



Original article

3D QSAR studies on ketoamides of human cathepsin K inhibitors based on two different alignment methods

Xulin Pan^a, Ninghua Tan^{a,*}, Guangzhi Zeng^a, Huoqiang Huang^{a,b}, He Yan^{a,b}^a State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, No. 132, Lanhei Road, Kunming 650204, PR China^b Graduate School of the Chinese Academy of Sciences, Beijing 10049, PR China

ARTICLE INFO

Article history:

Received 31 May 2009

Received in revised form

28 October 2009

Accepted 3 November 2009

Available online 11 November 2009

Keywords:

Cathepsin K

CoMFA

CoMSIA

Ketoamides

ABSTRACT

Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed on 64 ketoamides as human cathepsin K (CatK) inhibitors, using ROCS ligand-based alignment and receptor-based alignment. Results generated from the ligand-based model were found to be superior to those obtained by the receptor-based model. CoMFA and CoMSIA field distributions are in good agreement with the structural characteristics of the binding groove of CatK, suggesting moderate substitutes at the P1, P2, P3 and P1' may favor the inhibitory activity of ketoamides. These results provide useful information in understanding the structural and chemical features of CatK in designing and finding novel potential CatK inhibitors as osteoporosis therapeutic agents.

© 2009 Elsevier Masson SAS. All rights reserved.

1. Introduction

Osteoporosis is a skeletal disorder characterized by enhanced bone resorption relative to bone formation resulting in decreased bone mass and increased susceptibility to fracture. Cathepsin K (CatK, EC 3.4.22.38), a 27 kDa cysteine protease of the papain superfamily, was first discovered in a rabbit osteoclast cDNA library [1]. The crucial role played by CatK in bone matrix resorption is further demonstrated by a rare human skeletal dysplasia, pycnodysostosis, which is the consequence of mutations in the CatK gene [2]. With demographic trends toward increases in the elderly population and concomitant increases in the number of patients with osteoporosis, there is an urgent need for new anti-osteoporotic drugs that are safe and effective [3]. Therefore, efforts aimed at discovering effective inhibitors of CatK have led to two CatK inhibitors, odanacatib (MK-0822) and balicatib (AAE581) which have been demonstrated pharmacodynamically to act on reductions in bone resorption markers and increases in bone mineral density in humans [4]. Moreover, several classes of CatK inhibitors have been reported including odanacatib [5], 2-cyano-pyrimidines [6], relacatib [7], pyrrolopyrimidine [8], dipeptidyl nitriles [9], cyclohexanecarboxamide [10], acyclic cyanamides [11] and azepanone-based inhibitors [12].

As part of an effort to develop novel cathepsin K inhibitors from natural products, researchers from our laboratory have recently reported the discovery of some natural compounds, such as amentoflavone and 8,8''-biskoenigine [13,14] appeared to have an inhibitory activity against CatK. Apparently, it is difficult to collect inhibitors useful to elucidate the interaction and mechanism of CatK with its natural inhibitors. Furthermore, quantitative structure–activity relationships and inhibitory mechanism between a series of synthetic compounds and CatK would be greatly helpful in discovering more natural inhibitors of CatK. Moreover, the three-dimensional structures of CatK and its complex with different inhibitors have been experimentally determined. Much experimental works have been carried out on this family of proteins, and a number of QSAR studies have been carried out as well. These crystal structures provide not only insights into the interaction mechanisms of CatK with the inhibitors, but also valuable clues for designing new inhibitors. Recently, we applied a docking method to develop a CoMFA model in a data set of 59 aldehyde derivatives that inhibit CatK [15]. Moreover, based on the structural information, Deaton et al. synthesized a series of ketoamide analogues and also measured their inhibitory activities against CatK [16–21]. In this work we compared two different approaches to compute an alignment for those highly flexible structures. The first program was ROCS (Rapid Overlay of Chemical Structures) [22] which perceives inhibitor similarities based on their three-dimensional molecular shapes. ROCS approximates their volumes with Gaussian functions instead of hard spheres, thereby resulting in analytic and

* Corresponding author. Tel./fax: +86 871 5223800.

E-mail address: nhtan@mail.kib.ac.cn (N. Tan).

differentiable mathematical equations that allow for fast and robust global optimization of volume overlap by varying their relative orientations. A similarity function then measures the “shape distance” between the pair of molecules at optimal overlap of volumes. The second procedure was docking performed by Gold3.1 [23,24] which is a hybrid of a genetic algorithm and an adaptive local search method, named the Lamarkian genetic algorithm. Both alignments were respectively adopted in comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) and applied to 64 potent ketoamide inhibitors of CatK [16,17,20,21] to gain insights into how steric, electrostatic, hydrophobic, and hydrogen bonding interactions influence their activity. In addition, we have included sets of biologically diverse test compounds to confirm the validity of the current 3D QSAR investigation. This analysis provides a platform for the prediction of novel ketoamide inhibitors and subsequently enables the interpretation of interaction modes between natural inhibitors and CatK.

2. Results and discussion

2.1. CoMFA and CoMSIA statistical results

During 3D QSAR analyses, we selected 51 compounds (unasterisked molecules in Table 1) as the training set for model construction and 13 compounds (asterisked molecules in Table 1) as the testing set for model validation. Results of the PLS analyses of the two alignment approaches are summarized in Table 2. The ligand-based model yielded $r_{cv}^2 = 0.663$ and $r_{ncv}^2 = 0.980$, $r_{cv}^2 = 0.710$ and $r_{ncv}^2 = 0.974$ values for the CoMFA and CoMSIA models, respectively. The receptor-based model gave $r_{cv}^2 = 0.640$ and $r_{ncv}^2 = 0.963$ for CoMFA, $r_{cv}^2 = 0.622$ and $r_{ncv}^2 = 0.918$ for CoMSIA after removed the outlier (compounds **38**, **50** and **51**). The inhibitory activities (pIC_{50}), the calculated activities of the two alignment approaches using the CoMFA and CoMSIA models, and the residual values for the training set and the test set are listed in Table 1. Graphic representation of observed vs calculated inhibitory activity is shown in Fig. 3. It can be seen that the best model was obtained from the ligand-based model, whereas the receptor-based model showed lower statistical values and predictive ability because many conformations produced by docking have different binding modes compared to FSP (X-ray conformation of ketoamide binding to CatK, PDB entry code 1TU6). Hence, our attention was mainly focused on the ligand-based alignment model since it produced more satisfactory statistical data. The predictive abilities of the CoMFA and CoMSIA models were determined from a set of 13 test compounds. The predicted r^2 values from the CoMFA and CoMSIA models were found to be 0.848 and 0.833, respectively (Table 2). The correlation between observed activities and the predicted values depicted in Fig. 3 and Table 1 demonstrated that the predicted activities by the constructed 3D QSAR model were in good agreement with the experimental data, suggesting that a reliable CoMFA and CoMSIA model was successfully constructed.

2.2. CoMFA contour maps

In CoMFA steric field, the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80% and 20% contributions, respectively. Similarly in CoMFA electrostatic field, the red (electronegative charge favorable) and blue (electropositive charge favorable) contours represent 80% and 20% contributions, respectively.

The steric field contribution accounts for 81.7%, while the electrostatic field contribution only accounts for 18.3% of variance, which suggests that the electrostatic interactions are not crucial in

explaining the variations in inhibition potency of these molecules, and the generated CoMFA models explain well the variations between molecules having differences in steric interactions. Fig. 4(a) shows the steric contour maps for the CoMFA model with the most active inhibitor **21** as a reference. Sterically favored regions (green polyhedra) around the P2 and P3 of compound **21** suggest that bulk groups are favorable at these points. This situation is the same for compounds **15**, **17**, **18**, **19**, **20**, **21**, **22**, **24**, **43**, **47** and **51**, all of which showed less than 1 nM binding affinity and have bulky groups mainly at the P2 and P3. For example, sterically favored tert-butyl group occupies the P2 of compounds **19**, **20**, **21**, **22** and **24** (pIC_{50} values are 10.14, 10.59, 10.60, 9.47 and 9.89, respectively). As for the P3, these compounds all have a sterically favored phenyl group, except for compound **51** with a comparatively smaller tert-butyl substituent. But for inhibitors **16** ($pIC_{50} = 6.34$) and **46** ($pIC_{50} = 6.04$), which orient their sterically favored tert-butyl group against the P2 because of configuration, they were less potent. In addition, compound **2** ($pIC_{50} = 7.60$) is more potent than compound **1** ($pIC_{50} = 6.20$) because cyclopentyl is larger than a methyl group. An important feature of the steric map is dominated by regions unfavorable to bulky substituents for whole ketoamide inhibitors. Such regions are mostly observed surrounding the P1' and P1, indicating that compounds with bulky substituents are unfavorable in the active binding pocket. This is consistent with the fact that compounds **33** ($pIC_{50} = 7.57$) and **38** ($pIC_{50} = 7.24$) show less potent activity due to the exposure of their large long phenyl group to the surface of these molecules. Detailed analysis indicates that the structural features of other inhibitors also concord with the contour maps. So, it is important to pay more attention to the steric characters of these active sites (S1', 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 ketoamide inhibitors of CatK.

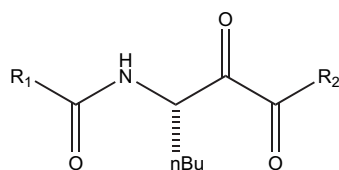
Electrostatic fields based on the PLS analysis of the CoMFA models are shown in Fig. 4(b) with compound **21** as a reference. A large blue contour surrounding the P2 indicates electropositive groups are beneficial to the activity. Isoxazole ring at the P2 of compound **41** ($pIC_{50} = 8.05$), bearing negative charges on the N and O atoms, decreases the activity, while dimethyl substituents of the pyrrole ring at the P2 of compound **28** ($pIC_{50} = 8.52$), increase the activity. Two small red polyhedrons near the P3 indicated electro-negative groups are favored at these positions. There are two different substituents at the R1 (those that include phenyl group and those that do not). The inhibitors which have a phenyl group are more potent than those inhibitors which have a non-phenyl group because the phenyl group is located at the P3 and favors the activity. For example, compounds **8–14** which have non-phenyl substituents at the R1 are poor inhibitors. There are two red contours near the P1' which are similar to the P3 inferring the aromatic substituents would favor the inhibitory activity. This observation is consistent with the experimental findings as compounds **8**, **9** and **10** are less active than compound **13** because the phenyl substituent is an electronegative group.

2.3. CoMSIA contour maps

The CoMSIA five fields based on the PLS analyses are presented as 3D contour plots in Fig. 5. The CoMSIA steric, electrostatic, hydrophobic, H-bond donor and acceptor fields explain variance of 20.2%, 13.8%, 38.3%, 13.5% and 14.3%, respectively. This demonstrated that hydrophobic interactions took an important role in describing the field properties of ketoamide inhibitors. Fig. 5(a) and (b) illustrated the CoMSIA contour maps of steric and electrostatic fields, which give the similar conclusions as the field distribution of the CoMFA model (Fig. 4).

Table 1

Structure, comparison of experimental (pIC_{50}) and predicted activities (PA) of 64 ketoamide derivatives by CoMFA and CoMSIA analyses using ligand-based and receptor-based methods. Compounds of the test set are labeled as asterisk.



Compd.	R ₁	R ₂	pIC_{50}	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
1			6.20	6.49	-0.29	6.56	-0.36	6.36	-0.16	6.36	-0.16
2			7.60	7.40	0.20	7.67	-0.07	7.33	0.27	7.26	0.34
3			6.82	6.61	0.21	6.84	-0.02	7.29	-0.47	7.52	-0.70
4			7.19	6.97	0.22	7.06	0.13	7.40	-0.21	7.34	-0.15
5			7.89	7.70	0.19	7.57	0.32	8.01	-0.12	7.66	0.23
6			7.62	7.52	0.10	7.76	-0.14	7.84	-0.22	8.02	-0.40
7			7.77	7.79	-0.02	8.05	-0.28	7.46	0.31	7.19	0.58
8			5.23	5.03	0.20	4.96	0.27	4.75	0.48	5.00	0.23

(continued on next page)

Table 1 (continued)

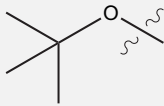
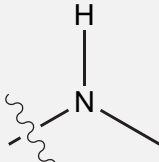
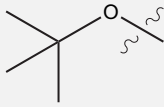
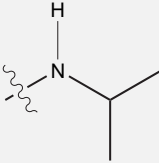
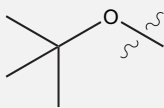
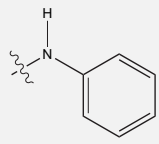
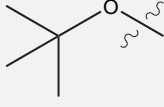
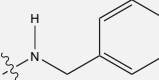
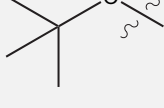
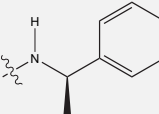
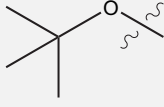
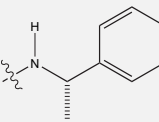
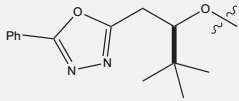
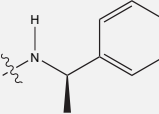
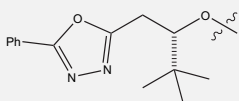
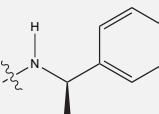
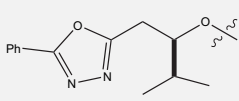
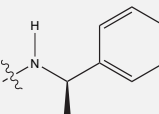
Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
9			4.68	4.90	−0.22	4.75	−0.07	4.65	0.03	5.16	−0.48
10			5.04	5.16	−0.12	5.40	−0.36	5.12	−0.08	5.60	−0.56
11			5.20	5.55	−0.35	5.36	−0.16	5.12	0.08	5.21	−0.01
12			5.54	5.65	−0.11	5.76	−0.22	5.78	−0.24	5.95	−0.41
13			5.92	6.33	−0.41	6.33	−0.41	5.88	0.04	5.74	0.18
14			5.38	5.29	0.09	5.18	0.20	5.89	−0.51	6.02	−0.64
15			9.62	9.46	0.16	9.41	0.21	9.21	0.41	9.45	0.17
16			6.34	6.22	0.12	5.99	0.35	6.62	−0.28	6.09	0.25
17			9.08	9.04	0.04	9.11	−0.03	8.41	0.67	8.95	0.13

Table 1 (continued)

Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
18			9.43	9.51	−0.08	9.52	−0.09	9.60	−0.17	10.02	−0.59
19			10.14	9.84	0.30	10.20	−0.06	9.97	0.17	9.70	0.44
20			10.59	10.91	−0.42	10.97	−0.48	10.67	−0.18	10.52	−0.03
21			10.60	10.69	−0.10	10.57	0.02	10.62	−0.03	10.37	0.22
22			9.47	9.55	−0.08	9.57	−0.10	9.29	0.18	9.58	−0.11
23			8.55	8.52	0.03	8.47	0.08	8.56	−0.01	8.39	0.16
24			9.89	10.08	−0.19	9.92	−0.03	10.03	−0.14	9.29	0.60
25			8.24	8.46	−0.22	8.27	−0.03	8.72	−0.48	8.86	−0.62
26			7.68	7.46	0.22	7.61	0.07	7.84	−0.16	7.86	−0.18

(continued on next page)

Table 1 (continued)

Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
27			8.08	8.03	0.05	7.83	0.25	7.69	0.39	7.28	0.80
28			8.52	8.44	0.08	8.13	0.39	8.21	0.31	8.19	0.33
29			7.54	7.36	0.18	7.28	0.26	7.74	−0.20	7.30	0.24
30			8.37	8.38	−0.01	8.48	−0.11	8.77	−0.40	8.95	−0.58
31			7.77	8.17	−0.40	7.93	−0.16	7.88	−0.11	8.15	−0.38
32			8.42	8.46	−0.04	8.61	−0.19	8.68	−0.26	8.62	−0.20
33			7.57	7.86	−0.29	7.72	−0.15	7.79	−0.22	8.03	−0.46
34			8.18	8.21	−0.03	8.27	−0.09	8.38	−0.20	8.72	−0.54
35			8.80	8.54	0.26	8.61	0.19	8.64	0.16	8.87	−0.07

Table 1 (continued)

Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
36			8.77	8.86	−0.09	8.81	−0.04	8.67	0.10	8.86	−0.09
37			8.16	7.92	0.24	8.02	0.14	8.14	0.02	8.23	−0.07
38			7.24	7.37	−0.13	7.19	0.05	outlier	–	–	–
39			8.30	8.58	−0.28	8.18	0.12	8.65	−0.35	8.96	−0.66
40			8.68	8.73	−0.05	8.98	−0.30	8.35	0.33	8.60	0.08
41			8.05	7.87	0.18	7.81	0.24	7.83	0.22	8.21	−0.16
42			7.85	7.89	−0.04	7.95	−0.10	7.84	0.01	7.67	0.18
43			9.38	9.29	0.09	9.12	0.26	9.32	0.06	9.01	0.37

(continued on next page)

Table 1 (continued)

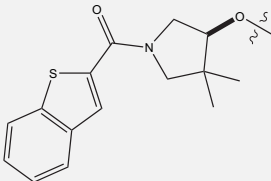
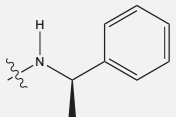
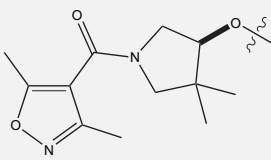
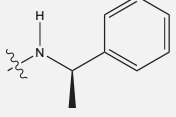
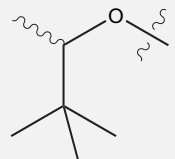
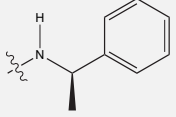
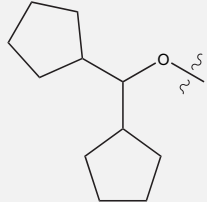
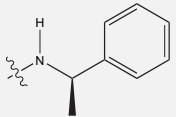
Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
44			7.47	7.33	0.14	7.31	0.16	7.05	0.42	6.79	0.68
45			8.96	8.98	−0.02	8.70	0.26	8.40	0.56	7.99	0.97
46			6.04	6.28	−0.24	6.71	−0.67	6.07	−0.03	6.10	−0.06
47			9.10	9.05	0.05	8.81	0.29	8.98	0.12	8.78	0.32
48			7.12	6.87	0.25	6.89	0.23	6.86	0.26	6.66	0.46
49			7.49	7.24	0.25	7.39	0.10	7.80	−0.31	7.10	0.39
50			8.60	8.66	−0.06	8.44	0.16	outlier	–	–	–
51			9.15	8.73	0.42	9.18	−0.03	outlier	–	–	–

Table 1 (continued)

Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
52*			8.07	7.89	0.18	7.88	0.19	8.02	0.05	7.37	0.70
53*			9.08	8.51	0.57	8.61	0.47	7.69	1.39	7.29	1.79
54*			4.74	5.11	−0.37	5.37	−0.63	3.93	0.81	4.03	0.71
55*			5.47	5.36	0.11	5.08	0.39	5.92	−0.45	6.06	−0.59
56*			9.39	8.79	0.60	9.49	−0.10	10.60	−1.21	10.19	−0.80
57*			10.54	10.12	0.42	11.46	−0.92	9.47	1.07	9.64	0.90
58*			9.19	8.87	0.32	9.79	−0.60	5.53	3.66	6.32	2.87
59*			7.19	8.05	−0.86	7.74	−0.55	7.38	−0.19	7.38	−0.19
60*			7.49	8.25	−0.76	8.16	−0.67	6.76	0.73	6.97	0.52

(continued on next page)

Table 1 (continued)

Compd.	R ₁	R ₂	pIC ₅₀	Ligand-based model				Receptor-based model			
				CoMFA		CoMSIA		CoMFA		CoMSIA	
				PA	resid.	PA	resid.	PA	resid.	PA	resid.
61*			8.55	8.55	0.00	8.70	−0.15	7.77	0.78	8.28	0.27
62*			7.82	8.12	−0.30	8.05	−0.23	8.57	−0.75	7.77	0.05
63*			7.70	7.39	0.31	7.73	−0.03	7.26	0.44	7.33	0.37
64*			8.20	8.52	−0.32	8.43	−0.23	7.95	0.25	7.90	0.30

The hydrophobic contour map of the CoMSIA model in the presence of compound **21** is displayed in Fig. 5(c). The white and yellow contour maps highlight areas where hydrophilic and hydrophobic properties are preferred. The hydrophobically unfavorable region overlaps the steric region. Indeed, many bulky substituents in ketoamide inhibitors are hydrophobic. Hydrophobically favored yellow regions are embedded in the hydrophobically disfavored white regions. There are two yellow polyhedrons near the P2 and P3, which are also consistent with the CoMFA steric contour map (Fig. 4(a)) in which the sterically more crowded substituents are necessary to enhance the biological activity at the same P2 and P3. The presence of a large white contour map surrounding the whole molecule suggested that its occupancy by hydrophilic groups would favor inhibitory activity. This observation is consistent with the facts that the active sites of CatK are shallow cavities and ketoamide inhibitors would be exposed out of the active sites.

The graphical interpretation of the hydrogen bond (H-bond) donor interaction in the CoMSIA model is represented in Fig. 5(d). It highlights areas beyond the molecules where putative functionality hydrogen bond acceptor in the enzyme can form H-bond with the molecule thereby influencing binding affinities. The cyan contours represent locations where a H-bond acceptor on the receptor will

improve the activity, that is H-bond donors in the ligand directing to these regions are favorable. For example, there are two small cyan areas: one near the S1' subsite which NH of the pyrazole at the R2 can form a H-bond with the peptide chain between Gly137 and Ser138, the other is near the NH group of the peptide chain of ketoamide forming a 1.817 Å strong hydrogen bond with the carbonyl group of Asn275 (Fig. 5(d)), which indicate the necessity of the hydrogen atom at this position for activity. Conversely, three big purple polyhedrons surround the α -keto position, indicating that compounds with hydrogen bond donor are unfavorable in the active binding site. These observation imply that α -keto and peptide chain have a major role in binding to the CatK active sites.

The hydrogen bond acceptor contour map of the CoMSIA model in the presence of compound **21** is depicted in Fig. 5(e). Magenta isopleths (85% contribution) encompass regions where hydrogen bond donors on the receptor site are expected, while proton acceptors in the ligand directing to red regions (15% contribution) decrease the binding affinities. A magenta contour near the S3 represents high activity of compounds having a hydrogen bond acceptor group at this position. Since the most active compounds (**19**, **20**, **21** and **22**) contain heterocyclic ring such as pyrazole and oxazol ring including N and O atoms. Moreover the least active compounds (**8–14**) have tert-butyl substituents at the R2. There are

Table 2

Summary of CoMFA and CoMSIA results based on ligand-based and receptor-based alignments.

PLS statistics	Ligand-based model				Receptor-based model			
	CoMFA		CoMSIA		CoMFA		CoMSIA	
	LOO ^a	LHO ^b	LOO	LHO	LOO	LHO	LOO	LHO
r_{cv}^2 ^c	0.663	0.647	0.710	0.711	0.640	0.646	0.622	0.610
SEE ^d	0.215		0.246		0.301		0.439	
r_{ncv}^2 ^e	0.980		0.974		0.963		0.918	
F (F-ratio)	440.718		278.887		215.850		120.487	
PLS components ^f	5	6	6	6	5	4	4	4
<i>Fields analysis</i>								
Steric	0.817		0.202		0.607		0.179	
Electrostatic	0.183		0.138		0.393		0.269	
Hydrophobic			0.383				0.308	
H-bond donor			0.135				0.099	
H-bond acceptor			0.143				0.145	
<i>Test set</i>								
r_{pred}^2 ^g	0.848		0.833		0.776		0.770	

^a Leave-one-out(LOO).^b Leave-half-out(LHO).^c Cross-validated correlation coefficient after the leave-one-out procedure.^d Standard error of estimate.^e Non-cross-validated correlation coefficient.^f Optimal number of principal components.^g Predicted correlation coefficient for the test set of compounds.

five large red polyhedrons found near the S1', S2 and S3 indicating that hydrogen bond acceptor groups are unfavorable. Moreover, the rigid main chain of ketoamides, which includes hydrogen donor group (NH) and hydrogen acceptor groups (carboxyl and α -keto), and covalent bond formed between carboxyl group of α -keto and the -SH group of Cys139 should limit the binding mode of ketoamides inhibitors. Consequently, a suitable substituent of the R2 may be the pyrazole ring because it can locate the N atom at the hydrogen acceptor area and NH group at hydrogen donor area.

3. Conclusion

CoMFA and CoMSIA studies have been carried out using the ligand-based and receptor-based alignments to build 3D QSAR models of 64 ketoamide inhibitors of CatK which identified the

regions of importance for steric, hydrophobic, and electronic interactions. The studies also demonstrated the comparative study among two different alignments in CoMFA and CoMSIA studies. Results generated from the ligand-based model were found to be superior (r_{pred}^2 values of 0.848 for CoMFA and 0.833 for CoMSIA) to those obtained by the receptor-based model (r_{pred}^2 values of 0.776 for CoMFA and 0.770 for CoMSIA). Therefore, it is reasonable that ligand-based conformers produced by the ROCS method would be used for alignments in those cases where the X-ray structure of protein is unknown. The predictive ability of the models was validated using a structurally and biologically diversified test set of 13 compounds that had not been included in a preliminary training set of 51 compounds. Despite the flexible molecular structure of the ketoamides, we were able to obtain statistically robust, predictive, and explanatory 3D QSAR models. The models have validated the pharmacophore hypothesis that the moderate bulky substituents at the P1, P2, P3 and P1' of ketoamide, may play a major role in its inhibitory activity and the bulky substituents of ketoamides were found to expose the active sites. Moreover, 3D QSAR results also demonstrated that the covalent bond of α -keto and peptide bond determine the ketoamide conformations in binding sites of CatK. Such insight provides means for designing new ligands, predicting the activities of new ketoamide candidates, and a rational basis for the discovery of other novel ketoamides that can act as CatK inhibitors via this mechanism.

4. Computational details

4.1. Data set

All computations were carried out on a Pentium IV personal computer with the RHEL 4.0 operating system using the molecular modeling package SYBYL 7.3 [25].

The data set used for 3D QSAR analysis contains 64 inhibitors of CatK, and was taken from the original work by Deaton et al. [16,17,20,21]. The IC₅₀ values, in nM, were converted into pIC₅₀(-logIC₅₀) values, which were used as dependent variables in 3D QSAR analysis. As a rule of thumb, the pIC₅₀ values of the training set should span approximately 3 log units. The pIC₅₀

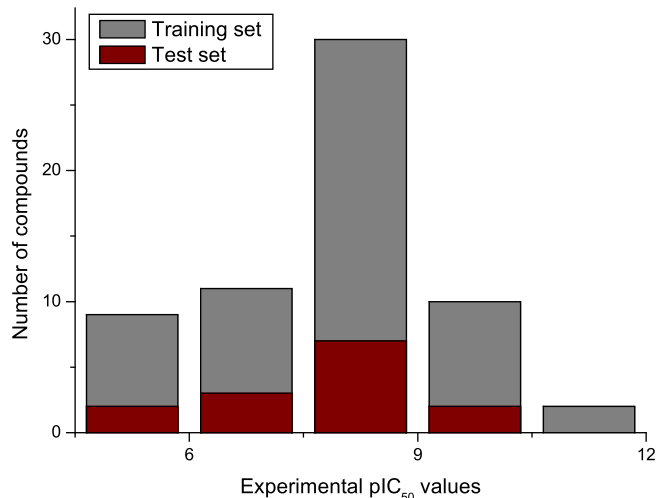


Fig. 1. Distribution of biological activities (pIC₅₀) for the training set and the test set versus numbers of compounds. The higher the column, the more compounds it contains (gray color, training set; wine color, test set). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

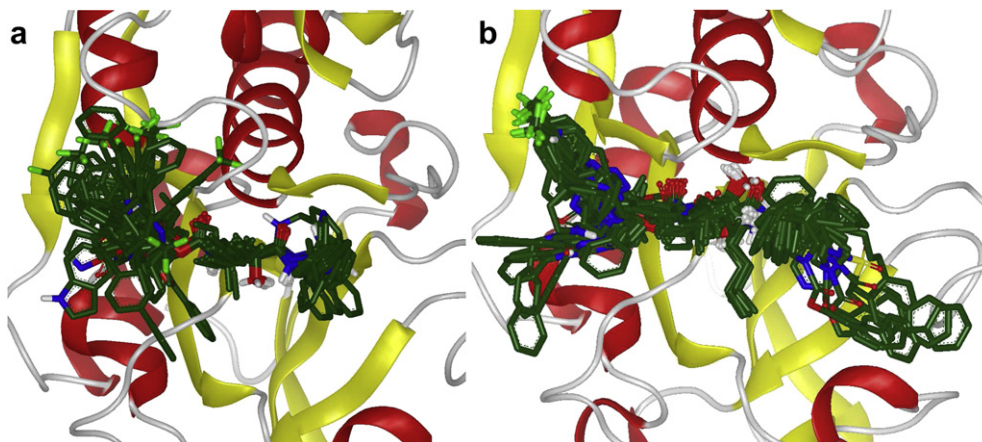


Fig. 2. Alignment of 64 ketoamides using two different alignment methods. (a) ligand-based model and (b) receptor-based model.

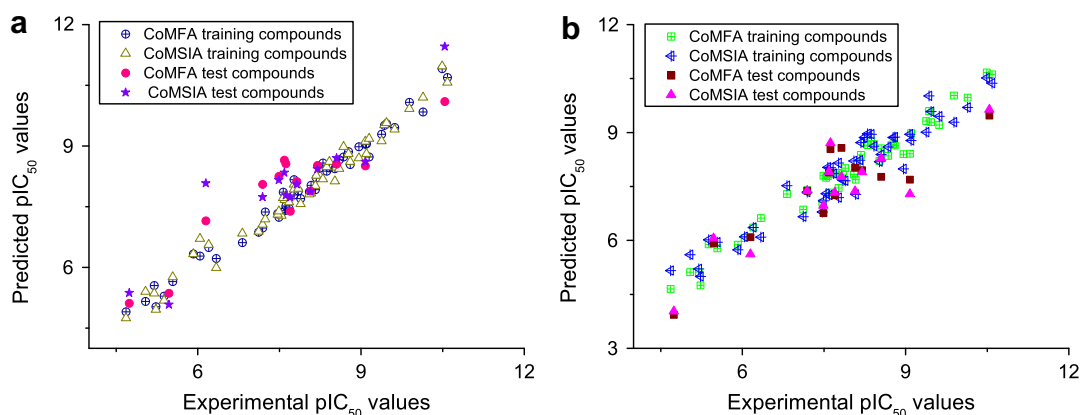


Fig. 3. Plots of the experimental versus predicted pIC_{50} values of 3D QSAR from both CoMFA and CoMSIA for the training and test compounds. (a) ligand-based model. (b) receptor-based model.

values of the training set described in this paper span 5.9 log units. Selection of training set and test set was done by considering the fact that the test set compounds represent structural diversity and a range of biological activities similar to those of the

training set. Accordingly, biological data distribution analysis of ketoamide derivatives has been carried out to confirm with the test set as a true representative of the training set (Fig. 1). The cluster method was employed to classify the compounds into 5 groups

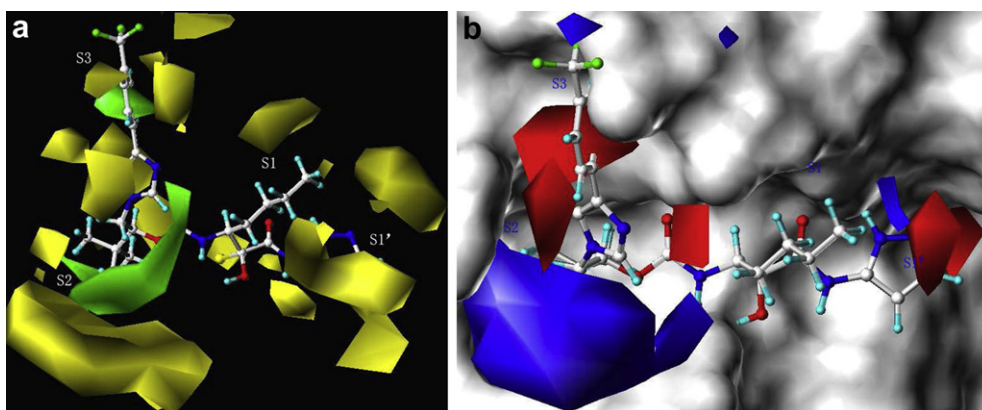


Fig. 4. CoMFA contour maps displayed in combination with compound **21**. (a) Steric field distribution; and (b) electrostatic field distribution. Sterically favored areas are in green; sterically unfavored areas are in yellow; positive potential favored areas are in blue; positive potential unfavored areas are in red. S1' (P1'), S1(P1), S2(P2) and S3(P3) represent the subsites of receptor (ligand). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

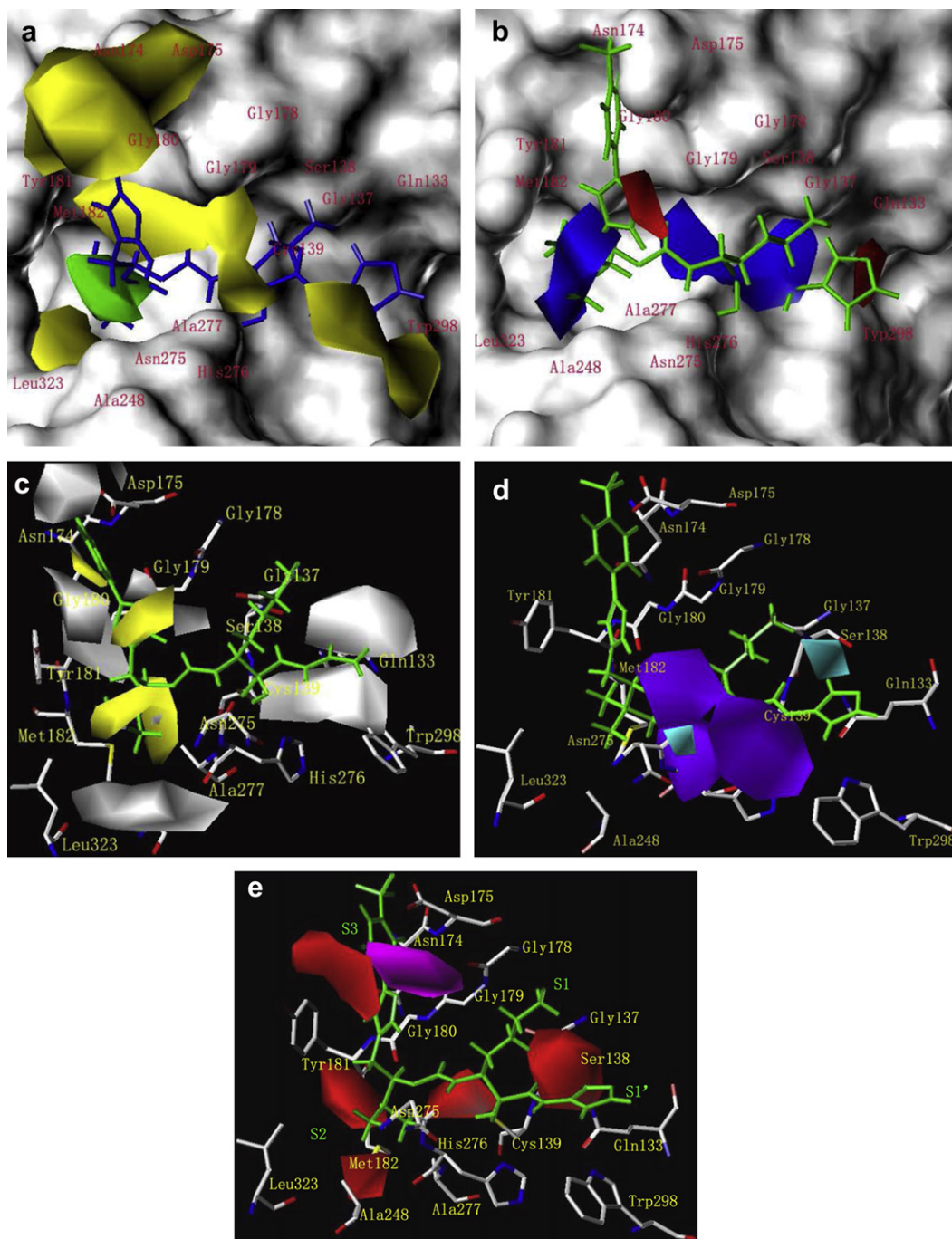


Fig. 5. CoMSIA contour maps displayed with compound **21**. (a) steric field distribution; (b) electrostatic field distribution; (c) hydrophobic field distribution; (d) hydrogen bond donor field distribution; (e) hydrogen bond acceptor field distribution. Sterically favored areas are in green (85% contribution); sterically unfavored areas are in yellow (15% contribution); positive potential favored areas are in blue; positive potential unfavored areas are in red; hydrophobic favored areas are in yellow (85% contribution); hydrophilic favored areas are in white (15% contribution); cyan and purple contours indicate favorable and unfavorable hydrogen bond donor group; hydrogen acceptor favored areas are in magenta (85% contribution) and hydrogen acceptor unfavored areas are in red (15% contribution). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

according to the one pIC_{50} value's interval, and the test set compounds were selected in proportion with the number of group compounds (the ratio's of the test set compounds and training set compounds is approximately equal to 0.20). Consequently, the data set was divided into the training set (51 compounds) and thirteen compounds were selected as the test set. The structures of these 64 compounds in the training and test sets are shown in Table 1.

4.2. Molecular models and alignment

The 1.75 Å resolution X-ray crystal structure of CatK complexing with ketoamide FSP (PDB entry code 1TU6, inhibitor **53** in Table 1) was retrieved from the Brookhaven Protein Data Bank (<http://www.rcsb.org/pdb>). The structures of these ketoamide inhibitors were constructed based on the structure of FSP and energetically minimized using Tripos force field with Gasteiger–Hückel charges

and the steepest descent followed by the conjugate gradient method with a convergence criterion of 0.001 kcal/mol Å.

4.2.1. Ligand-based alignment

The program OMEGA v3.1 [26], which uses a depth-first searching algorithm for generating conformational ensembles was employed to convert all compounds into 3D multiconformer structures. OMEGA dissects the molecules into fragments, reassembles and regenerates many possible combinations, and then submits each conformer to a simplified energy evaluation. Then, all conformers below an energy threshold are compared and those within a certain RMS distance are clustered into one single representation. The ligand in its receptor bound conformation from the X-ray crystal structure was used as the target structure for ROCS (Rapid Overlay of Chemical Structures) [22]. ROCS, a ligand-centric shape-based program was used to identify conformations that overlap in shape and chemistry with respect to the reference ligand. The overlays can be performed very quickly based on a description of the molecules as atom-centered Gaussian functions. ROCS maximizes the rigid overlap of these Gaussian functions and thereby maximizes the shared volume between a query molecule and a single conformation of a database molecule. The best 3D overlay of each molecule in the data set, obtained by ROCS, was the high-score ligands and was chosen to be input for CoMFA and CoMSIA calculations (Fig. 2(a)).

4.2.2. Receptor-based alignment

GOLD3.1 [23,24], which uses a powerful genetic algorithm (GA) method for conformational search, docking and being widely regarded as one of the best docking programs, was employed to generate an ensemble of docked conformations. The original ligand as well as the water molecules were removed from the coordinated set. The default GOLD parameters were used as well as the covalent bond between Cys139 and α -keto of ketoamides was considered. Firstly, cocrystallized inhibitors of FSP have been docked into CatK, which yielded slightly lower root-mean square deviation values (RMSD = 1.192 Å). Subsequently, we analyzed whether the docking scores can be correlated to the biological data of the inhibitors. Ten conformations were obtained for each ligand and the top scored conformation was used directly for CoMFA and CoMSIA analysis (Fig. 2(b)).

4.3. CoMFA and CoMSIA models

CoMFA calculations were carried out by applying the default settings. The standard CoMFA fields performing the Lennard-Jones potential for the electrostatic were used. A cut-off value for the fields was set at 30.0 kcal/mol. CoMFA descriptors were calculated using an sp³ carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric field energies and electrostatic (Coulombic potential) field with a dielectric constant at each lattice point.

The aligned molecules were placed in a 3D lattice with regular grid points separated by 2 Å similar to that of CoMFA studies. The five physicochemical properties for CoMSIA (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor) were evaluated using a common probe atom with 1 Å radius, +1.0 charge, and hydrophobic and hydrogen-bond property values of +1. The attenuation factor α , which determines the steepness of the Gaussian function, was assigned a default value of 0.3 [27].

4.4. PLS analysis and validation

In partial least-square (PLS) regression analyses, the CoMFA and CoMSIA descriptors were used as independent variables and the

pIC₅₀ values were used as dependent variables to derive 3D QSAR models. The predictive value of the models was evaluated by leave-one-out (LOO) and leave-half-out (LHO) cross-validation. The cross-validated coefficient, r_{cv}^2 , was calculated using Eq. (1):

$$r_{cv}^2 = 1 - \frac{\sum (Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum (Y_{\text{observed}} - Y_{\text{mean}})^2} \quad (1)$$

where $Y_{\text{predicted}}$, Y_{observed} , and Y_{mean} are the predicted, observed, and mean values of the target property (pIC₅₀), respectively. $\sum (Y_{\text{observed}} - Y_{\text{mean}})^2$ is the predictive residual sum of squares (PRESS).

The optimal number of components (ONC) obtained from the cross-validated PLS analysis were used to derive the final QSAR model using the compounds in the training set (without cross-validation). The non-cross-validated correlation coefficient (r_{ncv}^2) served as a measure of the quality of the model. To test the utility of the model as a predictive tool, an external set of compounds (the test set) with known activities but not used in model generation (the training set) was predicted. The predictive r^2 , calculated by using Eq. (2), was based on molecules in the test set and was used to evaluate the predictive power of the CoMFA and CoMSIA models.

$$\text{predictive } r^2 = 1 - (\text{"press"}/\text{SD}) \quad (2)$$

where SD is the sum of the squared deviations between the actual activities of the compounds in the test set and the mean activity of the compounds in the training set, "press" is the sum of the squared deviations between predicted and actual activities for every compound in the test set.

Acknowledgement

The present work was supported by the National Natural Science Foundation of China (30725048), National Basic Research Program of China (2009 CB522300) and the Fund of State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy Sciences. The authors would like to thank OpenEye Scientific Software for kindly providing us free academic use of their software packages. We thank Dr. Abiodun H. Adebayo for proof-reading the manuscript.

References

- [1] E. Altmann, J. Green, M. Tintelnot-Blomley, Arylaminoethyl amides as inhibitors of the cysteine protease cathepsin K—investigating p1' substituents. *Bioorg. Med. Chem. Lett.* 13 (2003) 1997–2001.
- [2] B.D. Gelb, G.P. Shi, H.A. Chapman, R.J. Desnick, Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. *Science* 273 (1996) 1236–1238.
- [3] M.F.M. Alves, L. Puzer, S.S. Cotrin, M.A. Juliano, L. Juliano, D. Brömme, A.K. Carmona, S3 to S3' subsite specificity of recombinant human cathepsin K and development of selective internally quenched fluorescent substrates. *Biochem. J.* 373 (Pt 3) (2003) 981–986.
- [4] S.A. Stoch, J.A. Wagner, Cathepsin K inhibitors: a novel target for osteoporosis therapy. *Clin. Pharmacol. Ther.* 83 (2008) 172–176.
- [5] J.Y. Gauthier, N. Chauret, W. Cromlish, S. Desmarais, L.T. Duong, J.-P. Falgoutret, D.B. Kimmel, S. Lamontagne, S. Lger, T. LeRiche, C.S. Li, F. Mass, D.J. McKay, D.A. Nicoll-Griffith, R.M. Oballa, J.T. Palmer, M.D. Percival, D. Riendeau, J. Robichaud, G.A. Rodan, S.B. Rodan, C. Seto, M. Thrien, V.-L. Truong, M.C. Venuti, G. Wesolowski, R.N. Young, R. Zamboni, W.C. Black, The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorg. Med. Chem. Lett.* 18 (2008) 923–928.
- [6] E. Altmann, R. Aichholz, C. Betschart, T. Buhl, J. Green, O. Irie, N. Teno, R. Lattmann, M. Tintelnot-Blomley, M. Missbach, 2-Cyano-pyrimidines: a new chemotype for inhibitors of the cysteine protease cathepsin K. *J. Med. Chem.* 50 (2007) 591–594.
- [7] S. Kumar, L. Dare, J.A. Vasko-Moser, I.E. James, S.M. Blake, D.J. Rickard, S.-M. Hwang, T. Tomaszek, D.S. Yamashita, R.W. Marquis, H. Oh, J.U. Jeong, D.F. Veber, M. Gowen, M.W. Lark, G. Stroup, A highly potent inhibitor of cathepsin K (relacatib) reduces biomarkers of bone resorption both in vitro

- and in an acute model of elevated bone turnover in vivo in monkeys. *Bone* 40 (2007) 122–131.
- [8] N. Teno, T. Miyake, T. Ehara, O. Irie, J. Sakaki, O. Ohmori, H. Gunji, N. Matsuura, K. Masuya, Y. Hitomi, K. Nonomura, M. Horiuchi, K. Gohda, A. Iwasaki, I. Umemura, S. Tada, M. Kometani, G. Iwasaki, S.W. Cowan-Jacob, M. Missbach, R. Lattmann, C. Betschart, Novel scaffold for cathepsin K inhibitors. *Bioorg. Med. Chem. Lett.* 17 (2007) 6096–6100.
 - [9] E. Altmann, R. Aichholz, C. Betschart, T. Buhl, J. Green, R. Lattmann, M. Missbach, Dipeptide nitrile inhibitors of cathepsin K. *Bioorg. Med. Chem. Lett.* 16 (2006) 2549–2554.
 - [10] S.N. Crane, W.C. Black, J.T. Palmer, D.E. Davis, E. Setti, J. Robichaud, J. Paquet, R.M. Oballa, C.I. Bayly, D.J. McKay, J.R. Somoza, N. Chauret, C. Seto, J. Scheigetz, G. Wesolowski, F. Mass, S. Desmarais, M. Ouellet, Beta-substituted cyclohexanecarboxamide: a nonpeptidic framework for the design of potent inhibitors of cathepsin K. *J. Med. Chem.* 49 (2006) 1066–1079.
 - [11] D.G. Barrett, D.N. Deaton, A.M. Hassell, R.B. McFadyen, A.B. Miller, L.R. Miller, J.A. Payne, L.M. Shewchuk, D.H. Willard, L.L. Wright, Acyclic cyanamide-based inhibitors of cathepsin K. *Bioorg. Med. Chem. Lett.* 15 (2005) 3039–3043.
 - [12] R.W. Marquis, Y. Ru, S.M. LoCastro, J. Zeng, D.S. Yamashita, H.J. Oh, K.F. Erhard, L.D. Davis, T.A. Tomaszek, D. Tew, K. Salyers, J. Proksch, K. Ward, B. Smith, M. Levy, M.D. Cummings, R.C. Haltiwanger, G. Trescher, B. Wang, M.E. Hemling, C.J. Quinn, H.Y. Cheng, F. Lin, W.W. Smith, C.A. Janson, B. Zhao, M.S. McQueney, K. D'Alessio, C.P. Lee, A. Marzulli, R.A. Dodds, S. Blake, S.M. Hwang, I.E. James, C.J. Gress, B.R. Bradley, M.W. Lark, M. Gowen, D.F. Veber, Azepanone-based inhibitors of human and rat cathepsin K. *J. Med. Chem.* 44 (2001) 1380–1395.
 - [13] G.-Z. Zeng, X.-L. Pan, N.-H. Tan, J. Xiong, Y.-M. Zhang, Natural biflavones as novel inhibitors of cathepsin B and K. *Eur. J. Med. Chem.* 41 (2006) 1247–1252.
 - [14] G.-Z. Zeng, N.-H. Tan, X.-J. Hao, Q.-Z. Mu, R.-T. Li, Natural inhibitors targeting osteoclast-mediated bone resorption. *Bioorg. Med. Chem. Lett.* 16 (2006) 6178–6180.
 - [15] X.-H. Pan, N.-H. Tan, G.-Z. Zeng, H.-J. Han, H.-Q. Huang, 3D-QSAR and docking studies of aldehyde inhibitors of human cathepsin K. *Bioorg. Med. Chem.* 14 (2006) 2771–2778.
 - [16] D.G. Barrett, V.M. Boncek, J.G. Catalano, D.N. Deaton, A.M. Hassell, C.H. Jurgensen, S.T. Long, R.B. McFadyen, A.B. Miller, L.R. Miller, J.A. Payne, J.A. Ray, V. Samano, L.M. Shewchuk, F.X. Tavares, K.J. Wells-Knecht, D.H. Willard, L.L. Wright, H.-Q.Q. Zhou, P2–p3 conformationally constrained ketoamide-based inhibitors of cathepsin K. *Bioorg. Med. Chem. Lett.* 15 (2005) 3540–3546.
 - [17] D.G. Barrett, J.G. Catalano, D.N. Deaton, A.M. Hassell, S.T. Long, A.B. Miller, L.R. Miller, L.M. Shewchuk, K.J. Wells-Knecht, D.H. Willard, L.L. Wright, Potent and selective p2–p3 ketoamide inhibitors of cathepsin K with good pharmacokinetic properties via favorable p1', p1, and/or p3 substitutions. *Bioorg. Med. Chem. Lett.* 14 (2004) 4897–4902.
 - [18] D.G. Barrett, J.G. Catalano, D.N. Deaton, S.T. Long, R.B. McFadyen, A.B. Miller, L.R. Miller, K.J. Wells-Knecht, L.L. Wright, A structural screening approach to ketoamide-based inhibitors of cathepsin K. *Bioorg. Med. Chem. Lett.* 15 (2005) 2209–2213.
 - [19] D.G. Barrett, J.G. Catalano, D.N. Deaton, S.T. Long, L.R. Miller, F.X. Tavares, K.J. Wells-Knecht, L.L. Wright, H.-Q.Q. Zhou, Orally bioavailable small molecule ketoamide-based inhibitors of cathepsin K. *Bioorg. Med. Chem. Lett.* 14 (2004) 2543–2546.
 - [20] J.G. Catalano, D.N. Deaton, S.T. Long, R.B. McFadyen, L.R. Miller, J.A. Payne, K.J. Wells-Knecht, L.L. Wright, Design of small molecule ketoamide-based inhibitors of cathepsin K. *Bioorg. Med. Chem. Lett.* 14 (2004) 719–722.
 - [21] D.N. Deaton, A.M. Hassell, R.B. McFadyen, A.B. Miller, L.R. Miller, L.M. Shewchuk, F.X. Tavares, D.H. Willard, L.L. Wright, Novel and potent cyclic cyanamide-based cathepsin K inhibitors. *Bioorg. Med. Chem. Lett.* 15 (2005) 1815–1819.
 - [22] T.S. Rush, J.A. Grant, L. Mosyak, A. Nicholls, A shape-based 3-D scaffold hopping method and its application to a bacterial protein–protein interaction. *J. Med. Chem.* 48 (2005) 1489–1495.
 - [23] G. Jones, P. Willett, R.C. Glen, A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aided Mol. Des.* 9 (2005) 532–549.
 - [24] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267 (1997) 727–748.
 - [25] SYBYL Molecular Modeling System, Version 7.3, Tripos Inc., St. Louis, MO, 63144–2913.
 - [26] J. Bostrom, J.R. Greenwood, J. Gottfries, Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graph. Model.* 21 (2003) 449–462.
 - [27] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 37 (1994) 4130–4146.