J Mol Model (2001) 7:223–230
DOI 10.1007/s008940100017

ORIGINAL PAPER

Lili Zhu · Tingjun Hou · Xiaojie Xu

Three-dimensional quantitative structure–activity relationship study on paullones as CDK inhibitors using CoMSIA and CoMFA

Received: 6 November 2000 / Accepted: 6 March 2001 / Published online: 12 July 2001

Abstract 3D-QSAR studies were conducted on a series of paullones as CDK inhibitors using three-dimensional quantitative structure–activity relationship (3D-QSAR) methods. Two methods were compared: the widely used comparative molecular field analysis (CoMFA) and the recently reported comparative molecular similarity indices analysis (CoMSIA). Systematic variations of some parameters in CoMSIA and CoMFA were performed to search for the best 3D-QSAR model. The computed results showed that the 3D-QSAR models from CoMSIA were clearly superior to those from CoMFA. Using the best model from CoMSIA analysis, a significant cross-validated q^2 was obtained and the predicted biological activities of the five compounds in the test set were in good agreement with the experimental values. The correlation results obtained from CoMSIA were graphically interpreted in terms of field contribution maps allowing physicochemical properties relevant for binding to be easily mapped back onto molecular structures. The features in the CoMSIA contour maps intuitively suggested where to modify a molecular structure in terms of physicochemical properties and functional groups in order to improve its binding affinity, which is very important for improving our understanding of the ligand–receptor interactions and in helping to design compounds with improved activity.

Keywords CoMSIA · CoMFA · Paullones · 3D-QSAR

Introduction

The cyclin-dependent kinases (CDKs) are a group of serine threonine kinases that play a crucial role among the manifold molecular entities that are involved in the surveillance of the cell cycle. The CDKs control the transmission between successive stages of the cell cycle. [1] The oscillating concentration of the cyclins during the cell cycle is the basis for the stage-dependent activity of the CDKs. Binding to CDK inhibitory proteins (CKIs) results in deactivations of CDKs. In various human tumors, deregulations of CDK-related mechanisms have been found, such as overexpression of cyclins or deletion of genes encoding for CKIs. [2, 3] Considering these observations, CDKs are attractive targets for the development of antitumor drugs. [4, 5]

According to a few reports, [6, 7, 8] three types of CDK inhibitors, such as lactones, flavonoids, and several purine derivatives have been found to show antiproliferative activity for colon and pancreatic carcinoma cell lines. However, the number of chemical agents that act selectively as CDK inhibitors is limited. Recently, inhibitory experiments revealed that the 9-bromo-7,12-dihydroindolo[3,2-d] benzazepin-6 (5H)-one (kenpaullone) is a potent inhibitor of CDKs with selectively for CDK_1, CDK_2, and CDK_5. [9] However, compared to the in vitro antitumor potency of fluopiridol, kenpaullone exhibits only a modest antiproliferative activity in the in vitro cancer cell line screening. In order to counter this problem, Schultz et al. [9] designed a synthesis study with a view to searching for kenpaullone-related CDK inhibitors with improved potency and antitumor activity. Although the results of the molecular modification experiments have been published, quantitative information on structure–activity relationships is needed for further rational development and direction of selective synthesis. Until now, however, very few studies on the relationship between the chemical structures and the biological functions of this kind of compound have been reported. In this paper, a detailed correlation study was accomplished on the basis of a newly developed 3D-QSAR technique: comparative molecular similarity indices analysis (CoMSIA). Moreover, comparative molecular field analyses (CoMFA) were also conducted as a comparison.

L. Zhu · T. Hou · X. Xu (✉)
College of Chemistry and Molecular Engineering,
Peking University, Beijing 100871, P. R. China
e-mail: xiaojxu@chem.pku.edu.cn

Materials and methods

The principles of CoMSIA

From its advent in 1988, CoMFA has been developed as one of the most powerful tools in 3D-QSAR. [10] CoMFA quantifies the statistical relationship between the 3-D properties of a set of small molecules and a global property, such as their potency in a particular biological assay. This method combines pharmacophore mapping, quantitative structure–activity relationships (QSARs), intermolecular energy calculations, and multivariate statistics to produce a 3D-QSAR model that appears to have wide applicability.

Recently, another alternative molecular field analysis, CoMSIA, based on molecular similarity indices, has been reported. [11, 12, 13] The approach can avoid some inherent deficiencies arising from the functional form of Lennard-Jones and Coulomb potentials used in the conventional CoMFA. 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 these grids in the proximity of the surface. Also no arbitrary definition of cutoff limits is required in CoMSIA. Moreover, using molecular similarity indices analysis, the contour maps of the relative spatial contributions of the different fields can be substantially improved. These are very intuitive for interpretation in terms of separate property fields. Compared with usual CoMFA, the CoMSIA has the better ability to visualize and interpret the correlations obtained in terms of field contributions. The unique difference between conventional CoMFA and CoMSIA is the field type and the field calculation 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-dependence function has been used to calculate different kinds of physicochemical properties. The indices $A_{F.K}$ between the compounds of interest and a probe atom are calculated according to:

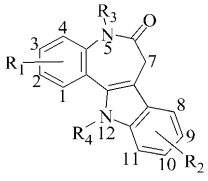
$$A_{F.K}{}^{q}(j) = -\sum_{i=1}^{n} \omega_{\text{probe},k}\,\omega_{ik}\,e^{-\alpha r_{iq}^{2}}$$

where i: summation index over all atoms of the molecule j under investigation; ω_{ik}: actual value of the physicochemical property k of atom, $\omega_{\text{probe},k}$: probe atom with charge +1, radius 1 Å, hydrophobicity +1, H-bond donor and acceptor property +1; a: attenuation factor; r_{iq}: mutual distance between probe atom at grid point q and atom i of the investigated molecule.

Data set and structural alignment

The paullone derivatives' inhibitory data, represented by IC_{50} (μM) values of CDK1 inhibition, were taken from the literature [9] and are listed in Table 1 (see also Fig. 1). The log(1/*C*) (*C* repre-

Table 1 Structures of the molecules studied, experimental and calculated biological activity by the best 3D-QSAR model from the CoMSIA analyses



No.[a]	R_1	R_2	log(1/*C*) (obsd.)	log(1/*C*) (calcd.)	Residue
1	H	9-Br	0.3980	0.3573	0.0407
2	H	10-Br	–0.1140	–0.2819	0.1679
3	H	H	–0.8450	–0.6930	–0.1520
4	H	9-Cl	0.2220	0.3584	–0.1364
5	H	9-F	–0.2040	0.0455	–0.2495
6	H	9-OCH_3	0.0460	–0.0757	0.1217
7	H	9-CH_3	–0.3010	–0.3947	0.0937
8	H	9-CF_3	0.3980	0.3995	–0.0015
9	H	9-CN	1.6200	1.6636	–0.0436
10	H	9-NO_3	1.4560	1.3355	0.1205
11	H	11-Cl	–0.1460	–0.1637	0.0177
12	H	11-Br	–0.1140	–0.1454	0.0314
13	H	11-Et	–0.5800	–0.6768	0.0968
14	H	8,10-di-Cl	–0.3980	–0.4407	0.0427
15	2-Br	H	–0.5190	–0.4982	–0.0208
16	2-Br	9-CF_3	0.6200	0.5962	0.0238
17	2,3-di-OCH_3	H	–0.6330	–0.5703	–0.0627
18	2,3-di-OCH_3	9-Br	0.6990	0.5765	0.1225
19	2,3-di-OCH_3	9-CF_3	0.5530	0.6196	–0.0666
20	4-OCH_3	H	–2.6330	–2.7694	0.1364
21	2,3-di-OH	9-Br	–0.4770	–0.4353	–0.0417
22	4-OH	9-Br	–1.6020	–1.5604	–0.0416
23[b]	H	11-Me	–0.4770	–0.6462	0.1692
24[b]	2-Br	9-Br	0.5230	0.5515	–0.0285
25[b]	4-OCH_3	9-Br	–2.3980	–1.9147	–0.4833
26	Boc	Boc	–3.0000	–3.1322	0.1322
27	CH_3	H	–1.3010	–1.6575	0.3565
28	C_2H_5	H	–2.6720	–2.2229	–0.4491
29	CH_2-Ph	H	–1.5440	–1.4834	–0.0606
30	CH_2OCH_3	H	–0.8060	–0.9137	0.1077
31	H	CH_3	–0.7920	–0.8516	0.0596
32	H	C_2H_5	–1.3620	–1.1189	–0.2431
33	H	CH_2-CH=CH_2	–1.7780	–1.7916	0.0136
34	H	$CH_2CO_2CH_3$	–0.1460	–0.1016	–0.0444
35[c]			–0.3620	–0.3770	0.0150
36[c]			–1.6330	–1.5925	–0.0405
37[c]			0.0000	0.0464	–0.0464
38[b]	H	Boc	–1.8450	–1.1687	–0.6763
39[b]	H	CH_2-CH_2-OH	–0.4770	–1.0224	0.5454

[a] In compounds **1**–**25**, R_3=R_4=H, in compounds **26**–**39**, R_1=R_2=H,
[b] These compounds were used as a test set and not included in the derivation of equations.
[c] The structures of **35**, **36**, and **37** are shown in the chart in Fig. 1.

sents IC_{50}) values were used as the dependent variable to derive 3D-QSAR models. A training set of 34 compounds was used for CoMSIA and CoMFA analyses. In addition, five other compounds, selected from various ranges of biological activity, were used as the test set to verify the actual prediction of the model.

The molecular geometries of all compounds in Table 1 were modeled using the SYBYL molecular simulation package. [14] The initial structures were first minimized using molecular mechanics with the MMFF94 force field. [15] Then these structures were fully optimized using the semiempirical AM1 method, available in MOPAC 7.0. [16] From the 3D structures, it was found that this series of compounds possessed relatively rigid core structures constituting a large conjugated system. So a rigid alignment was applied to superimpose all 39 compounds onto an unsubstituted template shown in the chart in Fig. 2 using an atom-by-atom least-square fit as implemented in the SYBYL FIT option. Compound **9** with the highest biological activity was selected as the reference molecule. Figure 3 shows the stereoview of aligned molecules (including the test set) within the grid box (grid spacing 2.0 Å) used to generate the CoMSIA columns.

Determination of the 3D-QSAR models

In the present CoMSIA analyses, five physicochemical properties *k* (steric, electrostatic, hydrophobic, and hydrogen-bond acceptor and donor) were evaluated, using a common probe with 1 Å radius and charge, hydrophobicity, and hydrogen-bond property of +1. The value of the attenuation factor *a* was initially set to 0.3. A lattice of 2 Å was generated to surround the whole molecular aggregates after alignment. The dimension of the surrounding lattice was selected with a sufficiently large margin (=4 Å) to enclose all aligned molecules. The effect of grid point spacing on the CoMSIA analysis was also investigated at 1.0, 1.5, 2.5, and 3.0 Å with the same orientation.

To check the statistical significance of the models, cross-validation was done by means of the leave-one-out procedure using the enhanced version of PLS, the SAMPLS method. [17] The optimal number of components was determined by selecting the highest q^2 value. Then the same number of components was subsequently used to derive the final 3D-QSAR model using the no cross-validation calculations. The CoMSIA results were interpreted graphically by field contribution maps using the field type "stdev*coeff". The contour level was chosen iteratively to produce the best interpretable contour maps.

In the CoMFA calculations, steric and electrostatic fields were calculated as implemented in SYBYL using Lennard-Jones and Coulomb potentials, respectively. MMFF94 charges were used 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 probe atom with a charge of +1.0. The extent and the orientation of the grids surrounding the molecules tested are the same as those in the CoMSIA analyses, and the grid spacing was set to 2 Å. All calculations in this study were performed in SYBYL on a SGI Octane 2-CPU workstation.

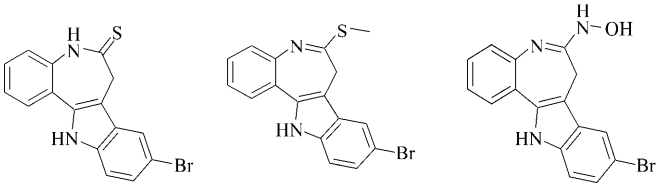


Fig. 1 Chart showing the structures of **35**, **36**, and **37**

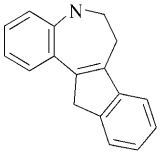


Fig. 2 Chart of the unsubstituted template

Results and discussion

CoMSIA and CoMFA studies

Three CoMSIA models using different fields are shown in Table 2. Only using the steric and the electrostatic fields, the 3D-QSAR model from the CoMSIA analyses (q^2=0.604) is acceptable. After the hydrophobic field was added, the predictive power of the 3D-QSAR model from the CoMSIA analyses was improved (q^2=0.659), which meant that the biological activity essentially exhibited a significant relationship with the hydrophobic field. After considering all five kinds of fields applied in CoMSIA including hydrogen-bonding fields (hydrogen-bonding donor and hydrogen-bonding acceptor fields), a CoMSIA model with the best statistical significance (q^2=0.704) was obtained. Obviously, the biological activity was also closely concerned with hydrogen-bonding properties, and all of the following CoMSIA models were concerned not only with the steric, electrostatic, and hydrophobic fields, but also with the two hydrogen-

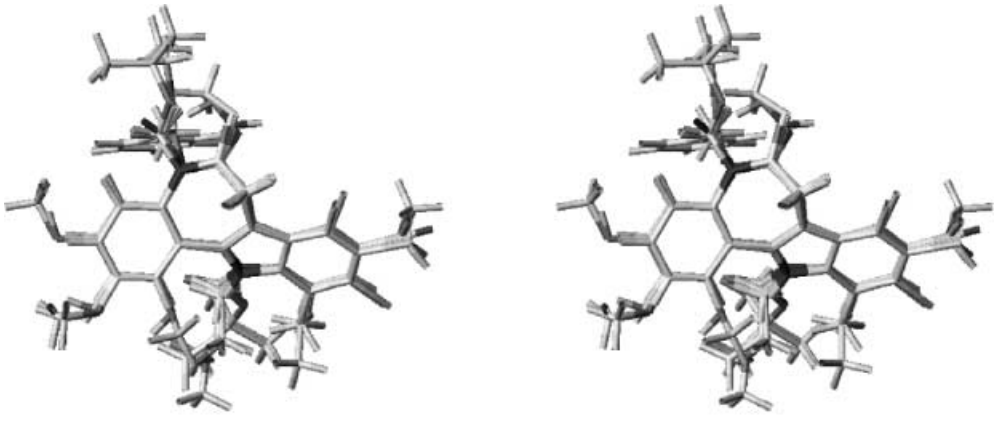


Fig. 3 Stereoview of all aligned molecules in the training and test set

Table 2 Results of the CoMFA and CoMSIA analyses of several different field combinations[a]

	CoMFA(1)	CoMFA(2)	CoMSIA(1)	CoMSIA(2)	CoMSIA(3)
q^2	0.307	0.638	0.604	0.659	0.704
r^2	0.563	0.999	0.974	0.983	0.983
F	41.224	1327.218	169.859	178.839	208.746
No. of compounds	1	12	6	8	7
Std error of estimate	0.720	0.049	0.190	0.161	0.160
Fraction					
Steric	0.270	0.300	0.275	0.190	0.086
Electrostatic	0.730	0.166	0.725	0.537	0.320
H-bond acceptor		0.342			0.176
H-bond donor		0.191			0.239
Hydrophobic				0.273	0.174

[a] In these models, all parameters concerned with CoMSIA and CoMFA analyses are defined as default values. (The attenuation factor used in the CoMSIA analyses above is 0.3, and the grid spacing used is 2.0 Å)

Table 3 Effect on the results of the CoMFA analyses using different grid spacings

	CoMFA(1)	CoMFA(2)	CoMFA(3)	CoMFA(4)	CoMFA(5)
Grid spacing (Å)	1.0	1.5	2.0	2.5	3.0
q^2	0.566	0.690	0.638	0.313	0.443
r^2	1.000	1.000	0.999	0.466	0.977
F	5301.393	2665.049	1327.218	27.914	97.449
No. of compounds	19	17	12	1	10
Std error of estimate	0.019	0.029	0.049	0.796	0.195
Fraction					
Steric	0.313	0.368	0.342	0.279	0.445
Electrostatic	0.161	0.186	0.191	0.069	0.291
H-bond acceptor	0.281	0.252	0.300	0.316	0.264
H-bond donor	0.244	0.195	0.166	0.336	0.000

Table 4 Effect on the results of the CoMSIA analyses using different grid spacings

	CoMSIA(1)	CoMSIA(2)	CoMSIA(3)	CoMSIA(4)	CoMSIA(5)
Grid spacing (Å)	1.0	1.5	2.0	2.5	3.0
q^2	0.696	0.697	0.704	0.661	0.609
r^2	0.982	0.982	0.983	0.979	0.991
F	205.619	206.156	208.746	174.580	250.947
No. of compounds	7	7	7	7	10
Std error of estimate	0.161	0.161	0.160	0.174	0.122
Fraction					
Steric	0.085	0.084	0.086	0.095	0.132
Electrostatic	0.322	0.328	0.320	0.300	0.239
Hydrophobic	0.180	0.174	0.174	0.156	0.227
H-bond acceptor	0.172	0.178	0.179	0.199	0.179
H-bond donor	0.241	0.237	0.239	0.250	0.223

[a] The attenuation factor used in the CoMSIA analyses above is 0.3

bonding fields. Two CoMFA models were developed using the different field combinations (Table 2). Only using steric and electrostatic fields, the CoMFA model was statistically unacceptable (q^2=0.307). After adding the hydrogen-bonding fields, the predictive power of the CoMFA model was significantly improved (q^2=0.638), but it is still worse than the best CoMSIA model.

In CoMSIA, besides the usual steric and electrostatic fields in CoMFA, three other kinds of fields, including a hydrophobic field, a hydrogen-bonding donor field, and a hydrogen-bonding acceptor field, were provided. Should we consider all these five fields at once in a CoMSIA model? We think that different researchers would give different answers according to the different systems studied. For example, some researchers consider that the hydrogen-bonding interactions should be distinguished from electrostatic interactions; consequently, the hydrogen-bonding fields should be considered in a CoMFA or CoMSIA analysis. However, others consider that hydrogen bonds are totally electrostatic in nature, which suggests no substantial angle dependence and a very slow falloff with distance. In practical studies, we think that the properties considered probably intercorrelate in a complicated manner. In the present work, we have considered five fields to produce the best 3D-QSAR model. It is unlikely that they are completely independent of each other; the degree of the interdependence is difficult to estimate. Focusing solely on the predictive power of a QSAR model cannot justify the consideration of five different kinds of fields. However, considering five fields provides an opportunity to partition the variance analysis with respect to particular physicochemical properties associated with the molecules. This aspect is of utmost importance if a targeted optimization of molecules in a design program is anticipated and 3D-QSAR is intended to support this step.

Optimization of 3D-QSAR models

At the start of this study, we used a default grid spacing of 2.0 Å, as suggested in SYBYL. Moreover, four additional grid spacings with 1.0, 1.5, 2.5, and 3.0 Å were used for CoMSIA and CoMFA analyses (see Tables 3 and 4). From the results of CoMSIA analyses, it seems that, 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 marked, and the field properties around the molecular aggregate could be well expressed. When the grid spacing is defined as 3.0 Å, some information on the field properties in some regions is lost, but the 3D-QSAR model (q^2=0.609) still has convincing predictive power. The 2.0 Å grid spacing resulted in the best model in terms of both cross-validated and non-cross-validated statistics. The model with the 2.0 Å grid spacing was selected as the model of choice for predictive purposes. In our previous work, we also found that, when the grid spacing is smaller than 3.0 Å, the statistical significance of the CoMSIA models did not show any obvious fluctuation. [18]

For comparison, the shift of the q^2 values for the 3D-QSAR models from CoMFA was evident as the grid spacing was changed. The 2.0 Å grid was shifted by +0.5 Å and +1.0 Å to determine the effect of altered lattice point location on the results of the CoMFA study. This resulted in significantly lower correlation (q^2=0.313 and q^2=0.443, respectively) than in the original study with 2 Å (q^2=0.638). Thus, at the large grid resolution in CoMFA, some important information in some regions may be lost. At lower grid spacing (1.0 Å), the 3D-QSAR model produced reduced statistical significance (q^2=0.566). This is because the increase in the number of lattice points also increases the noise in the 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 increased computation time and decreased predictivity. The best significant CoMFA model (q^2=0.690) was obtained with the grid spacing defined as 1.5 Å. However, its statistical significance seemed to be slightly poorer than the best model from the CoMSIA analyses (Table 2).

If we compare the values of q^2 for the CoMSIA and CoMFA models in Tables 3 and 4, we find that the changes of q^2 with the grid resolution in CoMSIA and CoMFA are quite similar. Neither of the best CoMSIA and CoMFA models prefers a specific grid spacing. Researchers generally select the model that works best when the resolution is set properly. Too high or too low a grid resolution will reduce the statistical significance of the CoMSIA or CoMFA models. However, for CoMSIA and CoMFA, the changeable amplitude of q^2 in different grid resolutions is quite different. For example, the difference of the q^2 of the CoMSIA models in 2.0 and 2.5 Å grid spacings is only 0.043, while this value for CoMFA is 0.327. In CoMSIA and CoMFA, a discrete grid must be used to represent the continuous molecular fields, and

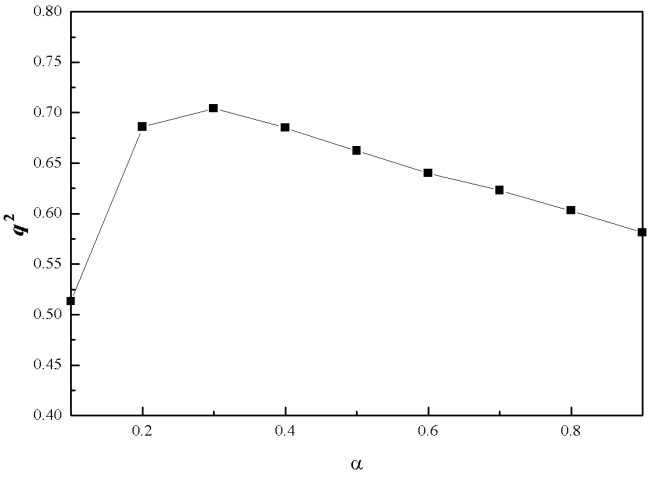


Fig. 4 Variation of q^2 upon changing the attenuation factor α used in the distance dependence between the probe atom and the atoms of the molecules in CoMSIA

the steric and electrostatic fields on each lattice point are calculated with distance-sensitive functions. In CoMFA, the Lennard-Jones steric field decays significantly with the increase of the distance. For CoMFA calculations, if we use a relatively large grid spacing, some important information in some regions may be lost. In CoMSIA, the Gaussian-type molecular fields will decay more slowly and gently with the increase of the distance, so the influence of the large grid spacing is not so obvious.

At 2.0 Å grid spacing, another important factor contributing to the CoMSIA analyses, the attenuation factor α, was also investigated. A series of α were selected from 0.1 to 0.9. In CoMSIA, a Gaussian-type distance dependence was applied. Reducing α to smaller values 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 α implied a more localized evaluation of similarity. Figure 4 shows the change of q^2 of different 3D-QSAR models with different α. For the data set, the values of α=0.2, 0.3, and 0.4 produced similar statistical models, but the statistical significance of the training set showed that the model of α=0.3 was much better than the other two models using α=0.2 and 0.4.

Through careful optimization of different models with different grid spacings and attenuation factors, the model from CoMSIA analyses with 2.0 Å grid spacing (α=0.3) was selected as the best judged by the cross-validation correlation (q^2=0.704, F=208.746, SD=0.160), and the following discussions refer only to this model. As a whole, the 3D-QSAR models from CoMSIA analyses were not very sensitive to the change of the grid spacing and the attenuation factor, but careful calibrations are needed to get the best model. As a comparison, the 3D-QSAR models from CoMFA analyses changed greatly with the shift of the grid spacing.

Figure 5 shows a plot of the observed versus calculated biological activities of the five candidates in the test

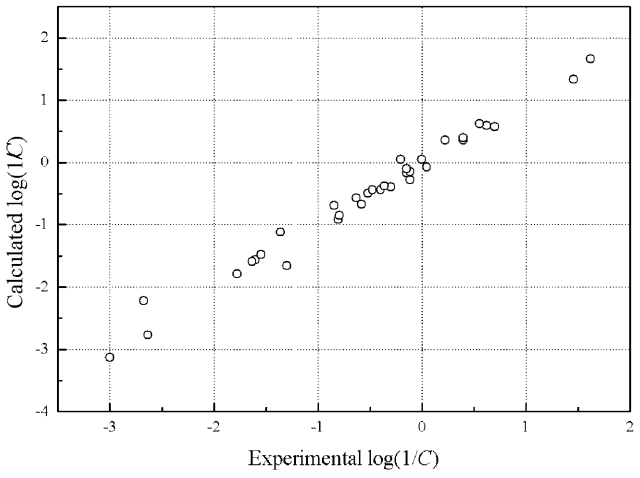


Fig. 5 Comparison of experimental log(1/*C*) with calculated log(1/*C*) using the best CoMSIA model

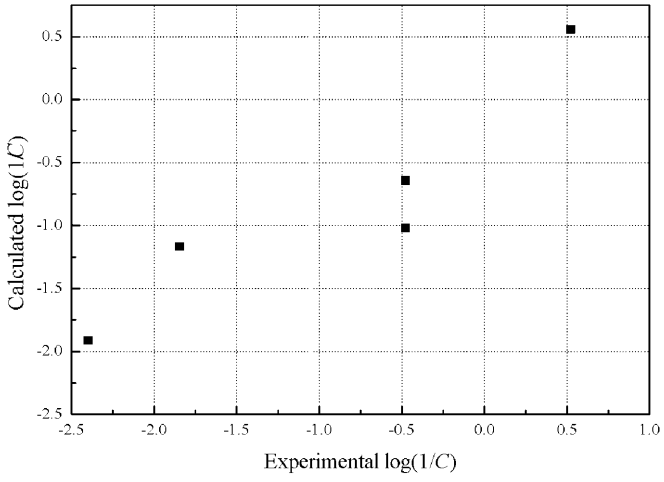


Fig. 6 Plot of the actual prediction of 3D-QSAR from CoMSIA(3) model in Table 4

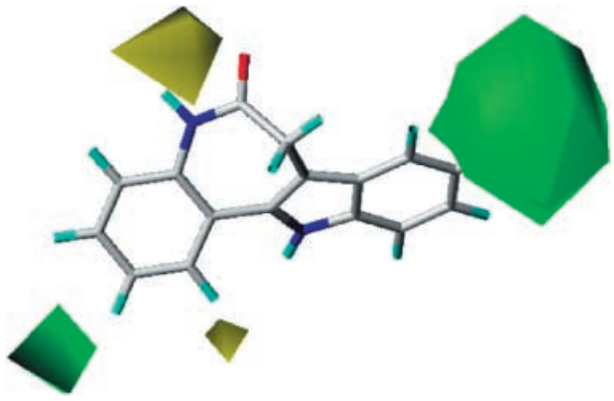


Fig. 7 The contour plot of the CoMSIA steric fields (stedv*coeff). The favorable steric areas with more bulk are indicated by *green* isopleths, whereas the disfavored steric areas are shown by *yellow* isopleths. The most active compound **9** is shown as the reference compound

Fig. 8 The contour plot of the CoMSIA electrostatic fields (stedv*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 **9** is shown as the reference compound

set. The calculated biological activities using the best CoMSIA model (grid spacing=2.0 Å, α=0.3) and the residue values from the observed values for the training set are shown in Table 1. The best 3D-QSAR model predicted well for the five tested compounds in terms of actual prediction, and the predictions were even better than for some molecules in the training set (see Fig. 6). So the derived model was satisfactory from the viewpoint of statistical significance and actual predictive ability. Combination of combinatorial library design, followed by application of the predictive tool developed in this research, could lead to a new compound with the activity of interest.

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 variables. At this point, the major advantage of CoMSIA compared to standard CoMFA becomes important: its better ability to visualize and interpret the correlations obtained in terms of field contributions. Strictly speaking, the plots represent contours of the coefficients obtained from PLS. They indicate those lattice points where a particular property contributes significantly and thus explain the variation in affinity data. They give an excellent insight into the relationship between structure and activity for the different physicochemical properties of the considered structures. The coefficient contour plots may help to determine the important regions where any changes of some kind of property may affect the biological activity and identify the important features contributing to interactions between ligand and receptor in the active site. Graphical representations of the best CoMSIA model are displayed in Figs. 7, 8, 9, 10, and 11.

Figures 7 and 8 show the steric and electrostatic coefficient contour maps, respectively. The compound **9**

with the highest inhibitory activity was taken as the reference compound for specifying 3D space. In Fig. 7, the CoMSIA 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. One bulky unfavorable area was near the 5-position, while two bulky favorable areas were near the 2- and 9-positions on the benzene ring. From the size of the two green areas, the green region near the 9-position is seen to be more important than that near the 2-position. Different substituted groups near these areas would affect the binding affinities by the steric complementarity between the receptor and the ligand. For example, the 9-bromo-substituted compounds **1**, **24**, **18**, and **25** showed higher activity than their unsubstituted counterparts **3**, **15**, **17**, and **20**. The shift of the bromo substituent from the 9-position in kenpaullone (compound 1) to the 10-position (compound 2), 11-position (compound 12) resulted in decreased activity. The different inhibition pattern caused by the shift of the substituents shows that the shape of the binding site means that R_2 group must be substituted at the specific position in order to produce good surface complementarity between ligand and enzyme. The unfavorable bulky region is located near the 5-position. Large groups in this area will decrease the activity. For example, none of the kenpaullone derivatives with substitution at the indole nitrogen atom was superior to the lead structure with respect to CDK1 inhibition.

In Fig. 8, regions where increased positive charge is favorable for affinity are indicated in blue, while regions where increased negative charge is favorable for affinity are indicated in red. From the fraction of field, the electrostatic field (0.320) seems more important than the steric field (0.086). There are two red areas near the 9- and 5-positions. In these areas, introduction of stronger electron-withdrawing substituents will be favorable for biological activity. For example, replacement of the 9-bromo substituent of kenpaullone with the electron-withdrawing 9-cyano group (compound 9) and 9-nitro group (compound 10) leads to a more than 10-fold higher potency than kenpaullone. We were interested to see that, near the 9-position on the benzene ring, there existed a relatively small blue area neighboring the red area. As to the blue region near the benzene ring, more positive charge is preferred on the benzene ring. The strong charge-withdrawing groups linked to the benzene ring will make the charge distributed on the benzene ring relatively positive, which is expressed by the blue region.

Figure 9 reveals regions colored in gray that correspond to a favorable influence of hydrophilicity and regions colored in yellow which correspond to a favorable influence of hydrophobicity. From Fig. 9, it can be seen that in this group of compounds more hydrophilic regions were preferred. The relatively excessive area colored in gray suggested that the more hydrophilic substituents near these regions would enhance the biological activity. It seems that the residues in the active site directly interacting with the inhibitors are mainly hydrophilic, and the introduction of hydropholic groups will be favorable. There is only one yellow area near the 1-position. We think that this area may be generated by the hydrophobic benzene ring.

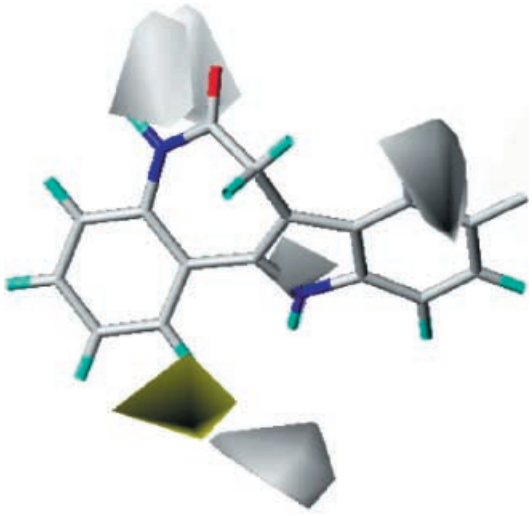


Fig. 9 The contour plot of the CoMSIA hydrophobic fields (stedv*coeff). The favorable hydrophobic areas are indicated by *yellow* isopleths, whereas the disfavored hydrophobic areas are shown by *white* isopleths. The most active compound **9** is shown as the reference compound

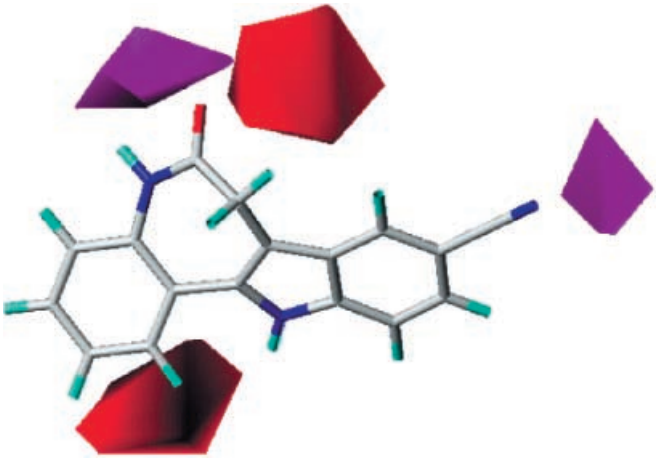


Fig. 10 The contour plot of the CoMSIA H-bond acceptor fields (stedv*coeff). *Magenta* isopleth contours maps beyond the ligands where H-bond donor groups in the receptor will be favorable for biological activity, while *red* isopleths represents H-bond donor groups in the receptor unfavorable for biological activity. The most active compound **9** is shown as the reference compound

The graphical interpretations of the field contributions of the H-bond properties are shown in Fig. 10 (H-bond acceptor field) and Fig. 11 (H-bond donor field). In principle, they should highlight areas beyond the ligands where putative hydrogen partners in the enzyme could form H-bonds to influence binding affinity. Two magenta isopleth contour maps in the H-bond acceptor field

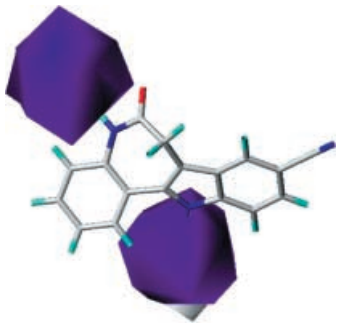


Fig. 11 The contour plot of the CoMSIA H-bond donor fields (stedv*coeff). *Blue* isopleth contours maps beyond the ligands where H-bond acceptor groups in the receptor will be favorable for biological activity, while *gray* isopleths represents H-bond acceptor groups in the receptor unfavorable for biological activity. The most active compound **9** is shown as the reference compound

(Fig. 10) surround the R_2 and R_3 sites, which may be attributed to the lone pair electrons on the indole oxygen atom and the O atoms on the substituents in the 9-position with the corresponding H-bond donors in receptor. The other two regions contoured in red should avoid H-bond acceptor capabilities. In the H-bond donor field (Fig. 11), there are two blue contour maps around the R_3 and R_4 sites. We believe that these two areas are obviously caused by the lactam and the indole nitrogen atoms. Two hydrogen atoms linked to these two nitrogen atoms acted as optimal hydrogen donors. However, when these hydrogen atoms are replaced by other groups, the hydrogen-donor capabilities are lost. Thus, none of the kenpaullone derivatives with substitution at either the lactam or the indole nitrogen atoms were superior to the lead structure with respect to the CDK1 inhibitions (compounds 26 – 34.

The previous discussion of the graphical results of the different field contributions has demonstrated that many of the features in these maps can be interpreted in terms of properties reflected in the surrounding environment. The modeling study demonstrates that the CoMSIA method can be used to rank possible candidates from a library in a predictive manner. Such information is important in designing and selection of components for a computational combinatorial library.

Conclusions

Multiple CoMSIA and CoMFA 3D-QSAR models have been developed from a data set of paullones as CDK inhibitors and validated with a test set. Five different types of fields, including steric, electrostatic, hydrophobic, and H-bond acceptor and donor fields were considered in CoMSIA analyses. In order to search for the best 3D-QSAR model, systematic variations of some parameters in CoMSIA and CoMFA were performed. The best model obtained from the CoMSIA analyses possesses promising predictive ability as indicated by the high cross-validated correlation and the prediction on the external test set. The characteristics of the CoMSIA 3D contour plots derived in this study may be 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.

Acknowledgements This project is supported by National Natural Science Foundations of China (NSFC) 29573095.

References

1. Morgan, D. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 261–291.
2. Draetta, G.; Pagano, M. *Annu. Rep. Med. Chem.* **1996**, *31*, 241–248.
3. Hartwell, L.; Pines H. *J. Cell* **1994**, *79*, 573–582.
4. Meijer, L. *Trends Cell Biol.* **1996**, *6*, 393–397.
5. Coleman, K. G.; Lyssikatos, J. P.; Ynag, B. V. *Annu. Rep. Med. Chem*. **1997**, *32*, 171–179.
6. Yamamoto, H.; Monden, T.; Miyoshi, H.; Izawa, H.; Ikeda, K.; Tsujie, M.; Ohnishi, T.; Sekimoto, M.; Tomita, N.; Monden, M. *Int. J. Oncol*. **1998**, *13*, 233–239.
7. Wada, M.; Hosotani, R.; Lee, J. U.; Doi, R.; Koshiba, T.; Fujimoto, K.; Miyamoto, Y.; Tsuji, S.; Nakajima, S.; Okuyama, A.; Imamura, M. *Anticancer Res.* **1998**, *18*, 2559–2566.
8. Sedlacek, H. H.; Czech, J.; Naik, R.; Kaur, G.; Worland, P.; Losiewicz, M.; Parker, B.; Carlson, B.; Smith, A.; Senderowicz, A.; Sausville, E. *Int. J. Oncol*. **1996**, *9*, 1143–1168.
9. Schultz, C.; Andreas, L.; Maryse, L.; Daniel, W. Z.; Rick, G.; Edward, A. S.; Laurent, M.; Conrad, K. *J. Med. Chem.* **1999**, *42*, 2909–2919.
10. Cramer, R. D.; Patterson, D. E.; Bunce, J. D.; *J.Am.Chem.Soc.* **1988**, *110*, 5959–5967.
11. Klebe, G.; Abraham, U.; Mietzner, T.; *J. Med. Chem.* **1994**, 37, 4130–4146.
12. Klebe, G..; Abraham, U. *J. Comput. Aid. Mol. Des.*, **1999**, 13, 1–10.
13. Bohm M.; Sturzebecher, J.; Klebe, G. *J. Med. Chem.*, **1999**, 42 (3), 458–477.
14. SYBYL, version 6.5; Tripos Associates: St. Louis, MO, USA, 1999; http://www.tripos.com/.
15. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.
16. Mopac, version 7.0; QCPE 688; Quantum Chemistry Program Exchange: Bloomington, USA, 1995; http://qcpe.chem.indiana.edu/.
17. Bush, B. L.; Nachbar, R. B. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.
18. Hou, T. J., Li, Z. M., Li, Z., Liu, J., Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1002–1009.