Journal of Computer-Aided Molecular Design, **12:** 397–409, 1998.
KLUWER/ESCOM

Comparative molecular field analysis of artemisinin derivatives: Ab initio versus semiempirical optimized structures

Somsak Tonmunphean[a], Sirirat Kokpol[a], Vudhichai Parasuk[a], Peter Wolschann[b], Rudolf H. Winger[c,*], Klaus R. Liedl[c] & Bernd M. Rode[c]

[a]*Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand;* [b]*Institute of Theoretical Chemistry and Radiation Chemistry, University of Vienna, Waehringer Strasse 17, A-1090 Wien, Austria;* [c]*Department of Theoretical Chemistry, Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 52a, A-6020 Innsbruck, Austria*

Received 13 August 1997; Accepted 22 January 1998

Key words: AM1 method, antimalarial drug, CoMFA, Hartree-Fock (HF), QSAR

Abstract

Based on the belief that structural optimization methods, producing structures more closely to the experimental ones, should give better, i.e. more relevant, steric fields and hence more predictive CoMFA models, comparative molecular field analyses of artemisinin derivatives were performed based on semiempirical AM1 and HF/3-21G optimized geometries. Using these optimized geometries, the CoMFA results derived from the HF/3-21G method are found to be usually but not drastically better than those from AM1. Additional calculations were performed to investigate the electrostatic field difference using the Gasteiger and Marsili charges, the electrostatic potential fit charges at the AM1 level, and the natural population analysis charges at the HF/3-21G level of theory. For the HF/3-21G optimized structures no difference in predictability was observed, whereas for AM1 optimized structures such differences were found. Interestingly, if ionic compounds are omitted, differences between the various HF/3-21G optimized structure models using these electrostatic fields were found.

Introduction

Malaria is one of the most widespread and prevalent endemic diseases, a risk for 40% of the world's population, which causes around 300 to 500 million illnesses and around 1.5 to 2.7 million deaths each year [1,2]. In many parts of the world, malarial parasites have developed resistance to most of the common chemotherapeutic agents, such as chloroquine, quinine, sulfa/pyrimethamine combination, and mefloquine [3–5]. This led to the introduction of artemisinin, a sesquiterpene endoperoxide compound isolated from a Chinese herb in 1972 [6]. However, its poor solubility stimulated scientists to synthesize more soluble and active derivatives such as dihydroartemisinin, artesunate, and artemether [7]. These derivatives are now being increasingly used for malaria treatment due to their low toxicity and high potency against drug-resistant strains of *Plasmodium falciparum* [8–10]. Nevertheless, there is still a need for new and more potent antimalarial drugs.

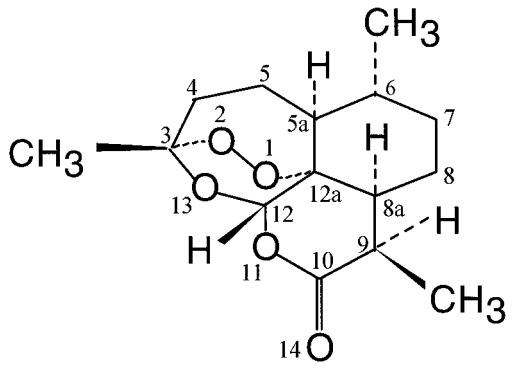


Figure 1. Stereochemistry and atomic numbering scheme of artemisinin.

*To whom correspondence should be addressed.

Table 1. Comparison of the most important structural parameters of artemisinin between X-ray structure, semiempirical, and ab initio optimized structures

Structural parameter	X-ray	CNDO	AM1	HF/STO-3G	HF/3-21G	HF/6-31G(d,p)
Bond length (Å)						
O1-O2	1.475	1.229	1.289	1.397	1.462	1.390
O1-C12a	1.450	1.410	1.468	1.467	1.477	1.430
O2-C3	1.417	1.394	1.447	1.450	1.441	1.396
O11-C12	1.455	1.401	1.421	1.442	1.428	1.408
C10-O14	1.201	1.340	1.231	1.214	1.197	1.183
Bond angle (deg)						
O1-C12a-C5a	106.4	107.4	106.0	105.9	105.1	105.6
O2-C3-C4	113.2	113.1	111.6	111.6	111.0	111.3
O11-C12-O13	105.9	105.5	101.0	105.8	108.8	107.7
O14-C10-O11	117.2	112.1	111.6	119.2	119.2	118.9
Torsional angle (deg)						
C12-O13-C3-C4	272.7	283.9	281.4	266.5	271.7	269.7
O1-C12a-C5a-C5	69.0	73.7	73.2	67.8	68.8	67.5
O2-C3-C4-C5	265.8	273.3	269.7	264.8	263.3	267.2
O11-C12-O13-C3	258.1	241.5	245.7	267.7	262.1	260.7

For drug discovery and development, quantitative structure–activity relationships (QSARs) are often employed because the bioactivity of new compounds can be predicted with strongly reduced efforts, and thus time and money can be saved [11]. Comparative Molecular Field Analysis (CoMFA), a three-dimensional quantitative structure–activity relationship (3D-QSAR) method [12], is based on the assumption that drug–receptor interactions, which are responsible for biological activities, are usually noncovalent. Therefore, the steric and electrostatic fields (noncovalent interactions) surrounding the analogues of drug molecules should correlate with their biological activities. From geometry optimization of artemisinin structures (Figure 1) with semiempirical (CNDO and AM1 methods) and ab initio molecular orbital methods (Hartree–Fock with STO-3G, 3-21G, and 6-31G(d,p) basis sets) we found that the ab initio level, using Hartree–Fock theory with the 3-21G basis set (HF/3-21G), reproduced most of the structural parameters very reliably in comparison to the X-ray structures [13] (see Table 1). This applies especially to the bond length of the endoperoxide linkage, which is held responsible for the antimalarial activity [7–9] and which is believed to be attacked by ferrous ions (Fe^{2+}) and then to be broken to form an oxy-radical in the key step of the mechanism of action [14]. The 1,2,4-trioxane compounds show that bulky groups at the C_4 position, which block the ferrous ions from attacking the peroxide linkage, significantly decrease the biological activity [15]. Therefore, structural optimization methods that give structures more closely to the experimental ones should give better, i.e. more relevant, steric fields and hence a better, i.e. more predictive, CoMFA model. This suggested the choice of the HF/3-21G level of theory for the optimization. In most of the CoMFA studies, molecules have been geometry optimized with empirical or semiempirical quantum mechanical methods [12,16–21]. For this reason, we additionally used the AM1 method to optimize the structures in order to compare these results with those obtained at the HF/3-21G level of theory.

Biological data

The structures and biological data are listed in Figure 2 and Table 2. Because the biological data arise from different sources [22–26], the relative activity, IC_{50} of artemisinin over IC_{50} of the compounds, was used to reduce inconsistencies due to individual experimental environments. The activities were measured against two strains of *Plasmodium falciparum*, i.e. the W-2 clone and the D-6 clone. The W-2 clone is chloroquine-resistant but mefloquine-sensitive, while

Table 2. Experimental biological activities and values predicted from the HF/3-21G optimized structures

No.	log(relative activity)					
	D-6 clone			W-2 clone		
	Expt.	Cal.[a]	Residual	Expt.	Cal.[a]	Residual
1	0.854	0.737	0.117	−0.019	0.364	−0.383
2	0.689	0.666	0.023	0.405	0.494	−0.089
3	0.202	0.221	−0.019	0.013	0.021	−0.008
4	0.580	0.531	0.049	0.251	0.203	0.048
5	−1.264	−1.266	0.002	−1.599	−1.599	0.000
6	−1.463	−1.353	−0.110	−1.547	−1.610	0.063
7	−1.411	−1.451	0.040	−1.202	−1.238	0.035
8	0.226	0.238	−0.012	−0.144	−0.154	0.010
9	−0.786	−0.656	−0.130	−1.086	−0.972	−0.114
10	−0.786	−0.937	0.151	−0.014	−0.101	0.087
11	−1.139	−1.121	−0.018	−0.290	−0.329	0.039
12	−1.666	−1.701	0.035	−0.960	−0.981	0.021
13	0.423	0.252	0.171	0.529	0.115	0.414
14	−0.122	−0.100	−0.022	−0.276	−0.393	0.117
15	0.375	0.411	−0.036	0.163	0.145	0.018
16	0.904	0.907	−0.003	1.034	0.926	0.108
17	0.655	0.785	−0.130	0.651	0.786	−0.134
18	0.199	−0.119	0.318	−0.053	−0.254	0.201
19	0.667	0.691	−0.024	0.677	0.679	−0.002
20	0.700	0.722	−0.022	0.849	0.888	−0.039
21	0.612	0.685	−0.073	0.645	0.677	−0.032
22	0.971	1.001	−0.030	1.319	1.240	0.079
23	0.522	0.541	−0.019	0.481	0.445	0.036
24	−0.399	−0.277	−0.122	−0.398	−0.423	0.025
25	−0.105	−0.121	0.016	−0.084	−0.037	−0.467
26	−0.094	0.078	−0.172	−0.123	−0.068	−0.055
27	0.079	0.098	−0.019	0.146	0.037	0.109
28	0.146	0.202	−0.056	0.079	0.100	−0.021
29	0.255	0.354	−0.099	0.176	0.199	−0.023
30	−0.824	−0.985	0.161	−0.854	−1.058	0.204
31	−0.347	−0.245	−0.102	−0.523	−0.230	−0.293
32	−1.097	−0.826	−0.271	−1.222	−0.852	−0.370
33	0.423	0.210	0.213	0.209	0.136	0.074
34	0.411	0.460	−0.049	0.782	0.599	0.182
35	−0.018	−0.153	0.135	−0.260	−0.296	0.036
36	−1.052	−1.062	0.010	−1.830	−1.754	−0.076
37	0.174	0.115	0.059	0.234	0.367	−0.133
38	1.302	1.191	0.111	0.807	0.968	−0.161
39	0.186	0.184	0.002	0.331	0.339	−0.008
40	0.841	0.917	−0.076	0.776	0.697	0.078

[a] Calculated from alignment 4 with O.sp^3 (−1.0) as the probe atom using the HF/3-21G optimized structures (Table 7).

the D-6 clone is mefloquine-resistant but chloroquine-sensitive.

Table 3. Atoms selected for the definition of alignment rules

Alignment	Selected atoms
1	O1-O2-C3-C4-C5-C5a-C12a-C12-O13
2	C5a-C6-C7-C8-C8a-C12a
3	O1-O2-O11
4	C12a-O1-O2-C3
5	C12a-O1-O2-C3-O13-C12-O11

Computational methods

Structural optimization

All structures were built using SYBYL 6.3 [27] and optimized at the HF/3-21G and AM1 level of theory using GAUSSIAN 94 [28]. The quantum mechanically optimized structures with their Mulliken population analysis (MPA) atomic charges were reimported into SYBYL. The charges according to Gasteiger and Marsili [29], the electrostatic potential fit (ESPFIT) charges [30] at the AM1 level, and the natural population analysis (NPA) charges [31] at the HF/3-21G level were calculated using SYBYL, MOPAC 6.0 [32], and GAUSSIAN 94, respectively.

Alignment rules

Five alignment rules were selected for this work in order to study the influence of different alignments. We also tested the alignment of Ref. 17 (denoted as alignment 3 in the following). For all alignments, the structures were adjusted using the 'Fit Atom' option in SYBYL which minimizes the root-mean-square (rms) differences between selected atoms of the structure in question and dihydroartemisinin. Dihydroartemisinin was chosen as the reference molecule due to its most common structure within the analogues (see Table 3). Data (structures used) are available upon request at the following e-mail address: Rudolf.Winger@uibk.ac.at.

CoMFA calculations

A regular 3D lattice with 2 Å spacing was created extending beyond the molecular dimensions of molecule number 22 by 4.0 Å in all directions. Two probe atoms, namely an sp^3 carbon with a charge of +1.0 and an sp^3 oxygen with a charge of −1.0, were used. Steric (Lennard-Jones 6–12 function) and electrostatic (Coulombic) interactions were calculated using the Tripos force field [33] with a distance-dependent dielectric constant. The cutoff was set to 30 kcal/mol. An

Table 4. CoMFA results for the 32 dihydroartemisinin analogues optimized with AM1

Activity	Alignment	AutoCoMFA			O.sp^3 (−1.0)			C.sp^3 (+1.0)		
		q^2	noc[a]	r^2	q^2	noc	r^2	q^2	noc	r^2
D-6	1	0.675	3	0.932	0.657	3	0.921	0.678	3	0.929
	2	0.665	3	0.917	0.667	4	0.947	0.627	3	0.899
	3	0.648	5	0.964	0.638	5	0.978	0.653	5	0.977
	4	0.619	4	0.940	0.609	5	0.977	0.652	5	0.976
	5	0.624	3	0.913	0.709[b]	5	0.977	0.604	3	0.899
W-2	1	0.580	2	0.768	0.599	2	0.786	0.580	2	0.774
	2	0.587	2	0.762	0.583	2	0.773	0.594	2	0.775
	3	0.586	2	0.755	0.567	2	0.764	0.609[b]	5	0.960
	4	0.576	2	0.750	0.563	2	0.762	0.564	2	0.754
	5	0.598	2	0.777	0.586	2	0.772	0.581	2	0.764

[a]Number of components.
[b]The best model.

Table 5. CoMFA results for the 32 dihydroartemisinin analogues optimized with HF/3-21G

Activity	Alignment	AutoCoMFA			O.sp^3 (−1.0)			C.sp^3 (+1.0)		
		q^2	noc[a]	r^2	q^2	noc	r^2	q^2	noc	r^2
D-6	1	0.698	4	0.926	0.703	3	0.932	0.679	3	0.932
	2	0.700	3	0.926	0.683	3	0.926	0.679	3	0.930
	3	0.647	3	0.923	0.706	3	0.933	0.692	3	0.932
	4	0.649	3	0.921	0.730[b]	3	0.940	0.681	3	0.934
	5	0.699	4	0.965	0.711	3	0.935	0.683	3	0.931
W-2	1	0.606	2	0.812	0.646	2	0.836	0.623	2	0.825
	2	0.585	2	0.803	0.654	2	0.840	0.632	2	0.824
	3	0.591	2	0.808	0.648	2	0.838	0.630	2	0.825
	4	0.599	2	0.809	0.686[b]	3	0.927	0.637	2	0.831
	5	0.607	2	0.810	0.649	2	0.837	0.633	2	0.827

[a,b]As in Table 4

Table 6. CoMFA results for the 40 artemisinin analogues optimized with AM1

Activity	Alignment	AutoCoMFA			O.sp^3 (−1.0)			C.sp^3 (+1.0)		
		q^2	noc[a]	r^2	q^2	noc	r^2	q^2	noc	r^2
D-6	1	0.663[b]	4	0.948	0.646	4	0.944	0.661	4	0.946
	2	0.656	4	0.928	0.626	4	0.932	0.631	4	0.929
	3	0.573	4	0.922	0.559	4	0.933	0.608	5	0.966
	4	0.577	4	0.923	0.545	4	0.935	0.598	5	0.962
	5	0.563	4	0.925	0.592	4	0.926	0.576	4	0.918
W-2	1	0.480	4	0.908	0.515	4	0.920	0.502	4	0.921
	2	0.548[b]	4	0.912	0.487	4	0.911	0.497	4	0.907
	3	0.490	4	0.899	0.377	4	0.897	0.468	5	0.940
	4	0.477	4	0.902	0.381	4	0.901	0.403	4	0.888
	5	0.493	4	0.909	0.465	4	0.906	0.477	4	0.907

[a,b]As in Table 4

Table 7. CoMFA results for the 40 artemisinin analogues optimized with HF/3-21G

Activity	Alignment	AutoCoMFA			O.sp^3 (−1.0)			C.sp^3 (+1.0)		
		q^2	noc[a]	r^2	q^2	noc	r^2	q^2	noc	r^2
D-6	1	0.610	4	0.953	0.708	5	0.977	0.662	5	0.978
	2	0.640	3	0.925	0.679	5	0.975	0.630	3	0.922
	3	0.569	4	0.944	0.709	5	0.979	0.668	5	0.976
	4	0.592	4	0.954	0.723[b]	5	0.979	0.673	5	0.980
	5	0.602	4	0.946	0.714	5	0.980	0.658	5	0.976
W-2	1	0.528	6	0.972	0.578	5	0.963	0.434	3	0.889
	2	0.494	4	0.935	0.551	5	0.970	0.458	3	0.884
	3	0.436	4	0.921	0.583[b]	5	0.961	0.479	4	0.930
	4	0.476	4	0.931	0.581	5	0.963	0.515	5	0.960
	5	0.556	6	0.970	0.572	5	0.961	0.485	4	0.934

[a,b]As in Table 4

Table 8. CoMFA results for the 32 dihydroartemisinin analogues optimized with AM1 and adjusted to be in the same orientation as those resulting from HF/3-21G

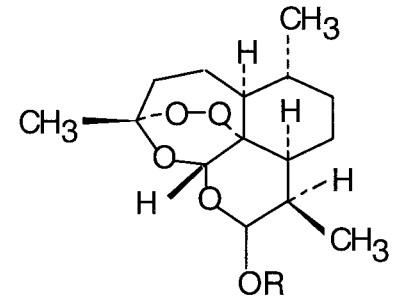
Activity	Alignment	AutoCoMFA			O.sp^3 (−1.0)			C.sp^3 (+1.0)		
		q^2	noc[a]	r^2	q^2	noc	r^2	q^2	noc	r^2
D-6	1	0.637	4	0.944	0.718[b]	5	0.980	0.639	4	0.946
	2	0.696	5	0.974	0.702	5	0.976	0.640	3	0.904
	3	0.631	4	0.945	0.713	5	0.980	0.608	3	0.905
	4	0.650	4	0.945	0.710	5	0.979	0.642	4	0.947
	5	0.682	4	0.954	0.707	5	0.980	0.610	3	0.907
W-2	1	0.587	2	0.763	0.605	2	0.780	0.600	2	0.775
	2	0.559	2	0.761	0.613[b]	2	0.791	0.608	2	0.777
	3	0.572	2	0.769	0.598	2	0.782	0.602	2	0.776
	4	0.592	2	0.766	0.609	2	0.785	0.599	2	0.775
	5	0.580	2	0.773	0.600	2	0.780	0.601	2	0.776

[a,b]As in Table 4

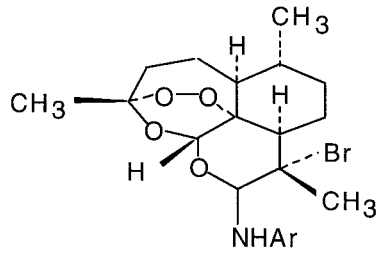
Table 9. CoMFA results for the 40 artemisinin analogues optimized with AM1 and adjusted to be in the same orientation as those resulting from HF/3-21G

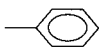
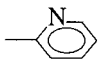
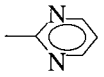
Activity	Alignment	AutoCoMFA			O.sp^3 (−1.0)			C.sp^3 (+1.0)		
		q^2	noc[a]	r^2	q^2	noc	r^2	q^2	noc	r^2
D-6	1	0.570	4	0.925	0.596	4	0.929	0.593	4	0.926
	2	0.566	5	0.953	0.583	6	0.974	0.583	5	0.941
	3	0.497	4	0.892	0.535	4	0.900	0.529	4	0.902
	4	0.567	4	0.925	0.589	4	0.929	0.642[b]	4	0.947
	5	0.585	4	0.929	0.613	4	0.929	0.618	4	0.932
W-2	1	0.497	4	0.911	0.509	4	0.909	0.491	4	0.906
	2	0.376	4	0.866	0.445	4	0.853	0.442	4	0.849
	3	0.341	4	0.872	0.397	4	0.876	0.382	4	0.877
	4	0.472	4	0.908	0.506	4	0.909	0.488	4	0.906
	5	0.490	4	0.910	0.521[b]	4	0.909	0.505	4	0.910

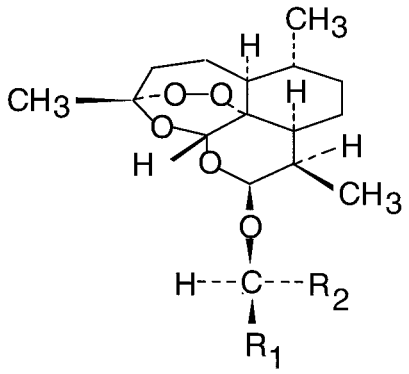
[a,b]As in Table 4



1 R = H
2 R = $CH_2COOCH_2CH_3$
3 R = $(CH_2)_2COOCH_3$
4 R = $CH_2C_6H_4COOCH_3$
5 R = CH_2COO^-
6 R = $(CH_2)_2COO^-$
7 R = $(CH_2)_3COO^-$
8 R = $H_2C_6H_4COO^-$
9 R = $(CH_2)_3COOH$
33 R = $Si(CH_3)_3$
13 R = CH_3



10 Ar = 
11 Ar = 
12 Ar = 



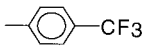
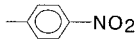
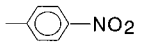
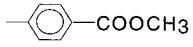
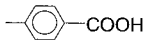
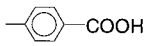
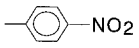
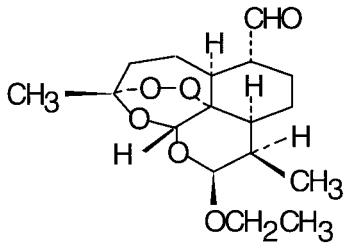
	R_1	R_2
14	$CH_2CH_2CH_3$	$CH_2C_6H_5$
15	$CH_2C_6H_5$	$CH_2CH_2CH_3$
16	C_6H_5	$COOCH_2CH_3$
17	$COOCH_2CH_3$	C_6H_5
18	CH_3	
19	$CH_2COOCH_2CH_3$	C_6H_5
20	C_6H_5	$CH_2COOCH_2CH_3$
21	$CH_2COOCH_2CH_3$	
22		$CH_2COOCH_2CH_3$
23		CH_3
24	CH_3	
25		CH_3
26	CH_2COOH	

Figure 2. Structures of artemisinin derivates used for the CoMFA studies.

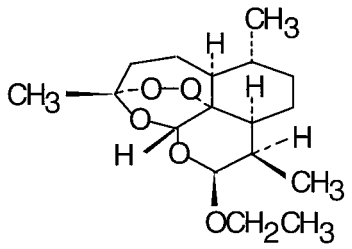
additional 'AutoCoMFA' column was created by using the automatically created region, which extends the region beyond the molecular dimension of the largest compound by 4.0 Å in all directions and uses C.sp^3 with a charge of +1.0 as the probe atom, in order to see the influence of grid box position and dimension.

Partial least-squares regression analysis

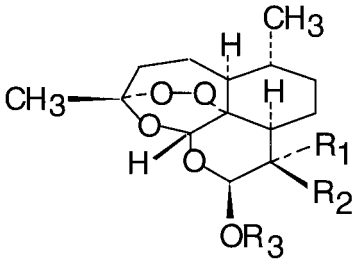
All models were investigated using the full cross-validated partial least-squares (PLS) method [34,35] (leave-one-out) with CoMFA standard options for scaling of variables. Minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation is below this threshold. To avoid an excessive number of components, the optimal number



27



28

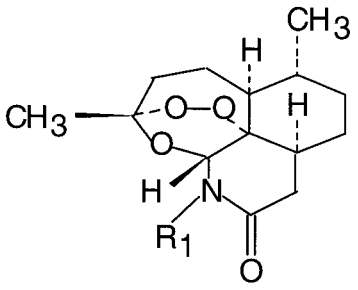


29 $R_1 = OH$, $R_2 = CH_3$, $R_3 = CH_2CF_3$

30 $R_1 = CH_3$, $R_2 = OH$, $R_3 = CH_2CF_3$

31 $R_1 = OH$, $R_2 = CH_3$, $R_3 = CH_2CH_3$

32 $R_1 = CH_3$, $R_2 = OH$, $R_3 = CH_2CH_3$

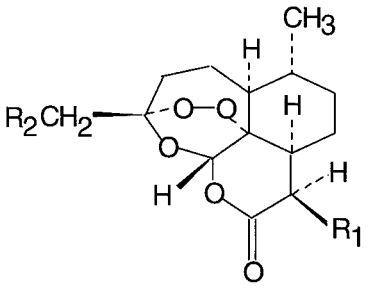


34 $R = CH_3$

35 $R = n\text{-}C_5H_{11}$

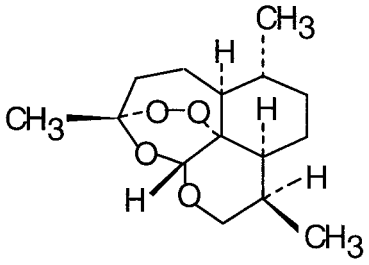
36 $R = (CH_2)_5COOH$

37 $R = CH_2C_6H_5$



38 $R_1 = H$, $R_2 = CH_2CH_3$

39 $R_1 = (CH_2)_3CH_3$, $R_2 = CH_3$



40

Figure 2. Continued.

of components (noc) was selected as the one which results in an increase of the cross-validated r^2 (also called r^2_{cv} or q^2), which indicates the predictive ability of the CoMFA model, of more than 5% [36] compared to the model with fewer components. Subsequently, it was used for the PLS without cross-validation and with scaling of variables.

Two sets of compounds were chosen: the dihydroartemisinin analogues (compounds **1**–**32** that have an -OR or -NR group at the C10 position) and the artemisinin analogues (compounds **1**–**40** that have either an -OR or -NR or =O group at the C10 position).

Results and discussion

The CoMFA results obtained using O.sp^3 (−1.0) and C.sp^3 (+1.0) as probe atoms and the five alignment rules for the dihydroartemisinin analogues (32 compounds) optimized at the AM1 and HF/3-21G level of theory are listed in Tables 4 and 5, respectively.

Table 10. CoMFA results for the 32 dihydroartemisinin and 40 artemisinin analogues optimized with AM1 using O.sp^3 with charge −1.0 as the probe atom and alignment 4 and translated grid position along the x, y, and z axes showing the q^2 value and the optimum number of components in parentheses

Activity	Set[a]	x-axis				y-axis				z-axis			
		+0.5	−0.5	+1.0	−1.0	+0.5	−0.5	+1.0	−1.0	+0.5	−0.5	+1.0	−1.0
D-6	1	0.585(3)	0.610(5)	0.670(4)	0.674(5)	0.531(5)	0.687(5)	0.458(2)	0.458(2)	0.676(4)	0.607(3)	0.618(5)	0.618(5)
	2	0.579(4)	0.557(4)	0.658(4)	0.645(4)	0.462(4)	0.592(4)	0.421(4)	0.421(4)	0.641(4)	0.603(4)	0.603(5)	0.603(5)
W-2	1	0.614(2)	0.592(2)	0.594(2)	0.580(2)	0.594(2)	0.630(4)	0.570(2)	0.570(2)	0.601(2)	0.627(3)	0.660(5)	0.660(5)
	2	0.536(4)	0.497(4)	0.585(4)	0.577(4)	0.399(4)	0.504(4)	0.501(6)	0.501(6)	0.545(5)	0.521(4)	0.611(5)	0.611(5)

[a] 1: dihydroartemisinin analogues (32 compounds);
2: artemisinin analogues (40 compounds).

Table 11. CoMFA results for the 32 dihydroartemisinin and 40 artemisinin analogues optimized with HF/3-21G using O.sp^3 with charge −1.0 as the probe atom and alignment 4 and translated grid position along the x, y, and z axes showing the q^2 value and the optimum number of components in parentheses

Activity	Set[a]	x-axis				y-axis				z-axis			
		+0.5	−0.5	+1.0	−1.0	+0.5	−0.5	+1.0	−1.0	+0.5	−0.5	+1.0	−1.0
D-6	1	0.723(3)	0.682(2)	0.714(3)	0.721(3)	0.674(3)	0.700(3)	0.610(3)	0.610(3)	0.706(3)	0.625(3)	0.655(3)	0.655(3)
	2	0.655(3)	0.628(5)	0.646(4)	0.656(4)	0.554(3)	0.677(5)	0.521(4)	0.524(4)	0.628(3)	0.612(5)	0.632(5)	0.632(5)
W-2	1	0.612(2)	0.654(2)	0.610(2)	0.616(2)	0.616(2)	0.659(3)	0.584(2)	0.584(2)	0.670(3)	0.613(2)	0.647(3)	0.647(3)
	2	0.533(5)	0.389(2)	0.477(4)	0.460(3)	0.396(3)	0.598(7)	0.418(4)	0.424(4)	0.526(3)	0.517(5)	0.494(5)	0.494(5)

[a] 1: dihydroartemisinin analogues (32 compounds);
2: artemisinin analogues (40 compounds).

With AM1, there are slight differences in q^2 when comparing the five alignment rules. The reason may be that the backbone structures in all compounds are nearly the same. Alignment 3, which is the one from Ref. 17, also gave acceptable results, although all the structures in Ref. 17 have =O at position C10 while our compounds are dihydroartemisinin analogues with -OR at C10. With HF/3-21G, there is also a slight difference in q^2 when comparing the five alignment rules. Comparing AM1 and HF/3-21G optimized structures, some significant differences can be observed: for example, for D-6 activity using alignment 4 with O.sp^3 as the probe atom, HF/3-21G gives $q^2 = 0.730$ (noc = 3) compared to 0.609 (noc = 5) from AM1, and for W-2 activity using the same alignment and probe atom, HF/3-21G gives $q^2 = 0.686$ (noc = 3) compared to 0.563 (noc = 2) resulting from AM1. However, the comparison of the highest q^2 resulting from AM1 for D-6 (0.709, noc = 5) and W-2 (0.609, noc = 5) and from HF/3-21G for D-6 (0.730, noc = 3) and W-2 (0.686, noc = 3) shows that HF/3-21G yielded only slightly better results.

The CoMFA results obtained using O.sp^3 (−1.0) and C.sp^3 (+1.0) as probe atoms and the five alignment rules for the artemisinin analogues (40 compounds) optimized at the AM1 and HF/3-21G level of theory are listed in Tables 6 and 7, respectively. Comparing the five alignment rules for both AM1 and HF/3-21G results, larger differences in q^2 are observed than for the dihydroartemisinin analogues. For W-2 activity using AM1 and O.sp^3 as the probe atom, alignment 3 gives $q^2 = 0.377$ (noc = 4) compared to 0.515 (noc = 4) of alignment 1 and for D-6 activity using AutoCoMFA, alignment 5 gives $q^2 = 0.563$ (noc = 4) compared to 0.663 (noc = 4) of alignment 1. The reason may be that the geometry of the lactone ring for the artemisinin analogues, with =O at C10, and for the dihydroartemisinin analogues, with -OR at that position, is quite different. As in the case of the dihydroartemisinin analogues, HF/3-21G optimizations usually give better results than AM1 optimizations. For D-6 activity, using alignment 4 with O.sp^3 as the probe atom, HF/3-21G gives q^2 = 0.723 (noc = 5) in comparison to 0.545 (noc = 4) for AM1, and for W-2 activity, using the same alignment and probe atom, HF/3-21G gives $q^2 = 0.581$ (noc = 5) in comparison to 0.381 (noc = 4) of AM1. Comparing the highest q^2 resulting from HF/3-21G and AM1, the difference

Table 12. Predicted and measured activities of five additional dihydroartemisinin analogues from HF/3-21G (MPA, NPA, AM1, and Gasteiger charges) and AM1 optimized structures using the best corresponding alignment rule for each model

Activity	Set[a]	Steric Field	Electrostatic Field	Compound no. **41**	**42**	**43**	**44**	**45**
D-6	Measured	–	–	0.0739	0.3130	−1.6162	−0.6499	0.0584
	1	HF/3-21G	HF/3-21G MPA	0.8850	0.4107	−0.5078	−0.2603	−0.0377
		HF/3-21G	HF/3-21G NPA	0.9136	0.4190	−0.4264	−0.2048	−0.0302
		HF/3-21G	AM1	0.8091	0.5036	−0.3964	−0.2149	0.0379
		HF/3-21G	Gasteiger	0.6944	0.4686	−0.3914	−0.2006	0.1019
		AM1	AM1	1.1251	0.9170	−0.2673	−0.0794	0.5192
	2	HF/3-21G	HF/3-21G MPA	1.0587	0.3936	−0.6243	−0.4481	−0.0733
		HF/3-21G	HF/3-21G NPA	1.0413	0.3423	−0.5477	−0.3925	−0.0650
		HF/3-21G	AM1	0.9060	0.4347	−0.5043	−0.4411	−0.0231
		HF/3-21G	Gasteiger	1.0840	0.6225	−0.3795	−0.2970	0.1416
		AM1	AM1	1.1563	0.7775	−0.3223	−0.2719	0.4504
W-2	Measured	–	–	0.5538	0.3411	−1.6999	−1.7189	−0.0813
	1	HF/3-21G	HF/3-21G MPA	0.8500	0.5505	−0.4057	−0.0978	0.2128
		HF/3-21G	HF/3-21G NPA	0.8180	0.3954	−0.1821	0.0947	0.0817
		HF/3-21G	AM1	0.7807	0.5820	−0.2582	−0.0117	0.1816
		HF/3-21G	Gasteiger	0.7232	0.6093	−0.0609	0.1257	0.2119
		AM1	AM1	0.8700	0.5964	0.0415	0.1199	0.2538
	2	HF/3-21G	HF/3-21G MPA	1.1120	0.5649	−0.6100	−0.3382	0.3549
		HF/3-21G	HF/3-21G NPA	1.1489	0.5582	−0.5607	−0.2716	0.2524
		HF/3-21G	AM1	1.0705	0.7389	−0.4692	−0.3184	0.2250
		HF/3-21G	Gasteiger	1.1359	0.7177	−0.3815	−0.2012	0.2633
		AM1	AM1	1.0071	0.8233	−0.2255	−0.1491	0.3351

[a] As in Table 10.

between D-6 activity, 0.723 (HF, noc = 5) compared to 0.663 (AM1, noc = 4), and W-2 activity, 0.583 (HF, noc = 5) compared to 0.548 (AM1, noc = 4), is nearly the same. The comparison between dihydroartemisinin analogues and artemisinin analogues reveals that the latter give slightly worse results.

It is well known that the orientation of the compounds within the predefined grid significantly affects the CoMFA results [12,37]. Therefore, we adjusted the AM1 structures to be in the same orientation as those resulting from HF/3-21G and recalculated the CoMFA. The CoMFA results using O.sp^3 (−1.0) and C.sp^3 (+1.0) as probe atoms and the five alignment rules of dihydroartemisinin and artemisinin analogues which were optimized at the AM1 level of theory and then oriented in the same way as the HF/3-21G optimized structures are shown in Tables 8 and 9, respectively. Interestingly, some results for the dihydroartemisinin analogues are remarkably improved, especially those using O.sp^3 as the probe atom. For D-6 activity using O.sp^3 as the probe atom in alignment 4, q^2 rose from 0.609 (noc = 5) to 0.710 (noc = 5) and for W-2 activity with the same probe atom and alignment q^2 rose from 0.563 (noc = 2) to 0.609 (noc = 2). In contrast to the dihydroartemisinin analogues, the results for artemisinin analogues are often worse than the original ones. For W-2 activity with AutoCoMFA in alignment 2, q^2 decreased from 0.548 (noc = 4) to 0.376 (noc=4); on the other hand in some models the results were improved, e.g. for W-2 activity using alignment 4 with O.sp^3 as the probe atom, where q^2 rose from 0.381 (noc = 4) to 0.506 (noc = 4).

We also translated the grid in the AM1 models by ±0.5 and ±1.0 Å in the x, y, and z directions. The CoMFA results using O.sp^3 as the probe atom and alignment 4 of both the dihydroartemisinin and artemisinin analogues, which were optimized with AM1 and HF/3-21G, are shown in Tables 10 and 11,

Table 13. Comparison of different electrostatic fields using HF/3-21G and AM1 optimized structures with alignment 4 of 40 artemisinin analogues showing the q^2 value, the optimum number of components in parentheses and the steric contribution

Optimization method			HF/3-21G				AM1	
Charges method			HF/MPA	HF/NPA	AM1/MPA	GM[a]	AM1/MPA	GM[a]
Activity	Type	Field						
D-6	Auto	Both	0.592 (4)	0.577 (4)	0.586 (4)	0.645 (5)	0.567 (4)	0.255 (4)
		Steric	0.495 (4)	0.495 (4)	0.495 (4)	0.495 (4)	0.239 (3)	0.239 (3)
		Electrostatic	0.596 (4)	0.579 (4)	0.547 (4)	0.417 (4)	0.604 (2)	0.060 (1)
		sc[b]	0.604	0.597	0.602	0.621	0.538	0.605
	O.sp^3	Both	0.723 (5)	0.718 (5)	0.725 (5)	0.702 (5)	0.589 (4)	0.345 (5)
		Steric	0.509 (2)	0.509 (2)	0.509 (2)	0.509 (2)	0.302 (3)	0.302 (3)
		Electrostatic	0.236 (4)	0.263 (4)	0.566 (4)	0.568 (6)	0.613 (3)	0.072 (1)
		sc[b]	0.562	0.565	0.568	0.603	0.529	0.591
	C.sp^3	Both	0.673 (5)	0.619 (4)	0.659 (5)	0.644 (5)	0.590 (4)	0.344 (5)
		Steric	0.452 (2)	0.452 (2)	0.452 (2)	0.452 (2)	0.352 (4)	0.352 (4)
		Electrostatic	0.236 (4)	0.263 (4)	0.566 (4)	0.568 (6)	0.613 (3)	0.072 (1)
		sc[b]	0.579	0.582	0.574	0.596	0.518	0.610
W-2	Auto	Both	0.476 (4)	0.503 (7)	0.463 (6)	0.443 (5)	0.427 (4)	0.088 (4)
		Steric	0.314 (4)	0.314 (4)	0.314 (4)	0.314 (4)	0.057 (1)	0.057 (1)
		Electrostatic	0.340 (2)	0.352 (2)	0.354 (1)	0.398 (5)	0.480 (2)	−0.030 (4)
		sc[b]	0.561	0.540	0.566	0.551	0.491	0.575
	O.sp^3	Both	0.581 (5)	0.550 (5)	0.521 (4)	0.487 (5)	0.506 (4)	0.224 (5)
		Steric	0.290 (4)	0.290 (4)	0.290 (4)	0.290 (4)	0.092 (2)	0.092 (2)
		Electrostatic	0.347 (1)	0.328 (1)	0.416 (2)	0.433 (4)	0.477 (2)	0.064 (4)
		sc[b]	0.520	0.515	0.523	0.539	0.500	0.573
	C.sp^3	Both	0.515 (5)	0.512 (6)	0.517 (6)	0.453 (5)	0.488 (4)	0.216 (5)
		Steric	0.283 (4)	0.283 (4)	0.283 (4)	0.283 (4)	0.088 (1)	0.088 (1)
		Electrostatic	0.347 (1)	0.328 (1)	0.416 (2)	0.433 (4)	0.477 (2)	0.064 (4)
		sc[b]	0.536	0.536	0.525	0.544	0.498	0.575

[a] Gasteiger and Marsili charges.
[b] Steric contribution.

respectively. Although the results show a strong influence of the CoMFA study on the grid position, HF/3-21G usually gives better results than AM1.

To check whether the addition of a smaller gridded box in a specific region significantly improves the results, we performed a CoMFA calculation of both HF/3-21G and AM1 optimized structures with an O.sp^3 probe atom using an additional small grid of dimensions $10 \times 9 \times 4$ Å and a grid spacing of 0.6 Å. Although there is slight improvement of the HF model, the enormous cost does not justify the use of this smaller grid.

To test the validity of the CoMFA models, we predicted the bioactivity of five additional compounds [38] using the best alignment rules of the dihydroartemisinin and artemisinin analogues. The structures, logarithmic relative biological activities and results are shown in Figure 3 and Table 12, and they confirm that HF/3-21G models give better results than those based on AM1.

To study the influence of atomic charges, we performed CoMFA studies using either only the steric or the electrostatic field interaction with alignment 4 (data not shown). Interestingly, the AM1 method using only electrostatic field interaction gives slightly better results than the use of both fields. This is in agreement with some previous reports [21]. The HF/3-21G method using both fields gives, however, better results than the use of only one field.

In order to investigate further which field is responsible for the difference between the HF/3-21G and AM1 results, we performed single point calculations to obtain AM1 charges using HF/3-21G optimized structures and then did the CoMFA using alignment 4

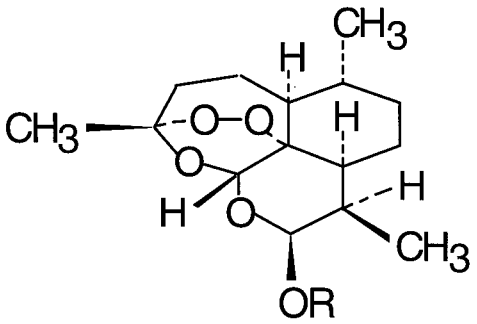
Table 14. Comparison of different electrostatic fields using HF/3-21G optimized structures with alignment 4 of 36 artemisinin analogues showing the q^2 value, the optimum number of components in parentheses and the steric contribution

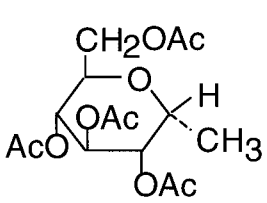
Activity	Type	Field	HF/MPA	HF/NPA	AM1/MPA	AM1/ESPFIT	GM[a]
D-6	Auto	Both	0.580 (5)	0.602 (5)	0.509 (5)	0.489 (5)	0.594 (4)
		Steric	0.609 (4)	0.609 (4)	0.609 (4)	0.609 (4)	0.609 (4)
		Electrostatic	0.633 (3)	0.631 (3)	0.516 (3)	0.414 (2)	0.439 (4)
		sc[b]	0.562	0.550	0.510	0.516	0.641
	O.sp^3	Both	0.659 (5)	0.674 (5)	0.623 (5)	0.589 (5)	0.710 (4)
		Steric	0.685 (4)	0.685 (4)	0.685 (4)	0.685 (4)	0.685 (4)
		Electrostatic	0.339 (5)	0.345 (5)	0.493 (3)	0.557 (5)	0.644 (4)
		sc[b]	0.545	0.553	0.516	0.528	0.612
	C.sp^3	Both	0.663 (5)	0.670 (5)	0.609 (5)	0.531 (5)	0.669 (4)
		Steric	0.616 (3)	0.616 (3)	0.616 (3)	0.616 (3)	0.616 (3)
		Electrostatic	0.339 (5)	0.345 (5)	0.493 (3)	0.557 (5)	0.644 (4)
		sc[b]	0.545	0.573	0.503	0.516	0.618
W-2	Auto	Both	0.514 (5)	0.550 (5)	0.428 (6)	0.477 (6)	0.631 (5)
		Steric	0.408 (4)	0.408 (4)	0.408 (4)	0.408 (4)	0.408 (4)
		Electrostatic	0.544 (5)	0.560 (5)	0.400 (5)	0.254 (2)	0.436 (4)
		sc[b]	0.533	0.527	0.510	0.515	0.600
	O.sp^3	Both	0.546 (5)	0.559 (5)	0.524 (5)	0.521 (5)	0.629 (4)
		Steric	0.465 (4)	0.465 (4)	0.465 (4)	0.465 (4)	0.465 (4)
		Electrostatic	0.320 (3)	0.313 (3)	0.397 (3)	0.374 (3)	0.557 (4)
		sc[b]	0.514	0.513	0.481	0.501	0.568
	C.sp^3	Both	0.530 (5)	0.563 (5)	0.493 (5)	0.498 (5)	0.646 (5)
		Steric	0.456 (4)	0.456 (4)	0.456 (4)	0.456 (4)	0.456 (4)
		Electrostatic	0.320 (3)	0.313 (3)	0.397 (3)	0.374 (3)	0.557 (4)
		sc[b]	0.509	0.536	0.482	0.500	0.588

[a,b]As in Table 13.

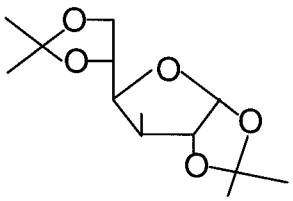
with the artemisinin analogues. The results are shown in Table 13. We found no significant difference between electrostatic fields built with different charges, as long as the HF/3-21G optimized structures were maintained. We also compared the AM1 charge model with HF/3-21G and AM1 optimized structures and found that almost all the models from HF/3-21G optimized structures give better results than the AM1 models.

In a previous study [39], a strong influence of different electrostatic field descriptors on CoMFA results was found. We, therefore, additionally used NPA charges at the HF/3-21G level. CoMFA was performed using alignment 4, and the results are shown in Table 13. For the HF/3-21G optimized structures, for the D-6 clone there is no significant difference among the electrostatic fields, whereas for the W-2 clone the Gasteiger and Marsili charge model is substantially worse than the others. Using O.sp^3 as the probe atom with both fields, the HF/3-21G/MPA charge model gives $q^2 = 0.581$ (noc = 5) compared to 0.487 (noc = 5) from the Gasteiger and Marsili charge model. Using the AM1 optimized structures, the Gasteiger and Marsili charges give much worse results than AM1/MPA. The AM1/MPA charge models with HF/3-21G optimized structures using both fields give better results than the use of only one field, while with AM1 optimized structures the use of only the electrostatic field gives better results. Among all methodical variations, HF/3-21G optimized structures give the best results in almost every model. We also calculated the ESPFIT charges at the AM1 level of the HF/3-21G optimized structures without compounds **7**–**10** in order to avoid problems arising from anionic compounds and then performed the CoMFA calculations using alignment 4. The results are shown in Table 14. Interestingly, different electrostatic fields significantly affect the CoMFA results. The Gasteiger and Marsili charge model yields substantially better results than the others, and the AM1/ESPFIT charge model usually





41 R =



42 R =

43 R = D-glucose

44 R = D-galactose

45 R = 5,6-isopropylidene-D-glucose

Figure 3. Structures of the five compounds in the test set.

yields worse results than the others: e.g. for D-6 activity with C.sp^3 as the probe atom and using both fields, the AM1/ESPFIT charge model gives $q^2 = 0.531$ (noc = 5) compared to 0.663 (noc = 5) resulting from HF/3-21G/MPA.

Conclusions

In CoMFA studies, there are many adjustable parameters which sometimes significantly affect the results [40]. Therefore, a systematic variation of some parameters can significantly increase their quality. In this study, we have examined the influence of the steric field, the electrostatic field, and the orientation of the molecules. For the steric field, HF/3-21G geometry optimization usually gives better results than AM1, albeit the computational demand is much higher. For the electrostatic field, different charge calculation methods sometimes significantly affect the CoMFA results, depending both on the level of theory used for optimization and on the data set.

Acknowledgements

The authors would like to thank the Computer Center of the University of Vienna, Austria for providing computing resources at their workstation clusters. V.P. would like to thank Prof. C.J. Cramer, Department of Chemistry, University of Minnesota, for allowing some calculations to be carried out at his facilities. S.T. gratefully acknowledges a scholarship for his visit to the University of Innsbruck from the Austrian Federal Ministry for Foreign Affairs.

References

1. World Health Organization, Bull. WHO, 71 (1993) 281.
2. Butter, D., Maurice, J. and O'Brien, C., Nature, 386 (1997) 535.
3. Olliaro, P.L. and Trigg, P.I., Bull. WHO, 73 (1995) 565.
4. World Health Organization, WHO Drug Inf., 7 (1993) 7.
5. Mockenhaupt, F.P., Parasitol. Today, 11 (1995) 248.
6. Qinghaosu Antimalaria Coordinating Research Group, Chin. Med. J., 92 (1979) 811.
7. China Cooperative Research Group on Qinghaosu and its Derivatives as Antimalarials, J. Tradit. Chin. Med., 2 (1982) 9.
8. Klayman, D.L., Science, 228 (1985) 1049.
9. Luo, X.D. and Shen, C.-C., Med. Res. Rev., 7 (1987) 29.
10. Meshnick, S.R., Taylor, T.E. and Kamchonwongpaisan, S., Microbiol. Rev., 60 (1996) 301.

11. Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
12. Cramer, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
13. Leban, I. and Golic, L., Acta Pharm. Jugosl., 38 (1988) 71.
14. Posner, G.H. and Oh, C.H., J. Am. Chem. Soc., 114 (1992) 8328.
15. Posner, G.H., Wang, D., Cumming, J.N., Oh, C.H., French, A.N., Bodley, A.L. and Shapiro, T.A., J. Med. Chem., 38 (1995) 2273.
16. Jiang, H.-L., Chen, K.-X., Wang, H.-W., Tang, Y., Chen, J.-Z. and Ji, R.-Y., Acta Pharmacol. Sin., 15 (1994) 481.
17. Avery, M.A., Gao, F., Chong, W.K.M., Mehrotra, S. and Milhous, W.K., J. Med. Chem., 36 (1993) 4264.
18. Kim, K.H., In Dean, P.M. (Ed.) Molecular Similarity in Drug Design, London, 1995, pp. 291–331.
19. Kroemer, R.T. and Hecht, P., J. Comput.-Aided Mol. Design, 9 (1995) 396.
20. Horwitz, J.P., Massova, I., Wiese, T.E., Besler, B.H. and Corbett, T.H., J. Med. Chem., 37 (1994) 781.
21. Hannongbua, S., Lawtrakul, L., Sotriffer, C.A. and Rode, B.M., Quant. Struct.–Act. Relatsh., 15 (1996) 1.
22. Lin, A.J., Klayman, D.L. and Milhous, W.K., J. Med. Chem., 30 (1987) 2147.
23. Lin, A.J., Li, L.Q., Klayman, D.L., George, C.F. and Anderson, J.L.F., J. Med. Chem., 33 (1990) 2610.
24. Lin, A.J. and Miller, R.E., J. Med. Chem., 38 (1995) 764.
25. Pu, Y.M., Torok, D.S., Ziffer, H., Pan, X.-Q. and Meshnick, S.R., J. Med. Chem., 38 (1995) 4120.
26. Avery, M.A., Mehrotra, S., Johnson, T.L., Bonk, J.D., Vromn, J.A. and Miller, R., J. Med. Chem., 39 (1996) 4149.
27. SYBYL Molecular Modeling Software, v. 6.3, Tripos Associates Inc., St. Louis, MO, U.S.A., 1996.
28. GAUSSIAN 94, Gaussian Inc., Pittsburgh, PA, U.S.A., 1995.
29. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
30. Singh, U.C. and Kollman, P.A., J. Comput. Chem., 5 (1984) 129.
31. Reed, A.E. and Weinhold, F.J., J. Am. Chem. Soc., 108 (1986) 3586.
32. Stewart, J.J.P., MOPAC 6.0, Quantum Chemical Program Exchange 455, 1990.
33. Clark, M., Cramer, R.D. and VanOpdenbosch, N., J. Comput. Chem., 10 (1989) 982.
34. Dunn, W.J., Wold, S., Edlund, U., Hellberg, S. and Gasteiger, J., Quant. Struct.–Act. Relatsh., 3 (1984) 131.
35. Geladi, P., J. Chemometrics., 2 (1988) 231.
36. Kubinyi, H. and Abraham, U., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 712–728.
37. Cramer, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 443–485.
38. Lin, A.J., Li, L.-Q., Andersen, S.L. and Klayman, D.L., J. Med. Chem., 35 (1992) 1639.
39. Kroemer, R.T., Hecht, P. and Liedl, K.R., J. Comput. Chem., 17 (1996) 1296.
40. Folkers, G., Merz, A. and Rognan, D., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 583–618.