3D QSAR Analyses of Novel Tyrosine Kinase Inhibitors Based on Pharmacophore Alignment

L. L. Zhu,[†] T. J. Hou,[†] L. R. Chen,[‡] and X. J. Xu*,[†]

College of Chemistry and Molecular Engineering and Department of Technical Physics, Peking University,
Beijing 100871, People's Republic of China

In an effort to develop a quantitative ligand-binding model for the receptor tyrosine kinases, a pharmacophore search was first used to identify structural features that are common in two novel sets of 12 molecules of the 3-substituted indolin-2-ones and 19 compounds of the benzylidene malononitriles with low-to-high affinity for HER2, a kind of receptor tyrosine kinase. The common pharmacophore model based on these 31 compounds was used as a template to obtain the aligned molecular aggregate, which provided a good starting point for 3D-QSAR analysis of only the 19 benzylidene malononitriles. Two molecular field analysis (MFA) techniques, including CoMFA and CoMSIA, were used to derive the quantitative structure−activity relationships of the studied molecules. From the studied results, it was obvious that the 3D-QSAR models based on the pharmacophore alignment were superior to those based on the simple atom-by-atom fits. Considering the flexibility of the studied molecules and the difference between the active conformers and the energy-lowest conformers, the pharmacophore model can usually provide the common features for the flexible regions. Moreover, the best CoMSIA model based on the pharmacophore hypothesis gave good statistical measure from partial least-squares analysis (PLS) ($q^2 = 0.71$), which was slightly better than the CoMFA one. Our study demonstrated that pharmacophore modeling and CoMSIA research could be effectively combined. Results obtained from both methods helped with understanding the specific activity of some compounds and designing new specific HER2 inhibitors.

INTRODUCTION

The human epidermal growth factor receptor-2 (HER2) proto-oncogene encodes a 185-kDa glycoprotein, often simply called the HER2 or c-erbB2 protein or receptor. HER2 plays a key role in one of the best studied growth factor receptor systems in breast cancer, the HER (or erbB, or type 1) tyrosine kinase receptor family. This family comprises four homologous epidermal growth factor (EGF) receptors: HER1 (EGFR/erbB1), HER2 (erbB2), HER3 (erbB3), and HER4 (erbB4).[1] Each of these receptors comprises an extracellular ligand-binding domain, a transmembrane lipophilic segment, and an intracellular protein kinase domain with a carboxyl terminal segment containing sites of phosphorylation or tyrosine residues.[2] Amplification of the HER2 gene has been found in about 30% of primary breast, ovary, and gastric carcinomas.[3] Survival rates and tumor aggressiveness can be directly correlated to the level of HER2 expression. In this regard, regulating the cellular signal transduction via inhibition of HER2 has been considered a promising way of controlling malignant tumors.

Recently, two groups of small molecules have been found to possess obvious inhibitory activity to HER2, including 19 compounds belonging to the benzylidene malononitrile family[4] and 12 compounds belonging to the 3-substituted indolin-2-one family.[5] Since no structure information is available about the inhibitors in complex with HER2 receptor, a ligand-based approach was implemented in an attempt to understand the important interactions necessary for ligand binding. Until now, however, studies on the pharmacophore modeling and the relationship between the chemical structures and the biological functions of these two kinds of compounds have never been reported. The first aim of the present research is to obtain the possible pharmacophore model based on these two groups of available compounds and to determine the possible important functional groups or chemical elements crucial for ligand binding. A further important role is to develop a QSAR model and take insight into the main intermolecular interactions involved in the HER2 inhibitions at the 3D-level. Moreover, in the current work we want to compare two alignment rules based on pharmacophore model and atom-by-atom fits. To reach our research objectives three modeling techniques have been used, including the pharmacophore modeling procedure implemented in the CATALYST 4.0,[6] the comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) in the SYBYL 6.5 molecular simulations package.[7]

CoMFA is a widely used technique for the study of quantitative structure−activity relationship (QSAR) at the 3-D level. Unlike the traditional 2D-QSAR methods, for example, the Hansch analysis, which relies on substituent parameters, CoMFA uses the active conformer and superposition rule for a set of molecules, supplied by pharmacophore mapping, to provide quantitative forecasts of

---

* Corresponding author phone: 8610-62757456; e-mail: xiaojxu@chemms.chem.pku.edu.cn.
† College of Chemistry and Molecular Engineering.
‡ Department of Technical Physics.

Table 1. Structure and Experimental Biological Activity of Benzylidene Malononitrile Family
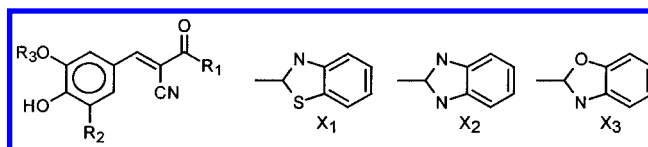
no.	R_1	R_2	R_3	$IC_{50}, \mu M$
1	-NH-Ph	-H	-H	45 ± 4.3
2	-NH-CH$_2$-Ph	-H	-H	12.1 ± 2.2
3	-NH-(CH$_2$)$_2$-Ph	-H	-H	9.4 ± 0.4
4	-NH-(CH$_2$)$_3$-Ph	-H	-H	33 ± 5
5	-NH-(CH$_2$)$_4$-Ph	-H	-H	22 ± 6
6	-Ph	-H	-H	20 ± 3.5
7	-NH-CH$_2$-(4'-OH)-Ph	-H	-H	2.9 ± 0.3
8	-NH-(4'-Cl)-Ph	-H	-H	62 ± 7
9	-NH-(2',4'-di-OMe)-Ph	-H	-H	20 ± 1.5
10	-NH-(2',6'-di-Me)-Ph	-H	-H	44 ± 13
11	-NH-(2',4',6'-tri-Me)-Ph	-H	-H	21 ± 6
12	-NH-cyclo-C$_6$H$_{11}$	-H	-H	19 ± 3
13	-NH$_2$	-CH$_2$-S-Ph	-Me	0.13 ± 0.007
14	-NH$_2$	-CH$_2$-S-(2'-COOH)-Ph	-Me	0.45 ± 0.07
15	-NH$_2$	-CH$_2$-S-(4'-Me)-Ph	-Me	1.65 ± 0.19
16	-NH$_2$	-CH$_2$-S-CH$_2$-Ph	-Me	0.2 ± 0.03
17	-NH$_2$	-CH$_2$-S-X$_1$	-Me	0.35 ± 0.07
18	-NH$_2$	-CH$_2$-S-X$_2$	-Me	1.5 ± 0.14
19	-NH$_2$	-CH$_2$-S-X$_3$	-Me	6.1 ± 0.1

potency.[8] CoMFA combines features of QSAR and pharmacophore mapping with insights gained from protein crystallography and theoretical investigations of intermolecular interactions.[9,10] The 3D-QSAR model from a CoMFA procedure can be summarized into three steps as following: (1) The first, all investigated molecules are structure-based or field-based aligned. (2) Then, an evenly spaced, rectangular or circular grids are generated to enclose the molecule aggregate. A probe atom, e.g. sp^3 carbon with $+1$ or -1 charge, will be placed at every lattice point to measure the electrostatic or steric field by using molecular mechanics. (3) Last, the results from the field samplings combined with the biological activities from the tested compounds are put into a table, and the partial least-squares (PLS) is applied to get the final CoMFA model. Generally, a leave-one-out cross-validated r^2 (q^2) is used as a quantitative measure for a CoMFA model. The graphical representations of CoMFA results are indicated by the regions where the variation in steric and electrostatic properties of different molecules in a data set is correlated with the variation of biological activity.

Recently, another 3D-QSAR technique, comparative molecular similarity indices analysis[11−13] (CoMSIA), has been reported, which was devised to overcome the problems in CoMFA of very rapidly changing steric fields near the atomic nucleic and scaling the two fields for PLS analysis. In CoMSIA analysis, 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 these grids in the proximity of the surface, thus no arbitrary definition of cutoff limits is required in CoMSIA. The result of CoMSIA is less allergic to the relative orientation of the aligned molecules respecting to the lattice, which is another well-known problem in CoMFA. Moreover, CoMSIA encodes hydrogen bonds and hydrophobic interactions those are not sufficiently described by the steric and electrostatic fields and includes also an entropy component. For these reasons, the CoMSIA methodology

was selected as an appropriate tool to study the QSAR of tyrphostins.

COMPUTATIONAL DETAILS

In the present study, two groups of tyrosine kinase inhibitors with different structures, including 19 compounds belonging to the benzylidene malononitrile family and 12 compounds belonging to the 3-substituted indolin-2-one family, were investigated to generate the pharmacophore model. And then only the 19 benzylidene malononitriles were used to perform a 3D-QSAR study. The HER2 inhibitory activities were selected from the recently published reports.[4,5] The $\log(1/C)$ (C represents IC_{50}) values were used to derive the 3D-QSAR models. The structures and inhibition activities of these 31 compounds are listed in Table 1 and Table 2. All molecular simulations were accomplished in a Silicon Graphics Octane 2-CPU workstation.

Pharmacophore Modeling. All compounds were built in the CATALYST View Compound Workbench and optimized by using a CHARMm-like force field.[14] The CATALYST model treats molecular structures as templates consisting of strategically positioned chemical functions that will bind effectively with complementary functions on receptor. The biologically most important binding functions are deduced from a small set of compounds that cover a broad range of activity. In this paper, considering the relatively high flexibility of ligands under investigation, the conformational analyses are carried out with an energy window high enough to include the bioactive conformation. CATALYST automatically generated conformational models for each compound, using the Poling algorithm.[15,16] Diverse conformational models for each compound were generated such that the conformers covered accessible conformational space defined within 20 kcal/mol of the estimated global minimum. The estimation of conformational energy was performed based on the CHARMm force field. CATALYST provides

1034 *J. Chem. Inf. Comput. Sci., Vol. 41, No. 4, 2001*

ZHU ET AL.

Table 2. Structures and Experimental Biological Activities (no SE) of 3-Substituted Indolin-2-ones
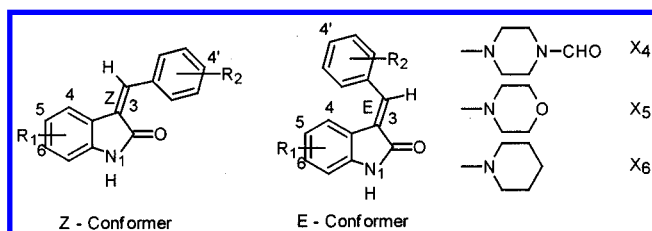
no.	conf	R_1	R_2	$IC_{50}, \mu M$
20	E	H	4'-X4	90.2
21	Z	4-Me	4'-X5	66.6
22	E	H	4'-X6	92.6
23	E	H	3',5'-di-CMe3, 4'-OH	64.8
24	E	5-Cl	3',5'-di-CHMe2, 4'-OH	8.2
25	E	H	3'-CMe3, 4'-OMe, 5'-Br	19.0
26	Z	5-Cl	3'-CMe3, 4'-OMe	16.2
27	E	H	4'-CHMe2	16.9
28	E	H	3',5'-di-CHMe2, 4'-OH	7.0
29	E	H	3'-CHMe2, 4'-OMe	13.2
30	Z	1-Me	4'-Br	22.5
31	Z	5-Cl	3'-CHMe2, 4'-OMe, 5'-Br	15.2

Figure 1. Correlation line of actual activity vs estimated activity ($r = 0.956$) of the best pharmacophore model.

two types of conformational analysis: fast and best quality. The best option was used, specifying 250 as the maximum number of conformers. The molecules associated with their conformational models were submitted to CATALYST hypothesis generation.

Hypothesis generation was performed in the CATALYST Hypothesis Generation Workbench using the generated conformers. Hypotheses approximating the pharmacophore were described as a set of features distributed within 3D space. In this paper, five kinds of surface-accessible functions were considered, including hydrogen bond acceptor (HBA), hydrogen bond donor (HDB), hydrophobic center (HP), aliphatic hydrophobic center (aliphatic HP), and aromatic hydrophobic center (aromatic HP). Furthermore, to emphasize the importance of an aromatic group of the studied molecules, aromatic ring (AR), which consists of directionality, was also included in the subsequent run. Considering the complexity of the studied molecules, the hypothesis generator was restricted to select only five or less features.

During a hypothesis generation run, CATALYST considers and discards many thousands of models. It attempts to minimize a score function consisting of two terms. One term penalizes the deviation between the estimated activities of the molecules in data set and their experimental values; the other term penalizes the complexity of the hypothesis. After assessing all generated hypotheses, the most plausible one is considered the best. Each hypothesis carries two regression lines which correlate the activity of each molecule with the geometric fit onto the pharmacophore model (activity estimation tool) and the estimated activities with the actual activities (correlation tool), the latter being useful to test the validity of each model.

Molecular Alignment. However, for most studies, the reliability and efficiency of a CoMFA or CoMSIA analysis depend on the correct molecular alignment of the ligands. The user must perform a pharmacophore mapping study to select both a proposed bioactive conformer and a superposition rule.[17] It is naturally deduced that the pharmacophore model thus obtained provides a good starting point for
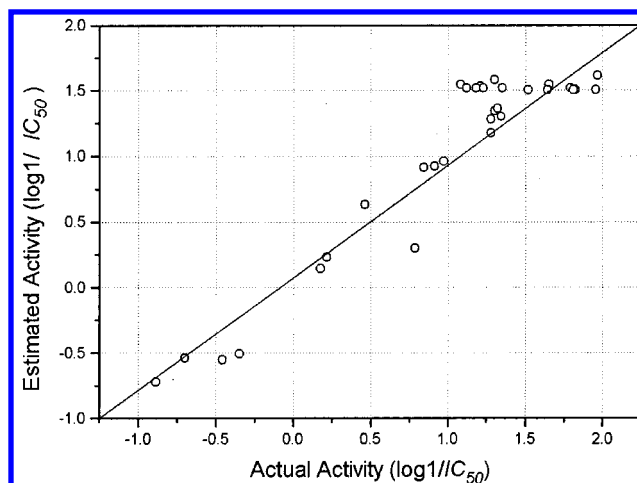
CoMFA and CoMSIA studies. In CATALYST, the studied molecules were automatically superimposed on the best pharmacorphore model in the Generation Hypothesis Workbench. For each ligand, one aligned conformer based on the lowest RMS deviation of feature atom coordinates from those of the corresponding reference features was superimposed on the hypothesis. In the studying process, we found that the best alignment based on the lowest RMS sometimes was not the most appropriate for a subsequent CoMFA or CoMSIA analysis, and artificial adjustment was usually needed to get the reasonable alignment.

In most cases of the CoMFA or CoMSIA calculations, the alignment from the simple atom-by-atom fits may be the most widely used procedure. So a rigid alignment was applied to superimpose 19 compounds onto an unsubstituented template shown in Figure 3 using an atom-by-atom least-squares fit as implemented in the SYBYL FIT option, and compound 13 with the best biological activity was treated as the reference molecule. In Figure 3, the atoms labeled with asterisks were selected as fit centers.

CoMFA and CoMSIA Calculations. 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 atomic Gasteiger charges, hydrophobicities by atom-based hydrophobic parameters, and hydrogen-bonding properties by suitably placed pseudoatoms, using a common probe with 1 Å radius, +1 charge, +1 hydrophobicity, and hydrogen-bond property of +1. The value of the attenuation factor a was defined as 0.3. A lattice of 2 Å was generated to surround the whole molecular aggregate. The dimensions of the surrounding lattice were selected with a sufficiently large margin (= 4 Å) to enclose all aligned molecules.

To choose the appropriated components and check the statistical significance of the models, leave-one-out cross-validations were used by the enhanced version of PLS, the SAMPLS method.[18] Then the final 3D-QSAR model is derived from the non-cross-validation calculations. The CoMSIA results were graphically interpreted by field contribution maps using the field type "stdev*coeff". The contour level was iteratively chosen to produce the best interpretable contour maps.

As a comparison, a conventional CoMFA was performed with the usually used steric and electrostatic fields imple-
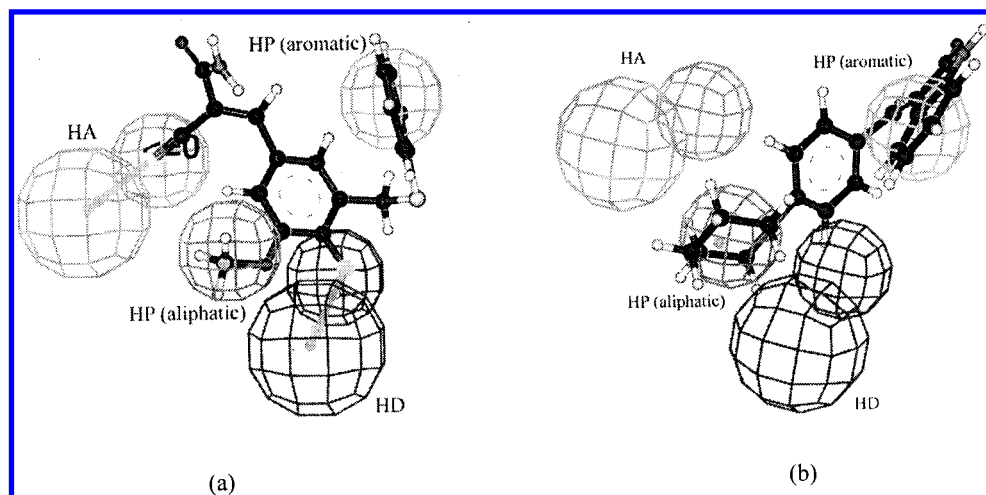
NOVEL TYROSINE KINASE INHIBITORS

J. Chem. Inf. Comput. Sci., Vol. 41, No. 4, 2001 **1035**

Figure 2. Superimposition of compound 13 and 22 with the best pharmacophore model.
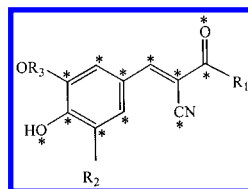

Figure 3. The common structure labeled with asterisks used in the alignment of the energy-lowest conformers.

mented in SYBYL. The atomic Gasteiger charges were applied in the determination of the electrostatic field. All CoMFA calculations were performed with 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 $+1.0$. The extent and the orientation of the grids surrounding the tested molecules were the same as those in CoMSIA analysis, and the grid spacing was set to 2 Å.

RESULTS AND DISCUSSION

Pharmacophore Modeling. From the calculated results, it can be found that the best hypothesis bears good correlation with four features which include HBA, HBD, aliphatic HP, and aromatic HP. The HP features are drawn as globes. HBA and HBD features are shown by two globes due to the directional nature of these chemical functions. In Figure 1, the good statistical significance of the model is indicated by the high correlation coefficient ($r = 0.956$). The number of the active conformers and the predicted biological activities for the studied molecules are listed in Table 3.

Figure 2a depicts the most active conformation of compound 13 superimposed on the best hypothesis. It can be found that the hypothesis can be well mapped by the corresponding functional groups of compound 13. It seems that the hydrophobic interactions and the H-bond interactions are crucial to ligand binding. The methyl group on R_3 site linked to the benzene ring overlaps with the aliphatic HP feature, while the phenyl group on the R_2 site serves as an aromatic HP feature. The nitrile group of the benzylidene malononitrile family is represented as the HA feature. Moreover, the hydroxy group on the benzene ring of the core structure is also very important, which corresponds to the HB feature. In Figure 2b, the mapping of the most active conformation of compound 22 on the best hypothesis is

Table 3. Energies of Conformers for Benzylidene Malononitriles and 3-Substituted Indolin-2-ones and Their Estimated vs Observed Biological Activities from the Best Pharmacophore Model

no.	confs[a]	conf/no.[b]	ΔE^c (kcal/mol)	estimate (IC$_{50}$, μM)	error
1	17	6	6.19	36	−1.2
2	23	6	0.00	34	2.8
3	31	4	6.69	9.8	1.0
4	42	2	5.01	33	1.0
5	46	9	3.24	28	1.3
6	16	9	19.16	37	1.9
7	42	2	5.78	4.3	1.5
8	21	12	1.09	36	−1.7
9	33	6	13.55	15	−1.3
10	24	17	17.99	36	−1.2
11	24	9	6.85	28	1.3
12	26	3	12.78	17	−1.1
13	50	4	16.28	0.17	1.3
14	141	138	5.69	0.33	−1.4
15	68	6	8.65	2	1.2
16	66	11	16.17	0.31	1.5
17	100	98	1.94	0.33	−1.1
18	101	55	4.15	1.3	−1.2
19	133	87	3.02	1.9	−3.2
20	13	4	10.73	33	−2.7
21	12	6	19.14	34	−2.0
22	7	5	11.43	58	−1.6
23	7	3	3.98	34	−1.9
24	16	10	18.48	7.4	−1.1
25	13	2	16.71	21	1.1
26	8	6	7.88	36	2.1
27	5	2	12.15	36	2.2
28	7	3	13.75	7.7	1.1
29	11	10	7.34	15	1.1
30	8	8	0.00	33	1.5
31	15	10	6.02	33	2.2

[a] Within 20 kcal.mol^{-1} above that of the lowest energy conformation found. [b] The number of the active conformer in the total conformers. [c] Energy difference between the active conformer and the energy-lowest conformer.

represented. It is clear that compound 22 only possesses two features fitted on the best hypothesis. From the mappings of compounds on hypothesis, the quantity of the inhibitors can be qualitatively determined. For these compounds with good biological activities, they can produce good fits with all four features in the best hypothesis. While for those compounds with poor biological activity, they can only produce relatively good fits with two or three features.

**Table 4.** CoMFA and CoMSIA Analyses Based on the Automatic Alignment on the Best Pharmacophore Model Provided by CATALYST

	CoMFA		CoMSIA			
	1	2	1	2	3	4
q^2	0.408	0.343	0.467	0.398	0.402	0.462
r^2	0.877	0.890	0.866	0.945	0.801	1.00
SE of estimate	0.348	0.398	0.379	0.170	0.405	0.007
F	47.423	79.990	59.616	297.366	32.103	6763.180
N	4	4	4	6	2	10
fraction						
steric	0.661	0.203	0.467	0.218	0.381	0.144
elelectrostatic	0.339	0.106	0.533	0.212	0.141	0.154
hydrophobic					0.478	0.280
H-acceptor		0.304		0.229		0.169
H-donor		0.386		0.342		0.254

Table 3 shows that the active conformers of the studied molecules usually do not adopt the energy-lowest conformers. For all the 31 molecules, only compounds 2 and 30 prefer to the energy-lowest conformers, while the other compounds adopt higher energetic ones. The results are not surprising, because in drug-receptor recognition process, the conformations of receptor and ligands will be adjusted in order to gain excellent steric and energetic complementarity. In some cases, the ligands prefer the conformers with a bit higher energy. For example, compound 13 with the highest affinity for HER2 possesses 50 conformers, while the active conformer is not the energy-lowest one but conformer 4 whose energy is 16.3 kcal/mol.

CoMFA and CoMSIA Calculations. Undoubtedly, a modest quantitative performance can only be expected on a structurally diverse series of compounds. Considering the structural diversity between benzylidene malononitriles and 3-substituted indolin-2-ones, after having performed the pharmacophore modeling, only 19 benzylidene malononitriles were aligned according to the most representative hypothesis for CoMFA and CoMSIA calculations. For comparison, the molecular aggregate of the energy-lowest conformations from conformational analyses by using simple atom-by-atom fits was also treated as input for CoMFA and CoMSIA analyses.

(1) Alignment Based on the Pharmacophore Model. Just as described above, when we used the 31 compounds to perform pharmacophore search, a model with one hydrogen bond acceptor, one hydrogen bond donor, one aliphatic hydrophobic group, and one aromatic hydrophobic group was obtained. After that, we superimposed the 19 molecules of the benzylidene malononitrile family onto the pharmacophore model automatically by using the "Show selected compounds/ Mapping" tool in the CATALYST Hypothesis Generation Workbench. The automatically aligned molecules were saved and imported into SYBYL for CoMFA and CoMSIA analyses.

Table 4 shows several CoMFA and CoMSIA models generated at 2.0 Å grid spacing using different field combinations. Using only steric and electrostatic fields, the CoMFA model is not good ($q^2 = 0.408$). After considering the H-bond fields, the statistical significance of the obtained CoMFA model is decreased significantly. The CoMSIA model with the steric and electrostatic fields ($q^2 = 0.467$) is obviously better than the CoMFA one. After considering the hydrophobic and H-bond fields, the values of q^2 of the CoMSIA models are decreased obviously.
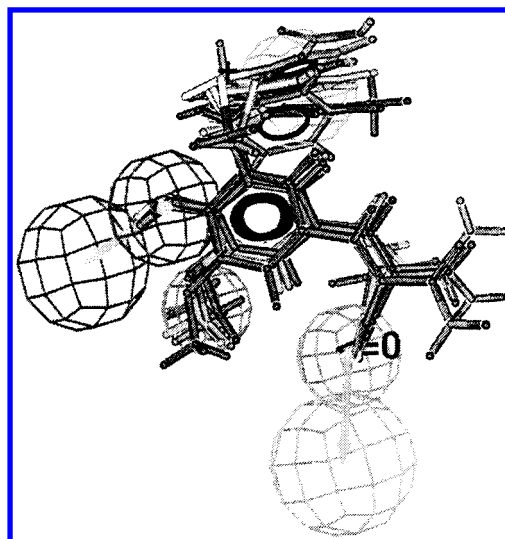
Figure 4. Alignment of compounds 13−19 with the best hypothesis.

It seems that the alignment is not perfect as superimposing only the active conformers automatically to the best hypothesis. If we carefully investigate the superimposition between the pharmocophore model and the studied compounds, we can find that for several compounds with relatively low biological activities, the best superimposed conformers searched by CATALYST are obvious not reasonable. Figure 4 shows the superimposition of compounds 13−19 on the best pharmacophore model. It is obvious that these seven compounds including other several compounds (not shown in Figure 4) can produce perfect mapping with the best hypothesis. From the chemical intuition, in the active site of receptor, the common core structures of benzylidene malononitriles should adopt similar orientation; that is to say, the other benzylidene malononitriles should adopt the similar orientation with compounds 13−19. Figure 5 shows the alignment of compounds 4 and 13 with the hypothesis. From Figure 5a, it is clear that the common structures of compounds 4 and 13 do not adopt a similar orientation. It is naturally deduced that the selected conformer by CATA- LYST of compound 4 possesses irrational space orientation and cannot be aligned to the pharmacophore model well. The reasons for the generation of the incorrect active conformer and the irrational pharmacophore alignment of compound 4 by CATALYST are not unreasonable. For CATALYST, all functional groups in a compound are equivalent. The CATALYST always attempts to find one conformer which can produce the best alignment with the pharmacophore model with lowest RMS value, and it does not consider any chemical restricts. So in some cases, the active conformer and pharmacophore mapping should be manually adjusted. In CATALYST, all possible alignments between all conformers and the pharmacophore model are provided. In the current work, for several compounds, the active conformers and the corresponding alignment were manually adjusted. In the adjusting process, two criteria were adopted. The first, the nitrile group and the hydroxy group on the benzene ring were superimposed or located near the HA and HD features in order to manually force all compounds to adopt similar space orientations. The second, the active conformer bears conformational energy as low as possible. While after being changed to another conformer
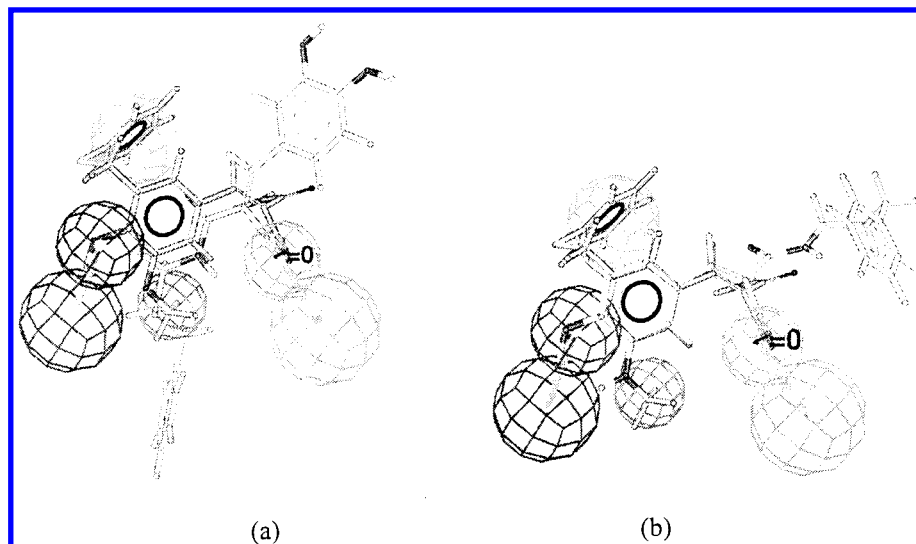
Figure 5. Superimposition of the compounds 4 and 13 on the pharmacophore model (a: superimposed automatically and b: superimposed manually).

Table 5. Energies of Conformers for the Benzylidene Malononitrile Family Superimposed to the Pharmacophore Model after Manual Adjustment

no.	conf/no.[a]	$\Delta E_1{}^a$ (kcal/mol)	conf/no.[a]	$\Delta E_2{}^b$ (kcal/mol)
1	6	6.19	6	6.19
2	6	0.000.	5	17.94
3	4	6.69	4	6.69
4	2	5.01	20	12.17
5	9	3.24	17	9.18
6	9	19.16	5	10.76
7	2	5.78	23	16.99
8	12	1.09	12	1.09
9	6	13.55	3	9.44
10	17	17.99	17	17.99
11	9	6.85	3	2.65
12	3	12.78	23	9.87
13	4	16.28	4	16.28
14	138	5.69	138	5.69
15	6	8.65	6	8.65
16	11	16.17	11	16.17
17	98	1.94	98	1.94
18	55	4.15	55	4.15
19	87	3.02	87	3.02

[a] The conformations selected by CATALYST automatically by mapping on the best hypothesis. [b] The conformations selected manually by mapping on the best hypothesis.

manually, the space orientation of each compound is consistent, although the energy of the manually selected conformer may be higher than that of the conformer selected automatically. The adjustment provides a correct alignment rule for the subsequent 3D-QSAR study. Table 5 shows the energies of active conformers for the benzylidene malononitriles by manual selections. From this table, it can be found that the pharmacophore mappings of eight compounds were adjusted.

We believe that after manual adjustment the 3D-QSAR would be surely more proper by substitution of some irrational conformers with more appropriate ones. From Table 5 we can find that the selected conformers of some molecules by hand are not the "bioactive conformers" evaluated by CATALYST although the latter possess lower energy.

The results of CoMFA and CoMSIA analyses using the pharmacophore-based alignment after manual adjustment are shown in Table 6. Table 6 shows that CoMFA model with steric and electrostatic fields bears good statistical significance ($q^2 = 0.655$). After the hydrogen bond acceptor and donor properties are added, the predictive power is decreased a little ($q^2 = 0.575$). Similarly, the CoMSIA model using only steric and electrostatic fields is more acceptable than that added in hydrogen bond fields. It seems that the quantitative structure–activity relationships cannot be well illustrated by using both steric, electrostatic and hydrogen bond properties together. But when hydrophobic property is considered, the predictive power is increased greatly, which may be consistent with the pharmacophore model in which there are one aliphatic hydrophobic group and one aromatic hydrophobic group. The best 3D-QSAR is derived from using

Table 6. Results of the CoMFA and CoMSIA Analyses of Several Different Field Combinations Based on the Alignment on the Best Pharmacophore Model after Manual Adjustment

	CoMFA		CoMSIA			
	1	2	1	2	3	4
q^2	0.656	0.575	0.569	0.522	0.615	0.710
r^2	1.00	0.915	0.801	1.000	0.992	1.00
SE of estimate	0.002	0.264	0.405	0.008	0.087	0.005
F	5254.148	86.427	32.103	22766.548	426.628	11689.361
N	10	2	2	10	4	7
fraction						
steric	0.779	0.272	0.590	0.263	0.172	0.226
elelectrostatic	0.221	0.123	0.410	0.223	0.143	0.195
hydrophobic					0.322	0.404
H-acceptor		0.338		0.153	0.106	0.176
H-donor		0.267		0.360	0.258	

Table 7. Results of the CoMFA and CoMSIA Analyses Based on the Atom-by-Atom Fits of the Energy-Lowest Conformations

	CoMFA		CoMSIA			
	1	2	1	2	3	4
q^2	0.529	0.472	0.523	0.542	0.490	0.503
r^2	0.861	0.880	0.883	0.911	0.823	0.887
SE of estimate	0.323	0.337	0.310	0.279	0.410	0.307
F	47.147	53.639	60.410	51.425	39.693	62.792
N	4	4	2	3	3	2
fraction						
steric	0.663	0.252	0.524	0.237	0.167	0.266
elelectrostatic	0.337	0.175	0.476	0.272	0.192	0.254
hydrophobic					0.288	0.380
H-acceptor		0.236		0.178	0.130	0.100
H-donor		0.337		0.313	0.223	

Figure 6. Superimposition of all compounds using the energy-lowest conformers by atom-by-atom fits.

steric, electrostatic, hydrogen bond acceptor, and hydrophobic fields afforded by CoMSIA, which possesses the best predictive power ($q^2 = 0.710$). From the comparison of these several PLS analyses results of different field combinations in Table 7, it is likely that the properties are considered intercorrelated however in a complicated way. The inter-correlations of these numerically intensive grid fields are difficult to detect, so sometimes it is relatively very difficult to determine whether some kind of field is more important or not. The advantages of using four different fields of well-defined molecular properties have to be seen in the straight-forward partitioning of these properties into spatial locations where they take a determining role on biological activity. The best CoMSIA model gives good statistical measure ($q^2 = 0.710$ by using seven latent variables, non-cross-validated $r^2 = 1.00$, $F = 11689.361$, $SD = 0.005$).

Although the pharmacophore modeling and CoMFA technique have been developed for more than 10 years. But CoMFA based on pharmacophore model is really a new approach to 3D QSAR analysis. In the papers cited in "SCI" from 1989 to Dec 7, 2000, the total number of papers with the keyword "CoMFA" are more than 5000, but only 51 have both "CoMFA" and "pharmacophore". After careful examination, we found that only several papers were related with CoMFA base on pharmacophore model using DISCO, Apex-3D, or CATALYST software.[19−26] We believe that the inherent withdraws of pharmacophore modeling may be the main reason for restricting the combined applications of these two techniques. First, for compounds with relatively low activity, a pharmcophore modeling technique such as CATA-LYST usually cannot produce good fitting with hypothesis. Second, pharmophore modeling cannot deal with these functional groups that are not included in the hypothesis. From the compounds studied in this paper, these two questions are also partly expressed. For example, for 19 benzylidene malononitriles, both of the R_1 and the R_2 groups in some compounds possess flexibility. From the experimental biological activities, the R_2 groups are obviously more important than the R_1 groups. Consequently, R_2 groups are revealed by the aromatic HP feature in the hypothesis, while the R_1 groups are not. The real space stretch of the flexible R_2 group can be perfectly treated by the best hypothesis, but the information from the flexible R_1 groups seems not to be directly revealed by the hypothesis. For these 19 benzylidene malononitriles, the conformations of R_1 groups are not crucial to ligand binding, and selection of the conformations with low energy is accepted. But to other
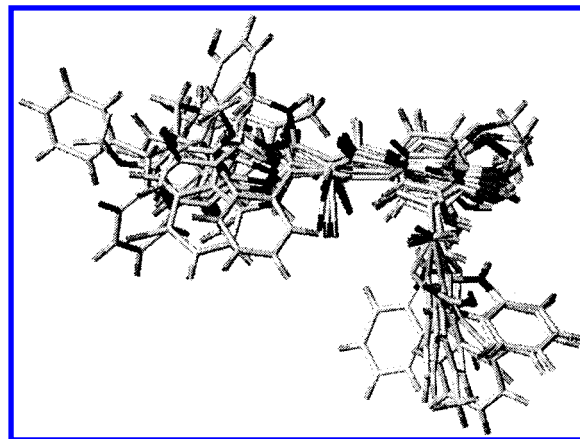
kinds of compounds, the proper treatment of the flexible substituents not considered by the pharmacophore model may be very critical to the consequent 3D-QSAR analyses. For these groups of compounds, we mentioned above that some fittings between the hypothesis and several compounds with relatively low biological activities are obviously questionable, and manual adjustments should be needed. To our astonishment, it seems that the authors of these papers related to CoMFA based on the pharmacophore model did not care about these questions.

(2) Alignment Based on the Energy-Lowest Conformers. In many cases, the active conformers of the studied molecules cannot be easily determined, so we usually treat the energy-lowest conformers as the active conformers. In the current work, the CoMFA and CoMSIA calculations based on atom-by-atom fits of the energy-lowest conformers were also performed as comparative fashion.

In CATALYST, the energy-lowest conformers of the studied molecules from the conformational analyses could be determined. These energy-lowest conformers were imported into SYBYL for CoMFA and CoMSIA analyses. Using the energy-lowest conformers, we superimposed these 19 compounds onto the template compound 13 with atom-by-atom least-squares fit. Figure 4 shows the plot of the aligned molecules within the grid box (grid spacing 2.0 Å) applied to generate the CoMFA and CoMSIA columns.

Table 7 shows several CoMFA and CoMSIA models generated at 2.0 Å grid spacing using different field combinations. Obviously, using only steric and electrostatic fields, the 3D-QSAR model from CoMFA analysis is not very good ($q^2 = 0.529$). After considering the hydrogen bond field, the predictive power of the CoMFA model ($q^2 = 0.472$) is reduced apparently. The model from CoMSIA is also not satisfactory ($q^2 = 0.523$) using only steric and electrostatic fields, which is comparative with that from CoMFA. After considering the H-bond fields, the quality of the CoMSIA model is improved a little ($q^2 = 0.542$). After hydrophobic field is introduced in the CoMSIA analysis, the predictive power of the 3D-QSAR model ($q^2 = 0.558$) is further increased, but it is still not very satisfactory. Obviously, simply using the energy-lowest conformer is unsuitable since the active conformer is often not the energy-lowest conformer, and by doing so the flexibility of compounds is inevitably overlooked. Thus it can be considered that the alignment based on the energy-lowest conformers is not
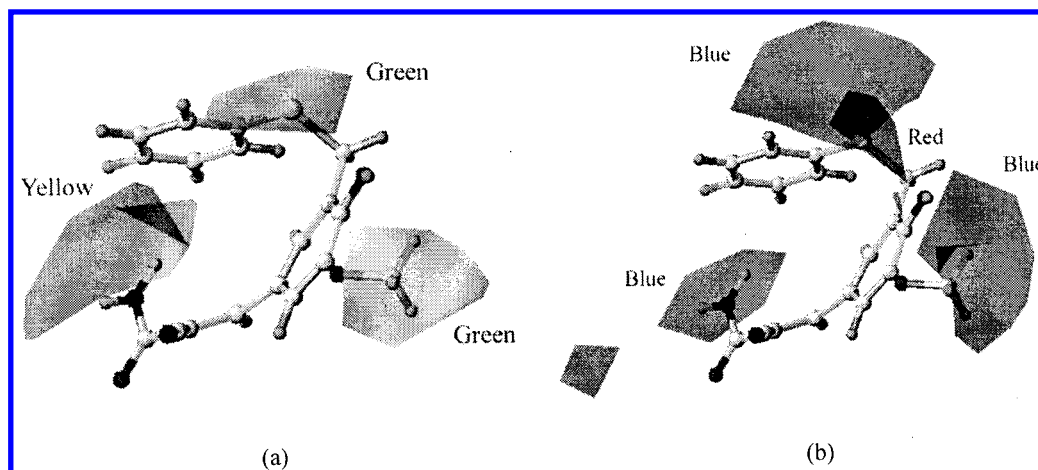
Figure 7. Steric and electrostatic contour plots for the best CoMSIA model (a: the steric contour maps and b: the electrostatic contour maps).

acceptable, which can be approved further by Figure 4. Although the core structures of these 19 compounds are forced to produce good fitting, the alignment of the flexible substituents on R_1 and R_2 groups are really not acceptable.

CoMSIA Contour Maps. The results of CoMSIA are best interpreted as field graphics. These graphs show regions in the space around the molecules, where specific field interactions enhance or detract from the activity. A 3D steric and electrostatic fields graphs around the reference conformer of compound 13 are shown in Figure 7. These two contour maps depict regions around the molecules where enhanced receptor binding affinity is associated with increasing (green) and decreasing (yellow) steric bulk and with increasing (red) and decreasing (blue) negative charge. The relative contribution of these steric factors is comparative with the electrostatic one (0.226 vs 0.195). The CoMSIA steric and electrostatic fields are consistent with the known structure−activity relationships for the benzylidene malononitriles from experimental data. For example, the green contour in the region around the R_3 group for this CoMSIA model indicates that the presence of bulky groups, such as the methyl group found in compounds 13−19, is expected to enhance binding affinity, compared with the corresponding compounds with one hydrogen atom found in compounds 1−12. Consistent with the presence of the rather large yellow contour closely surrounding the R_1 group, enhanced binding affinity is shown for compounds 13−19 in which amino substitution appears close to the yellow contour, while diminished binding is found for compounds 1−12 in which bulky groups are substituted at the R_1 site. The position of two blue zones near the R_3 and R_1 groups indicate that the enhanced binding affinity is associated with decreasing negative charge groups such as methyl substitution and amino substitution found in compounds 13−19, respectively.

The map of hydrophobic properties indicates two very distinct hydrophobic favored sites colored with yellow (Figure 8): one larger region near the R_2 group on the benzene ring and the other smaller one near the R_3 group connected with the oxygen atom, which means these groups with high hydrophobicity will be favorable to biological activity. It can be reasonably presumed that the benzene ring combined with these substituents on the R_2 site and the methyl group at the R_3 site can produce strong hydrophobic interaction with the receptor. From the faction of hydrophobic
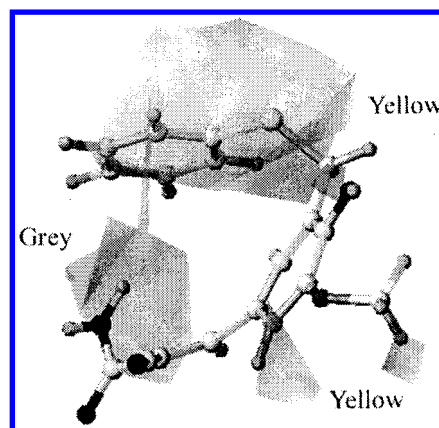

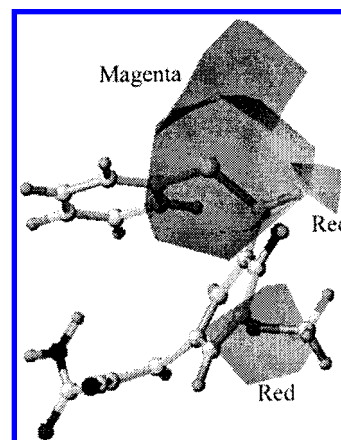**Figure 8.** Hydrophobic contour plots for the best CoMSIA model.

Figure 9. H-bond acceptor contour plots for the best CoMSIA model.

field, it is obvious that the hydrophobic interactions may be dominant for ligand binding. The hydrophobic contour map is perfectly consistent with the best pharmacophore model. Two hydrophobic favorable areas correspond to the aromatic HP feature located at the R_2 site and the aliphatic HP feature located at the R_3 site, respectively.

The graphical interpretation of the field contributions of the H-bond properties is shown in Figure 9 (the H-bond acceptor field). In principle, it should highlight areas beyond the ligands where putative hydrogen bond partners in the enzyme can form H-bonds that influence binding affinity significantly. Areas indicated by magenta contours cor-

respond to regions where hydrogen bond acceptor substitutions at the ligands will increase biological activity. Areas encompassed by red isopleths should be avoided from hydrogen bond acceptors; otherwise reduced affinity can be expected. Two magenta isopleths in Figure 9 around the oxygen atom on amido connected with the R_1 group and the S atom in some R_2 substitutions on the benzene ring are indicated as favorable H-bond acceptor sites. From the fraction of the fields, it seems that the hydrogen bond acceptor field contributes less compared with the hydrophobic field.

CONCLUSIONS

This work demonstrated that the pharmacophore modeling and MFA technquies (CoMFA or CoMSIA) could be effectively combined. The best model identified by pharmacophore search served as suitable modes of superimposition for subsequent CoMFA and CoMSIA analyses. For comparison, a conventional superimposition based on the energy-lowest conformations was used to perform a 3D-QSAR research. And better results (higher q^2 values) were obtained by superimposing the molecules onto the pharmacophore model than onto the common structure used in the energy-lowest conformers. The CoMSIA model, with steric, electrostatic, hydrophobic, and hydrogen bond acceptor properties contributing simultaneously, was proven to be the best model ($q^2 = 0.710$). Moreover, from our calculations, it can be found that the active conformers and the pharmacophore mapping automatically afforded by CATALYST are usually not the optimal ones. The manual adjustments are usually needed to obtain the best results.

As illustrated by the contour maps, our 3D-QSAR model, which is based on a homogeneous set of ligands, is expected to correctly predict affinities of structurally related compounds. These results will provide useful information in understanding the structural and chemical features of tyrphostins and in designing new potential compounds.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Riese, D. J.; Stern, D. F. Specificity within the EGF family ErbB receptor family signaling network. *Bioessays* **1998**, *20*, 41−48.

(2) van der Geer, P.; Hunter, T. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* **1994**, *10*, 251−337.

(3) Slamon, D. J.; Godolphin, W.; Jones, L. A.; Holt, J. A.; Wong, S. G.; Keith, D. E.; Levin, W. J.; Stuart, S. G.; Udove, J.; Ullrich, A.; Press, M. F. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **1989**, *244*, 702−712.

(4) Osherov, N.; Gazit, A.; Gilon, C.; Levitzki, A. Selective inhibition of the epidermal growth factor and HER2/neu receptors by tyrphostins. *J. Bio. Chem.* **1993**, *268*, 11134−11142.

(5) Sun, L.; Tran, N.; Tang, F.; App, H.; Hirth, P.; McMahon, G.; Tang, C. Synthesis and biological evaluations of 3-substituted indolin-2-

(6) CATALYST 4.0 User Guide; Molecular Simulations Inc.: Burlington, MA, 1998.

(7) SYBYL 4.5; Tripos Inc.: St. Louis, MO, 1999.

(8) 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.

(9) Wade, R. C. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 486−506.

(10) Wold, S.; Johansson, E.; Cocchi, M. In *3D QSAR in Drug Design, Theory Methods and Applications*; Kubinyi, H., Ed., ESCOM: Leiden, 1993; pp 523−550.

(11) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(12) Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1−10.

(13) Bohm, M.; Sturzebecher, J.; Klebe, G. Three-dimensional quantitative structure−activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477.

(14) Brooks, B. R.; Brucolleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(15) Smellie, A.; Teig, S. L.; Towbin, P. Poling − promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171−187.

(16) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of conformational coverage 0.1. Validation and estimation of coverage. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285−294.

(17) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83−102.

(18) Bush, B. L.; Nachbar, R. B. Sample Distance Partial Least Squares: PLS Optimized for Many Variables, with Application to CoMFA. *J. Comput.-Aided Mol.* **1993**, *7*, 587−619.

(19) Myers, A. M.; Charifson, P. S.; Owens, C. E.; Kula, N. S.; McPhail, A. T.; Baldessarini, R. J.; Booth, R. G.; Wyrick, S. D. Conformational analysis, pharmacophore identification, and comparative molecular field analysis of ligands for the neuromodulatory sigma 3 receptor. *J. Med. Chem.* **1994**, *37*, 4109−4117.

(20) Kamano, Y.; Kotake, A.; Hashima, H.; Inoue, M.; Morita, H.; Takeya, K.; Itokawa, H.; Nandachi, N.; Segawa, T.; Yukita, A.; Saitou, K.; Katsuyama, M.; Pettit, G. R. Structure-cytotoxic activity relationship for the toad poison bufadienolides. *Bioorgan. Med. Chem.* **1998**, *6*, 1103−1115.

(21) Langer, T.; Hoffmann, R. On the use of chemical function-based alignments as input for 3D-QSAR. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 325−330.

(22) Talele, T. T.; Kulkarni, S. S.; Kulkarni, V. M. Development of pharmacophore alignment models as input for comparative molecular field analysis of a diverse set of azole antifungal agents. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 958−966.

(23) Borosy, A. P.; Morvay, M.; Matyus, P. 3D QSAR analysis of novel 5-HT1A receptor ligands. *Chemometr. Intell. Lab.* **1999**, *47*, 239−252.

(24) Bursi, R.; Grootenhuis, P. D. J. Comparative molecular field analysis and energy interaction studies of thrombin-inhibitor complexes. *J. Comput. Aid. Mol. Des.* **1999**, *13*, 221−232.

(25) Palomer, A.; Pascual, J.; Cabre, F.; Garcia, M. L.; Mauleon, D. Derivation of pharmacophore and CoMFA models for leukotriene D(4) receptor antagonists of the quinolinyl(bridged)aryl series. *J. Med. Chem.* **2000**, *43*, 392−400.

(26) Borosy, A. P.; Keseru, K.; Penzes, I.; Matyus, P. 3D QSAR study of class I antiarrhythmics. *J. Mol. Struct.-Theochem.* **2000**, *503*, 113−129.

ones: A novel class of tyrosine kinase inhibitors that exhibit selectivity toward particular receptor tyrosine kinases. *J. Med. Chem.* **1998**, *41*, 2588−2603.