

## Three-Dimensional Quantitative Structure–Activity Relationship Analysis of the New Potent Sulfonylureas Using Comparative Molecular Similarity Indices Analysis

T. J. Hou,<sup>†</sup> Z. M. Li,<sup>‡</sup> Z. Li,<sup>†</sup> J. Liu,<sup>‡</sup> and X. J. Xu<sup>\*,†</sup>

Department of Chemistry and Molecular Engineering, Peking University, Beijing 100871, People's Republic of China, and Elemento-Organic Chemistry Institute, State Laboratory of Elemento-Organic Chemistry, Nankai University, Tianjin 300071, People's Republic of China

Received January 11, 2000

The present study describes the implementation of a new three-dimensional quantitative structure–activity relationship (3D-QSAR) technique: comparative molecular similarity indices analysis (CoMSIA) to a set of novel herbicidal sulfonylureas targeted acetolactate synthase. Field expressions in terms of similarity indices in CoMSIA were applied instead of the usually used Lennard-Jones and Coulomb-type potentials in CoMFA. Two different kinds of alignment techniques including field-fit alignment and atom-by-atom fits were used to produce the molecular aggregate. The results indicated that those two alignment rules generated comparative 3D-QSAR models with similar statistical significance. However, from the predictive ability of the test set, the models from the alignment after maximal steric and electrostatic optimization were slightly better than those from the simple atom-by-atom fits. Moreover, systematic variations of some parameters in CoMSIA were performed to search the best 3D-QSAR model. A significant cross-validated  $q^2$  was obtained, indicating the predictive potential of the model for the untested compounds; meanwhile the predicted biological activities of the five compounds in the test set were in good agreement with the experimental values. The CoMSIA coefficient contour plots identified several key features explaining the wide range of activities, which were very valuable for us in tracing the properties that really matter and getting insight into the potential mechanisms of the intermolecular interactions between inhibitor and receptor, especially with respect to the design of new compounds.

### INTRODUCTION

It is well-known that inhibition of essential amino acid biosynthesis in plants is one of the most prominent and attractive principles of herbicidal action.<sup>1</sup> Acetolactate synthase (ALS), the first common enzyme in the biosynthetic route to the branched chain amino acids valine, leucine, and isoleucine, has been identified as the target of action for several structurally distinct classes of compounds with high herbicidal activity, such as sulfonylureas, sulfonamides, imidazolinones, and pyrimidylsalicylates. The four classes of herbicides were all obtained by traditional screening methods and have been developed as new weapons in weed control. The attributes of low application rates, good crop selectivity, environmental safety, and compatibility with the trend toward postemergence weed control exhibited by these compounds are important characteristics for modern agrochemical application, which has led to the rapid success of ALS inhibitors as herbicidal products and attracted a worldwide research commitment.<sup>1</sup>

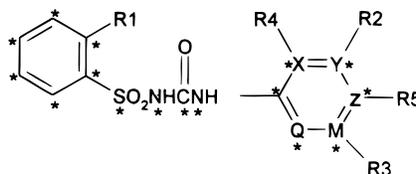
Sulfonylureas, one of the most important ALS inhibitors, have been developed for many years, which possess ultralow-dosage, low mammalian toxicity, and desirable environmental properties. They kill weeds by inhibiting the activity of the ALS enzyme of the weeds and destroying the synthesis of branch-chain amino acid. Sulfonylurea herbicides were first discovered by Levitt at Dupont.<sup>2</sup> Fused heterocyclic

sulfonamides were the product of a research effort initially aimed at preparing bioisosteric analogues of sulfonylures. Due to the exact 3D-structure target receptor ALS enzyme not being successfully established, 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 the receptor.<sup>3</sup> Under this condition, the quantitative structure–activity relationship (QSAR) analysis may be the most direct and effective method for optimizing a leading compound and designing new potential herbicidal sulfonylureas. Until now, however, very few studies on the relationship between the chemical structures and the biological functions of this kind of compounds have been reported. In this paper, a profound correlation study was accomplished on the basis of a new developed 3D-QSAR technique: comparative molecular similarity indices analysis (CoMSIA), which was expected to make prediction concerning the quantitative structure–effect relationships based on an ensemble of ligands.

### MATERIAL

**Experimental Data.** The compounds studied in this paper were designed and synthesized in our laboratory according to the conventional method<sup>4</sup> to cover the potential range as widely as possible. All compounds were purified by repeated recrystallization, and their chemical structures were confirmed by <sup>1</sup>H NMR spectra and elemental analyses. The herbicidal activities of the compounds were measured by the biological screening system using the rape-root growth

\* To whom correspondence should be addressed. Telephone: 62754198. E-mail: xiaojxu@chemms.chem.pku.edu.cn.

**Table 1.** Structures of Sulfonylurea Derivatives, Experimental and Calculated Biological Activity by the Best 3D-QSAR Model from the CoMSIA Analyses

compd	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	X	Y	Z	M	Q	pI <sub>50</sub>		residue
											obsd	calcd <sup>b</sup>	
1	COOEt	Cl	OCH <sub>3</sub>		H	N	C	C	C	N	7.5751	7.572	0.003
2	CH <sub>3</sub>	OCH <sub>3</sub>	OCH <sub>3</sub>		H	N	C	C	C	N	7.0700	7.131	-0.061
3	COOCH <sub>3</sub>	OCH <sub>3</sub>	H		H	N	C	C	C	N	6.5638	6.484	0.080
4	COOCH <sub>3</sub>	CH <sub>3</sub>	OCH <sub>3</sub>			N	C	N	C	N	6.4353	6.542	-0.107
5	COOEt	OCH <sub>3</sub>	H		H	N	C	C	C	N	6.2700	6.262	0.008
6	Cl	CH <sub>3</sub>	OCH <sub>3</sub>			N	C	N	C	N	6.1040	5.967	0.137
7	COOEt	OCH <sub>3</sub>	H		H	N	C	C	C	N	6.0991	6.096	0.003
8	NO <sub>2</sub>	OCH <sub>3</sub>	H		H	N	C	C	C	N	5.8961	5.823	0.073
9	CH <sub>3</sub>	CH <sub>3</sub>	H		H	N	C	C	C	N	5.1475	5.202	-0.056
10	COOEt	CH <sub>3</sub>	H		H	N	C	C	C	N	5.1290	5.047	0.082
11	Cl	CH <sub>3</sub>	H		H	N	C	C	C	N	4.9066	5.033	-0.126
12	CH <sub>3</sub>	Cl	H		H	CH	C	C	C	CH	4.2874	4.336	-0.048
13	COOEt		H		H	N	N	C	C	N	4.2373	4.262	-0.025
14	COOEt	H			H	N	C	C	N	CH	4.1349	4.156	-0.021
15	CH <sub>3</sub>	H	H		H	N	C	C	C	CH	4.0726	3.858	0.215
16	CH <sub>3</sub>	H	H	Cl	H	C	C	C	C	CH	4.0538	3.991	0.063
17	Cl	H	H		H	N	C	C	C	N	3.9666	3.967	-0.001
18	COOEt	H	H		Cl	N	C	C	C	CH	3.9547	3.915	0.040
19	CH <sub>3</sub>	H	H		Cl	N	C	C	C	CH	3.8447	4.046	-0.201
20	CH <sub>3</sub>	H	CH <sub>3</sub>		H	N	C	C	C	CH	3.7696	3.771	-0.001
21	COOEt	H	CH <sub>3</sub>		H	N	C	C	C	CH	3.4685	3.510	-0.041
22	COOEt	OCH <sub>3</sub>	H		H	N	C	N	C	N	3.4389	3.458	-0.019
23	NO <sub>2</sub>	OCH <sub>3</sub>	H		H	N	C	N	C	N	3.4293	3.476	-0.046
24	CH <sub>3</sub>	OCH <sub>3</sub>	H		H	N	C	N	C	N	3.3925	3.363	0.029
26	Cl	OCH <sub>3</sub>	H		H	N	C	N	C	N	3.1278	3.032	0.096
27	Cl	N(CH <sub>3</sub> ) <sub>2</sub>			H	N	C	N	C	N	3.4293	3.150	-0.075
28 <sup>b</sup>	NO <sub>2</sub>	CH <sub>3</sub>	H		H	N	C	C	C	N	6.3840	5.801	-0.583
29 <sup>b</sup>	Cl	OCH <sub>3</sub>	H		H	N	C	C	C	N	5.3915	5.179	-0.213
30 <sup>b</sup>	COOEt	H	H		H	N	C	C	C	N	4.2757	5.153	0.877
31 <sup>b</sup>	COOEt	H	H	Cl	H	C	C	C	C	CH	3.9706	4.100	0.129
32 <sup>b</sup>	COOEt	H	H		H	N	C	C	C	CH	3.8327	3.531	-0.302

<sup>a</sup> The values of pI<sub>50</sub> were calculated using the CoMSIA(4) model in Table 4 based on field-fit alignment. <sup>b</sup> These compounds were used as a test set and not included in the derivation of equations.

method. The PI<sub>50</sub> potency was used as a dependent variable in the QSAR study (see Table 1).

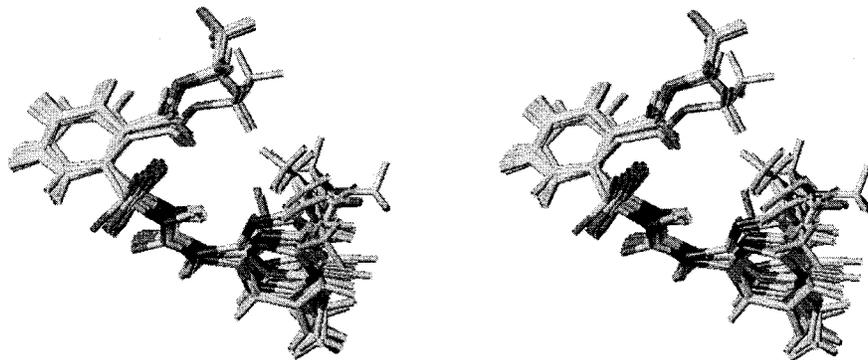
## METHODS

**Principle of CoMSIA.** Three-dimensional QSAR (3D-QSAR) analysis mathematically relates the three-dimensional physicochemical and biological properties of a set of small molecules. Since its advent in 1988, the comparative molecular field analysis (CoMFA) has become one of the most powerful tools for 3D-QSAR studies.<sup>5</sup> The basic assumption for CoMFA is that the observed biological properties can be well-understood or -correlated with the suitable samplings of the steric and electrostatic fields surrounding a set of ligands. Recently, another 3D-QSAR technique, comparative molecular similarity indices analysis (CoMSIA), has been reported.<sup>6</sup> This technique was devised to overcome the problems in CoMFA of very rapidly changing steric fields near the atomic nuclei, and scaling the two fields for PLS analysis, which can avoid some inherent deficiencies arising from the functional forms of the Lennard-Jones and Coulomb potentials used in CoMFA. In CoMSIA, a distance-dependent Gaussian-type functional form is introduced, which can avoid

singularities at the atomic positions and the dramatic changes of potential energy for these grids in the proximity of the surface. Meanwhile, no arbitrary definition of cut-off limits is required in CoMSIA. Moreover, using CoMSIA, the contour maps of the relative spatial contributions can be substantially improved, which are intuitively interpreted in terms of the separate property. The unique differences between conventional CoMFA and CoMSIA are the field-type and the potential functional forms. 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. The indices  $A_{F,K}$  between the compounds of interest and a probe atom are calculated according to

$$A_{F,K}^q(j) = - \sum_{i=1}^n \omega_{\text{probe},k} \omega_{ik} e^{-\alpha r_{iq}^2}$$

where  $i$  is the summation index over all atoms of the molecule  $j$  under investigation;  $\omega_{ik}$  is the actual value of the physicochemical property  $k$  of the atom,  $\omega_{\text{probe},k}$  is the probe atom with charge +1, radius 1 Å, hydrophobicity +1, and



**Figure 1.** Stereoview of all aligned compounds in the training and test set.

H-bond donor and acceptor property +1;  $a$  is the attenuation factor; and  $r_{iq}$  is the mutual distance between the probe atom at grid point  $q$  and atom  $i$  of the investigated molecule.

**Molecular Modeling.** All structures in Table 1 were constructed using the SYBYL molecular simulation package.<sup>7</sup> It is well-known that the selection of bioactive conformer and the alignment rule are often the most important parts in the 3D-QSAR studies. In principle, the real bioactive conformation and the best alignments can only be obtained from the complex structure of ligand and receptor. Sometimes, the complex structure, even the receptor structure, cannot be obtained. But in most cases, the bioactive conformer will adopt a relatively low-energy conformation, so it is suitable to utilize the crystal structure of the compound as the starting geometry. Considering the compound in the solvent interacts with the receptor in terms of the free molecule, we submitted the X-ray structure of no. 10 in Table 1 as the template.<sup>8</sup> Other sulfonyleureas were constructed by modifying the template. The initial structures were first minimized using the molecular mechanism with the Tripos force field,<sup>9</sup> followed by full geometry optimizations using the PM3 method, available in MOPAC 7.0.<sup>10</sup> It seemed that by adopting the technique of semiempirical geometry optimizations, reasonable structures would be obtained for the molecules revised from an X-ray crystal structure.

**Molecular Alignments.** Besides the selection of bioactive conformer, the alignment rule may be another important factor affecting the 3D-QSAR analysis because different alignment will produce different field samplings surrounding the molecular aggregate. Different alignment rules have been compared in many previous studies. It seems that there does not really exist an alignment rule which is better than the other ones. We think that the alignment rule should be carefully selected according to different cases. In this paper, two different kinds of alignment rules, including field-fit alignment and atom-by-atom fits, were applied.

Atom-by-atom fits may be the most widely used alignment rule. In this paper, a rigid alignment was applied to superimpose all 32 compounds onto the fit centers (atoms labeled with asterisks) shown in Table 1 using an atom-by-atom least-squares fit implemented in the "SYBYL FIT" option in SYBYL. Compound 1 with the best biological activity was selected as the reference molecule.

Another molecular aggregate was produced by sterically and electrostatically aligning the molecules using the Align module in Cerius2,<sup>11</sup> and compound 1 was treated as the reference molecule. The field alignment fixed the position of the reference molecule and moved the second molecule

to a random starting position with respect to the first molecule. Optimal electrostatic and steric complementarity between the molecules was achieved by the implementation of a simple two-component force-field algorithm. In order for this procedure to be carried out, partial charges were required for each molecule. The PM3 molecular electrostatic fitted atomic charges by MOPAC calculations were assigned, and van der Waals parameters available in the CVFF force field<sup>12</sup> were used to estimate the steric field.

**CoMSIA Analyses.** In this paper, a training set of 27 sulfonyleureas was used for CoMSIA analyses (Figure 1). In addition, five compounds, randomly selected from various ranges of biological activity, were kept to test the actual prediction of the obtained 3D-QSAR models. The CoMSIA analyses were carried out on the aligned molecules, within the QSAR module of SYBYL running on a Silicon Graphics Octane 2-CPU workstation. In the present CoMSIA analyses, five kinds of physicochemical properties were evaluated, including steric contributions by the third power of the atomic radii, electrostatics by PM3 molecular electrostatic potential fitted atomic charges, hydrophobicities by atom-based hydrophobic parameters and hydrogen bond properties by suitably placed pseudoatoms, using a common probe with a 1 Å radius, +1 charge, +1 hydrophobicity, and H-bond property of +1. The dimensions of the surrounding lattice were selected with a sufficiently large margin (=4 Å) to enclose all the aligned molecules.

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 SAMPLES method.<sup>13</sup> Subsequently, the final 3D-QSAR models were derived from the non-cross-validated calculations. The CoMSIA results were graphically interpreted by the field contribution maps using the field type "stdev\*coeff".

A series of CoMSIA analyses were carried out to calibrate two important parameters including the grid spacing and the attenuation factor, which were closely concerned with the quality of the final models. At first, a regular 3D lattice with 2 Å grid spacing was created automatically, extending beyond the molecular dimensions of the molecular aggregates by 4.0 Å in all directions. An additional five lattices with 1.0, 1.5, 2.5, 3.0, and 3.5 Å grid spacing with the same orientation were subsequently generated to select the most optimal grid spacing. Then, for the best grid spacing, the attenuation factor was varied in a parameter study within the range from 0.1 to 0.8 in steps of 0.1; subsequently, the similarity indices and the  $q^2$  values were computed each time.

**Table 2.** Results of the CoMSIA Analyses of Several Different Field Combinations at 2.0 Å Grid Spacing

	CoMFA	steric + electrostatic	steric + electrostatic + hydrophobic	steric + electrostatic + hydrophobic + H-bond <sup>a</sup>	steric + electrostatic + hydrophobic + H-bond <sup>b</sup>
$q^2$	0.487	0.558	0.741	0.909	0.900
$r^2$	0.846	0.876	0.951	1.000	1.000
std error of estimate	0.558	0.512	0.331	0.037	0.015
$F$	40.141	36.492	76.922	1979.680	9799.900
no. of compds	3	4	5	16	19
fraction					
steric	0.864	0.566	0.240	0.120	0.144
electrostatic	0.136	0.434	0.152	0.117	0.088
hydrophobic			0.608	0.408	0.451
H-bond donor				0.300	0.267
H-bond acceptor				0.055	0.050

<sup>a</sup> The 3D-QSAR model from field-fit alignment. <sup>b</sup> The 3D-QSAR model from atom-by-atom fits.

Comparatively, a conventional CoMFA was performed with the usually used steric and electrostatic fields implemented in SYBYL. The PM3 molecular electrostatic potential fitted atomic charges were applied in the determination of the electrostatic field. All CoMFA calculations were performed with the SYBYL standard setup (steric and electrostatic fields with Lennard-Jones and Coulomb-type potentials; dielectric constant,  $1/r$ ; cutoff, 30 kcal/mol) using an  $sp^3$  carbon atom with a charge of  $+1.0|e|$ . The extent and orientation of the grids surrounding the tested molecules were the same as those in the CoMSIA analyses, and the grid spacing was set to 2.0 Å.

## RESULTS AND DISCUSSION

**CoMSIA Results from Two Different Kinds of Alignment Rules.** First, for the molecular aggregate after field-fit alignment, several CoMSIA models were generated at 2.0 Å grid spacing using several combinations of different fields (see Table 2). Only using the steric and the electrostatic fields, the 3D-QSAR model from the CoMSIA analyses did not bear good statistical significance ( $q^2 = 0.558$ ); meanwhile the model from CoMFA was also very dissatisfactory ( $q^2 = 0.487$ ). So, it could be concluded that the biological activity could not be well-expressed by only using the steric and the electrostatic fields. After the hydrophobic field was added, the predictive power of the 3D-QSAR model ( $q^2 = 0.741$ ) was greatly improved, which meant that the biological activity essentially exhibited a significant relationship with the hydrophobic field. Moreover, after considering the H-bond fields, the quality of the 3D-QSAR model was promoted significantly, indicating that the biological activity was closely concerned with H-bond properties. The best 3D-QSAR was derived from use of all five fields afforded by CoMSIA, which yielded the best statistical significance ( $q^2 = 0.909$ ,  $n = 16$ ,  $SD = 0.037$ ). So, in the present work, we considered five fields to produce the best 3D-QSAR model. From the statistical results of these models using different field combinations (Table 2), it was unlikely that they were completely independent from each other; however, the degree of the interdependence was difficult to estimate. Focusing solely on the predictive power of the QSAR model could not justify the consideration of five different kinds of fields. However, considering five fields opened the opportunity to partition the variance analysis with respect to particular physicochemical properties associated with the molecules. This aspect gained importance if results of comparative molecular fields analyses were used for the design of improved ligands.

**Table 3.** Difference between Predicted and Actual Activities for Five Molecules in the Test Set Using Three Models in Table 2

compd	CoMFA $pI_{50}(\text{calcd})$	CoMSIA(1) <sup>a</sup> $pI_{50}(\text{calcd})$	CoMSIA(2) <sup>b</sup> $pI_{50}(\text{calcd})$
<b>28</b>	4.364	5.801	5.225
<b>29</b>	5.497	5.179	5.627
<b>30</b>	4.399	5.153	5.033
<b>31</b>	3.958	4.100	3.643
<b>32</b>	4.083	3.531	4.191
$r_{\text{pred}}^2$	0.230	0.735	0.536
SSE <sup>c</sup>	4.170	1.262	2.208

<sup>a</sup> The CoMSIA model derived from field-fit alignment using five different kinds of fields. <sup>b</sup> The CoMSIA model derived from atom-by-atom fits using five different kinds of fields. <sup>c</sup> Sum of square error of predictions for five tested compounds.

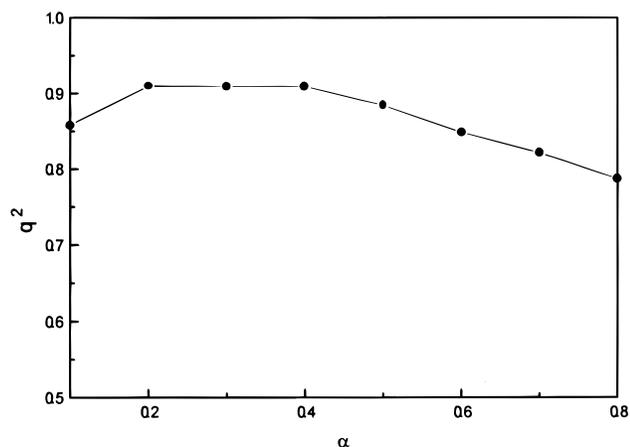
From Table 2, it could be found that the 3D-QSAR model based on field-fit alignment yielded a very good correlation with a cross-validated  $r^2$  of 0.909 ( $F = 1979.680$ ) using 16 principle components (PCs). This model expressed good predictive ability (Table 3) for the external test set ( $r_{\text{pred}}^2 = 0.735$ ,  $SSE = 1.262$ ) with an average absolute error of 0.452 log units across a range of 4.771 log units.

The analysis from simple atom-by-atom fits yielded a little worse model ( $q^2 = 0.900$ ,  $F = 9799.9$ ); meanwhile the predictive result for the test-set compounds indicated poor predictive ability ( $r_{\text{pred}}^2 = 0.536$ ,  $SSE = 2.208$ ). These results were possibly indicative of poor molecular alignment caused by the structural differences represented by different charged substituents and the resultant changes in charge distribution. It seemed that in using field-fit alignment, the actual field properties around the molecular aggregate could be expressed better than those using simple atom-by-atom fits.

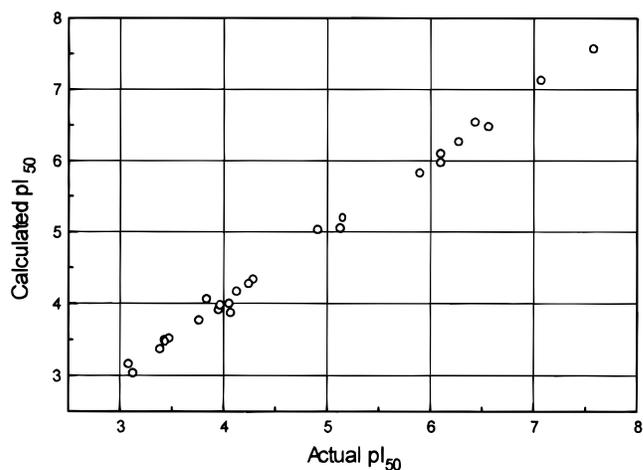
**Influences of Grid Spacing and Attenuation Factor.** At the start of this study, we accepted a default grid spacing of 2.0 Å, as suggested in SYBYL. Moreover, five additional grid spacings with 1.0, 1.5, 2.5, 3.0, and 3.5 Å were used for CoMSIA analyses (see Table 4). When the grid spacing was smaller than 3.0 Å, the shift of the  $q^2$  values for the 3D-QSAR models was not very obvious. But when grid spacing was defined as 3.5 Å, the statistical significance and the predictive ability were very poor. It was evident that when the grid spacing was equal to or smaller than 3.0 Å, the field properties around the molecular aggregate could be well-expressed, while when the grid spacing was defined as 3.5 Å, some important information of the field properties in some regions would be lost. From the predictive ability of the external test set, the model with 3.0 Å grid

**Table 4.** Results of the CoMSIA Analysis Using Different Grid Spacings

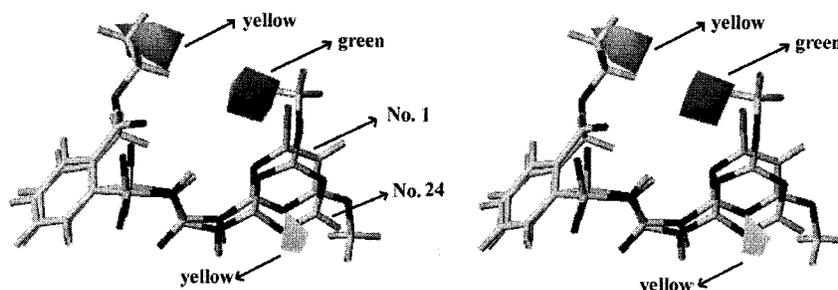
	CoMSIA(1)	CoMSIA(2)	CoMSIA(3)	CoMSIA(4)	CoMSIA(5)	CoMSIA(6)
Field-Fit-Alignment						
$q^2$	0.890	0.873	0.909	0.873	0.897	0.603
$r^2$	1.000	1.000	1.000	1.000	0.996	0.973
std error of estimate	0.074	0.034	0.037	0.037	0.062	0.265
$F$	2564.942	2423.942	1979.680	1365.312	353.32	76.585
no. of compds	16	16	16	16	10	8
fraction						
steric	0.133	0.118	0.120	0.120	0.123	0.211
electrostatic	0.074	0.080	0.117	0.078	0.096	0.079
hydrophobic	0.365	0.402	0.409	0.400	0.390	0.373
H-bond donor	0.374	0.350	0.300	0.348	0.339	0.276
H-bond acceptor	0.054	0.051	0.055	0.054	0.053	0.061
grid spacing (Å)	1.0	1.5	2.0	2.5	3.0	3.5
pI <sub>50</sub> (calcd)						
compd 28	5.803	5.882	5.801	5.812	5.801	6.272
compd 29	5.192	5.123	5.179	5.120	5.179	4.572
compd 30	5.109	5.171	5.153	5.200	5.133	5.260
compd 31	4.795	4.009	4.100	4.009	4.100	2.490
compd 32	4.050	4.009	3.531	4.107	3.531	3.640
$r_{\text{pred}}^2$	0.750	0.750	0.735	0.752	0.743	0.605
SSE	1.799	1.799	1.262	1.332	1.228	3.882
Atom-by-Atom Fits						
$q^2$	0.901	0.884	0.900	0.872	0.879	0.660
$r^2$	1.000	1.000	1.000	0.983	0.998	0.952
std error of estimate	0.044	0.037	0.015	0.212	0.078	0.342
$F$	2587.332	1857.942	1979.999	120.731	560.812	515.559
no. of compds	16	16	19	8	13	7
fraction						
steric	0.131	0.142	0.144	0.155	0.118	0.170
electrostatic	0.083	0.086	0.088	0.114	0.107	0.117
hydrophobic	0.372	0.447	0.451	0.407	0.397	0.403
H-bond donor	0.331	0.265	0.267	0.270	0.339	0.259
H-bond acceptor	0.083	0.048	0.050	0.054	0.039	0.051
grid spacing (Å)	1.0	1.5	2.0	2.5	3.0	3.5
pI <sub>50</sub> (calcd)						
compd 28	5.425	5.325	5.225	5.454	5.426	5.161
compd 29	5.337	5.513	5.627	5.337	5.445	5.207
compd 30	4.907	5.031	5.033	4.830	4.891	4.614
compd 31	3.615	3.378	3.643	3.650	3.373	3.821
compd 32	4.195	3.987	4.191	3.863	4.196	3.821
$r_{\text{pred}}^2$	0.687	0.585	0.536	0.766	0.635	0.778
SSE	1.579	2.082	2.208	1.279	1.788	1.667

**Figure 2.** Variation of  $q^2$  upon changes of the attenuation factor  $\alpha$  used in the distance dependence between the probe atom and the atoms of the molecules in CoMSIA.

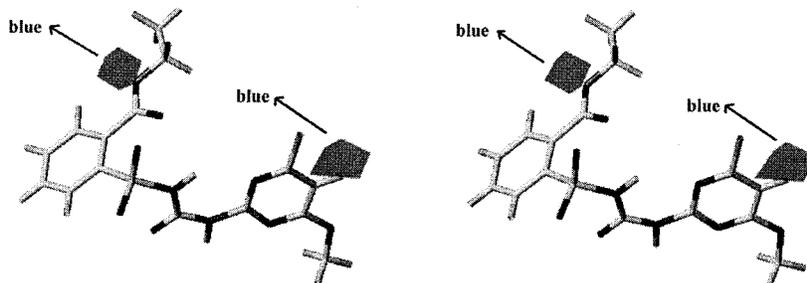
spacing based on field-fit alignment could be selected as the best model. In Table 4, the influence of different grid spacings to these models based on atom-by-atom fits were also investigated, and their statistical significances were

**Figure 3.** Comparison of experimental pI<sub>50</sub> with calculated pI<sub>50</sub> obtained using the 3D-QSAR model from CoMSIA(3) based on field-fit alignment in Table 3.

comparative with those from field-fit alignment, but their predictive abilities seemed to be slightly poorer than  $r_{\text{pred}}^2$  and SSE of the test set.



**Figure 4.** Stereoview of the contour plots of the CoMSIA 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. Compounds **1** and **24** are shown as the reference compounds.



**Figure 5.** Stereoview of the contour plots of the CoMSIA 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 **1** is shown as the reference compound.

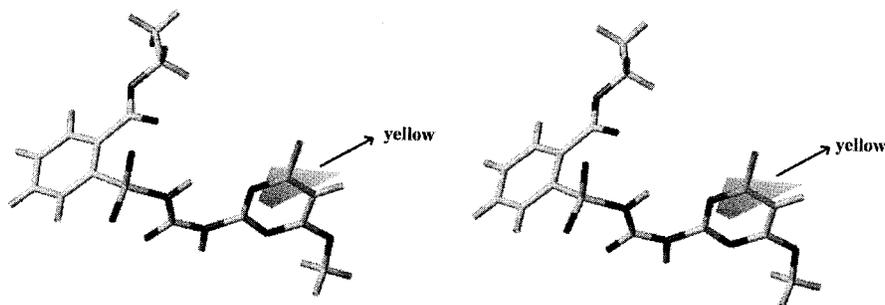
At 3.0 Å grid spacing, another important factor contributing to the CoMSIA analysis, the attenuation factor  $\alpha$ , was also investigated. A series of  $\alpha$  were selected from 0.1 to 0.8 (see Figure 2). In CoMSIA, a Gaussian-type distance dependence was applied. Reducing  $\alpha$  to smaller value meant that a probe placed at a particular lattice point detected molecular similarity in its neighborhood more globally. On the other hand, larger values of  $\alpha$  implied a more localized evaluation of similarity. Figure 2 shows the changes of  $q^2$  of different 3D-QSAR models with different  $\alpha$ . For the data set, the values of  $\alpha = 0.2, 0.3,$  and  $0.4$  produced similar statistical models, but the predictive ability of the test set showed that the model of  $\alpha = 0.3$  ( $r_{\text{pred}}^2 = 0.735$ , SSE = 1.262) was much better than other two models using  $\alpha = 0.2$  ( $r_{\text{pred}}^2 = 0.700$ , SSE = 1.446) and  $\alpha = 0.4$  ( $r_{\text{pred}}^2 = 0.699$ , SSE = 1.436).

As a whole, the 3D-QSAR models from the CoMSIA analyses were not very sensitive to the changes of the grid spacing and the attenuation factor, but careful calibrations should be needed to get the best model. Through careful comparisons of different grid spacings and attenuation factors, the model with 3.0 Å grid spacing ( $\alpha = 0.3$ ) was selected as the best model judged by the cross-validation correlation ( $q^2 = 0.897$ ,  $F = 353.32$ ,  $SD = 0.062$ ) and the predictive ability of the test set ( $r_{\text{pred}}^2 = 0.743$ , SSE = 1.228), and the following discussions would only refer to this model. The biological activities ( $\text{pI}_{50}$ ), the calculated activities using the CoMSIA(5) model in Table 4 based on field-fit alignment and the residue values from the observed values for the training set, were shown in Table 1. Figure 3 shows a plot of the observed vs the calculated biological activities. The predicted biological activities of the test-set compounds were listed in Table 4. The best 3D-QSAR model predicted well for the five tested compounds in terms of actual prediction, and the predictions were even better than some molecules

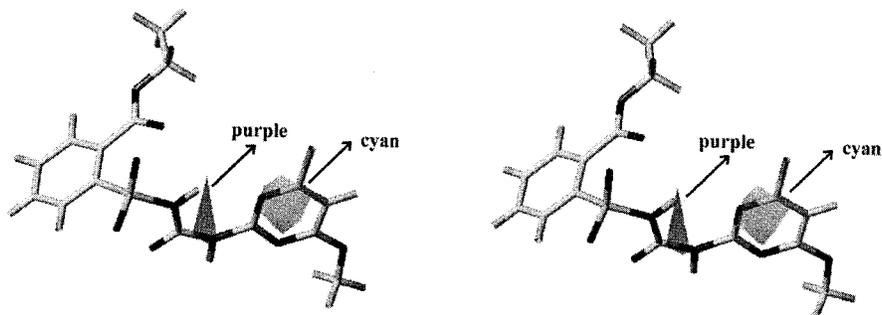
in the training set. So the derived model was very satisfactory from the viewpoint of statistical significance and actual predictive ability. On the basis of the best 3D-QSAR model obtained, we expected to find more potential compounds with the aid of the computational combinatorial chemistry method.

**CoMSIA Contour Maps.** The field type "stdev\*coeff" was used to obtain contours from a CoMSIA analysis that elucidated the relationship between differences in the fields and variations in the dependent variable. Graphical representations of the best CoMSIA model are displayed in Figures 4–8. They show regions where variations of the steric, electrostatic, hydrophobic, and H-bond natures in the structural features of different molecules contained in the training set led to increase or decrease the biological activity.

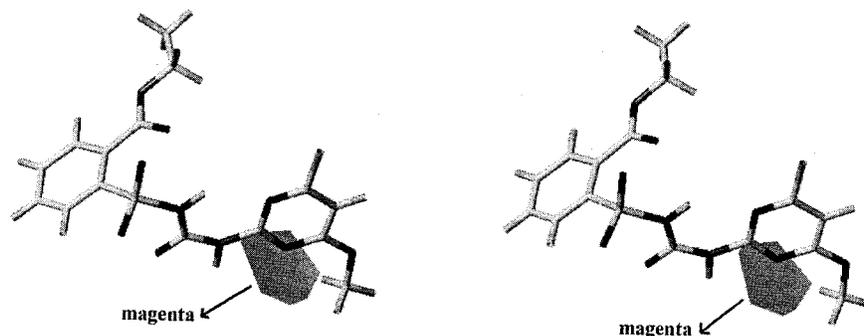
Figure 4 shows that the molecules orienting groups with increasing bulk occupancy into areas contoured in green would enhance the biological activity, as well as the groups with a smaller group placed into areas indicated in yellow. One bulky favorable area was near the  $R_2$  group on the pyrimidinyl moiety, while two bulky unfavorable areas were near the  $R_1$  group on the benzene ring and the Q group on pyrimidinyl part, respectively. The different substituted groups would introduce a different influence on the conformation of one molecule and the steric complementarity between the receptor and the ligand. For example, in Table 1, we could see that when the Z group was substituted from C atom to N atom (see compounds **22–27**), the biological activity would decrease substantially, which could be partly interpreted from the distribution of the steric field. In Figure 4, compounds **1** and **24** were superimposed with the steric contour together. Compared with the conformations of compounds **1** and **24**, it can be seen that their conformations near the benzene ring were similar, but their conformations near the fused heterocycle moieties were quite different. The



**Figure 6.** Stereoview of the contour plots of the CoMSIA hydrophobic fields (stdev\*coeff). The favorable hydrophobic areas are indicated by yellow isopleths, whereas the disfavored hydrophobic areas are shown by white isopleths. The most active compound **1** is shown as the reference compound.



**Figure 7.** Stereoview of the contour plots of the CoMSIA H-bond acceptor fields (stdev\*coeff). Magenta isopleth contours maps beyond the ligands where an H-bond acceptor group in the receptor will be favorable for biological activity, while red isopleths represent an H-bond donor in the ligands unfavorable for biological activity. The most active compound **1** is shown as the reference compound.



**Figure 8.** Stereoview of the contours plots of the CoMSIA H-bond acceptor fields (stdev\*coeff). Magenta isopleth contours maps beyond the ligands, where an H-bond acceptor group in the receptor will be favorable for biological activity, while red isopleths represent an H-bond donor in the receptor unfavorable for biological activity. The most active compound **1** is shown as the reference compound.

introduced N atom in the Z group made some parts of the pyrimidinyl moiety contact with the smaller unfavorable bulky area, which would introduce bad surface complementarity between the ligand and the receptor, and subsequently, it would decrease the biological activity.

From the fraction of the fields, it seemed that the electrostatic field contributed less compared with the steric field. Close inspection of the electrostatic contour plots revealed that the tested molecules were more favorable to positive charge (see Figure 5). Two blue areas indicated that more positive charge substituents near the R<sub>2</sub> and R<sub>5</sub> groups would enhance the biological activity.

The maps of the hydrophobic properties revealed that the very distinct hydrophobic favored site was near the Y substituted site. It seemed that the group near the yellow area with high hydrophobicity would be favorable to biological activity. It could be reasonably presumed that the group substituted at R<sub>2</sub> would produce hydrophobic interactions with the receptor. From the fraction of fields, the hydrophobic

effects seemed to be the most important factor contributing to biological activity.

The graphical interpretations of the field contributions of the H-bond properties were shown in Figure 7 (H-bond donor field) and Figure 8 (H-bond acceptor field). In principle, they should highlight areas beyond the ligands where putative hydrogen partners in the enzyme could form H-bonds that influence binding affinity. In Figure 7, cyan isopleth contours map beyond the ligands where an H-bond donor group in the receptor would be favorable for biological activity, while purple isopleths represented H-bond donor in the receptor unfavorable for biological activity. Some areas near the X and Y groups, which were represented with cyan areas, were indicated as favorable H-bond donor sites. In fact, the N atoms in the X or Y groups were frequently involved as a H-bond acceptor in a hydrogen bond, which perhaps could form hydrogen bonds with H-bond donors in the receptor. The region contoured in purple should lack H-bond donor capabilities, which was near two amide groups in the

connecting areas between the benzene ring and the pyrimidinyl moiety. For a putative partner in the ligand, the amide groups near the purple area should act as H-bond donors, which certainly lacked H-bond acceptor capabilities. Compared with the H-bond donor field, the influence of the H-bond acceptor field seemed to be not very important, because its fraction of the field was very small. In the H-bond acceptor field, the magenta isopleth was around the Q group in the pyrimidinyl moiety, which was difficult to interpret because the nearby atoms in the Q group could not act as a H-bond donor and it could only be assumed that this region might be produced by the amide group and the R<sub>3</sub> substituents near the Q group.

### CONCLUSIONS

In this study, the CoMSIA approach was applied to correlate the herbicidal activities against rape with the steric, electrostatic, hydrophobic, and H-bond fields of a set of sulfonylurea analogues. With these five types of fields provided by CoMSIA, two sets of 3D-QSAR models were constructed by using two different kinds of alignment rules: field-fit alignment and atom-by-atom fits. The results indicated that the two alignment rules used could generate comparative 3D-QSAR models with similar statistical significance. However, from the predictive ability of the test set, the models from maximal steric and electrostatic alignment were slightly better than those from the simple atom-by-atom fits. The calculation results from the systematic variation of grid spacing and attenuation factor revealed that the quality of the 3D-QSAR model was not very sensitive to the changes of these two parameters.

The best model possesses promising predictive ability as indicated by the high cross-validated correlation and the prediction on the external test set, which was significantly superior to the model constructed by conventional CoMFA. Moreover, some important factors contributing to the biological activity were reflected by the contour maps of different properties. The steric, hydrophobic, and H-bond donor fields seemed to affect the biological activity significantly. The characteristics of the CoMSIA 3D contour plots derived in this study were very helpful for us to understand

the underlying mechanism of the receptor–drug interaction and are expected to provide significant information for designing new potential drugs.

### ACKNOWLEDGMENT

This project was supported by NCSF Grants 29992590-2 and 29573095 of the People's Republic of China, and the state laboratory of elemento-organic chemistry of Nankai University.

### REFERENCES AND NOTES

- (1) Stetter, J. *Herbicides Inhibitors Branched-Chain Amino Acid Biosynthesis: Recent Developments*; Springer-Verlag: New York, 1994.
- (2) Kleschick, W. A.; Costales, M. J.; Dunbar, J. E.; Meikle, R. W.; Monte, W. T.; Peason, N. R.; Snider, S. W.; Vinogradoff, A. P. New Herbicidal Derivatives of 1,2,4-Trizolo[1,5-A] Pyrimidine. *Pestic. Sci.* **1990**, *29*, 341–355.
- (3) Li, Z. M. Chemistry of novel bio-regulating substances. *J. Pesticide Sci.* **1996**, *21*, 124–128.
- (4) Levitt, G. U.S. Pat. 4 127 405, 1978.
- (5) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular-Field Analysis (Comfa). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (6) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indexes In A Comparative-Analysis (Comsia) Of Drug Molecules To Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- (7) *SYBYL 6.5 User Guide*; Tripos Inc., St. Louis, MO, 1999.
- (8) Li, Z. M.; Jia, G. F.; Wang, X. L.; Lai, C. M. Syntheses, Structure and SAR Study on Sulfonylurea Compounds (III)—Crystal and Molecular Structure of N-[2'-(4,6-Dimethyl)pyrimidinyl]-2-ethyl Carboxylate-benzene Sulfonylurea. *Chem. J. Chin. Univ.* **1994**, *15*, 227–229.
- (9) Clark, M.; Cramer, R. D.; Vanopdenbosch, H. Validation Of The General-Purpose Tripos 5.2 Force-Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (10) *MOPAC 7.0 User Guide*; Quantum Chemistry Program Exchange (QCPE), Indiana University: Bloomington, IN, 1993.
- (11) *Cerius2 User Guide*; Molecular Simulation, Inc.: San Deigo, CA, 1998.
- (12) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and Energetics of Ligand-Binding to Proteins—Escherichia-Coli Dihydrofolate Reductase Trimethoprim, A Drug-Receptor System. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 31–47.
- (13) Bush, B. L.; Nachbar, R. B. Sample Distance Partial Least Squares: PLS Optimized for Many Variables, with Application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.

CI000323E