

## 3D-QSAR and docking studies of aldehyde inhibitors of human cathepsin K

Xulin Pan,<sup>a,b</sup> Ninghua Tan,<sup>a,\*</sup> Guangzhi Zeng,<sup>a,b</sup> Hongjin Han<sup>a</sup> and Huoqiang Huang<sup>a,b</sup>

<sup>a</sup>State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China

<sup>b</sup>Graduate School of the Chinese Academy of Sciences, Beijing 100039, China

Received 2 November 2005; revised 24 November 2005; accepted 29 November 2005

Available online 27 December 2005

**Abstract**—In order to better understand the structural and chemical features of human cathepsin K (CatK), which is an important cysteine protease in the pathogenesis of osteoporosis, the 3D-QSAR (CoMFA) studies were conducted on recently explored aldehyde compounds with known CatK inhibitory activities. The genetic algorithm of GOLD2.2 has been employed to position 59 aldehyde compounds into the active sites of CatK to determine the probable binding conformation. Good correlations between the predicted binding free energies and the experimental inhibitory activities suggested that the identified binding conformations of these potential inhibitors are reliable. The docking results also provided a reliable conformational alignment scheme for 3D-QSAR model. Based on the docking conformations, highly predictive comparative molecular field analysis (CoMFA) was performed with  $q^2$  value of 0.723. The predictive ability was validated by some compounds that were not included in the training set. Furthermore, the CoMFA model was mapped back to the binding sites of CatK, to get a better understanding of vital interactions between the aldehyde compounds and the protease. The CoMFA field distributions are in good agreement with the structural characteristics of the binding groove of the CatK, which suggested that the *n*-Bu in R4 position is the favor group substitute at P1 and moderate groups in R2 group are required on P2 substitute. In addition, 3D-QSAR results also demonstrated that aldehyde is an important pharmacophore because of electrostatic effect. These results, together with the good correlations between the inhibitory activities and the binding free energies predicted by GOLD2.2, demonstrated the power of combining docking/QSAR approach to explore the probable binding conformations of compounds at the active sites of the protein target, and further provided useful information in understanding the structural and chemical features of CatK in designing and finding new potential inhibitors.

© 2005 Elsevier Ltd. All rights reserved.

### 1. Introduction

Cathepsin K (CatK, EC 3.4.22.38), a 27-kDa cysteine protease of the papain superfamily, is an important extracellular matrix-degrading protease. It is abundantly and selectively expressed in bone-resorbing osteoclasts. In the acidic lacunae of bone resorption, a Cys25-S<sup>−</sup>/His162-ImH<sup>+</sup> ion pair will be formed by the thiol and imidazole side chains of Cys25 and His162. For the nucleophilic attack of S<sup>−</sup> from the ion pair, CatK cleaves type I collagen at multiple sites, which is the predominant component of the extracellular matrix of bone. So, CatK is a pivotal protease in osteoclast-mediated bone resorption, and it highlights the

attractiveness of this cysteine protease as a target for inhibition in diseases characterized by elevated levels of bone turnover such as osteoporosis.<sup>1–4</sup> For this reason, many kinds of inhibitors against CatK have been designed, which include nonpeptidic biaryl compounds, aldehydes and their derivatives, acyclic and cyclic ketones, nonpeptidyl nitriles, epoxysuccinyl analogues,  $\beta$ -lactams, vinyl sulfones, and so on. Some of them inhibited bone resorption well in vivo.<sup>1,5</sup>

Currently, within the research program on cathepsin run in our laboratory during past few years, preliminary studies showed that some natural compounds, such as biflavones, appeared to have an interesting affinity.<sup>6</sup> Furthermore, quantum chemistry and dock method were employed to elucidate their structure–activity relationship (SAR) as well as binding mode with protein.<sup>7</sup> Apparently, the interaction mechanism of the CatK with its inhibitors would be greatly helpful in discovering

**Keywords:** 3D-QSAR; Docking; Cathepsin K.

\*Corresponding author. Tel.: +86 871 5223800; fax: +86 871 5223800; e-mail: [nhtan@mail.kib.ac.cn](mailto:nhtan@mail.kib.ac.cn)

natural inhibitors for ceasing the function of CatK. Moreover, some three-dimensional structures of the CatK itself and its complex with different inhibitors have been experimentally determined.<sup>8,9</sup> These crystal structures provided not only insights into the interaction mechanisms of CatK with the inhibitors, but also valuable clues for designing new inhibitors.<sup>10–14</sup> Recently, based on the structural information, Deaton and co-workers designed and synthesized a series of aldehyde analogues which mainly interact with nonprimed active sites of CatK, their inhibitory activities to CatK also measured.<sup>15,16</sup> In this paper, with the molecular docking and 3D-QSAR (CoMFA) analysis it is possible to get new insights into the relationship between the structural information of the series of 59 aldehyde inhibitors<sup>15,16</sup> and the inhibitory potency, aimed at identifying structural features in CatK that can be used to find new inhibitors.

## 2. Computational details

### 2.1. Biological data and molecular structures

Tables 1 and 3 showed a series of CatK aldehyde inhibitors published by Deaton et al. in 2004,<sup>15,16</sup> which were divided into a training set and a test set. The training set consists of 50 compounds and the test set, which was selected randomly, comprises nine compounds. The CatK IC<sub>50</sub> were converted to pIC<sub>50</sub> (–log IC<sub>50s</sub>) and used as dependent variables in the CoMFA calculations.

The 3D structures of these compounds were constructed by using molecular modeling software package Sybyl6.9.<sup>17</sup> Partial atomic charges were calculated by the Gasteiger–Huckel method, and energy minimizations were performed using the Tripos force field<sup>18</sup> with a distance-dependent dielectric and the Powell conjugate gradient algorithm (convergence criterion of 0.001 kcal/mol Å). Atomic coordinates for the CatK complex with ligand TCO, used for our modeling study, have been deposited in the Brookhaven Protein DataBank<sup>16</sup> with a resolution of 1.8 Å (PDB ID:1Q6K).

### 2.2. Docking studies

To locate the appropriate binding orientations and conformations of these aldehyde analogue inhibitors interacting with CatK, a powerful computational searching method is needed. The advanced molecular docking program GOLD, version 2.2,<sup>19</sup> which uses a powerful genetic algorithm (GA) method for conformational search and docking and being widely regarded as one of the best docking programs, was employed to generate an ensemble of docked conformations. The structures of CatK and aldehyde analogue inhibitors were built using the Sybyl6.9 molecular modeling software. The original ligand as well as the water molecules were removed from the coordinated set. The genetic operators were 100 for the population size, 1.1 for the selection, 5 for the number of subpopulations, 100,000 for the maximum number of genetic applications, and 2 for the size of the niche used to increase population diversity. The weights

were chosen so that crossover mutations were applied with equal probability (95/95 for the values) and migration was applied 5% of the time.

### 2.3. Prediction of binding free energies

ChemScoring function for predicting binding free energy encoded in GOLD was applied to evaluate the binding affinities between CatK and 59 inhibitors. This scoring function described by Eldridge et al.<sup>20,21</sup> estimates the free energy of binding of a ligand to protein as follows:

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} S_{\text{hbond}} + \Delta G_{\text{metal}} S_{\text{metal}} + \Delta G_{\text{lipo}} S_{\text{lipo}} + \Delta G_{\text{rot}} H_{\text{rot}}, \quad (1)$$

where  $S_{\text{hbond}}$ ,  $S_{\text{metal}}$ , and  $S_{\text{lipo}}$  are scores for hydrogen bonding, acceptor metal, and lipophilic interactions, respectively.  $H_{\text{rot}}$  is a score representing the loss of conformational entropy of the ligand upon binding to the protein. The  $\Delta G$  terms are coefficients derived from a multiple linear regression analysis on a training set of 82 protein–ligand complexes from the PDB.<sup>22</sup>

### 2.4. Structural alignment

Ten conformations were obtained through GOLD for each ligand and that conformation with the strongest predicted binding affinity to the CatK was extracted from the optimized inhibitor–CatK complex. These conformations were aligned together inside the binding pocket of CatK and used directly for CoFMAs to explore the specific contributions of electrostatic and steric effects to the molecular bioactivities.

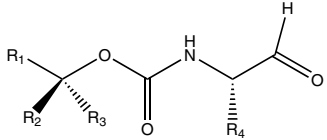
### 2.5. CoMFA

Steric and electrostatic interactions were calculated using the Tripos force field with a distance-dependent dielectric constant at all intersections in a regular space (2 Å) grid taking an sp<sup>3</sup> carbon atom as steric probe and a +1 charge as electrostatic probe. The cutoff was set to 30 kcal/mol. With standard options for scaling of variables, the regression analysis was carried out using the full cross-validated partial least-squares (PLS) methods of LOO (leave-one-out) and leave-20%-out. The minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. The final model, noncross-validated conventional analysis, was developed with the optimum number of components to yield a noncross-validated  $r^2$  value.

## 3. Results and discussion

### 3.1. Binding conformations of aldehyde inhibitors

In order to determine the probable binding conformations of these aldehyde inhibitors, the GOLD was used to dock all compounds into the active sites of CatK. The docking reliability was validated using the known

**Table 1.** Structures, experimental activities (pIC<sub>50</sub>), predicted activities (PAs), residuals by CoMFA model, and docking energies of compounds in the training set


Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	pIC <sub>50</sub>	Predict	Residue	ChemScore (kJ/mol)
1	Me	Me	Me	H	2.24	2.54	−0.30	−20.43
2	Me	Me	Me	Me	3.70	3.77	−0.07	−21.49
3	Me	Me	Me	Et	4.24	3.70	0.54	−22.52
4	Me	Me	Me	<i>i</i> -Pr	3.80	3.94	−0.14	−22.52
5	Me	Me	Me	Pr	4.42	4.35	0.07	−21.46
6	Me	Me	Me	CH(Me)Et( <i>S</i> )	3.60	3.93	−0.33	−22.86
7	Me	Me	Me	CH(Me)Et( <i>R</i> )	3.89	4.04	−0.15	−22.38
8	Me	Me	Me	CH <sub>2</sub> C(CH <sub>3</sub> =CH <sub>2</sub> )	4.10	3.93	0.17	−21.85
9	Me	Me	Me	CH <sub>2</sub> - <i>i</i> -Pr	4.51	4.36	0.15	−22.63
10	Me	Me	Me	CH <sub>2</sub> tBu	3.60	3.52	0.08	−20.89
11	Me	Me	Me	Bu	4.29	4.24	0.05	−21.71
12	Me	Me	Me	( <i>Z</i> )CH <sub>2</sub> CH=CHCH <sub>3</sub>	4.39	4.24	0.15	−22.20
13	Me	Me	Me	( <i>E</i> )CH <sub>2</sub> CH=CHCH <sub>3</sub>	3.82	3.73	0.09	−21.93
14	Me	Me	Me	CH <sub>2</sub> C≡CCH <sub>3</sub>	3.39	3.47	−0.08	−22.77
15	Me	Me	Me	<i>n</i> -Pentyl	3.96	4.13	−0.17	−23.41
16	Me	Me	Me	<i>n</i> -Hexyl	4.64	4.78	−0.14	−23.19
17	Me	Me	Me	CH <sub>2</sub> SEt	4.08	4.12	−0.04	−23.15
18	Me	Me	Me	(CH <sub>2</sub> ) <sub>2</sub> OMe	3.17	3.21	−0.04	−21.18
19	Me	Me	Me	(CH <sub>2</sub> ) <sub>2</sub> SMe	4.11	3.99	0.12	−22.26
20	Me	Me	Me	Ph	4.29	4.28	0.01	−23.05
21	Me	Me	Me	Benzyl	3.96	3.96	0.00	−23.12
22	Me	Me	Me	CH <sub>2</sub> cyclohexyl	4.44	4.57	−0.13	−23.44
23	Me	Me	Me	(CH <sub>2</sub> ) <sub>2</sub> cyclohexyl	4.59	4.60	−0.01	−25.52
24	Me	Me	Me	(CH <sub>2</sub> ) <sub>3</sub> Ph	4.82	4.81	0.01	−26.04
25	Me	Me	Me	(CH <sub>2</sub> ) <sub>3</sub> cyclohexyl	4.80	4.72	0.08	−26.98
26	Me	Me	Me	CH <sub>2</sub> SCH <sub>2</sub>	4.40	4.44	−0.04	−26.17
27	Me	Me	Me	CH <sub>2</sub> NHCO <sub>2</sub> Me	3.43	3.51	−0.08	−20.61
28	Me	Me	Me	(CH <sub>2</sub> ) <sub>4</sub> NHCO <sub>2</sub> Me	4.82	4.80	0.02	−22.56
29	Me	Me	Me	(CH <sub>2</sub> ) <sub>4</sub> N(H)COCF <sub>3</sub>	4.39	4.31	0.08	−21.22
30	Me	Me	Me	(CH <sub>2</sub> ) <sub>4</sub> N(Me)COCF <sub>3</sub>	4.30	4.31	−0.01	−21.54
31	PhCH <sub>2</sub>	Me	Me	<i>n</i> -Bu	4.92	5.08	−0.16	−26.95
32	PhCH <sub>2</sub>	(CH <sub>2</sub> ) <sub>3</sub>		<i>n</i> -Bu	5.62	5.48	0.14	−26.86
33	PhCH <sub>2</sub>	Et	Et	<i>n</i> -Bu	5.68	5.76	−0.08	−26.71
34	PhCH <sub>2</sub>	(CH <sub>2</sub> ) <sub>4</sub>		<i>n</i> -Bu	6.46	6.63	−0.17	−26.62
35	PhCH <sub>2</sub>	(CH <sub>2</sub> ) <sub>5</sub>		<i>n</i> -Bu	5.70	5.53	0.17	−26.25
36	PhCH <sub>2</sub>	Me	H	<i>n</i> -Bu	5.74	5.75	−0.01	−26.66
37	C <sub>6</sub> H <sub>11</sub> CH <sub>2</sub>	Me	H	<i>n</i> -Bu	5.57	5.49	0.08	−26.97
38	PhCH <sub>2</sub>	Et	H	<i>n</i> -Bu	6.89	6.67	−0.85	−26.85
39	PhCH <sub>2</sub>	<i>n</i> -Pr	H	<i>n</i> -Bu	5.82	5.82	0.00	−27.88
40	PhCH <sub>2</sub>	<i>i</i> -Pr	H	<i>n</i> -Bu	6.30	6.41	−0.11	−28.34
41	PhCH <sub>2</sub>	<i>i</i> -Bu	H	<i>n</i> -Bu	5.19	5.19	0.00	−27.33
42	H	Et	Et	<i>n</i> -Bu	5.40	5.45	−0.05	−23.46
43	H	<i>n</i> -Pr	<i>n</i> -Pr	<i>n</i> -Bu	6.06	6.10	−0.04	−25.46
44	H	<i>i</i> -Pr	<i>i</i> -Pr	<i>n</i> -Bu	6.25	6.44	−0.19	−25.39
45	H	<i>i</i> -Bu	<i>i</i> -Bu	<i>n</i> -Bu	5.96	6.18	−0.22	−25.94
46	Me	<i>i</i> -Pr	<i>i</i> -Pr	<i>n</i> -Bu	5.33	5.32	0.01	−23.20
47	3-MeO-Ph-CH <sub>2</sub>	Me	Me	<i>n</i> -Bu	5.27	5.19	0.08	−27.97
48	2-Cl-Ph-CH <sub>2</sub>	Me	Me	<i>n</i> -Bu	5.51	5.37	0.14	−27.64
49	4-Cl-Ph-CH <sub>2</sub>	Me	Me	<i>n</i> -Bu	5.34	5.23	0.11	−26.88
50	3-Thiophenyl-CH <sub>2</sub>	Me	Me	<i>n</i> -Bu	5.12	4.92	0.20	−27.73

X-ray structure of CatK in complex with a small molecular ligand TCO (Fig. 1). The ligand TCO was redocked to the binding sites of CatK and the docked conformation corresponding to the lowest free energies was selected as the most probable binding conformation.

The GOLD predicted conformation of TCO was shown in Figure 1 with the superposition of X-ray crystallographic one in active sites of CatK. The root-mean-square deviation (RMSD) between these two conformations is equal to 0.57 Å, suggesting that a high

**Table 2.** Summary of CoMFA results

	CoMFA	
	LOO	Leave-20%-out
<i>PLS statistics</i>		
$q^2$ (CV correlation coefficient)	0.723	0.713
$N$ (number of components)	6	5
$S$ (standard error of prediction)	0.159	0.189
$r^2$ (correlation coefficient)	0.976	0.966
$F$ ( $F$ -ratio)	294.58	247
<i>Field distribution (%)</i>		
Steric	59.9	65.2
Electrostatic	40.1	34.8
<i>Testing set</i>		
$r^2$ (correlation coefficient)	0.708	
$S$ (standard error of prediction)	0.188	

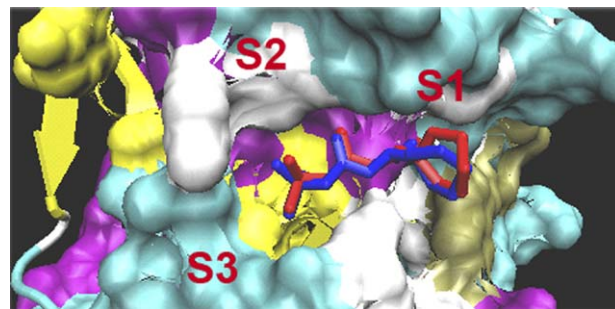
CV, cross-validation; LOO, leave-one-out.

docking reliability of GOLD in reproducing the experimentally observed binding mode for CatK inhibitors and the parameter set for the GOLD simulation is reasonable to reproduce the X-ray structure. The GOLD method and the parameter set could be extended to search the CatK binding conformations for other inhibitors accordingly.

Figure 2a shows the 3D model of aldehyde analogues, and Figure 2b illustrates the probable binding conformational alignment for the 59 aldehyde compounds chosen from the docked conformations according to their ChemScores. All of the aldehyde compounds are bonded in the active sites of CatK in a similar conformation of TCO (52) in the X-ray structure cocrystallized with CatK (Fig. 2), and the common chain structures superimposed each other rather well. Based on this set of binding conformations and their alignment, CoMFA was performed.

### 3.2. Docking study

The predicted binding free energies ( $\Delta G_{\text{binding}}$ ) of the aldehyde analogue to the CatK and the corresponding



**Figure 1.** Superimposition of docked compound calculated by GOLD (blue) and X-ray crystallographic (red) conformations of TCO in active sites of CatK. This figure was rendered with the program VMD1.83.<sup>23</sup>

experimental  $\text{pIC}_{50}$  ( $-\log \text{IC}_{50}$ ) values are also listed in Table 1. A correlation was found between  $\Delta G_{\text{binding}}$  and the  $\text{pIC}_{50}$ s via a linear regression analysis, as shown in Eq. 2 and Figure 3a, the correlation coefficient ( $r^2$ ) being 0.628. This rather good correlation demonstrated that the binding conformations and binding models of the aldehyde inhibitors to CatK are reasonable and reliable. Therefore, they should be more accurate conformations for CoMFA study than the usual ones by a manual structural alignment.

$$\Delta G_{\text{binding}} = -14.802 - 1.996 \times \text{pIC}_{50}, \quad (2)$$

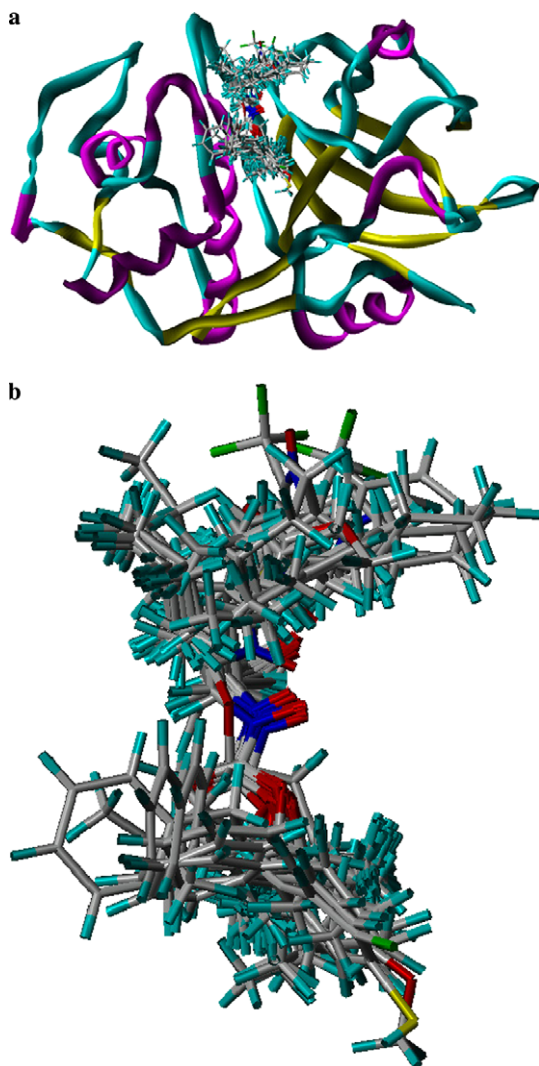
( $N = 50$ ,  $r^2 = 0.628$ ,  $F = 81.05$ ,  $S = 0.221$ ) where  $S$  is the standard error,  $r^2$  is the correlation coefficient, and  $F$  is the testing factor of the reliability.

In addition, Figure 3b represents the interaction model of the docked inhibitor 38 with CatK, as generated with the program Ligplot.<sup>24</sup> There are three hydrogen bonds formed between inhibitor 38, and the residues Cly66, Asn161, and His162 with bond lengths of 2.99, 2.68, and 2.99 Å, supporting the experimental conclusion that the nonprimed active sites of CatK are sterically very restricted. More attention should be paid to the phenyl substitute at R1 which interacts via  $\pi$ - $\pi$  stacking interaction with the amino acid residue Tyr67 of CatK. So,

**Table 3.** Structures, experimental activities ( $\text{pIC}_{50}$ s), predicted activities (PAs), residuals by CoMFA model, and docking energies of compounds in the test set

Compound					$\text{pIC}_{50}$	Predict	Residue	ChemScore (kJ/mol)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>				
51	Me	Me	Me	(CH <sub>2</sub> ) <sub>3</sub> CF <sub>3</sub>	4.06	4.09	−0.03	−20.45
52	Me	Me	Me	Cyclohexyl	4.59	3.97	0.62	−23.32
53	Me	Me	Me	(CH <sub>2</sub> ) <sub>2</sub> Ph	4.52	4.52	0.00	−23.83
54	Me	Me	Me	CH <sub>2</sub> OCH <sub>2</sub>	4.42	4.42	0.00	−23.88
55	Me	Me	Me	(CH <sub>2</sub> ) <sub>3</sub> N(Me)COCF <sub>3</sub>	4.54	4.50	0.04	−21.29
56	Ph(CH <sub>2</sub> ) <sub>2</sub>	H	H	<i>n</i> -Bu	3.22	3.16	0.06	−27.42
57	PhCH <sub>2</sub>	H	Me	<i>n</i> -Bu	4.00	4.14	−0.14	−27.37
58	PhCH <sub>2</sub>	PhCH <sub>2</sub>	H	<i>n</i> -Bu	4.15	4.05	0.10	−27.50
59	4-MeO-Ph-CH <sub>2</sub>	Me	Me	<i>n</i> -Bu	5.54	5.42	0.12	−27.29





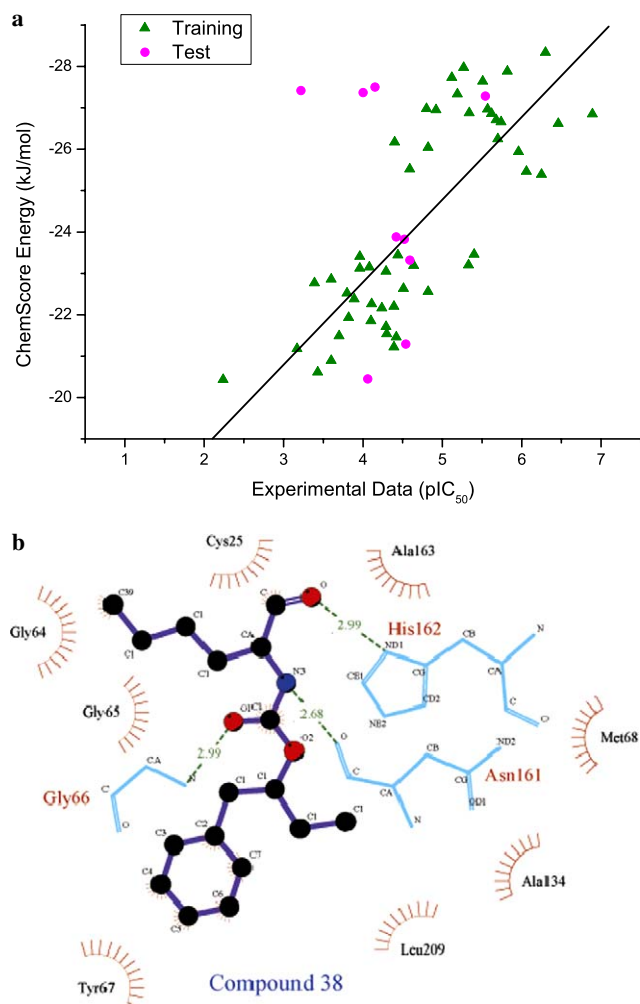
**Figure 2.** (a) Binding conformations of docked compounds at the active sites of CatK. (b) Superimposition of 59 aldehyde analogues for 3D-QSAR studies.

three outliers (**58**, **57**, and **56**) in the test set (Fig. 3a) probably exist in the  $\pi$ - $\pi$  interaction which may be optimistically calculated by ChemScore.

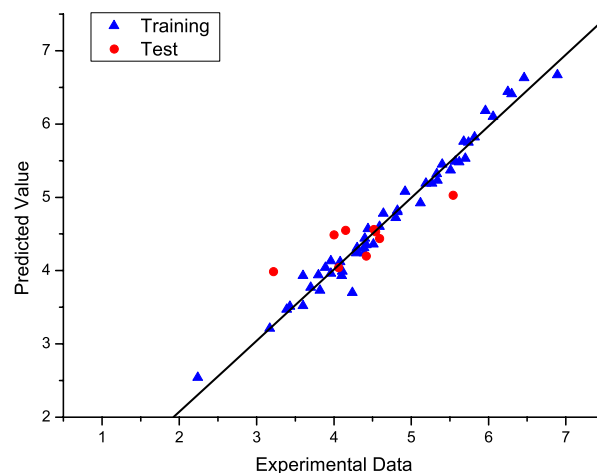
### 3.3. CoMFA model

Fifty of the 59 aldehyde CatK inhibitors were randomly picked up as training set for constructing CoMFA models, the remaining nine used as test set for the model validation.

PLS analysis was carried out for the 59 binding conformations, and the result is listed in Table 2, which showed that a CoMFA model with a cross-validated  $q^2$  of 0.723 for six components was obtained. The non-cross-validated PLS analysis with the optimum components of 6 revealed a conventional  $r^2$  value of 0.976,  $F = 294.58$ , and an estimated standard error of 0.159. The steric field descriptors explain 59.9% of the variance, while the electrostatic descriptors explain 40.1%. The predicted activities for the 50 inhibitors versus their experimental activities with their residues are listed in



**Figure 3.** (a) Plots of pIC<sub>50</sub> versus ChemScore energy for compounds from the training and the test sets. Training set predictions are represented with olive filled triangles and test set predictions with magenta filled circles. (b) Proposed interaction model of inhibitor **38** in the active sites of CatK, drawn using the Ligplot<sup>24</sup> program. Broken lines represent hydrogen bonds and spiked residues formed van der Waals contacts with the inhibitors.



**Figure 4.** Correlation between predicted activities (PA) by CoMFA models and the experimental pIC<sub>50</sub> values of training and test sets, blue filled triangles represent predictions for the training set, while red filled circles represent predictions for the test set.

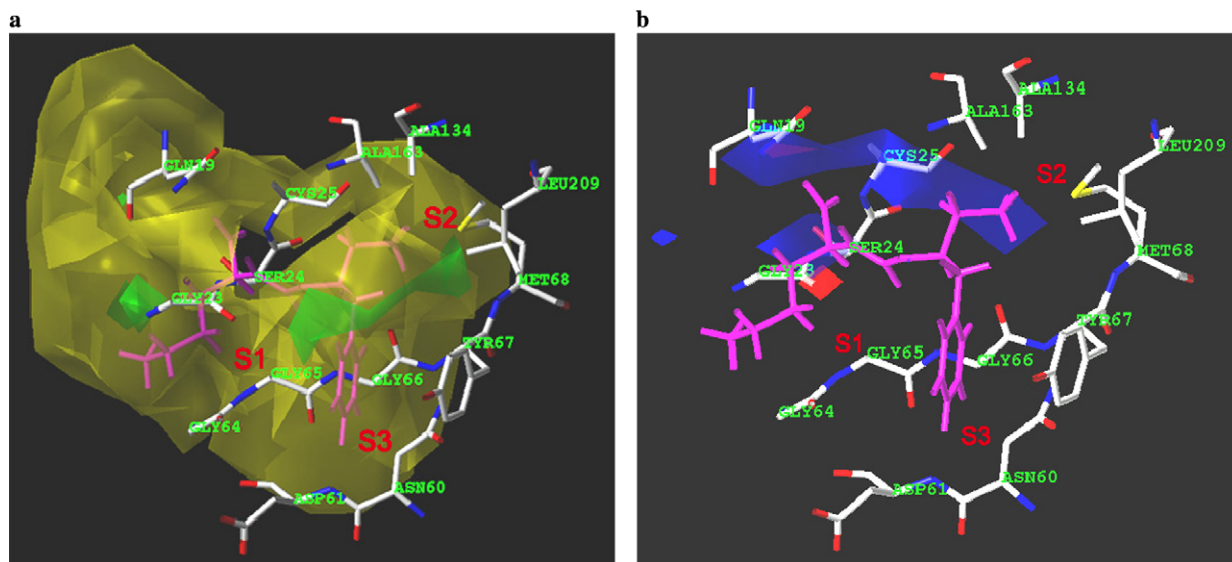
Table 1, the correlation between the predicted activities and the experimental activities depicted in Figure 4. Table 1 and Figure 4 demonstrate that the predicted activities by the constructed CoMFA model are in good agreement with the experimental data, suggesting that a reliable CoMFA model was successfully constructed. Furthermore, the fitness of the model (predictivity of the model) was computed by the leave-20%-out cross-validation method, which gave the similar results as LOO (Table 2).<sup>25</sup>

The CoMFA result is usually represented as 3D ‘coefficient contour.’ It shows regions where variations of steric and electrostatic nature in the structural features of the different molecules contained in the training set lead to increase or decrease in the activity. The CoMFA steric and electrostatic fields are presented as contour plots in Figure 5. To aid in visualization, compound **38** is displayed with magenta color in the map as reference structure. Green-colored regions indicate areas where increased steric bulk is associated with enhanced CatK inhibitory activity, and yellow regions suggest areas where increased steric bulk is unfavorable to activity. Blue-colored regions show areas where more positively charged groups are favorable to CatK inhibitory activity, while red regions represent areas where increased negatively charged groups are favorable to activity.

The important binding pocket or binding positions S1, S2, and S3 of CatK described by McGrath et al.<sup>8</sup> where corresponding groups from the aldehyde inhibitors can bind with are depicted in Figure 5. While S1 pocket that is formed by residues Gly23, Ser24, Gly64, and Gly65 is characterized as a wall (about 5 Å in length), which prefers linear hydrophobic substitutes in this position. The CoMFA contour maps showed that there are favorable regions for steric interaction (displayed as green color contour in Fig. 5a) near the position P1 of compound **38**, which explained the experimental results that the

*n*-Bu substitute on R4 (**11**) is better than other substitutes, such as Me (**2**), *i*-Pr (**4**), *n*-pentyl (**15**), and so on (Table 1). S2 pocket, which is composed by residues Tyr67, Met68, Ala134, and Ala163 is more likely a broad and shallow pocket (about 4 Å in depth). Similarly, a broad and shallow region of green contour is displayed near the P2 substructure of inhibitor **38** in R2 group, which suggested that more bulky substituents are favored in these positions. Exactly, the inhibitory activity of compounds **31** (substituted by Me group), **32** ((CH<sub>2</sub>)<sub>3</sub>), **33** (Et), and **34** ((CH<sub>2</sub>)<sub>4</sub>) is placed in the ordered as **31** < **32** < **33** < **34** which agrees with the order of substitutes on R2. Moreover, the *i*-Pr substitute (**40**) is more favored than the *n*-Pr substitute (**39**) because *i*-Pr is more broad than *n*-Pr. The S3 pocket consists of a groove that is formed by residues Gly66 and Tyr67, where yellow regions are surrounded, suggesting that bulk substitute groups at R1 are disfavored, which agrees with compound **44** being more active than **46**. However, considering the  $\pi$ – $\pi$  interaction between Tyr67 and aryl substitute at R1 position of ligand, the aryl substitute on R1 position is better than other substitutes. Moreover, the whole active sites are embedded in the unfavorable yellow region, suggesting that too bulky substitutes on R1, R2, R3, and R4 are disfavored, such as compounds **47**–**50**. So, it is important to pay more attention to the steric characters of active sites (S1, S2, and S3 pockets), suggesting there is a definite requirement of a substructure with an appropriate shape to exhibit high activity when we design novel aldehyde inhibitors of CatK.

The electrostatic contribution to the whole molecular field is 40.1%. Figure 5b displays the electrostatic contour map of the CoMFA model in combination with compound **38** and superposition to the active sites of CatK. An important feature of the CoMFA model is that the electrostatic contour map is dominated by regions favorable to positive charges. Such large blue



**Figure 5.** CoMFA contour maps in combination with inhibitor **38**: (a) the steric field distribution; (b) the electrostatic field distribution. The inhibitor is shown in cyan stick. Sterically favored areas in green; sterically disfavored areas in yellow. Positive potential favored areas in blue; positive potential unfavored areas in red.

polyhedra regions are mainly observed to be distributed along the backbone and P2 of compound **38**, indicating that compounds with low electron density are preferred in the active binding pocket. This is consistent with the fact that the inhibitor is stabilized by the hydrogen bond formed by the residues of Gln19 and Gly66, respectively. It is also explained that substitute of R2 with Et (**38**) group is more active than those of *i*-Pr group (**40**) and *i*-Bu group (**45**). In addition, there is another blue contour surrounding the conserved catalytic binding cavity of CatK which is composed by residues of Cys25 and Gln19. This is consistent with the fact that main active sites of CatK are oxy-anion hole. There is a red contour region near the Gly23, which facilitates the hydrogen bond formed between hydroxyl group of Ser24 and inhibitors. There is another small red contour region near His162, which favors the hydrogen bond between NH group of His162 and inhibitors.

### 3.4. Validation of the 3D-QSAR models

The nine randomly selected compounds (Table 3) were used as the test set to verify the constructed CoMFA models. The calculated results are listed in Table 3 and displayed in Figure 4 (red circle). The predicted  $\text{pIC}_{50}$  with the QSAR models are in good agreement with the experimental data with a statistically tolerable error range, with a correlation coefficient of  $r^2 = 0.708$  (Table 2). The test results indicated that the CoMFA model would be reliably used in new aldehyde inhibitor design for developing drug leads against osteoporosis.

## 4. Conclusion

In this work, using the alignment scheme generated from the docking study, a highly predictive CoMFA model was developed on aldehyde CatK inhibitors. The satisfactory model was obtained with leave-one-out (LOO) cross-validation  $q^2$  and conventional  $r^2$  values of 0.723 and 0.976, respectively. Both the binding conformations of 59 aldehyde analogues and their binding free energies were determined and predicted by molecular docking. The reliability of the model was verified by the compounds in the test set. The consistency between the CoMFA field distributions and the 3D topology of the protein structure showed the robustness of the 3D-QSAR model. Moreover, the 3D-QSAR results suggested that the *n*-Bu in R4 group is the favor group substitute at P1 and moderate groups in R2 group are required on P2 substitute in order to enhance the inhibitory activity of CatK. In addition, 3D-QSAR results also demonstrated that aldehyde is an important pharmacophore because of electrostatic effect. These results, together with the good correlations between the inhibitory activities and the binding free energies predicted by GOLD, demonstrated the power of combining docking/QSAR approach to explore the probable binding conformations of compounds at the active sites of the protein target, and further provided useful information in understanding the structural and chemical features of CatK in designing and finding new potential inhibitors.

## Acknowledgments

The present work was supported by the Foundation of Chinese Academy Sciences (West Light Program, KSCX1-09-03-1 and KSCX-SW-11) and the National Natural Science Foundation of China (30572258).

## Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.bmc.2005.11.061](https://doi.org/10.1016/j.bmc.2005.11.061).

## References and notes

- Robichaud, J.; Oballa, R.; Prasit, P.; Falgoutret, J.-P.; Percival, M. D.; Wesolowski, G.; Rodan, S. B.; Kimmel, D.; Johnson, C.; Bryant, C.; Venkatraman, S.; Setti, E.; Mendonca, R.; Palmer, J. T. *J. Med. Chem.* **2003**, *46*, 3709–3727.
- Alves, M. F. M.; Puzer, L.; Cotrin, S. S.; Juliano, M. A.; Juliano, L.; Bromme, D.; Carmona, A. K. *Biochem. J.* **2003**, *373*, 981–986.
- Zeng, G.-Z.; Tan, N.-H.; Jia, R.-R.; Pan, X.-L. *Acta Bot. Yunnanica* **2005**, *27*, 337–354.
- Marquis, R. W.; Ru, Y.; LoCastro, S. M.; Zeng, J.; Yamashita, D. S.; Oh, H.-J.; Erhard, K. F.; Davis, L. D.; Tomaszek, T. A.; Tew, D.; Salyers, K.; Proksch, J.; Ward, K.; Smith, B.; Levy, M.; Cummings, M. D.; Haltiwanger, R. C.; Trescher, G.; Wang, B.; Hemling, M. E.; Quinn, C. J.; Cheng, H.-Y.; Lin, F.; Smith, W. W.; Janson, C. A.; Zhao, B.; McQueney, M. S.; D'Alessio, K.; Lee, C.-P.; Marzulli, A.; Dodds, R. A.; Blake, S.; Hwang, S.-H.; James, I. E.; Gress, C. J.; Bradley, B. R.; Lark, M. W.; Gowen, M.; Veber, D. F. *J. Med. Chem.* **2001**, *44*, 1380–1395.
- Lecaille, F.; Kaleta, J.; Bromme, D. *Chem. Rev.* **2002**, *102*, 4459–4488.
- Zhang, Y.-M.; Tan, N.-H.; Huang, H.-Q.; Jia, R.-R.; Zeng, G.-Z.; Ji, C.-J. *Acta Bot. Yunnanica* **2005**, *27*, 107–110.
- Pan, X.-L.; Tan, N.-H.; Zeng, G.-Z.; Zhang, Y.-M.; Jia, R.-R. *Bioorg. Med. Chem.* **2005**, *13*, 5819–5825.
- McGrath, M. E.; Klaus, J. L.; Barnes, M. G.; Bromme, D. *Nat. Struct. Biol.* **1997**, *4*, 105–109.
- Zhao, B.; Janson, C. A.; Amegadzie, B. Y.; D'Alessio, K.; Griffin, C.; Hanning, C. R.; Jones, C.; Kurdyla, J.; McQueney, M.; Qiu, X.; Smith, W. W.; Abdel-Meguid, S. S. *Nat. Struct. Biol.* **1997**, *4*, 109–111.
- Thompson, S. K.; Halbert, S. M.; Bossard, M. J.; Tomaszek, T. A.; Levy, M. A.; Zhao, B.; Smith, W. W.; Abdel-Meguid, S. S.; Janson, C. A.; D'Alessio, K. J.; McQueney, M. S.; Amegadzie, B. Y.; Hanning, C. R.; DesJarlais, R. L.; Briand, J.; Sarkar, S. K.; Huddleston, M. J.; Ijames, C. F.; Carr, S. A.; Garnes, K. T.; Shu, A.; Heys, J. R.; Bradbeer, J.; Zembryki, D.; Lee-Ryckaczewski, L.; James, I. E.; Lark, M. W.; Drake, F. H.; Gowen, M.; Gleason, J. G.; Veber, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14249–14254.
- DesJarlais, R. L.; Yamashita, D. S.; Oh, H.-J.; Uzinskas, I. N.; Erhard, K. F.; Allen, A. C.; Haltiwanger, R. C.; Zhao, B.; Smith, W. W.; Abdel-Meguid, S. S.; D'Alessio, K.; Janson, C. A.; McQueney, M. S.; Tomaszek, T. A.; Levy, M. A.; Veber, D. F. *J. Am. Chem. Soc.* **1998**, *120*, 9114–9115.
- LaLonde, J. M.; Zhao, B.; Smith, W. W.; Janson, C. A.; DesJarlais, R. L.; Tomaszek, T. A.; Carr, T. J.; Thompson,

- S. K.; Oh, H.-J.; Yamashita, D. S.; Veber, D. F.; Abdel-Meguid, S. S. *J. Med. Chem.* **1998**, *41*, 4567–4576.
13. Barrett, D. G.; Catalano, J. G.; Deaton, D. N.; Hassell, A. M.; Long, S. T.; Miller, A. B.; Miller, L. R.; Shewchuk, L. M.; Wells-Knecht, K. J.; Willard, D. H., Jr.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4897–4902.
14. Deaton, D. N.; Hassell, A. M.; McFadyen, R. B.; Miller, A. B.; Miller, L. R.; Shewchuk, L. M.; Tavares, F. X.; Willard, D. H.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1815–1819.
15. Boros, E. E.; Deaton, D. N.; Hassell, A. M.; McFadyen, R. B.; Miller, A. B.; Miller, L. R.; Paulick, M. G.; Shewchuk, L. M.; Thompson, J. B.; Willard, D. H., Jr.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3425–3429.
16. Catalano, J. G.; Deaton, D. N.; Furfine, E. S.; Hassell, A. M.; McFadyen, R. B.; Miller, A. B.; Miller, L. R.; Shewchuk, L. M.; Willard, D. H., Jr.; Wright, L. L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 275–278.
17. Sybyl Version 6.9, St. Louis (MO), Tripos Associates, 2001.
18. Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J. Comput. Chem.* **1989**, *10*, 982–1012.
19. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.
20. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425–445.
21. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins* **1998**, *33*, 367–382.
22. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murrar, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609–623.
23. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
24. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 127–134.
25. Lagunin, A. A.; Dearden, J. C.; Filimonov, D. A.; Poroikov, V. V. *Mutat. Res.* **2005**, *586*, 138–146.