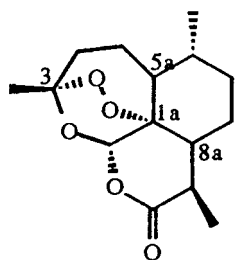


Figure 1. (a) Atoms selected for molecular alignment (C-1a, C-3, C-5a, and C-8a) and (b) the overlay of 154 artemisinin analogues.

with the associated biological activity of each molecule via partial least squares (pls) [5]. The results of a 'CoMFA run' can be evaluated by means of statistical parameters such as r^2 , q^2 , F, s, etc. The definition of the cross-validated correlation coefficient or q^2 is the quotient $(SD - PRESS)/SD$, where SD is the sum of squares of deviations of the observed values from their mean values and PRESS is the prediction error sum of squares. A high correlation constant or r^2 value indicates the ability of the model to reproduce the input data. A high q^2 value is associated with the ability to predict the activity of compounds left out of the dataset by a model made up of the remaining compounds in the dataset. The CoMFA model is then able to make activity predictions of compounds in the test set.

The HASL approach [6–9] is also based on working with an aligned set of compounds in a lattice space or grid of a chosen resolution. The molecular lattice is then designated as those points that are within van der Waals radii of each compound. Although molecules can be aligned through the intermediate superpositioning of their respective lattices, the present

investigation relied on molecules previously aligned using the fit atom protocol in SYBYL 6.3. These grid points are then assigned a 'HASL-type' value (1, 0, or –1) based on the electron density, where electron-rich substituents (e.g., O, N) are designated +1 and electron-poor substituents (e.g., C in C=O) are designated –1. Many of the lattice points are shared by more than one molecule, and partial bioactivity values are, at first, averaged over these points and, subsequently, adjusted by an iterative protocol to take into consideration the predicted (total of the lattice point activity values for a compound) versus observed activity.

Both HASL and CoMFA are powerful methods, and a comparison of the two approaches with respect to their abilities in a single system is of interest. This paper proposes one such system, the dataset of 154 analogues of the antimalarial compound artemisinin, and herein we describe the comparison of the predictive attributes of both 3D-QSAR methods.

Experimental

All molecular modeling and CoMFA-related calculations were carried out on a Silicon Graphics Indigo2 Impact (250 MHz Mips IP22 R4400/R4010). SYBYL 6.3 [26] and the included interfaces were used for all molecular modeling (drawing, minimization, charge calculation) and CoMFA methodology. All HASL-related calculations were run on a PC platform (486 66 MHz CPU) using HASL v. 3.28 [36], designed for 386 and higher CPU.

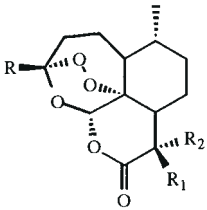
Materials and methods

CoMFA

An initial dataset of 171 compounds, all analogues of the antimalarial natural product artemisinin, was assembled using in vitro bioassay data from our laboratory [10–13], as well as published data from other laboratories [14–25]. Each of these compounds had associated in vitro bioactivity values (IC_{50} values reported in ng/ml) against the drug-resistant malarial 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

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

Table 1. Relative activities of artemisinin analogues

				
Compound	R	R ₁	R ₂	log(RA) W-2
1	CH ₃	CH ₃	H	1.00
6	C ₄ H ₈ Ph	H	H	0.45
9	CH ₃	H	2-Z-Butenyl	-1.10
10	CH ₃	H	H	0.79
12	CH ₃	H	CH ₃	-0.17
13	CH ₃	H	2-E-Butenyl	-0.60
16	CH ₃	Allyl	H	-0.10
24	CH ₃	C ₄ H ₉	H	0.17
26	C ₄ H ₈ Ph	C ₄ H ₉	H	-0.32
27	C ₃ H ₆ (<i>p</i> -Cl-Ph)	C ₄ H ₉	H	-0.28
28	CH ₂ CH ₂ CO ₂ Et	C ₄ H ₉	H	1.36
29	C ₄ H ₉	C ₄ H ₉	H	-0.48
30	CH ₃	C ₂ H ₅	H	1.40
35	CH ₃	C ₆ H ₁₃	H	0.86
36	CH ₃	<i>i</i> -C ₄ H ₉	H	-0.55
37	CH ₃	<i>i</i> -C ₆ H ₁₃	H	-0.04
38	CH ₃	<i>i</i> -C ₃ H ₇	H	-0.04
39	CH ₃	<i>i</i> -C ₅ H ₁₁	H	0.07
61	C ₃ H ₆ (<i>p</i> -Cl-Ph)	H	H	0.10
64	C ₄ H ₉	H	H	-0.74
66	CH ₂ CH ₂ CO ₂ Et	H	H	0.37
67	C ₂ H ₅	H	H	0.05
70	<i>i</i> -C ₄ H ₉	H	H	-0.35
72	C ₃ H ₇	H	H	0.83
76	CH ₃	C ₃ H ₆ (<i>P</i> -Cl-Ph)	H	1.37
85	CH ₃	Br	CH ₂ Br	-1.64
91	CH ₃		=CH ₂	-0.89
95	CH ₃	CH ₂ CH ₃	R ₁ =R ₂	-0.36
96	CH ₃		-CH ₂ CH ₂ -	-0.94
101	CH ₃	C ₅ H ₁₁	H	1.02
102	CH ₃	C ₄ H ₈ Ph	H	0.63
103	CH ₃	C ₂ H ₄ Ph	H	0.12
105	CH ₃	C ₃ H ₆ Ph	H	0.78
107	CH ₃	C ₃ H ₇	H	1.13

Molecular models of the artemisinin analogues were built using bond and angle standard geometries from Sybyl 6.3 [26] before being fully minimized by the Tripos force field. The database was aligned (Figure 1) with the fit atom protocol using a low-

energy conformer of artemisinin as a template. Atomic charges were calculated using the Gasteiger–Hückel protocol within SYBYL 6.3. The compounds underwent a CoMFA protocol using an sp³ carbon +1 probe and 1 Å lattice spacing. An initial CoMFA run us-

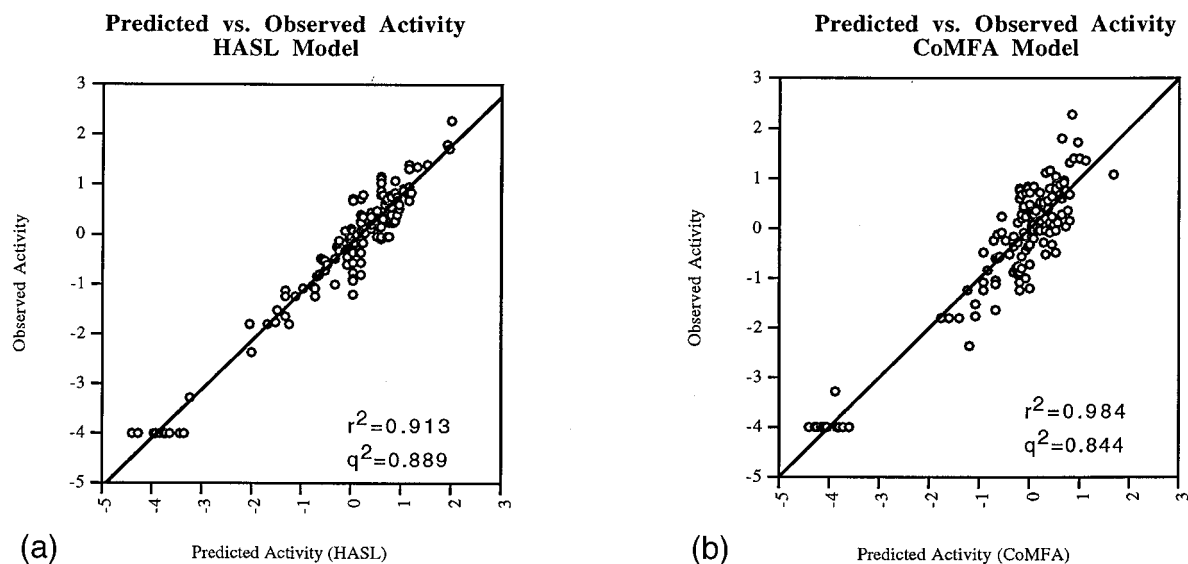


Figure 2. (a) HASL model (rec45, 31 points) predictions vs. observed relative *in vitro* antimalarial activity. (b) CoMFA model predictions vs. observed activity.

Table 1 (Continued). Deoxy-artemisinin analogues

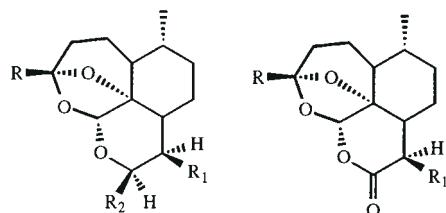
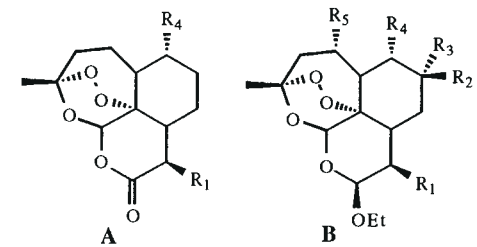
				
Compound	R	R ₁	R ₂	log(RA)
148	CH ₃	CH ₃	OEt	-4.00
149	CH ₃	CH ₃	OH	-4.00
150	CH ₃	C ₄ H ₈ Ph	-	-4.00
151	CH ₃	C ₃ H ₇	-	-4.00
152	CH ₃	C ₆ H ₁₃	-	-4.00
153	CH ₃	CH ₃	H	-4.00
154	CH ₃	C ₄ H ₉	H	-4.00
155	CH ₃	<i>i</i> -C ₅ H ₁₁	-	-4.00
156	CH ₂ CH ₂ CO ₂ Et	H	H	-4.00
157	C ₂ H ₄ Ph	H	-	-4.00
158	CH ₂ CH ₃	H	-	-4.00
159	<i>i</i> -C ₄ H ₉	H	-	-4.00
160	<i>i</i> -C ₄ H ₉	H	H	-4.00
161	CH ₃	C ₂ H ₄ Ph	-	-4.00
162	CH ₃	C ₃ H ₆ Ph	-	-4.00
163	CH ₃	CH ₃	-	-4.00

Table 1 (Continued). Various derivatives of artesimnin and arteether

							
Compound	Type	R ₁	R ₂	R ₃	R ₄	R ₅	log(RA)
8	A	H	-	-	H	-	-0.36
15	B	CH ₃	OH	H	CH ₃	H	0.34
118	A	H	-	-	H	-	-1.79
131	B	CH ₃	H	H	CH ₃	OH	-1.27

ing 10 components and leave-one-out cross-validation pls indicated that the optimal number of components was four. Cross-validated pls analyses were used to grade the predictive character of the model. Those compounds which had high residual values (difference between predicted and actual activity) were gradually pruned from the dataset until a satisfactory q^2 value was achieved at $n = 154$ compounds as shown in Table 1.

To obtain a more reproducible q^2 result, the recently published cross-validated r^2 guided region se-

lection (q^2 -grs) approach [28] was used with the final $n = 154$ dataset. This simple SYBYL programming language-based method divides the initial lattice into 125 'sublattices', and runs a CoMFA in each new lattice. Those areas in which the resulting q^2 is not above a certain cutoff, by default $q^2 \leq 0.1$, are not included in the final CoMFA region. The results of these CoMFA analyses are indicated in Table 2. A non-cross-validated pls analysis was run in order to generate the electrostatic and steric contour maps.

In attempting to more thoroughly compare the CoMFA and HASL methodologies, a hydrophilicity parameter was considered within the CoMFA model using HINT [37, 38]. Polar/lipophilic fields around the analogues were measured and compared to activity by use of pls analysis. Contour maps again reflect the outcome of the pls analysis, and statistical values are indicated in Table 2.

HASL

The aforementioned final dataset of 154 compounds yielded a 511-point HASL model, generated at a point spacing of 2.00 Å with H-value definitions (+1, 0, -1) [6]. This HASL-511 model provided a well-behaved self-consistent set of predictions for the entire set (Figure 2a). Here, as in CoMFA, compound activity was predicted by the model, and these predictions were compared with experimental *in vitro* values.

In an effort to highlight the most important molecular features for the artemisinin series, HASL-511 was subjected to a trimming process wherein repeated removal of 5% of the least significant points, i.e., those containing partial binding values (RA) nearest zero, resulted in a series of 65 HASL models, or 3D pharmacophores ranging from 484 (rec0) to 1 (rec65) points [4]. Each trimmed model (rec#) was evaluated for self-consistency by comparing predicted to actual $\log(\text{RA})$ by means of the correlation coefficient (r^2). This analysis is illustrated in Figure 3. As the trimming process is conducted from right to left, the number of points in the HASL model are reduced and the predictivity of the model falls off asymptotically, with a significant loss in predictivity as points critical to the model are removed.

The HASL program identifies molecular regions most critical in affecting activity, or binding, by summing the partial activity values associated with an atom by means of the lattice point(s) located within the van der Waals radius of that atom. In this way, submolecular features, favorable or unfavorable to activity, can be highlighted either visually or in a tab-

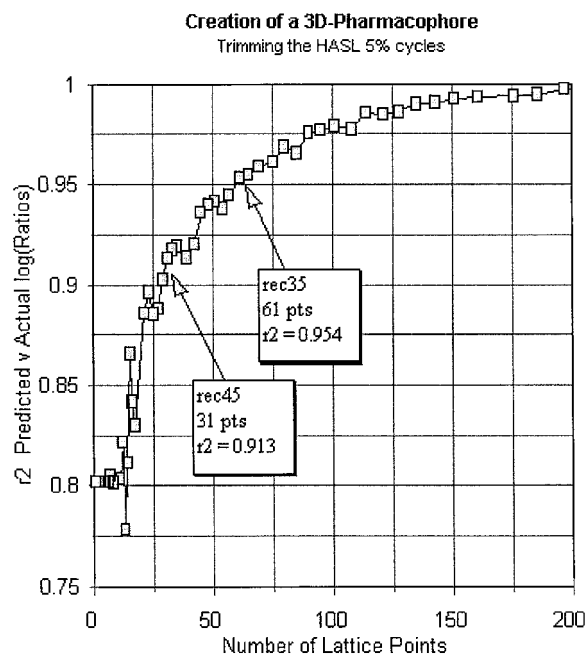


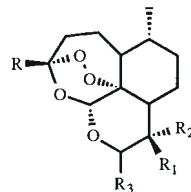
Figure 3. HASL trimming process: Repeated removal of 5% of the points containing partial binding values nearest zero resulted in 65 3D pharmacophores ranging from 484 (rec0) to 1 (rec 65) points. Correlation coefficients (predicted vs. actual activity) decrease as the number of points making up the pharmacophore decrease.

ularized format wherein $\Delta G(\text{binding})$ is estimated for each atom in the molecule. These estimates are made with the assumption that activity values are provided in units of $-\log K_i$ (pK_i) [6].

Three of the aforementioned models (511, rec35, and rec45) were used to predict the activities of 27 compounds (Table 3) either not included in the original ($n = 171$) dataset or compounds **87**, **97**, and **147** which were pruned from the dataset due to high initial residual values. The predictions, done in a blind protocol, were then compared to the CoMFA results of similar predictions done in parallel, as well as with the actual biological data (Table 4).

Finally, an HASL model including all 171 original analogues was developed to determine if the compounds indicated by the $n = 171$ CoMFA model as being 'outliers' would also have high residual values in an HASL analysis. Results from the statistical (pls) analysis are included in Table 2.

Table 1 (Continued). 10-Substituted artemisinin derivatives



Compound	R	R ₁	R ₂	R ₃	log(RA)
2	CH ₃	CH ₃	H	H	0.75
3	CH ₃	CH ₃	H	OH	0.55
4	CH ₃	CH ₃	H	OE _t	0.34
5	CH ₃	C ₄ H ₉	H	OH	0.96
7	CH ₃	CH ₃	H	OE _t	-1.08
11	CH ₃	H	Br	H	0.28
17	CH ₃	CH ₃	Br	NH-2-(1,3-thiazole)	0.66
18	CH ₃	CH ₃	Br	<i>p</i> -Cl-aniline	0.79
19	CH ₃	CH ₃	Br	Aniline	0.18
20	CH ₃	CH ₃	Br	NH-2-pyridine	-0.09
21	CH ₃	CH ₃	Br	NH-2-pyridine	-0.77
22	CH ₃	CH ₃	H	OMe	0.28
23	CH ₃	CH ₃	H	α-OE _t	0.32
25	CH ₃	C ₄ H ₉	H	H	1.32
31	CH ₃	C ₂ H ₅	H	H	0.67
32	CH ₃	C ₃ H ₇	H	OE _t	-0.04
33	CH ₃	H	H	OE _t	0.43
34	CH ₃	C ₂ H ₅	H	OE _t	0.50
46	CH ₃	CH ₃	H	C ₃ H ₆ OH	0.78
47	CH ₃	CH ₃	H	C ₄ H ₉	0.06
50	CH ₃	CH ₃	H	OCH ₂ CO ₂ Et	0.52
51	CH ₃	CH ₃	H	OC ₂ H ₄ CO ₂ Me	0.10
52	CH ₃	CH ₃	H	OC ₃ H ₆ CO ₂ Me	-0.03
53	CH ₃	CH ₃	H	OCH ₂ (4-PhCO ₂ Me)	-0.07
54	CH ₃	CH ₃	H	(<i>R</i>)-OCH ₂ CH(CH ₃)CO ₂ Me	1.79
55	CH ₃	CH ₃	H	(<i>S</i>)-OCH ₂ CH(CH ₃)CO ₂ Me	2.25
56	CH ₃	CH ₃	H	(<i>R</i>)-OCH(CH ₃)CH ₂ CO ₂ Me	0.87
57	CH ₃	CH ₃	H	(<i>S</i>)-OCH(CH ₃)CH ₂ CO ₂ Me	1.70
58	CH ₃	CH ₃	H	OCH ₂ -adamantyl	0.28
59	CH ₂ CH ₂ CO ₂ Et	H	H	H	0.70
60	C ₃ H ₆ (<i>p</i> -Cl-Ph)	H	H	H	-0.55
65	C ₄ H ₉	H	H	H	0.75
68	C ₂ H ₅	H	H	H	-1.00
71	<i>i</i> -C ₄ H ₉	H	H	H	0.40
73	C ₃ H ₇	H	H	H	0.84
74	C ₄ H ₈ Ph	H	H	H	0.58
89	CH ₃	-OCH ₂ -		OOH	-0.62
90	CH ₃	-CH ₂ O-		OOH	-0.57
92	CH ₃	=CH ₂		OOH	-0.990
100	CH ₃	C ₅ H ₁₁	H	H	0.16
104	CH ₃	C ₃ H ₆ Ph	H	H	1.40
106	CH ₃	C ₃ H ₇	H	H	0.74
120	CH ₃	CH ₃	H	OO <i>t</i> -C ₄ H ₉	0.92

All R₃ substituents are β except where noted.

Table 1 (Continued). *Seco*-artemisinin derivatives

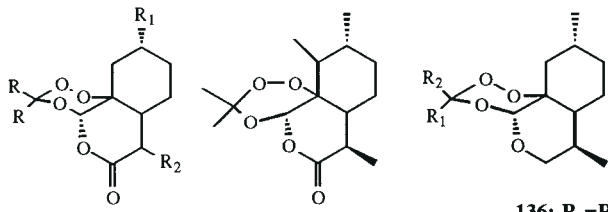
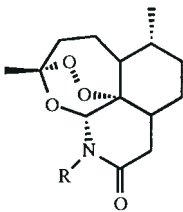
				
				136: R₁=R₂= -(CH₂)₄- 184: R₁=R₂= CH₃
		125		
Compound	R	R ₁	R ₂	log(RA)
121	CH ₃	H	H	-2.37
122	CH ₃	H	C ₄ H ₉	-1.13
123	C ₂ H ₅	H	H	-0.60
124	CH ₃	CH ₃	CH ₃	-0.15
125	-	-	-	-0.86
126	CH ₃	H	CH ₃	-1.27
136	-	-	-	-0.26

Table 1 (Continued). 11-Aza-artemisinin derivatives

		
Compound	R	log(RA)
40	C ₃ H ₆ Ph	0.02
41	C ₂ H ₄ Ph	0.16
42	C ₅ H ₁₁	-0.20
43	<i>i</i> -C ₅ H ₁₁	-0.04
44	CH ₂ (<i>p</i> -Cl-Ph)	-0.16
45	<i>i</i> -C ₄ H ₉	0.02
77	CH ₂ Ph	0.34
78	CH ₃	0.70
84	C ₃ H ₇	0.05

Results

CoMFA

Increasing the number of compounds in the database from previous models from this laboratory [1] has led to a corresponding increase in q^2 (q^2 -leave-one-out = 0.840, q^2 -grs = 0.844). Within the CoMFA q^2 -leave-one-out model, the residuals of these compounds ranged from 0 to ± 1.45 log units, indicating that the predicted activities of the compounds within the dataset are within 1.45 of the observed value. The residual values were smaller for the test set of 24 compounds, which fell within a range of 0 to ± 1.4 log units. Compounds with high residuals in the original $n = 174$ dataset that were excluded in the final $n = 154$ dataset were also predicted poorly by the $n = 154$ CoMFA model: compounds **97** and **147** had residual values of 2.11 and 1.89, respectively. Of the 24 test set compounds predicted using the q^2 -leave-one-out method, 10 compounds were within 0.5 log units, 14 within 0.66 log units, and 16 (of 24) within 0.8 log units. Within the later 'high residual' compounds, structure **190** is predicted to be much less active than artemisinin, though the magnitude of the residual value is still significant. A comparison of the final $n = 154$ CoMFA with the model described earlier [1] indicates that this new model is more robust in both predictivity of training and test set analogues. The predicted versus actual activities, including r^2 and q^2 , of

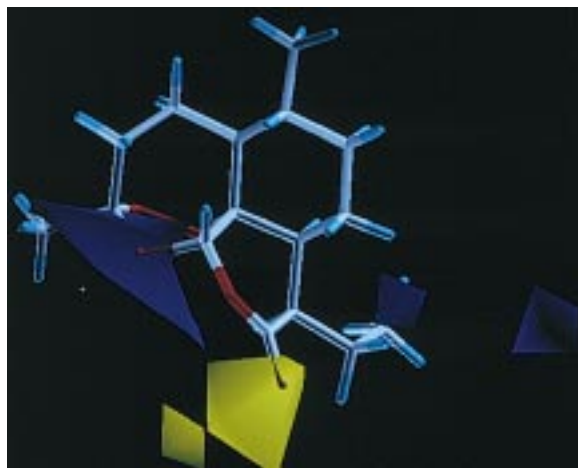


Figure 4. 9-Ethyl artemisinin within the HINT CoMFA contour map: yellow contours around regions where high lipophilicity is expected to increase activity, and blue contours represent areas where hydrophilicity is expected to increase activity.

the final CoMFA are presented in Figure 2b. It is interesting to note that the four most active analogues in the dataset (**55**, **54**, **57**, and **30**) and six of the top 10 had the highest residual values within the cross-validated leave-one-out analysis.

Using the q^2 -grs methodology, an improvement in both predictivity of training and test set analogues was noticed. This increase in q^2 value indicates a better ability of the model to predict the activity of analogues that have been left out of the model. Although the range of residual values (0 to ± 1.7) increased over that of the cross-validated leave-one-out analysis, better predictions were obtained for most compounds within the dataset. It is noteworthy that in this analysis, the most active analogues are much better predicted, and no longer represent the extreme high end of the residual range.

Small changes in the choice of atoms by which the molecules are aligned result in a significant reduction in q^2 values. However, once aligned, adjustment of lattice size from 1 to 4 Å made little difference in the predictivity or reliability of either the q^2 -leave-one-out or q^2 -grs models. To better understand the contribution each factor might play in the total model, it can be advantageous to run a CoMFA analysis based on either the electrostatic, steric, or hydrophilicity field data alone. In this particular case, though, models based on either field (electrostatic or steric) independent of the other were of lesser statistical validity than the model based on the electrostatic and steric field values combined by the standard CoMFA/pls protocol.

The inclusion of a hydrophilicity parameter into CoMFA (using HINT) resulted in a lower r^2 and q^2 (0.535 and 0.233, respectively) and high standard error, indicating poor correlation between molecular hydrophilicity/lipophilicity and activity (Table 2). Though not considered highly useful in the design of new compounds, examination of the contour maps (Figure 4) resulting from the HINT pls analysis does expose some interesting points. The blue (polar) contour in the peroxide region correlates well to the blue (electron-rich) contour in the HASL model, as well as to the red (electronegative) contour within CoMFA. It is also noteworthy that some of the more active compounds in the database are in the 10-deoxo (less polar) class (e.g., analogue **25**), which are described by the yellow (hydrophobic) contour near the C-10 position.

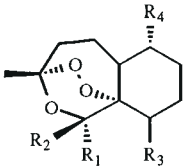
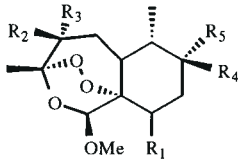
HASL

A detailed illustration of the trimming process is shown in Figure 3 wherein the 511-point HASL model was successively reduced to a number of pharmacophores from which two representative models were chosen for analysis: rec35, a HASL model containing 61 points reproducing the activities of its members (performance) with an r^2 of 0.954, and rec45 having 31 points and a performance r^2 of 0.913; the performance of rec45 is also shown in Figure 2 as a plot of predicted versus actual log(RA). The two models (rec35 and rec45) represent 3D pharmacophores which retain reasonable levels of performance ($r^2 > 0.9$). As can be seen from Figure 3, further removal of points from the rec45 model would result in a significant reduction of the r^2 .

Both HASL-derived rec35 and rec45 3D pharmacophores were subjected to leave-one-out cross-validation analysis wherein the activity of each artemisinin analogue was predicted by a model made up of the other 153 analogues. Surprisingly strong correlation coefficients (q^2) of 0.924 and 0.889 were obtained for pharmacophores rec35 and rec45, respectively (rec45 is illustrated in Figure 2).

The kcal/atom function associated with the HASL program results in a listing of estimated kcal/atom values obtained when a molecule is fitted to the model. As an illustration of this process, molecule **152** was fitted to both the rec35 and rec45 pharmacophores and the kcal/atom listings having values other than zero are shown in Table 5. Listed are the atom numbers (as found in the SYBYL file), the MM2 type (used by HASL), and the kcal/atom values determined by each atom's proximity to a HASL point.

Table 1 (Continued). Artemisinin derivatives lacking the D-ring

<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>A</p> </div> <div style="text-align: center;">  <p>B</p> </div> </div>						
Compound A	R ₁	R ₂	R ₃	R ₄	log(RA)	
128	-O ₂ CCH ₂ Ph	H	H	CH ₃	-0.51	
130	H	H	H	CH ₃	-0.32	
135	H	OCH ₃	H	H	-0.31	
137	OCH ₃	H	H	H	-1.04	
139	OCH ₂ Ph	H	H	H	-0.09	
145	OCH ₃	H	C ₂ H ₄ O ₂ CNEt ₂	H	0.65	
146	OCH ₃	H	C ₂ H ₄ O ₂ CNPh ₂	H	1.06	
164	H	OCH ₃	C ₂ H ₄ OPO ₃ Et ₂	H	0.37	
165	H	OCH ₃	C ₂ H ₄ OP(=S)O ₂ Et ₂	H	0.72	
166	H	OCH ₃	C ₂ H ₄ OS(O) ₂ (<i>p</i> -toluene)	H	0.20	
168	H	OCH ₃	C ₂ H ₄ OS(O) ₂ (5-NMe ₂ naphthalen-1-yl)	H	0.33	
169	H	OCH ₃	C ₂ H ₄ OCH ₃	H	-0.39	
170	H	OCH ₃	C ₂ H ₄ OCH ₂ Ph	H	0.75	
171	H	OCH ₃	C ₂ H ₄ O-allyl	H	0.40	
172	H	OCH ₃	C ₂ H ₄ OCH ₂ -(3,5-dimethyl-1,2-oxazole)	H	0.58	
173	H	OCH ₃	C ₂ H ₄ O ₂ CPh	H	-0.59	
174	H	OCH ₃	C ₂ H ₄ O ₂ C(4-PhCO ₂ Me)	H	0.27	
175	H	OCH ₃	C ₂ H ₄ O ₂ C(4-PhCO ₂ H)	H	-0.81	
176	H	OCH ₃	C ₂ H ₄ O ₂ C(4-PhCONEt ₂)	H	0.23	
177	H	OCH ₃	C ₂ H ₄ O ₂ C(4-PhCO ₂ C ₂ H ₄ NMe ₂)	H	-0.60	
178	H	OCH ₃	C ₂ H ₄ O ₂ CCH ₂ NCO ₂ -(<i>t</i> -C ₄ H ₉)	H	-0.04	
183	H	OCH ₃	C ₂ H ₄ OCH ₂ (4-N-Me-pyridine)	H	-0.90	
Compound B	R ₁	R ₂	R ₃	R ₄	R ₅	log(RA)
138 ^a	H	H	H	CH ₃	CH ₃	-0.41
140	C ₂ H ₄ OH	H	CH ₃	H	H	-1.80
141	C ₂ H ₄ OH	CH ₃	H	H	H	0.23
142	C ₂ H ₄ OH	CH ₃	CH ₃	H	H	-1.80
143	C ₂ H ₄ OCH ₂ Ph	CH ₃	CH ₃	H	H	-1.80

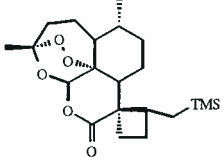
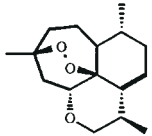
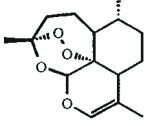
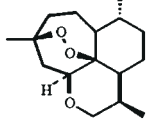
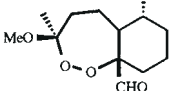
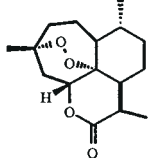
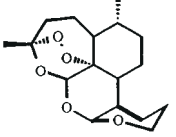
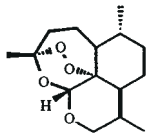
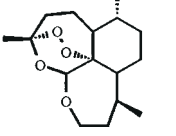
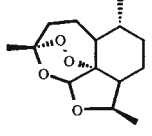
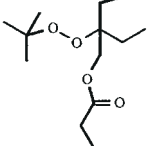
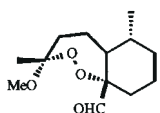
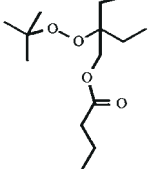
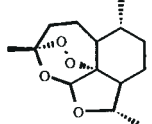
^aThe -OMe group is in the α -position in this compound.

The kcal/atom estimations are made on the assumption that binding values are of the type pK_i and not $\log(RA)$. The kcal/atom calculation is made and the partial pK_i values (ppK_i) associated with each point in the pharmacophore using the following relationship: $\Delta G(\text{kcal/atom}) = 2.303RT(ppK_i)$. Since the $\log(RA)$ values are assumed linearly proportional to pK_i , this exercise provides an insight into the relative importance of various features in a given molecule. Highlighted in the table are atoms which were found to be

associated with a consistent kcal/atom estimate in both pharmacophoric models. These atoms have also been highlighted in Figure 5, where positive steric contributions are indicated in green while negative steric contributions are shown in yellow.

To reproduce the method used in the development of the CoMFA pharmacophore, the $n = 171$ dataset was subjected to the HASL methodology. It is interesting to note that the HASL model resulted in lower residual values and a better prediction of all com-

Table 1 (Continued). Miscellaneous artemisinin derivatives

Compound	Structure	log(RA)	Compound	Structure	log(RA)
48		-1.24	191		-0.24
49		0.78	192		-2.59
127		-4.00	193		-0.96
129		0.23	194		-0.79
133		-1.20	196		-0.64
186		-4.00	198		-3.30
187		-4.00	199		0.353

pounds included in the $n = 171$ database. This can probably be attributed to the iterative nature of HASL. The two highest residual compounds determined by the HASL model (**30** and **102**) were among those compounds included in the $n = 154$ CoMFA model. The pls analysis of the $n = 171$ CoMFA indicated a consistent underprediction for the active compounds in the dataset, and an overprediction for the least (non-) ac-

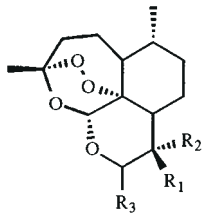
tive compounds. The two highest residual analogues in the original ($n = 171$) CoMFA run included the carboxylic acid **63**, potentially problematic in cell absorption, and the 3-desmethyl analogue **69**, considered unstable.

Table 2. Statistical results of CoMFA and HASL runs

Analysis	r^2	q^2	Components	F	S
HASL ($n = 171$)	0.902	—	—	—	—
HASL-511	0.923	—	—	—	—
HASL-31 (rec35)	0.913	0.889	—	—	—
HASL-61 (rec35)	0.954	0.924	—	—	—
CoMFA ($n = 171$)	0.908	0.644	5	0.937	1.012
CoMFA (HINT only)	0.423	0.144	3	21.71	1.118
CoMFA (electrostatic field only)		0.725	6		
CoMFA (steric field only)		0.815	6		
CoMFA (cross-validated leave-one-out)	0.972 ^a	0.840	4	281.78	0.625
CoMFA q^2 -grs	0.984 ^a	0.844	5	290.00	0.622

^aNon-cross-validated run using the same region as the cross-validated run.

Table 3. Test set structures: 10-substituted artemisinin derivatives

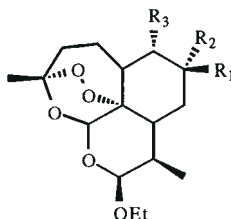
				
Compound	R ₁	R ₂	R ₃	log(RA)
98	CH ₃	OH	α-OH	−0.89
109	CH ₃	H	CH ₂ CHF ₂	0.11
111	CH ₃	OH	OCH ₂ CF ₃	0.33
112	OH	CH ₃	OCH ₂ CF ₃	−0.70
113	CH ₃	OH	OEt	−0.44
114	OH	CH ₃	OEt	−1.13

The R₃ substituent is β unless otherwise noted.

Discussion

The HASL model data for the 31-point rec45 3D pharmacophore are listed in Table 6 in terms of x , y , z , H-values, and partial log(RA). The scarcity of points with H-values other than zero suggests that the activities of this artemisinin series are largely explained by non-polar interactions, i.e., hydrophobic or steric contact. H-values of +1 and −1 refer to atom types with high and low electron density, respectively, and are suggestive of polar interaction.

Table 3 (Continued). Various derivatives of artemisinin and arteether

				
Compound	R ₁	R ₂	R ₃	log(RA)
108^a	F	—	CH ₃	−0.12
115	H	H	C(=O)H	0.21
116^b	=O	—	CH ₃	0.16
117	H	H	CHF ₂	0.41
119	F	F	CH ₃	0.19

^a A double bond exists between C-6 and C-7.

^b C-7 contains a carbonyl and no hydrogens.

The location of the 31 points contained in pharmacophore rec45 is illustrated along with the structure of molecule **152** in Figure 5 (positive and negative contributions). The pharmacophore appears to be in contact with the molecule at several largely aliphatic regions. As suggested earlier, the models indicate that binding is enhanced by the presence of the lipophilic propyl phenyl moiety as well as the methyl group on the opposite end of the molecule. Negative binding contributions appear to owe their presence to the aliphatic array near the center of the molecule.

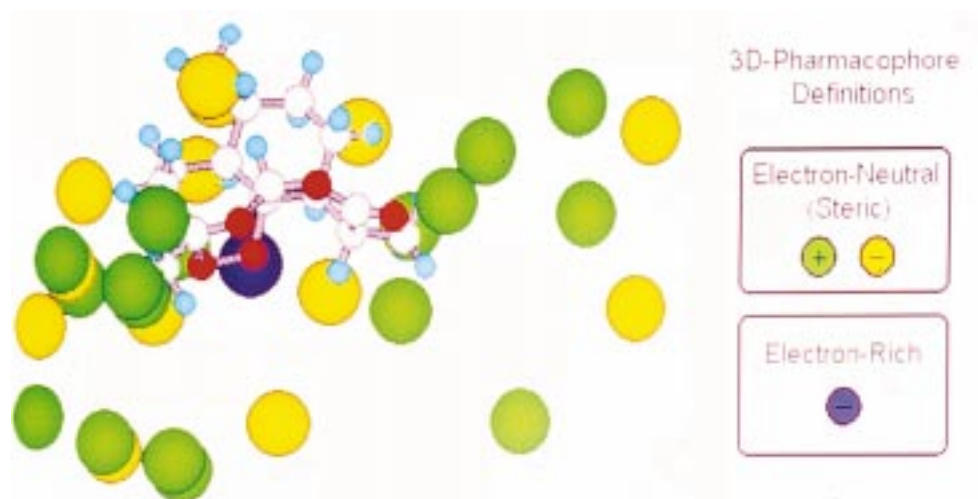


Figure 5. Artemisinin within the HASL pharmacophore. Green and yellow spheres represent positive and negative steric interactions, respectively; the blue sphere indicates an area of high electron density having a negative binding contribution.

Table 3 (Continued). 11-Aza-artemisinin derivatives

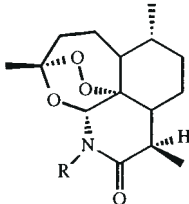
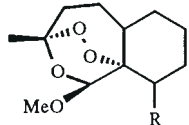
		
Compound	R	log(RA)
197	CH ₂ -(2-C ₅ H ₄ N)	1.46
200	2-Thiophene	0.17
201	Acetaldehyde	1.47
202	2-Furan	0.11

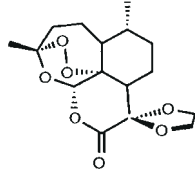
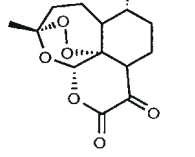
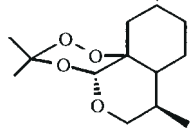
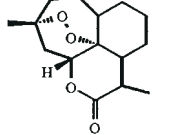
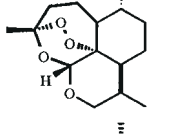
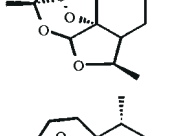
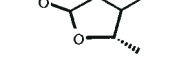
Table 3 (Continued). Artemisinin derivatives lacking the D-ring

		
Compound	R	log(RA)
147	C ₂ H ₄ OPO ₃ Ph ₂	1.10
180	C ₂ H ₄ OCH ₂ (4-F-Ph)	0.38
182	C ₂ H ₄ OCH ₂ (4-py)	0.14

A visual examination of the HASL results (Figure 5) with respect to the CoMFA contour map (Figure 6) makes for an interesting comparison. The HASL and CoMFA models strongly correlate on the influence substitution plays on activity. The CoMFA steric contour map indicates areas in which molecular steric bulk might have a favorable (green) or unfavorable (yellow) effect on the activity of an analogue. As is indicated by the higher activity of C-9 substituted analogues, the green contours which surround the C-9, C-8 and C-7 regions of the molecule reflect the increase in activity found by adding steric bulk in this region (e.g., **28** and **30**). This finding is essentially confirmed by the HASL model identifying the C-6, C-7, and C-8 region as sterically unfavorable, i.e., to substitution by electron-neutral atom types. The yellow CoMFA contours surrounding the peroxide bridge reinforce the importance of this part of the molecule in its mechanism of action: a sterically hindered peroxide would not be expected to be as active [29–35].

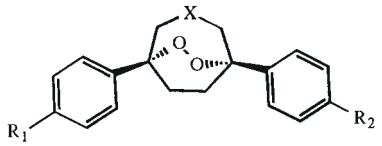
The CoMFA electrostatic map indicates red contours in regions where high electron density (partial or full negative charge) might play a favorable role in activity, and blue contours in areas in which this negative charge is predicted to decrease activity. With the artemisinin series, two major contours are located around the peroxide (red) and lactone carbonyl (blue) moieties. This indicates the importance of the negative charge associated with the peroxide to the activity of the artemisinin series. Earlier versions of the CoMFA model (smaller databases) were unable to gauge the importance of the peroxide bridge in

Table 3 (Continued). Miscellaneous artemisinin derivatives

Compound	Structure	log(RA)
87		-2.09
97		-2.49
184		-0.80
193		-0.96
194		-0.79
196		-0.64
199		-0.353

relation to activity because all active compounds included this peroxide functionality. However, inclusion of a series of inactive ether ('deoxy') analogues (**148–163**) has allowed the CoMFA model to differentiate between inclusion or removal of this moiety. The HASL model, although largely attributing steric effects (electron-neutral) to both positive and negative contributions, does identify a single, highly negative

Table 3 (Continued). Bicyclo artemisinin derivatives

			
Compound	X	R ₁ and R ₂	log(RA)
134	CH ₂	OCH ₃	-0.67
190	O	H	-2.26

contribution for electron-rich atoms in the region of the peroxide bridge, as indicated by CoMFA. The difference between the electrostatic aspects of each model is probably related to the type of charge associated with each portion of the molecule, especially the peroxide. The Gasteiger-Hückel protocol measures the atomic charges of the ether ('deoxy') oxygen as significantly more negative than either of the peroxide oxygens, leading to the differentiation of the two moieties. Another explanation of the difference can be traced to the nature of the HASL paradigm. Each molecule supplies to the model its own balance of negative and positive contribution to activity. Thus, there can be at least two reasons why a region of the model appears to contain negative contributions to activity: (i) molecular features existent at these sites are indeed repelled by the receptor; or (ii) the region is used by the model to attenuate the positive features present elsewhere, and are, therefore, a mathematical artifact created by the limited series of molecules. In addition, other significant binding may be occurring that was not detected by HASL modeling. This 3D-QSAR paradigm involves the overlapping of molecules, notably the overlapping of similar or identical molecular components. The model is necessarily blinded to such features unless an effort is made to include molecules that contain other features in their place and that can be detected as present at the resolution of the model.

Although the high q^2 obtained for the HASL-derived models indicates a better predictivity than the CoMFA model within the database, the predictivity of the HASL models is similar to that of the CoMFA model when applied to the 24 test set compounds. As the CoMFA model residual values might predict, the leave-one-out pls analysis does not predict as well for the more highly active in the series (**197** and **201**). The q^2 -grs protocol results in a better CoMFA model that

Table 4. Predicted activities of 27 non-database compounds

Name	HASL models ^a			CoMFA models ^b		Activity log(RA)
	511 pt	61 point (rec35)	31 point (rec45)	q^2	q^2 -rgs	
ART87 ^c	-0.0333	-0.6717	-0.6899	-0.280	-0.642	-2.09
ART97 ^c	0.3136	-0.9487	-0.6899	-0.360	-0.236	-2.47
APRT98 ^d	0.3328	-0.2401	0.0123	-0.250	-0.416	-0.89
ART108	0.0874	0.1835	0.0123	-0.369	0.042	-0.12
ART109	0.1109	0.9277	0.0123	-0.213	-0.386	0.11
ART111	0.5233	0.149	0.0123	-0.892	-0.448	0.28
ART112	0.2134	-0.3585	-0.6899	-0.343	-0.454	-0.70
ART113	0.4005	0.35	0.0123	-0.236	-0.605	-0.44
ART114	0.5044	-0.5595	-0.6899	-1.044	-0.734	-0.13
ART115	0.0921	0.1835	0.0123	-0.360	0.315	0.21
ART116	0.4818	2.0257	2.155	-0.823	-0.206	0.16
ART117	-0.4121	-0.3941	0.0123	-0.385	0.423	0.41
ART119	1.0175	0.6269	0.0123	-0.145	0.493	0.19
ART134	0.091	-0.6024	0.7211	-0.683	-1.225	-0.67
ART147 ^c	-2.3565	-1.8874	-1.1135	-0.797	-1.972	1.10
ART180	0.4852	-0.3812	0.2248	0.186	-0.894	0.379
ART182	-0.6181	-0.4065	0.0123	0.229	-0.573	0.138
ART184	0.0011	0.0643	-0.7381	-0.258	-0.229	-0.80
ART190	-1.2238	0.3047	1.3236	-3.474	-2.112	-2.26
ART193	-0.9897	0.0368	0.0123	-0.165	-0.539	-0.96
ART194	0.8449	0.6145	0.0123	0.100	-0.347	-0.19
ART196	1.1437	0.6145	0.0123	-0.089	0.227	-0.640
ART197	-0.2561	0.351	0.0123	0.133	0.129	1.46
ART199	0.287	0.4934	-1.2868	-0.657	-0.311	0.353
ART200	-0.6925	0.4595	0.0123	0.307	0.266	0.017
ART201	0.1898	-0.2401	0.0123	0.057	0.268	1.47
ART202	-0.2902	0.4595	0.0123	0.049	0.202	0.11

^a Predicted activities for compounds made by HASL models. The original untrimmed model contained 511 points; the trimmed versions, rec35 and rec45, contained 61 and 31 points, respectively.

^b Predicted activities for compounds based on the CoMFA models. The results for the default region CoMFA with cross-validated (leave-one-out) pls analysis are indicated in the column q^2 , while the same predictions using the region resulting from the guided region selection protocol are displayed in the column q^2 -rgs.

^c Compound in original $n = 172$ CoMFA database; removed because of high residual value in initial model.

^d Compound in original $n = 172$ CoMFA database; reasonable residual in initial model (-0.719).

predicts 14 (of 23) compounds (61%) within 0.5 log units, 18 (78%) within 0.66 log units, and 20 (87%) within 0.8 log units.

It is apparent that in this particular trial the HASL models are not as consistent between themselves as the two CoMFA runs, although one would expect this, for as the trimming process continues, each resulting model will be slightly different, with partial activity values redistributed into the remaining points available to the model. In most cases a relatively large range exists between the HASL-511, -61 (rec35), and -31

(rec45) models. For example, the HASL predictions for a given compound can be up to and more than 1 log unit different from another HASL prediction; this is indeed the case for 11 of the entries. In contrast, only eight of the CoMFA predictions (cross-validated leave-one-out versus q^2 -rgs) differed by more than 0.5 log units (and only two by more than 0.85 log units).

A further comparison of interest entails the HASL prediction of compounds initially pruned from the CoMFA database (high residuals). In this small collection of examples (compounds **147**, **87**, and **97**) HASL

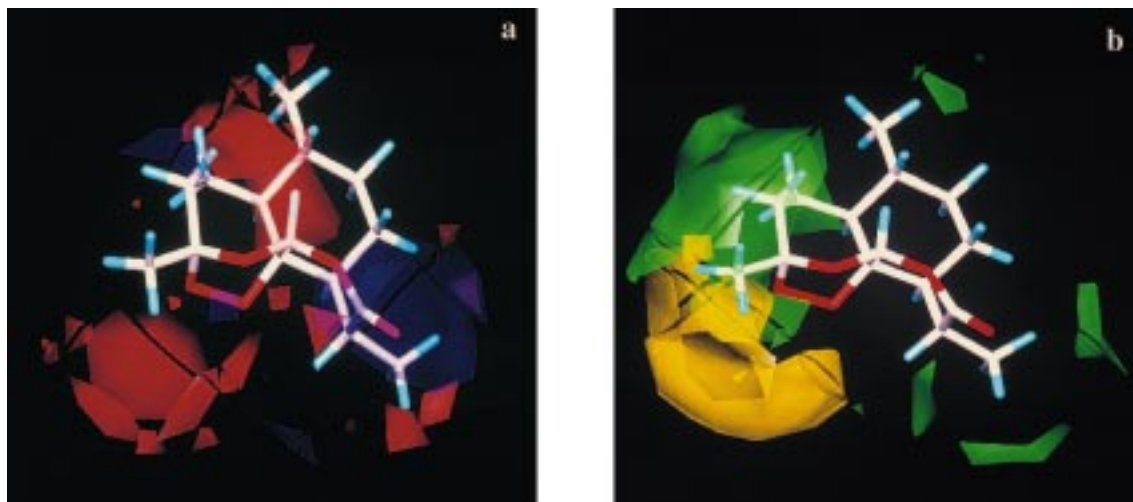


Figure 6. Artemisinin within the CoMFA contour map: (a) Electrostatic map indicating red contours around regions where high electron density (negative charge) is expected to increase activity, and blue contours represent areas where low electron density (partial positive charge) is expected to increase activity; and (b) steric map indicating areas where steric bulk is predicted to increase (green) or decrease (yellow) activity.

predictions ranged from as poor as CoMFA (147) to much better than CoMFA (97). Instances in which both models fail might indicate a different mechanism of action, a possible error in activity assay, or in fact may be simply due to the high variability associated with any modeling system attempting to correlate essentially 4D descriptors (x,y,z, and 'property') to activity using a limited dataset.

Conclusions

Both CoMFA and HASL represent viable and complementary approaches in the development of 3D QSAR for a series of molecules. The predictive and modeling capabilities of both methods were assessed by comparing their performances on the same training set of 154 artemisinin analogues and test set of 27 additional analogues. Training set correlation coefficients and cross-validation (leave-one-out) q^2 values were found to be significant and served to confirm each method's intrinsic attributes. In addition, physical models emerging from either method (CoMFA: steric/electrostatic fields; HASL: 3D pharmacophore) were found largely consistent, given that some interpretive bridging was needed to relate HASL atom types to either steric or electrostatic effects. Predictions of test set activities made by either method suffered similar limitations, since both methods rely on

the deconvolution of molecular interaction to regularly spaced sets of points.

HASL makes few assumptions in developing a 3D-QSAR model. The user can investigate potentially unique functional group/atom binding interactions by defining atom type appropriately and then assessing the resulting model's performance. In the present case, the use of three atom types appears to be sufficient to characterize the *in vitro* antimalarial activities of the artemisinin series.

An advantage of the HASL method is the availability of specific information regarding the visual output data; the specific point information related to the contour maps associated with a CoMFA/pls run is difficult to retrieve. Ease of use and accessibility (PC platform) may be considered as additional advantages of the HASL modeling system. The CoMFA methodology is advantageous in its ability to be reduced into solely its steric or electrostatic contributions toward activity.

Acknowledgements

Support of this work was funded by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and the University of Mississippi Research Institute of Pharmaceutical Sciences. The authors are indebted to Dr(s). Glen Kellogg and David Haney (eduSoft) for providing the HINT software along with technical support.

Table 5. Atoms in molecule 152 having non-zero kcal/atom estimates made by the two HASL-derived pharmacophores rec35 and rec45

Atom	HASL model	
	rec35 (61 points)	rec45 (31 points)
Positive contributions		
11	0.193	0.191
16	2.016	2.127
17	0.193	0.191
22	0.193	0.191
35	0.193	0.191
37	0.526	
44	1.008	0.813
45	0.245	1.104
48	0.193	0.191
49	0.507	0.290
52	0.203	
53	0.203	
57	0.203	
Negative contributions		
1	-0.280	-0.516
4	-0.187	-0.344
5	-0.187	-0.344
6	-0.674	-0.816
10	-0.979	
12	-0.674	-0.816
25	-0.280	-0.516
27	-0.187	-0.344
31	-0.488	-0.472
37		-0.472
38	-0.187	-0.344
43	-0.589	

Table 6. HASL model data for 31-point rec45 3D pharmacophore

*** Merged and iterated HASL ***					
31 MATRIX POINTS WITH RESOLUTION OF 2.00000 ANGS					
0.00000	2.00000	-2.00000	0.00000	-0.75768	1
2.00000	-2.00000	-2.00000	0.00000	-0.75768	1
2.00000	2.00000	0.00000	0.00000	-1.38507	1
0.00000	-4.00000	-2.00000	0.00000	0.70225	1
2.00000	2.00000	-2.00000	0.00000	-0.75768	1
-2.00000	2.00000	4.00000	0.00000	1.62080	1
-2.00000	0.00000	-2.00000	1.00000	-2.75355	1
-2.00000	2.00000	-2.00000	0.00000	1.21689	1
-4.00000	4.00000	2.00000	0.00000	0.59694	1
2.00000	-6.00000	2.00000	0.00000	0.88225	1
0.00000	-6.00000	8.00000	0.00000	1.37287	1
-2.00000	-4.00000	-4.00000	0.00000	0.21251	1
2.00000	-8.00000	-8.00000	0.00000	0.64331	1
-4.00000	2.00000	4.00000	0.00000	0.75043	1
-4.00000	2.00000	6.00000	0.00000	0.93928	1
-4.00000	4.00000	6.00000	0.00000	-0.83139	1
-4.00000	4.00000	8.00000	0.00000	0.91196	1
-4.00000	-8.00000	-4.00000	0.00000	0.21689	1
-2.00000	4.00000	6.00000	0.00000	-0.58655	1
4.00000	-8.00000	0.00000	0.00000	0.74372	1
0.00000	-10.00000	-8.00000	0.00000	-0.35664	1
-6.00000	4.00000	10.00000	0.00000	-0.86932	1
-8.00000	4.00000	6.00000	0.00000	1.16697	1
4.00000	-10.00000	-2.00000	0.00000	-0.42434	1
-2.00000	-2.00000	-4.00000	0.00000	-0.85248	1
-6.00000	-2.00000	0.00000	0.00000	-1.12590	1
-4.00000	2.00000	0.00000	0.00000	-2.40804	1
-8.00000	2.00000	2.00000	0.00000	-1.19845	1
-8.00000	0.00000	4.00000	0.00000	0.79357	1
-8.00000	2.00000	4.00000	0.00000	0.79357	1
-8.00000	0.00000	2.00000	0.00000	0.79357	1

References

- Avery, M.A., Gao, F., Chong, W.K.M., Mehrotra, S. and Milhaus, W.K., *J. Med. Chem.*, 36 (1993) 4264.
- Cramer III, R.D., Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
- Doweyko, A.M. and Mattes, W.B., *Biochemistry*, 31 (1992) 9388.
- Doweyko, A.M., *J. Med. Chem.*, 37 (1994) 1769.
- Cramer III, R.D., Bunce, J.D., Patterson, D.E. and Frank, I.E., *Quant. Struct.-Act. Relatsh.*, 7 (1988) 18.
- Doweyko, A.M., *J. Med. Chem.*, 31 (1988) 1396.
- Wiese, M., In Kubinyi, H. (Ed.) *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, 1993, p. 431.
- Doweyko, A.M., In Magee, P., Henry, D.R. and Block, J.H. (Eds.) *New Tool for the Study of Structure-Activity Relationships in Three Dimensions*, Vol. 413, American Chemical Society, Washington, DC, 1989, pp. 82.
- Doweyko, A.M., *J. Math. Chem.*, 7 (1991) 273.
- Avery, M.A., Mehrotra, S., Bonk, J.D., Vroman, J.A., Goins, K.D. and Miller, R., *J. Med. Chem.*, 39 (1996).
- Avery, M.A., Gao, F., Chong, W.K.M., Hendrickson, T.F., Inman, W.D. and Crews, P., *Tetrahedron*, 50 (1994) 957.
- Avery, M.A., Jennings-White, C. and Chong, W.K.M., *Tetrahedron Lett.*, 28 (1987) 4629.
- Avery, M.A., Bonk, J.D., Chong, W.K.M., Mehrotra, S., Miller, R., Milhaus, W., Goins, D.K., Venkatesan, S., Wyandt, C., et al., *J. Med. Chem.*, 38 (1995) 5038.
- Jung, M., Li, X., Bustos, D.A., ElSohly, H.N., McChesney, J.D. and Milhaus, W.K., *J. Med. Chem.*, 33 (1990) 1516.
- Brossi, A., Venugopalan, B., Gerpe, L.D., Yeh, H.J.C., Flippen-Anderson, J.L., Buchs, P., Luo, X.D., Milhaus, W. and Peters, W., *J. Med. Chem.*, 31 (1988) 645.
- Pu, Y.M., Torok, D.S., Ziffer, H., Pan, X.-Q. and Meshnick, S.R., *J. Med. Chem.*, 38 (1995) 4120.

17. Posner, G.H., Wang, D., González, L., Tao, X., Cummings, J.N., Klinedinst, D. and Shapiro, T.A., (1996).
18. Posner, G.H., Oh, C.H., Gerena, L. and Milhous, W.K., *J. Med. Chem.*, 35 (1992) 2459.
19. Posner, G.H., Oh, C.H., Webster, H.K., Ager Jr., A.L. and Rossan, R.N., *Am. J. Trop. Med. Hyg.*, 50 (1994) 522.
20. Lin, A.J., Lee, M. and Klayman, D.L., *J. Med. Chem.*, 32 (1989) 1249.
21. Khalifa, S.I., Baker, J.K., Rogers, R.D., El-Ferally, F.S. and Hufford, C.D., *Pharm. Res.*, 11 (1994) 990.
22. Jefford, C.W., Velarde, J.A., Bernardinelli, G., Bray, D.H., Warhurst, D.C. and Milhous, W.K., *Helv. Chim. Acta*, 76 (1993) 2775.
23. Posner, G.H., Oh, C.H., Gerena, L. and Milhous, W.K., *Heteroat. Chem.*, 6 (1995) 105.
24. Torok, D.S., Ziffer, H., Meshnick, S.R., Pan, X.-Q. and Ager, A., *J. Med. Chem.*, 38 (1995) 5045.
25. Acton, N., Karle, J.M. and Miller, R.E., *J. Med. Chem.*, 36 (1993) 2552.
26. SYBYL 6.3, Tripos Associates, St. Louis, MO, U.S.A.
27. eduSoft, Ashland, VA, U.S.A.
28. Cho, S.J. and Tropsha, A., *J. Med. Chem.*, 38 (1995) 1060.
29. Haynes, R.K. and Vonwiller, S.C., *Tetrahedron Lett.*, 37 (1996) 253.
30. Haynes, R.K. and Volwiller, S.C., *Tetrahedron Lett.*, 37 (1996) 257.
31. Jefford, C.W., Vicente, M.G.H., Jacquier, Y., Favarger, F., Mareda, J., Millasson-Schmidt, P., Brunner, G. and Burger, U., *Helv. Chim. Acta*, 79 (1996) 1475.
32. Posner, G.H., Cummings, J.N., Polypradith, P. and Oh, C.H., *J. Am. Chem. Soc.*, 117 (1995) 5885.
33. Posner, G. and Oh, C., *J. Am. Chem. Soc.*, 114 (1992) 8328.
34. Posner, G.H., Oh, C.H., Wang, D., Gerena, L., K., M.W., Meshnick, S.R. and Asawamahasadka, W., *J. Med. Chem.*, 37 (1994) 1256.
35. Kamchonwongpaisan, S. and Meshnick, S.R., *Gen. Pharmacol.*, 27 (1996) 587.
36. Hypothesis Software, Long Valley, NJ, U.S.A.
37. eduSoft, LC, Ashland, VA, U.S.A.
38. Kellogg, G.E., Semus, S.F. and Abraham, D.J., *J. Comput.-Aided Mol. Design*, 5 (1991) 545.