Three-dimensional Quantitative Structure-Activity Relationship Analyses of a Series of Butenolide ET_A Antagonists

ZHU, Li-Li(朱丽荔) XU, Xiao-Jie*(徐筱杰)
College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Two kinds of Three-dimensional Quantitative Structure-activity Relationship (3D-QSAR) methods, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), were applied to analyze the structure-activity relationship of a series of 63 butenolide ET_A selective antagonists with respect to their inhibition against human ET_A receptor. The CoMFA and CoMSIA models were developed for the conceivable alignment of the molecules based on a template structure from the crystallized data. The statistical results from the initial orientation of the aligned molecules show that the 3D-QSAR model from CoMFA ($q^2 = 0.543$) is obviously superior to that from the conventional CoMSIA ($q^2 = 0.407$). In order to refine the model, all-space search (ASS) was applied to minimize the field sampling process. By rotating and translating the molecular aggregate within the grid systematically, all the possible samplings of the molecular fields were tested and subsequently the one with the highest q^2 was picked out. The comparison of the sensitivity of CoMFA and CoMSIA to different space orientation shows that the CoMFA q^2 values are more sensitive to the translations and rotations of the aligned molecules with respect to the lattice than those of CoMSIA. The best CoMFA model from ASS was further refined by the region focused technique. The high quality of the best model is indicated by the high cross-validated correlation and the prediction on the external test set. The CoMFA coefficient contour plots identify several key features that explain the wide range of activities, which may help us to design new effective ETA selective antagonists.

Keywords CoMFA, CoMSIA, 3D-QSAR, butenolide ET_A selective antagonists, all-space search

Introduction

Since the discovery of endothelin (ET) in 1985, as the peptidic endothelial derived constricting factor, the human genome identified for ET was subsequently found to encode for three separate isoforms of the peptide, known as ET-1, ET-2 and ET-3.[1] ET exerts its physiological effects by acting upon specific G protein coupled receptors. Two distinct ET receptors, ET_A and ET_B, have been cloned and expressed in mammalian species.[2-5] ET_A is selective for ET-1, and ET_B has equal affinity for the three isoforms. Both receptors have been shown to mediate vasoconstriction in animal tissues depending upon the species and vascular bed under study.[6-10]

During the last decades, a large number of peptide and non-peptide ET antagonists have been reported.[11-15] The discovery of small non-peptide receptor antagonist may represent a new group of therapeutic agents to treat a variety of human diseases. Recently, it has been validated that a series of γ-hydroxy butenolide compounds are potential ET antagonists.[16] Lacking of the exact 3D-structure of ETA receptor, it is very difficult to clarify a model of action on the drug-receptor, which restrains further research to find any other leading compounds based on the structure of receptor. Under this condition, the quantitative structure-activity relationship (QSAR) analysis may be the most direct and effective tool for optimizing a leading compound and designing new potential ET_A antagonists. Until now, however, studies on the relationship between the chemical structures and the biological functions of this kind of compounds have never been reported. In this paper, a profound QSAR analysis was performed using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). It is expected that the research in this work can afford some useful information for further development of new effective ET_A selective antagonists.

CoMFA is the most widely used approach for the study of QSAR at the 3-D level.[17,18] Recently, another 3D-QSAR procedure, comparative molecular similarity indices analysis (CoMSIA), has been reported.[19-21] The approach is designed to avoid some inherent deficiencies arising from the functional form of Lennard-Jones and Coulomb potentials used in CoMFA.[22-25] In CoMSIA, a distance-dependent Gaussian-type functional form has been introduced, which can avoid singularities at the atomic positions and the dramatic changes of potential energy for those grids in the proximity of the surface. Meantime, no arbitrary definition of cut-off limits is required in CoMSIA. The unique difference between conventional CoMFA and CoMSIA is the field type and the potential function. In CoMSIA, similarity is expressed in terms of different physicochemical properties: steric occupancy, partial atomic charges, local hydrophobicity, and H-bond donor and acceptor properties. A Gaussian-type distance-dependent function has been used to calculate different kinds of physicochemical

* E-mail: xiaojxu@chem.pku.edu.cn
Received February 22, 2002; revised July 10, 2002; accepted October 14, 2002.
Project supported by the National Natural Science Foundation of China (No. 29992590-2).

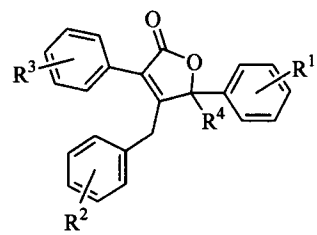
properties.

Computational methods

Molecular modeling

The paullone derivatives inhibitory data against human ET_A receptor, represented by $IC_{50}(\mu mol \cdot L^{-1})$ values, were taken from the literature[16] and listed in Table 1. The $\log(1/C)$ (C represents IC_{50}) values were used as the dependent variable to derive 3D-QSAR models. A training set of 54 γ-hydroxy butenolide analogues was used for the CoMFA and CoMSIA analyses. In addition, 9 compounds, randomly selected from various ranges of biological activity, were kept to test the actual prediction of the obtained 3D-QSAR models. Structural variations of the parent strucuture, present in all molecules, were allowed at positions R^1, R^2, R^3 and R^4 (Table 1). The considered substituents of the antagonists were summaried in Table 1.

Table 1 Structures of butenolide derivatives and the experimental and calculated biological activity by the best CoMFA model



Compd	R^1	R^2	R^3	R^4	log(1/C)		Residue
					Obsd	Calcd[a]	
1	4-OCH_3	3,4,5-$(OCH_3)_3$	3,4-OCH_2O	OH	0.523	0.442	0.081
2	4-OCH_3	4-OH	3,4-OCH_2O	OH	-0.903	-0.990	0.087
3	4-OCH_3	4-$N(CH_3)_2$	3,4-OCH_2O	OH	-1.447	-1.180	-0.267
4	4-OCH_3	4-SCH_3	3,4-OCH_2O	OH	-0.591	-1.024	0.433
6	4-OCH_3	4-NHAc	3,4-OCH_2O	OH	-0.079	-0.196	0.117
7	4-OCH_3	4-CO_2CH_3	3,4-OCH_2O	OH	-1.041	-0.866	-0.175
8	4-OCH_3	4-NO_2	3,4-OCH_2O	OH	-2.903	-2.700	-0.203
9	4-OCH_3	3-CO_2CH_3	3,4-OCH_2O	OH	-0.929	-0.905	-0.024
10	4-OCH_3	3-OCH_3	3,4-OCH_2O	OH	-1.699	-1.792	0.093
11	4-OCH_3	3-NO_2	3,4-OCH_2O	OH	-1.845	-1.736	-0.109
12	4-OCH_3	3-CF_3	3,4-OCH_2O	OH	-1.633	-1.618	-0.015
13	4-OCH_3	3-OCH_3	3,4-OCH_2O	OH	-0.477	-0.621	0.144
14	4-OCH_3	2,3-$(OCH_3)_2$	3,4-OCH_2O	OH	-1.041	-0.866	-0.175
15	4-OCH_3	2,5-$(OCH_3)_2$	3,4-OCH_2O	OH	-0.322	-0.498	0.176
16	4-OCH_3	3,4-$(OCH_3)_2$	3,4-OCH_2O	OH	-0.146	-0.024	-0.122
17	4-OCH_3	3,5-$(OCH_3)_2$	3,4-OCH_2O	OH	-0.230	-0.099	-0.131
19	4-OCH_3	3,4-OCH_2CH_2O	3,4-OCH_2O	OH	-0.875	-0.810	-0.065
20	4-OCH_3	2,3,4-$(OCH_3)_3$	3,4-OCH_2O	OH	-1.079	-1.021	-0.058
21	4-OCH_3	2,4,5-$(OCH_3)_3$	3,4-OCH_2O	OH	-0.255	-0.475	0.220
22	4-OCH_3	3,4,5-$(OC_2H_5)_3$	3,4-OCH_2O	OH	0.920	0.778	0.142
23	4-OCH_3	3-OCH_3, 4,5-OCH_2O	3,4-OCH_2O	OH	-1.146	-1.244	0.098
24	4-OCH_3	H	3,4-$(OCH_3)_2$	OH	-1.845	-1.719	-0.126
25	4-OCH_3	H	3-OCH_3	OH	-1.954	-2.070	0.116
26	4-OCH_3	H	2,4-$(OCH_3)_2$	OH	-2.477	-2.475	-0.002
28	4-OCH_3	H	3,5-$(OCH_3)_2$	OH	-1.662	-1.870	0.208
29	4-OCH_3	H	3,4,5-$(OCH_3)_3$	OH	-3.114	-2.925	-0.189
30	4-OCH_3	H	3-OCH_3, 4,5-OCH_2O	OH	-0.826	-0.914	0.088
31	4-OCH_3	3,4,5-$(OCH_3)_3$	3-OCH_3, 4,5-OCH_2O	OH	0.301	0.033	0.268
32	3-CH_3, 4-OCH_3	3,4,5-$(OCH_3)_3$	3-OCH_3, 4,5-OCH_2O	OH	0.046	0.387	-0.341
33	3-CH_3, 4-OCH_3	3,4,5-$(OCH_3)_3$	3,4-OCH_2O	OH	0.301	0.231	0.070
34	3-CH_3, 4-OCH_3	H	3-OCH_3, 4,5-OCH_2O	OH	-0.914	-0.545	-0.369

Continued

Compd	R^1	R^2	R^3	R^4	log (1/C) Obsd	log (1/C) Calcd[a]	Residue
35	4-OCH_3	H	3,4-OCH_2O	OH	-0.869	-0.685	-0.184
36	4-OCH_3	H	3,4-OCH_2O	OCH_3	-2.477	-2.344	-0.133
38	4-OCH_3	3,4,5-$(OCH_3)_3$	3,4-OCH_2O	OCH_3	-2.778	-2.750	-0.028
39	4-Cl	H	3,4-OCH_2O	OH	-2.633	-2.815	0.182
41	4-CH_3	H	3,4-OCH_2O	OH	-3.690	-3.261	-0.429
42	3,4-Cl_2	H	3,4-OCH_2O	OH	-2.602	-2.820	0.218
43	3-CH_3, 4-OCH_3	H	3,4-OCH_2O	OH	-1.079	-1.284	0.205
44	4-OCH_3	4-CH_3	3,4-OCH_2O	OH	-1.708	-1.490	-0.218
45	4-OCH_3	4-Cl	3,4-OCH_2O	OH	-2.041	-1.592	-0.449
46	4-OCH_3	4-OCH_3	3,4-OCH_2O	OH	-0.255	-1.117	0.862
47	4-OCH_3	3,4-Cl_2	3,4-OCH_2O	OH	-2.176	-1.719	-0.457
49	4-OCH_3	4-OCH_3	H	OH	-3.342	-3.070	-0.272
50	4-OCH_3	4-OCH_3	4-Cl	OH	-2.845	-3.211	0.366
51	4-OCH_3	4-OCH_3	4-CH_3	OH	-2.756	-2.971	0.215
52	4-OCH_3	4-OCH_3	3,4-Cl_2	OH	-2.230	-2.179	-0.051
54	2-OCH_3	H	3,4-OCH_2O	OH	-2.681	-2.773	0.092
56	4-OC_2H_5	H	3,4-OCH_2O	OH	-2.799	-3.072	0.273
57	4-SCH_3	H	3,4-OCH_2O	OH	-2.230	-2.577	0.347
58	4-OC_3H_7-*i*-	H	3,4-OCH_2O	OH	-3.061	-2.704	-0.357
59	4-OCH_2Ph	H	3,4-OCH_2O	OH	-2.978	-3.198	0.220
60	3,4-$(OCH_3)_2$	H	3,4-OCH_2O	OH	-1.041	-0.968	-0.073
61	4-CO_2CH_3	H	3,4-OCH_2O	OH	-2.602	-2.530	-0.072
62	3,4-OCH_2O-	H	3,4-OCH_2O	OH	-1.279	-1.065	-0.214
5[a]	4-OCH_3	4-SO_2CH_3	3,4-OCH_2O	OH	-0.602	-0.901	0.299
18[a]	4-OCH_3	3,4-OCH_2O	3,4-OCH_2O	OH	-1.204	-1.100	-0.104
27[a]	4-OCH_3	H	2,5-$(OCH_3)_2$	OH	-2.380	-2.900	0.520
37[a]	4-OCH_3	3,4,5-$(OCH_3)_3$	3,4-OCH_2O	H	-2.845	-2.993	0.148
40[a]	H	H	3,4-OCH_2O	OH	-2.778	-2.850	0.072
48[a]	4-OCH_3	3-CH_3, 4-OCH_3	3,4-OCH_2O	OH	-1.380	-1.001	-0.379
53[a]	4-OCH_3	4-OCH_3	4-OCH_3	OH	-2.279	-2.211	-0.068
55[a]	3-OCH_3	H	3,4-OCH_2O	OH	-2.699	-2.501	-0.198
63[a]	3,4,5-$(OCH_3)_3$	H	3,4-OCH_2O	OH	-1.398	-1.801	0.403

[a] The compounds were used as a test set and not included in the derivation of equations.

The 3D structures of compounds in this study were generated by the molecular modeling software package SYBYL ver 6.6.[26] As being well-known, the proposal of the bioactive conformer and the alignment rule may be the most important parts in the 3D-QSAR studies. In principle, if a prior knowledge is not known about the complex structure of ligand and receptor, proposing the real bioactive conformation is relatively difficult. But in most cases, the bioactive conformer will adopt a relatively low-energy conformation, so it is suitable to utilize the crystal structure of compound as the starting geometry. The X-ray structure of copmound **1** was treated as the template molecule. Other molecules in Table 1 were constructed by modifying the template. The initial structures were minimized using molecular mechanism calculation performed with MMFF force field.[27] All the structures after minimization were put into 3D lattice by superimpose all compounds onto the fit centers (the common parent structure shown in Table 1) using an atom-by-atom least-square fit implemented in the "SYBYL FIT" option in SYBYL. The compound **22** with the highest biological activity was selected as the reference molecule.

Determinations of CoMFA *and* CoMSIA *models*

The CoMFA analyses were done with the "Advanced CoMFA" module of SYBYL. A sp^3 carbon atom with + 1 charge was selected as probe atom for the calculations of the steric and electrostatic fields around the aligned molecules. Values of steric and electrostatic energies were truncated at 125 kJ/mol. The dimensions of the surrounding lattice were selected with a sufficiently large margin (= 0.4 nm) to enclose the aligned molecules. Considering the importance of

H-bond interaction in most ligand-receptor systems, besides the usually used steric and electrostatic fields, H-bond fields were also included in the CoMFA analyses. The scheme proposed by Bohacek and McMartin[28] was applied to generate the H-bond fields.

To choose the appropriate components and check the statistical significance of the models, leave-one-out cross-validations were used by the enhanced version of PLS, the SAMPELS method.[29] Subsequently, the final 3D-QSAR models were derived from the non-cross-validated calculations. The CoMFA results were graphically interpreted by the field contribution maps using the field type "stdev* coeff".

The CoMSIA analyses were done in the "QSAR" module in SYBYL with five kinds of physicochemical properties in CoMSIA implemented in SYBYL, which include steric contributions by the third power of the atomic radii, electrostatics by MMFF charges, hydrophobicities by atom-based hydrophobic parameters and H-bond properties by suitably placed pseudo atoms, using a common probe with 0.1 nm radius, +1 charge, +1 hydrophobicity and H-bond property of +1. The extent and orientation of the grids surrounding the tested molecules were the same as those in the CoMFA analyses. The attenuation factor, α, which is the coefficient of the squared mutual distance in the Gaussian-type function in the calculation of similarity indices, was set at 0.3. The statistical evaluation for the CoMSIA analyses was performed in the same manner as described for CoMFA.

All-space search procedure

It is well known that the 3D-QSAR models from conventional CoMFA are very sensitive to the different space orientations of the molecular aggregate with respect to the lattice.[30] In most cases, it is possible that the low q^2 value obtained from conventional CoMFA, which often frustrates the researcher, may be simply caused by the poor orientation of the molecular aggregate. In this paper, in order to investigate the sensitivity of CoMFA to different space orientation and get the best model, we translated and rotated the aligned molecules systematically to cover the full space. The translation and rotation procedures were performed using the STATIC TRANSLATE and STATIC ROTATE commands in SYBYL, respectively. For each orientation of the molecular aggregate, the CoMFA analyses were performed to get the q^2 of the model. A SPL script in SPL was written to do ASS automatically. As a comparison, the ASS was also performed for the CoMSIA analyses.

Results and discussion

3D-QSAR analyses adopting the initial orientation of the aligned molecules

The obtained QSAR equations for the inhibitory activity are listed in Table 2. In these equations, n represents the number of the compound, SD the standard error of estimate, r the correlation coefficient and q the correlation coefficient obtained from the leave-one-out cross-validation. Only using steric and electrostatic fields, the quality of the CoMFA-based model ($q^2 = 0.467$) in Table 2 was better than the CoMSIA-based model ($q^2 = 0.376$) in terms of the q^2 value, but both of these two models are not satisfactory. After adding the H-bond fields, the predictive power of the CoMFA model ($q^2 = 0.543$) was improved obviously, indicating that the biological activity essentially exhibited a significant relationship with the H-bond fields.

Table 2 Results of the CoMFA and CoMSIA calculations using several different field combinations adopting the initial orientation[a]

	CoMFA		CoMSIA		
	Model 1	Model 2	Model 3	Model 4	Model 5
	steric + electrostatic	steric + electrostatic + H-bond	steric + electrostatic	steric + electrostatic + hydrophobic	steric + electrostatic + hydrophobic + H-bond
q^2	0.467	0.543	0.376	0.407	0.351
r^2	0.920	0.932	0.548	0.611	0.665
SD	0.338	0.283	0.774	0.703	0.667
F	90.659	107.110	30.902	40.129	50.980
n	6	6	2	2	2
Fraction					
steric	0.549	0.245	0.215	0.148	0.102
elelectrostatic	0.451	0.160	0.785	0.553	0.390
hydrophobic					0.107
H-bond acceptor		0.457		0.144	0.139
H-bond donor		0.138		0.155	0.263

[a] The grid spacing for all CoMFA and CoMSIA analyses is set to 0.2 nm. [b] n represents the number of the compound, SD the standard error of estimate, r^2 the correlation coefficient and q^2 the correlation coefficient obtained from the leave-one-out cross-validation.

Table 2 also shows several sets of analysis results of different field combinations from CoMSIA. After adding the H-bond fields, the q^2 value of the CoMSIA model was improved, but it was also poorer than that of the CoMFA model. However, the hydrophobic regions, which were drawn according to model 5 in Table 2 failed to satisfactorily explain the variations in inhibitory activity. In a previous CoMFA paper about insecticides, Akamatsu *et al*. reported that fractions of the hydrophobic term were included within steric and electrostatic field descriptor terms.[31] It is also likely that the hydrophobic field effect is correlated partially with the steric and electrostatic contributions.

It is well-known that the shift of the q^2 values for the 3D-QSAR models from CoMFA is evident as the grid spacing changed. The 0.2 nm of grid was shifted by +0.05 nm and +0.1 nm to find out the effect of altered lattice point location on the results of the CoMFA study (see Table 3). This resulted in slightly lower correlation ($q^2 = 0.528$ and $q^2 = 0.537$, respectively) as compared to the original study with 0.2 nm ($q^2 = 0.543$). It is naturally deduced that at the large grid resolution in CoMFA, some important information in some regions may be lost. At lower grid spacing (0.1 nm), the 3D-QSAR model produced reduced statistical significance ($q^2 = 0.537$), because the increase in the number of lattice points also increases the noise in PLS analysis and leads to a less statistically significant model. Thus, if not incorporated with a variable selection procedure, increasing the grid resolution in CoMFA studies will generally result in increaseing computation time and decreasing predictivity. The best significant CoMFA model ($q^2 = 0.578$) was yielded as the grid spacing was defined as 0.15 nm. In 0.15 nm of grid spacing, the numbers of optimal components that produce the best cross-validation linear regression coefficient were used to produce the non-cross-validated model. The leave-one-out cross-validated PLS analysis results in a q^2 of 0.578 using 8 principle components. The non-cross-validated PLS analysis yields a higher r^2 of 0.961 with a low standard error of estimate (*SD*) 0.241 and a large *F* (139.791).

Table 3 Results of the different CoMFA analyses to the training data set using 0.1, 0.15, 0.25 and 0.30 nm grid spacing

	CoMFA(1)	CoMFA(2)	CoMFA(3)	CoMFA(4)
q^2	0.537	0.578	0.528	0.537
r^2	0.939	0.961	0.906	0.857
SD	0.397	0.241	0.372	0.448
F	119.870	139.791	63.159	57.592
n	6	8	7	5
Fraction				
steric	0.246	0.264	0.309	0.246
electrostatic	0.151	0.161	0.182	0.151
H-bond acceptor	0.492	0.471	0.457	0.492
H-bond donor	0.111	0.106	0.052	0.111
Grid spacing (nm)	0.10	0.15	0.25	0.30

All-space search strategy

All-space search procedure means that all aligned molecules adopting different orientations are sampled. First, starting from the initial orientation, the molecular aggregate was rotated around the *x*, *y* and *z* axes in an increment of 20 degree. For each orientation, CoMFA and CoMSIA fields were calculated and PLS analyses were subsequently performed to get the q^2 values. Thus, totally $18 \times 9 = 162$ orientations were explored for the data set. The results of CoMFA and CoMSIA using ASS are summarized in Table 4. The q^2 values using the molecular fields in CoMSIA are much more concentrated than those using the molecular fields in CoMFA. For the data set, the q^2 values for CoMFA may vary as much as 0.23, while those values for CoMSIA are only about 0.01. That is to say, the models from CoMSIA are much more stable than those from CoMFA.

Table 4 Variation of q^2 values observed in all-space sampling for the data set

Fields	Leave-one-out cross-validated r^2 (q^2)					
	Rotational operation			Translational operation		
	best	worst	span[a]	best	worst	span[a]
CoMFA	0.625	0.391	0.234	0.630	0.412	0.218
CoMSIA	0.429	0.422	0.008	0.430	0.422	0.009

[a] span = best. − worst.

Then, the molecular aggregate with the highest q^2 after rotations was selected for translations, which were systematically translated along the *x*, *y* and *z* axes. The original grid box with 0.15 nm was extended by 0.4 nm in *x*, *y* and *z* axes to guarantee adequate margin of the lattice surrounding all molecules during translations. The lattice surrounding all molecules along the *x*, *y* and *z* axes from 0 to 0.15 nm was in steps of 0.03 nm, therefore, totally $5 \times 5 \times 5 = 125$ orientations were explored. The frequency distributions of q^2 values among all translations using CoMFA fields are investigated, and the q^2 values from CoMFA seem to be dispersed in a relative larger region compared with those from CoMSIA. The span value for CoMSIA (Table 4) is only 0.008, while that value for CoMFA is 0.218.

After exploring the dependence of q^2 for CoMSIA and CoMFA on translation and rotation of the molecular aggregate, it is clear that the CoMFA q^2 is more sensitive to the translations and rotations of the aligned molecules with respect to the lattice, whereas CoMSIA produces less unsatisfactory changes in q^2. The instabilities of CoMFA can be attributed to the shape and steepness of the Lennard-Jones potential and in consequence to the required arbitrary fixation of cutoff values. For example, in some orientation, some atoms of the aligned molecules are near some grids. If we use the potentials in CoMFA, the Lennard-Jones potentials will increase to large values near these grids, and generally the large potentials will be truncated to the cutoff value. The

contributions of those grids using cutoff values will introduce certain unpredicted instability to the final models. So, in conventional CoMFA analysis, ASS must be performed to get the best QSAR model. While in CoMSIA, a distance-dependent Gaussian-type functional form has been introduced, which can avoid singularities at the atomic positions and the dramatic changes of potential energy for those grids in the proximity of the surface. The inherent limitation in CoMFA can be partly overcome by CoMSIA.

From the values of q^2, the best one from those models after ASS calculations was obtained using CoMFA fields with specific molecular orientation. The leave-one-out cross-validated PLS analysis of the best model in Table 5 results in a q^2 of 0.630 using seven principle components.

Table 5 Results of the CoMFA PLS analyses of the training set for the best CoMFA model after ASS and region focusing (grid spacing: 0.15 nm)

	CoMFA(1)[a]	CoMFA(2)[b]
q^2	0.630	0.653
r^2	0.950	0.951
SD	0.272	0.264
F	124.033	153.363
n	7	6
Fraction		
steric	0.237	0.361
electrostatic	0.159	0.107
H-bond acceptor	0.472	0.335
H-bond donor	0.132	0.197

[a] The best CoMFA model using steric, electrostatic and H-bond fields after ASS. [b] The best CoMFA model using steric, electrostatic and H-bond fields after ASS and region focusing.

In order to enhance the quality of the best CoMFA model from ASS procedure, the region focusing technique, available in the "Advanced CoMFA" module in SYBYL was used to refine the model by increasing the weights for those lattice points which were most pertinent to the model. Using the region focusing technique, the statistical significance of the model was improved obviously. The leave-one-out cross-validated PLS analysis of the model results in a q^2 of 0.653 using six principle components. The non-cross-validated PLS analysis yields a higher r^2 of 0.951 with a very low standard error of estimate (*SD*) 0.264 (Table 5). Moreover, this model expresses good predictive ability (Table 6) for the external test set ($r^2_{pred} = 0.929$, $SSE = 0.934$) with the average absolute error of 0.243 log units across a range of 2.029 log units, and the following discussions would only refer to this model. The biological activities ($\log 1/C$), the calculated activities using the best CoMFA model and the residue values from the observed values for training set, were shown in Table 1. Fig. 1 shows a plot of observed vs. the calculated biological activities. The predicted biological activities of the test-set compounds were listed in Table 1 and Fig. 2. The derived model was satisfactory from the viewpoint of statistical significance and actual predictive ability. On the basis of the best 3D-QSAR model obtained, we expect to find more potential compounds with the aid of the computational combinatorial chemistry method.

Table 6 Difference between predicted and actual activities for nine molecules in the test sets using four 3D-QSAR models

Compd.	CoMFA(1)[a]	CoMFA(2)[b]	CoMFA(3)[c]	CoMSIA[d]
5	0.973	0.913	0.901	1.897
18	1.705	1.101	1.100	1.001
27	3.232	2.914	2.900	1.857
37	3.618	3.021	2.993	3.167
40	2.896	2.877	2.850	2.296
48	1.280	1.031	1.001	1.709
53	2.165	2.172	2.211	2.050
55	2.321	2.412	2.501	2.160
63	1.840	1.794	1.801	2.083
r^2_{pred}	0.815	0.895	0.929	0.376
SSE[e]	2.567	1.108	0.934	4.906

[a] The CoMFA model using steric, electrostatic and H-bond fields adopting initial orientation of the aligned molecules. [b] The best CoMFA model using steric, electrostatic and H-bond fields after ASS. [c] The best CoMFA model using steric, electrostatic and H-bond fields after ASS and region focusing. [d] The CoMSIA model using steric, electrostatic and H-bond fields adopting initial orientation of the aligned molecules. [e] Sum of square error of predictions for nine tested compounds.

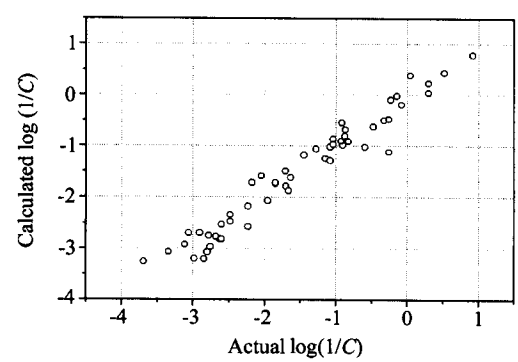


Fig. 1 Comparison of experimental $\log(1/C)$ with calculated $\log(1/C)$ obtained from the best CoMFA model after all-space searching and region focusing to those molecules in the training set.

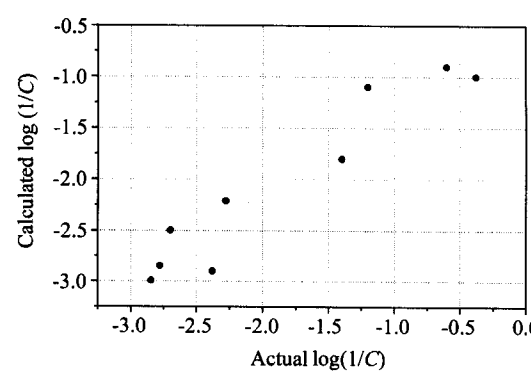


Fig. 2 Comparison of experimental $\log(1/C)$ with calculated $\log(1/C)$ obtained from the best CoMFA model after all-space searching and region focusing to those molecules in the test set.

Graphical interpretation of the results

One of the best advantages of CoMFA is that the effects of all kinds of properties contributing to biological activity can be divided and viewed as 3D coefficient contour plots. Fig. 3 shows an overlay of the structure of compound **22**, which is the most potent in this study, with the major steric potential contour maps drawn according to model 2 in Table 5. In Fig. 3, the CoMFA contour plot shows green colored regions where increased steric bulk is associated with enhanced affinity and yellow colored regions where increased steric bulk is associated with diminished affinity. Three small bulky unfavorable areas were near the R^1 and R^3 sites, while four bulky favorable areas were near the R^1, R^2 and R^3 sites. From the size of the four green areas, the green regions near the R^1 and R^3 sites are seen to be more important than that near the R^2 site. Different substituted groups near these areas would affect the binding affinities by the steric complementarity between the receptor and the ligand. For example, compound **1** showed higher activity than compounds **2**—**19**. In compound **1**, the R^2 site is substituted by the 3,4,5-$(OCH_3)_3$ group, while the R^2 sites in compounds **2**—**19** are substituted by other smaller groups, and these smaller groups can not form good steric complementarity with the receptor. The shift of the 3,4,5-$(OCH_3)_3$ group from the R^2 site in compound **1** to the 3,4,5-$(OC_2H_5)_3$ group from the R^2 site in compound **22** results in higher activity. Moreover, most groups in R^2 sites are aliphatic, which may produce good van der Waals or hydrophobic interactions with the nonpolar side chains of the receptor. The bulky groups near the sites R^1 and R^3 on the benzene rings are also favorable, but they must be restricted to some extent in order to avoid bad steric contacts with receptor in some orientations.

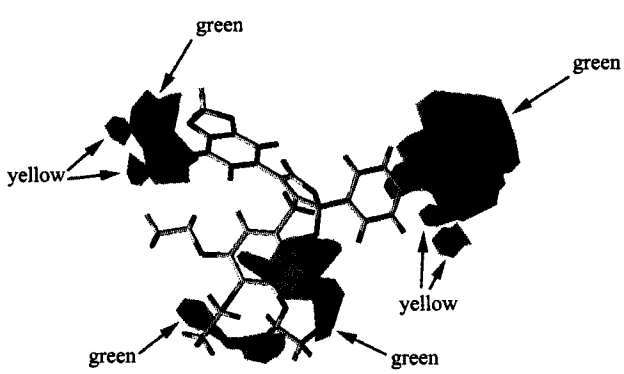


Fig. 3 Contour plots of the CoMFA steric fields (stdev * coeff). The favorable steric areas with more bulk are indicated by green isopleths, whereas the disfavored steric areas are shown by yellow isopleths. The most active compound **22** is shown as the reference compound.

Compared with the relative contributions of the steric field, the electrostatic factors were less important. Besides one small positive preferred contour near site R^3, the other positive preferred contour and two negative preferred contours are located near site R^2 (Fig. 4). Two negative preferred contours can be suggested by the charge withdrawing groups linked to benzene ring, such as groups including O or N atom. The charge withdrawing atom linked to benzene ring will make the charge distributed near this atom relatively positive, which is represented by this blue region.

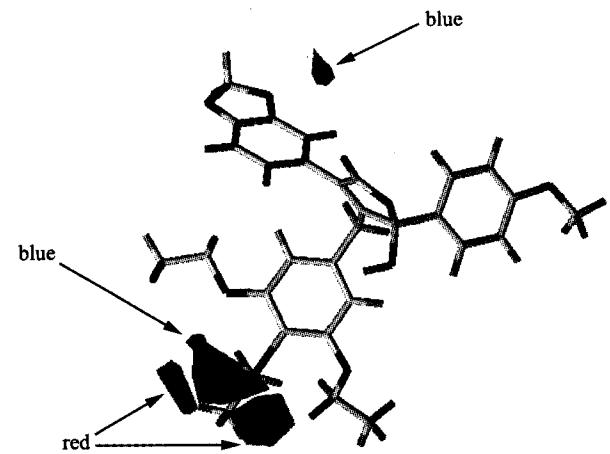


Fig. 4 Contour plots of the CoMFA electrostatic fields (stdev* coeff). The favorable electrostatic areas with positive charges are indicated by blue isopleths, whereas the favorable electrostatic areas with negative charges are shown by red isopleths. The most active compound **22** is shown as the reference compound.

The graphical interpretations of the field contributions of the H-bond are shown in Figs. 5 and 6. In principle, they should highlight the areas near these areas where H-bond donor or acceptor on the ligand can form H-bonds with the receptor to influence binding affinity. From the fraction of field, the H-bond acceptor field is very important. The contour shows that more H-bond acceptor groups in the green regions increase potency, whereas more H-bond acceptor groups in the yellow regions decrease potency. The largest green region

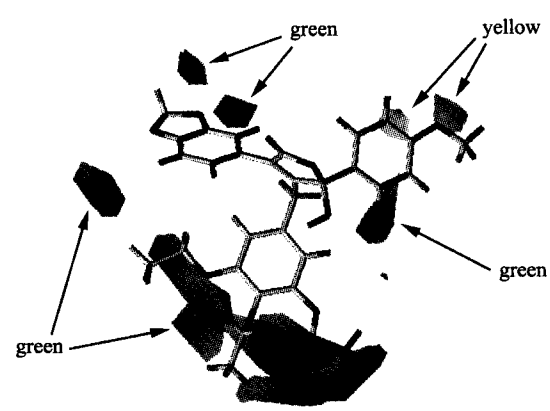


Fig. 5 Contour plots of the CoMFA H-bond acceptor fields (stdev* coeff). Green isopleth contours maps where an H-bond acceptor in the ligand will be favorable for biological activity, while yellow isopleths represents H-bond acceptor in the ligands unfavorable for biological activity. The most active compound **22** is shown as the reference compound.

near the R^2 site indicates that 2,3,4-R^3 group with H-bond acceptor linked to the benzene ring will improve the ligand binding. Near the R^2 site, there also exist two separate small green contours, showing that groups with H-bond acceptor linked to 2- or 3-R^2 position on the benzene ring will be favorable to biological activity. The H-bond acceptor fields near the R^1 site are different from those near the R^2 and R^3 sites. Except one green contour at the 1-R^1 position, there exist other two yellow contours located near the 3- or 4-R^1 position, indicating that the H-bond acceptors near the 3- or 4-R^1 position will decrease biological activity. But we must pay attention that in most cases the electrostatic interactions and the H-bonding interactions can not be completely separated, and so the contour maps of the H-bonding fields may not afford us complete explicit information about the H-bonding interactions between ligand and receptor.

Compared with H-bond acceptor field, the H-bond donor field seems not so important. In Fig. 6, the H-bond donor favorable areas are represented by blue contours, while the H-bond unfavorable areas by red contours. From Fig. 6, only one H-bond donor feature can be seen, which is indicated by one blue contour. The blue contour implies that the groups with H-bond donors linked at the R^4 site will enhance the biological activity. For example, compounds **35** and **36** have the same substituented groups at R^1, R^2 and R^3 sites, but their biological potencies are quit different. It is naturally deduced that the OH at the R^4 site of compounds **35** can form H-bonds with the receptor, which will significantly enhance the biological activity.

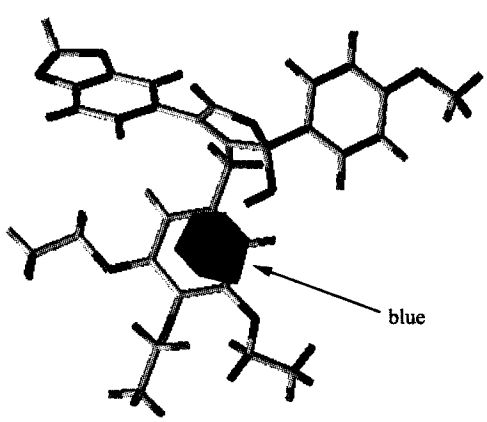


Fig. 6 Contour plots of the CoMFA H-bond donor fields (stdev* coeff). Blue isopleth contours maps where an H-bond donor group in the ligand will be favorable for biological activity, while red isopleths represents H-bond donor in the ligand unfavorable for biological activity. The most active compound **22** is shown as the reference compound.

Conclusions

Multiple CoMSIA and CoMFA 3D-QSAR models have been developed from a data set of γ-hydroxy butenolide endothelin antagonists. By using the four types of fields provided by CoMFA, the produced 3D-QSAR model was better than those from other fields combinations in CoMSIA and the conventional CoMFA. Based on the initial orientation of the molecular aggregate, the ASS strategy was used to investigate the sensitivity of q^2 values of different space orientation of the aligned molecules and determine the best model. The best CoMFA model from ASS possesses promising predictive ability as indicated by the high cross-validated correlation and the prediction on the external test set. Some important factors contributing to the biological activity are indicated by the contour maps of different properties, which are very helpful to understand the underlying mechanism of receptor-drug interaction.

References

1 Inoue, A.; Yanagisawa, M.; Kimura, S.; Kasuya, Y.; Miyauchi, T.; Goto, K.; Masaki, T. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2863.

2 Arai, H.; Hori, S.; Aramori, I.; Ohkubo, H.; Nakanishi. S. *Nature* **1990**, *348*, 730.

3 Sakurai, T.; Yanagisawa, M.; Takuwa, Y.; Miyazaki, H.; Kimura, S.; Goto, K.; Masaki, T. *Nature* **1990**, *348*, 732.

4 Sakamoto, A.; Yanagisawa, M.; Sakurai, T.; Takuwa, Y.; Yanagisawa, H.; Masaki, T. *Biochem. Biophys. Res. Commun.* **1991**, *178*, 656.

5 Hosoda, K.; Nakao, K.; Arai, H.; Suga, S.; Ogawa, Y.; Mukoyama, M.; Shirakami, G.; Saito, Y.; Nakanishi, S.; Imura, H. *FEBS Lett.* **1991**, *287*, 23.

6 Clozel, M.; Gray, G. A.; Breu, V.; Loffler, B. M.; Osterwalder, R. *Biochem. Biophys. Res. Commun.* **1992**, *186*, 867.

7 Sumner, M. J.; Canon, T. R.; Mundin, J. W.; White, D. G.; Watts, I. S. *J. Pharmacol.* **1992**, *107*, 858.

8 Davenport, A.; Maguire, J. J. *Trends Pharmacol. Sci.* **1994**, *15*, 136.

9 Bax, W. A.; Bos, E.; Saxena, P. R. *Eur. J. Pharmacol.* **1993**, *239*, 267.

10 Sudjarwo, S. A.; Hori, M.; Tanaka, T.; Matsuda, Y.; Okada, T.; Karaki, H. *Biochem. Biophys. Res. Commun.* **1994**, *200*, 627.

11 Ishikawa, K.; Fukami, T.; Nagase, T.; Fujita, K. *J. Med. Chem.* **1992**, *35*, 2139.

12 Ihara, M.; Noguchi, K.; Saeki, T.; Fukuroda, T.; Tsuchida, S.; Kimura, S.; Fukami, T.; Ishikawa, K.; Nishikibe, M.; Yano, M. *Life Sci.* **1992**, *50*, 247.

13 Sogabe, K.; Nirei, H.; Shoubo, M.; Nomoto, A.; Ao, S.; Notso, Y.; Ono, T. *J. Pharmacol. Exp. Ther.* **1993**, *264*, 1040.

14 Clozel, M.; Breu, V.; Burri, K.; Cassao, J. M.; Fischili, W.; Gray, G. A.; Hirth, G.; Loffler, B. M.; Muller, M.; Neidhart, W.; Ramuz, H. *Nature* **1993**, *365*, 759.

15 Elliott, J. D.; Lago, M. A.; Cousins, R. D.; Gao, A.; Leber, J. D.; Erhard, K. F.; Nambi, P.; Elshourbagy, N. A.; Kumar, C.; Lee, J. A.; Bean, J. W.; DeBrosse, C. W. *J. Med. Chem.* **1994**, *37*, 1553.

16 Patt, W.; Edmunds, J.; Repine, J.; Berryman, K. *J. Med. Chem.* **1997**, *40*, 1063.

17 Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am.*

Chem. Soc. **1988**, *110*, 5959.
18 Hou, T. J.; Xu, X. *Chin. J. Prog. Chem.* **2001**, *13*, 436 (in Chinese).
19 Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
20 Klebe, G.; Abraham, U. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1.
21 Bohm, M.; Sturzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458.
22 Hou, T. J.; Li, Y. Y.; Xu, X. J. *J. Mol. Model.* **2000**, *6*, 438.
23 Hou, T. J.; Li, Z. M.; Li, Z.; Liu, J.; Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1002.
24 Zhu, L. L.; Hou, T. J.; Xu, X. J. *J. Mol. Model.* **2001**, *7*, 223.
25 Zhu, L. L.; Hou, T. J.; Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1032.
26 *SYBYL* ver 6.5, Tripos Associates, St. Louis, MO, USA, **1999** (http://www.tripos.com/).
27 Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
28 Bohacek, R. S.; McMartin, C. *J. Med. Chem.* **1992**, *35*, 1671.
29 Bush, B. L.; Nachbar, R. B. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587.
30 Hou, T. J.; Xu, X. *Chemom. Intell. Lab. Syst.* **2001**, *56*, 123.
31 Akamatsu, M.; Nishimura, K.; Osabe, H.; Ueno, T.; Fujita, T. *Pestic. Bichem. Physiol.* **1994**, *48*, 15.

(E0202221 SONG, J. P.; DONG, H. Z.)