Med Chem Res (2009) 18:683–701
DOI 10.1007/s00044-008-9160-x

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

3D-QSAR study of pyrrolidine derivatives as matrix metalloproteinase-2 inhibitors

Huawei Zhu · Hao Fang · Xianchao Cheng · Qiang Wang · Lei Zhang · Jinhong Feng · Wenfang Xu

Received: 5 October 2008 / Accepted: 10 December 2008 / Published online: 10 January 2009

Abstract In order to develop potent inhibitors of matrix metalloproteinase-2(MMP-2) as anticancer agents, a three-dimensional quantitative structure–activity relationship (3D-QSAR) model was established by using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) methods. This study correlates the MMP-2 inhibitory activities of 67 pyrrolidine derivatives to steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor fields. After using two different molecular alignments, both CoMFA and CoMSIA models resulted in good statistical predictions, a case in point being their high q^2 values of between 0.757 and 0.843. The CoMFA and CoMSIA models established herein will be helpful in understanding the structure–activity relationship of pyrrolidine derivatives as well as in the design of novel derivatives with enhanced MMP-2 inhibitory activity.

Keywords Matrix metalloproteinase · Inhibitors · 3D-QSAR · CoMFA · CoMSIA · Molecular docking

Introduction

Matrix metalloproteinases (MMPs) are a family of zinc-dependent endopeptidases including more than 20 subtypes. The main function of MMPs is to degrade and re-establish the structural proteins in the extracellular matrix. MMP-2 is a member of the MMP family that is considered a tumor therapeutic target because of its key function in the invasion and metastasis process of the tumor cell (Murphy *et al*., 1991; Polette *et al*., 2004; Björklund and Koivunen, 2005). The MMP-2 crystal

H. Zhu · H. Fang · X. Cheng · Q. Wang · L. Zhang · J. Feng · W. Xu (✉)
Institute of Medicinal Chemistry, School of Pharmaceutical Sciences, Shandong University, Ji'nan, Shandong 250012, China
e-mail: xuwenf@sdu.edu.cn; wfxu@yahoo.cn

structure indicates that the binding site consists of the zinc ion and two hydrophobic domains called the S1′ and S2′ pockets (Nagase *et al*., 2006; Verma and Hansch, 2007). Therefore, MMP-2 inhibitors were designed to introduce corresponding fragments so as to provide interaction with the two pockets and the zinc ion (Kontogiorgis *et al*., 2005; Fisher and Mobashery, 2006). Our group has reported a series of pyrrolidine derivatives as MMP-2 inhibitors and some of these compounds exhibited highly inhibitory activity against MMP-2 with 50% inhibition concentration (IC_{50}) values in the nanomolar range (Table 1) (Cheng *et al*., 2008a, b, c).

In our ongoing project, further structural modification will be carried out as suggested by the structural–activity relationships of pyrrolidine derivatives. Therefore, three-dimensional quantitative structure–activity relationship (3D-QSAR) of these series of pyrrolidine derivatives will be of benefit for us to develop new potent MMP-2 inhibitors. Herein, 3D-QSAR studies for 67 pyrrolidine derivatives using comparative molecular field analysis (CoMFA) (Cramer *et al*., 1988a) and comparative molecular similarity indices analysis (CoMSIA) (Klebe *et al*., 1994) methods are described.

Materials and methods

Molecular construction, geometry optimization, and molecular dynamics simulated annealing (MDSA) were carried out using the Tripos Sybyl7.3 (Tripos, St. Louis, MO, USA) package on a Dell Precision 360 workstation. 3D-QSAR model generation, molecular docking study, and MOLCAD (Exner *et al*., 1998) surface construction were carried out using the Tripos Sybyl7.0 (Tripos, St. Louis, MO, USA) package on a Dell Precision 390 workstation. The parameters involved in the study were set to default values except those specifically mentioned.

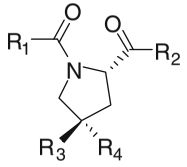
Data sets and biological activity

The compounds in the training set and test set for the QSAR analyses were 67 pyrrolidine derivatives reported by Cheng *et al*. (2008a, b, c) whose IC_{50} values for MMP-2 span approximate five orders of magnitude. Fifty compounds were randomly selected as the training set and the remaining 17 compounds were used as the test set. The number of samples in the test set was approximately 34% of the training set. The biological data obtained as IC_{50} (M) were converted to pIC_{50} ($-\log IC_{50}$) values and used as dependent variables in the 3D-QSAR analyses (Table 2).

Template selection and molecular alignment

Considering the important role of structure-based information in achieving reliable 3D-QSAR models, a crystal structure including both ligand and protein would reflect the most accurate binding environment for complex analysis. However, so far there is no available crystal structure of any pyrrolidine derivatives complex with

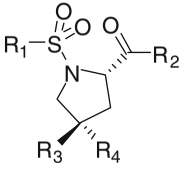
Table 1 Molecular structures and their MMP-2 inhibition data used in the 3D-QSAR model



No.	R_1	R_2	R_3	R_4	IC_{50} (μM)
A3a	CH_3	OCH_3	OH	H	2.0 ± 0.2
A3b	$CH{=}CHC_6H_3(OCH_3)_2$-3′,4′(E)	OCH_3	OH	H	2.9 ± 0.2
A4a	CH_3	OCH_3	$OSO_2C_6H_4CH_3$-p	H	0.3 ± 0.1
A4b	$CH{=}CHC_6H_3(OCH_3)_2$-3′,4′(E)	OCH_3	OSO_2CH_3	H	1.1 ± 0.04
A5a	CH_3	OCH_3	H	N_3	2.6 ± 0.3
A5b	$CH{=}CHC_6H_3(OCH_3)_2$-3′,4′(E)	OCH_3	H	N_3	3.4 ± 0.4
A6a	CH_3	OCH_3	H	NH_2	1.5 ± 0.1
A6b	$CH{=}CHC_6H_3(OCH_3)_2$-3′,4′(E)	OCH_3	H	NH_2	ND
A6c	$CH_2CH_2C_6H_3(OCH_3)_2$-3′,4′	OCH_3	H	NH_2	2.1 ± 0.3
A7a	CH_3	OCH_3	H	$NHCOC_6H_5$	2.7 ± 0.3
A7b	CH_3	OCH_3	H	$NHCOC_6H_4Cl$-p	4.2 ± 0.4
A7c	CH_3	OCH_3	H	$NHCOC_6H_4Br$-p	6.1 ± 1.0
A7d	CH_3	OCH_3	H	$NHCOC_6H_4$ NO_2-p	5.2 ± 0.6
A7e	CH_3	OCH_3	H	$NHCOC_6H_2(OCH_3)_3$-3′,4′,5′	0.1 ± 0.01
A7f	CH_3	OCH_3	H	$NHCOC_5H_4$ N-3′	1.0 ± 0.1
A7g	CH_3	OCH_3	H	$NHCOC_5H_4$ N-4′	2.2 ± 0.2
A7h	CH_3	OCH_3	H	$NHCOC_4H_3N_2$-2′,5′	2.1 ± 0.3
A7i	CH_3	OCH_3	H	$NHCOCH{=}CHC_6H_5$(E)	0.1 ± 0.01

Table 1 continued

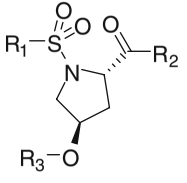
No.	R_1	R_2	R_3	R_4	IC_{50} (μM)
A7j	CH_3	OCH_3	H	NHCOCH=$CHC_6H_3(OCH_3)_2$-3′,4′(E)	0.8 ± 0.1
A7k	CH=$CHC_6H_3(OCH_3)_2$-3′,4′(E)	OCH_3	H	$NHCOC_5H_4$ N-3′	3.0 ± 0.04
A7l	$CH_2CH_2C_6H_3(OCH_3)_2$-3′,4′	OCH_3	H	$NHCOC_5H_4$ N-3′	2.2 ± 0.2
A7m	$CH_2CH_2C_6H_3(OCH_3)_2$-3′,4′	OCH_3	H	NHCOCH=CHC_6H_5(E)	1.6 ± 0.3
A8a	CH_3	NHOH	H	$NHCOC_6H_2(OCH_3)_3$-3′,4′,5′	0.001 ± 0.0001
A8b	CH_3	NHOH	H	$NHCOC_5H_4$ N-3′	0.005 ± 0.0004
A8c	CH_3	NHOH	H	NHCOCH=CHC_6H_5(E)	0.004 ± 0.0005
A8d	$CH_2CH_2C_6H_3(OCH_3)_2$-3′,4′	NHOH	H	$NHCOC_5H_4$ N-3′	0.012 ± 0.002
A8e	$CH_2CH_2C_6H_3(OCH_3)_2$-3′,4′	NHOH	H	NHCOCH=CHC_6H_5(E)	0.011 ± 0.001
A9	CH_3	OH	H	$NHCOC_6H_2(OCH_3)_3$-3′,4′,5′	0.016 ± 0.002
A10	CH_3	$NHCH_2COOCH_3$	H	$NHCOC_6H_2(OCH_3)_3$-3′,4′,5′	0.109 ± 0.012



No.	R_1	R_2	R_3	R_4	IC_{50} (μM)
B3a	p-$CH_3C_6H_4$	OCH_3	OH	H	0.1 ± 0.02
B3b	C_6H_5	OCH_3	OH	H	0.8 ± 0.1
B3c	CH_3	OCH_3	OH	H	7.1 ± 0.6
B4a	p-$CH_3C_6H_4$	OCH_3	C=O		0.4 ± 0.04

Table 1 continued

No.	R_1	R_2	R_3	R_4	IC_{50} (μM)
B4b	C_6H_5	OCH_3	C=O		2.0 ± 0.1
B4c	CH_3	OCH_3	C=O		9.5 ± 1.2
B5a	p-$CH_3C_6H_4$	OCH_3	O–$(CH_2)_2$–O		0.1 ± 0.01
B5b	p-$CH_3C_6H_4$	OCH_3	O–$(CH_2)_3$–O		0.2 ± 0.03
B5c	p-$CH_3C_6H_4$	OCH_3	O–$(CH_2)_4$–O		0.3 ± 0.04
B5d	C_6H_5	OCH_3	O–$(CH_2)_2$–O		0.2 ± 0.04
B5e	C_6H_5	OCH_3	O–$(CH_2)_3$–O		1.1 ± 0.1
B5f	C_6H_5	OCH_3	O–$(CH_2)_4$–O		1.5 ± 0.1
B6a	p-$CH_3C_6H_4$	NHOH	O–$(CH_2)_2$–O		0.003 ± 0.0004
B6b	p-$CH_3C_6H_4$	NHOH	O–$(CH_2)_3$–O		0.006 ± 0.0005
B6c	p-$CH_3C_6H_4$	NHOH	O–$(CH_2)_4$–O		0.008 ± 0.001
B6d	C_6H_5	NHOH	O–$(CH_2)_2$–O		0.008 ± 0.0008
B6e	C_6H_5	NHOH	O–$(CH_2)_3$–O		0.011 ± 0.001
B6f	C_6H_5	NHOH	O–$(CH_2)_4$–O		0.012 ± 0.001



No.	R_1	R_2	R_3	IC_{50} (μM)
C4a	p-$CH_3C_6H_4$	OCH_3	C_6H_5CO	3.2 ± 0.3

Table 1 continued

No.	R_1	R_2	R_3	IC_{50} (μM)
C4b	p-$CH_3C_6H_4$	OCH_3	p-ClC_6H_4CO	2.2 ± 0.2
C4c	p-$CH_3C_6H_4$	OCH_3	p-$CH_3C_6H_4SO_2$	0.007 ± 0.001
C4d	p-$CH_3C_6H_4$	OCH_3	$C_6H_5SO_2$	0.05 ± 0.006
C4e	p-$CH_3C_6H_4$	OCH_3	CH_3SO_2	0.2 ± 0.01
C4f	p-$CH_3C_6H_4$	OCH_3	(E)C_6H_5CH=CHCO	0.04 ± 0.008
C4g	C_6H_5	OCH_3	C_6H_5CO	3.5 ± 0.4
C4h	C_6H_5	OCH_3	p-ClC_6H_4CO	2.7 ± 0.2
C4i	C_6H_5	OCH_3	p-$CH_3C_6H_4SO_2$	0.04 ± 0.003
C4j	C_6H_5	OCH_3	$C_6H_5SO_2$	0.01 ± 0.001
C4k	C_6H_5	OCH_3	CH_3SO_2	1.3 ± 0.1
C4l	C_6H_5	OCH_3	(E)C_6H_5CH=CHCO	0.09 ± 0.01
C4m	CH_3	OCH_3	C_6H_5CO	5.2 ± 0.4
C4n	CH_3	OCH_3	p-ClC_6H_4CO	1.5 ± 0.2
C4o	CH_3	OCH_3	p-$CH_3C_6H_4SO_2$	0.2 ± 0.01
C4p	CH_3	OCH_3	$C_6H_5SO_2$	0.7 ± 0.06
C4q	CH_3	OCH_3	CH_3SO_2	8.9 ± 1.0
C4r	CH_3	OCH_3	(E)C_6H_5CH=CHCO	0.5 ± 0.07
C5a	p-$CH_3C_6H_4$	NHOH	H	0.002 ± 0.0002
C5b	C_6H_5	NHOH	H	0.004 ± 0.0003
C5c	CH_3	NHOH	H	0.02 ± 0.001

Table 2 Statistics of CoMFA and CoMSIA PLS analyses

Statistical parameters	CoMFA		CoMSIA	
	Model A	Model B	Model A	Model B
q^2	0.843	0.833	0.757	0.801
ONC	10	8	5	12
SEE	0.230	0.261	0.394	0.248
F value	114.306	109.943	72.081	81.786
R^2	0.694	0.723	0.699	0.793
R_0^2	0.693	0.695	0.678	0.727
K	1.037	0.994	0.991	1.011
$\frac{(R^2-R_0^2)}{R^2}$	0.001	0.039	0.030	0.083
Fraction of field contributions				
Steric	0.63	0.578	0.193	0.184
Electrostatic	0.37	0.422	0.173	0.200
Hydrophobic	–	–	0.258	0.243
Donor	–	–	0.113	0.201
Acceptor	–	–	0.263	0.172

MMP-2. Based on the nuclear magnetic resonance (NMR) conformation of the MMP-2 catalytic domain (PDB ID: 1HOV) (Feng *et al*., 2002), MDSA and flexible docking methods were employed to predict the ligand–protein binding pose. Because it had the highest MMP-2 inhibitory activity, compound **A8a** was selected as the template compound to provide a minimized conformation suggested by MDSA. In brief, compound **A8a** was constructed with Sybyl/Sketch module and optimized using Powell's (1977) method with the Tripos force field (Clark *et al*., 1989) with convergence criterion set at 0.05 kcal/(Å mol), and assigned with Gasteiger–Hückel method (Gasteiger and Marsili, 1980). This optimized conformation was then flexible-docked into the binding pocket of MMP-2.

Molecular alignment based on low-energy template

MDSA was used to obtain the global minimum-energy conformation of compound **A8a**. The optimized conformation of **A8a** was subjected to heat at 2,000 K for 2,000 fs followed by annealing of the molecule to 0 K in 5,000 fs. The same cycle was run 100 times to obtain the minimum-energy conformer as a template. The remaining molecules in both the training set and test set were constructed with the Sybyl/Sketch module based on this template, and aligned into the common scaffold using the Sybyl/alignment database method.

Molecular alignment based on docking template

Molecular docking method was often applied to predict the binding conformation of ligand when there was no available co-crystal structure (Awale and Mohan, 2008;

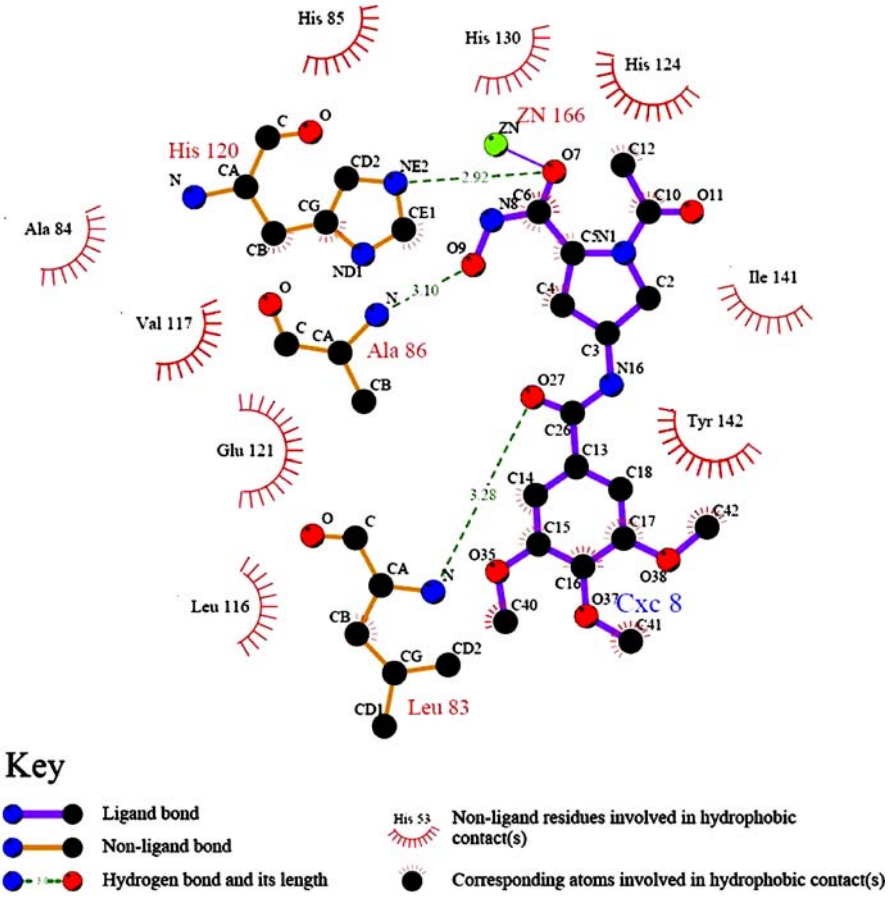


Fig. 1 Schematic representation of interaction between compound **A8a** and MMP-2 using Ligplot program to identify some specific contact between atoms of ligand and receptor. The template compound **A8a** is marked as **Cxc 8**

Nurbo *et al*., 2008). Here molecule docking was performed using the Sybyl/FlexX module based on the solution structure of a catalytic domain of MMP-2 (PDB ID: 1HOV) (Feng *et al*., 2002). Residues around the radius of 4 Å of SC-74020 (the ligand of MMP-2 in the crystal structure 1HOV) were considered as the active site of MMP-2, including the zinc ion. The optimized conformation of **A8a** was then docked into the active site. The binding conformation of compound **A8a** suggested by FlexX was treated as a reference scaffold. Other molecules were sketched and aligned using the same process as described above.

CoMFA and CoMSIA analysis

CoMFA and CoMSIA models were built separately for the aligned molecular dataset with two different alignment methods. CoMFA steric and electrostatic fields

were generated using Lennard–Jones and Columbic potential, respectively. An sp^3-hybridized carbon probe atom with a charge of +1.0 was used to calculate the steric field and electrostatic field of CoMFA. A grid spacing of 2 Å was used for the models. The CoMSIA fields (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were generated with the default parameters implemented in Sybyl. Partial least-squares (PLS) regression was separately performed on the data sets. Leave-one-out (LOO) cross-validation process, with SAMPLS method, was firstly used to seek the optimum number of components (ONC), and then the CoMFA and CoMSIA models were computed with non-cross-validation PLS at the obtained ONC. To avoid the influence of noise in the process, the column filtering value was set to 2.0. To clarify the results of the QSAR models, the alignment set based on the low-energy template is named model A, and the alignment set based on the docking template is named model B.

Result and discussion

Docking result of the most potent molecule

The two-dimensional representation for the interaction mode between compound **A8a** and MMP-2 generated by the Ligplot (Wallace *et al*., 1995) program is shown in Fig. 1. Three hydrogen bonds between compound **A8a** and active site of MMP-2 (His120, Ala86, and Leu83) are formed, as well as the chelation interaction between the hydroxamate group and the zinc ion. The trimethoxybenzene group also triggers hydrophobic interactions with some residues in the S1′ pocket (Ala84, Val117, Ile141, Tyr142, Leu183, Leu116, and Glu121). Meanwhile, the pyrrolidine group forms hydrophobic contacts with His85 and His120, and acetyl group also induces

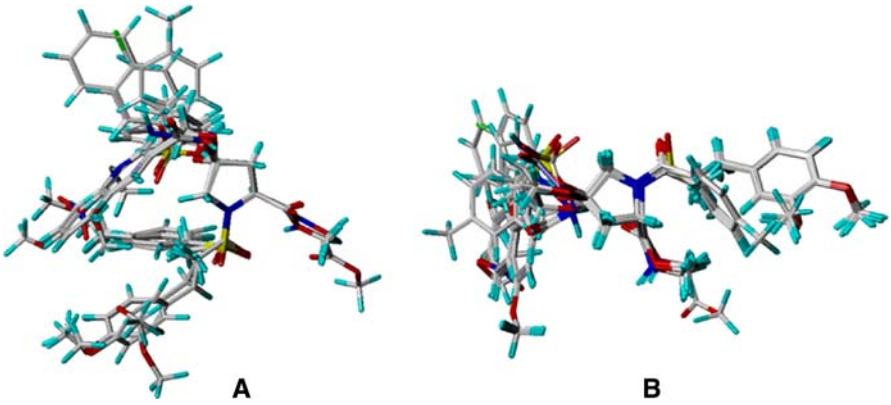


Fig. 2 Structural alignments of the compounds in the training set and test set for constructing 3D-QSAR CoMFA and CoMSIA models. **a** Structural alignments based on the global minimum-energy conformation of compound **A8a**. **b** Structural alignments based on the docking conformation of compound **A8a**

Fig. 3 Contour maps of CoMFA and CoMSIA models, with the most potent compound **A8a** taken as the reference molecule. All map regions are represented in transparent view mode. (**a**, **c**) Steric contour maps of CoMFA (**a**) and CoMSIA (**c**) model, in which green indicates regions where bulky groups increase activity and yellow indicates regions where bulky groups decrease activity. (**b**, **d**) Electrostatic contour maps of CoMFA (**b**) and CoMSIA (**d**) model, in which blue indicates regions where more positively charged groups increase activity and red indicates regions where more negatively charged groups increase activity. (**e**) Hydrophobic contour map of CoMSIA model, in which yellow indicates regions where hydrophobic groups increase activity and white indicates regions where hydrophilic groups increase activity. (**f**) Hydrogen-bond donor contour map of CoMSIA model, in which cyan indicates regions where hydrogen-bond donor groups increase activity and purple indicates regions of unfavorable contributions from hydrogen-bond donor. (**g**) Hydrogen-bond acceptor contour map of CoMSIA model, in which magenta indicates regions where hydrogen-bond acceptor groups increase activity and red indicates regions where they make unfavorable contribution to activity [refer color figure in online version] ▶

hydrophobic interaction with His130 and His124. It could be concluded that the chelation of the zinc ion, hydrogen bonding, and hydrophobic interactions form the structural basis of the MMP-2 inhibiting activity of compound **A8a**.

Statistical analysis of CoMFA and CoMSIA models

Two 3D-QSAR models, model A and model B, were constructed with the same training set by using the CoMFA and CoMSIA methods. The superposition of all the conformations for each model is presented in Fig. 2, and the statistical results of the PLS analysis are presented in Table 2.

To select the best CoMFA and CoMSIA models for further discussion, the standard error of estimate (SEE) and F values, which are usually used as criteria accompanying the q^2 value, were calculated. A reliable model was considered to have high q^2 and F and low SEE value. In our case the CoMFA statistical data of model A almost coincided with that of model B, but the CoMSIA data of model B was obviously better than that of model A. Based on these considerations, the CoMFA and CoMSIA results from model B appear better than those from model A. A value of q^2 above 0.3 indicates that the probability of chance correlation is less than 5% (Cramer *et al*., 1988b). In the present work, either CoMFA or CoMSIA analysis for these two models yields high q^2 values of more than 0.3. For model B, q^2 values were 0.833 and 0.801, respectively, for CoMFA and CoMSIA analysis. Statistical results indicate that the docking conformation could provide better molecular alignment than the MDSA method.

CoMFA and CoMSIA contour maps analysis

The contour maps for model B are subjected to detailed discussion because of their better statistical results. Figure 3 presents the contour maps of model B, in which the docking conformation of compound **A8a** was taken as the reference molecule. In addition, in order to discuss the 3D-QSAR contour maps further, combined with the surface properties of the MMP-2 active site, the solvent-accessible surface of the MMP-2 active site was mapped onto the corresponding contour map (Fig. 4).

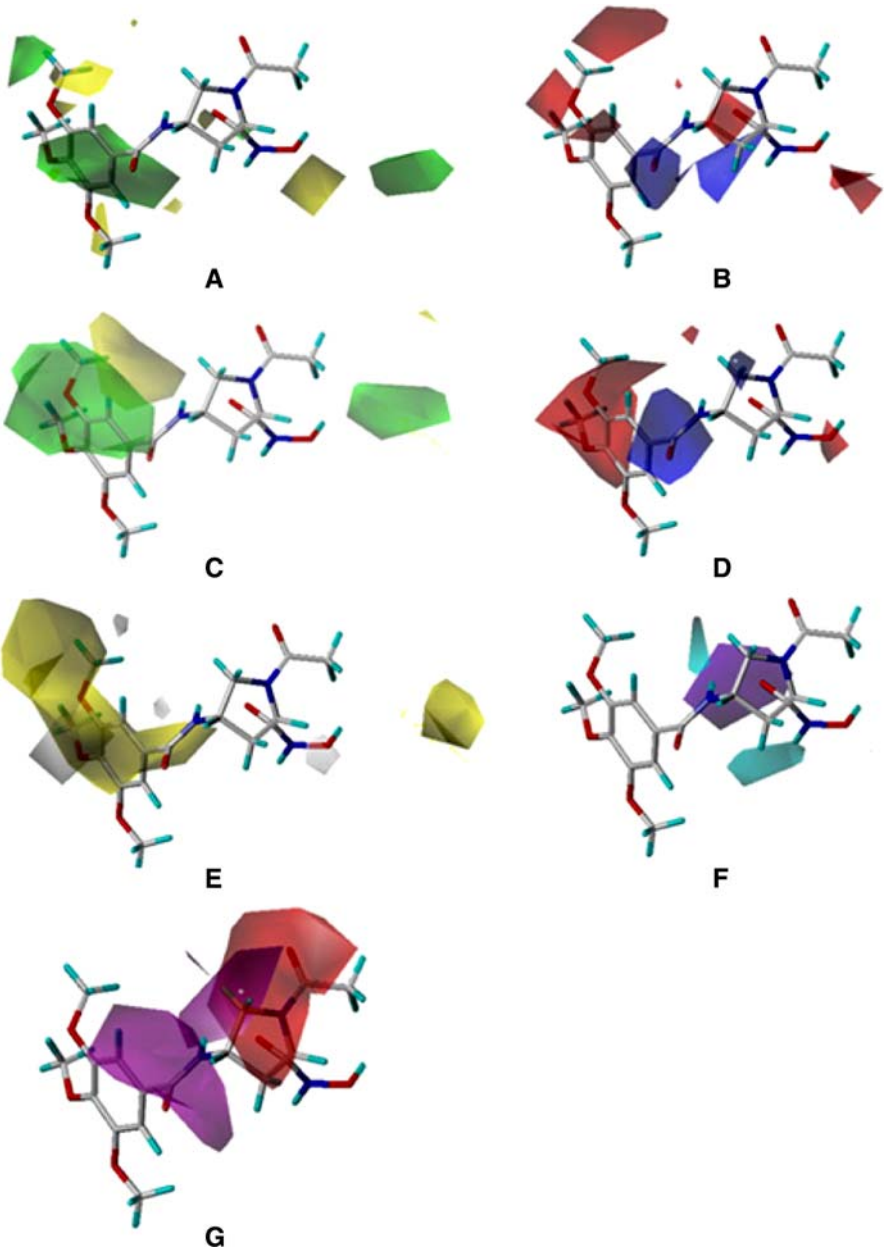


Figure 3a presents the CoMFA steric contour maps of model B, which consist of two green and two yellow regions. It could be observed from Fig. 4 that the two green regions correspond to the S1′ and S2′ pockets on the MMP-2 active site, which shows that substitution of bulky group at these positions would help increase

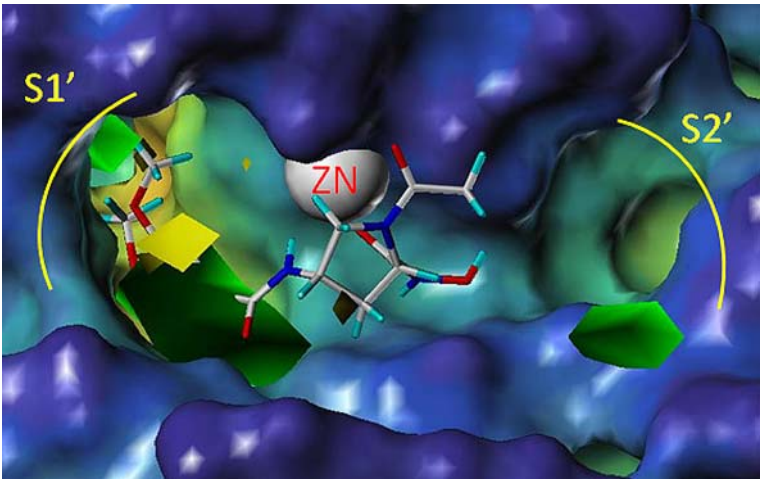


Fig. 4 CoMFA steric contour map from model B plus compound **A8a** on the cavity depth surface (blue: outside of the molecule; orange/yellow: cavities deep inside the molecule) of the MMP-2 active site [refer color figure in online version]

potency. The greater potency exhibited by compounds **A4a**, **A7e**, **B3a**, **B3b**, and **C4c** compared with compounds **A3a** and **A6a** may be due to the bulky group substitution at the R1 or R3 position. The yellow region of the CoMFA steric contour maps (Fig. 3a) around the R2 position cannot be seen in Fig. 4 because the yellow region is overlapped by the protein surface. This phenomenon suggests that substitution of bulky group at this position might result in steric hindrance with the protein; for example, compound **A10** exhibits less potency than compound **A9** because of its bulky R2 group. Another yellow region of the CoMFA steric contour map is located on the upside of the S1′ pocket (Fig. 4). Compounds with substituent at this position result in insufficient interaction with the S1′ pocket, which leads to lower potency (e.g., compounds **A5a** and **A5b**). CoMFA electrostatic contour maps, including three red parts (left, middle, and right regions) and one blue region, are shown in Fig. 3b. The red region adjacent to the R2 group (right region) indicates that negatively charged groups at this position would increase activity. Therefore, compounds with an NH–OH group at the R2 position (e.g., compounds **A8d** and **A8e**) will have greater potency than those with methyl ester (e.g., compounds **A7l** and **A7m**). The red regions in the middle and left parts were mapped with the amide and methoxyl groups of compound **A8a**, and compounds with negatively charged groups at this position usually exhibited good potency (e.g., compounds **A4a** and **A7j**). The large blue region in the middle indicates that the presence of positively charged groups would increase activity. There seemed to be no typical positively charged group found in the training set, but the electrostatic potential surface (Fig. 5) suggested that a blue surface area corresponds to the blue contour map, indicated the existence of electropositive residues in the protein and that electronegative groups are preferred at this position.

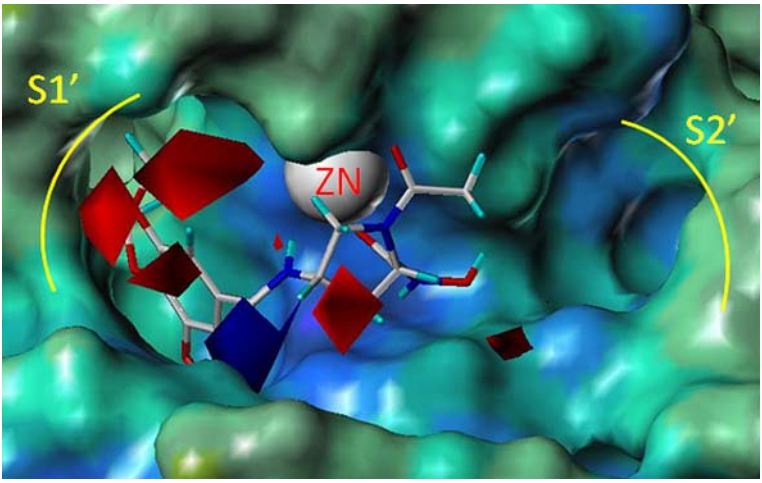


Fig. 5 CoMFA electrostatic contour map from model B plus compound **A8a** on the electrostatic potential surface (blue, negative potential; red/brown, positive potential) of the MMP-2 active site [refer color figure in online version]

The CoMSIA steric and electrostatic contour maps were found to be approximately identical to the corresponding CoMFA contour maps. Since these contour maps were discussed above, this paragraph will focus on the CoMSIA hydrophobic contour maps as well as the CoMSIA hydrogen-bond donor and acceptor contour maps. Figure 3c shows the CoMSIA hydrophobic contour maps of model B. There were two yellow regions close to the S1′ and S2′ pockets, and one white region at the R2 position. The yellow regions indicate that hydrophobic groups will be favored at these positions. The lipophilic potential surface of the active site of MMP-2 (Fig. 6) further confirmed the CoMSIA hydrophobic contour maps. The brown surface deep in the S1′ pocket shows that this area is highly hydrophobic so that compounds with hydrophobic substituent (e.g., compounds **A7i** and **A7j**) will possess better potency than those that do not (e.g., compound **A5a**). The small white region adjacent to the R2 position indicates that hydrophilic groups will be favorable, thus compounds that have an NH–OH group (e.g., compounds **C5a**, **C5b**, and **C5c**) at R2 will have much greater potency than the corresponding esters (e.g., compounds **B3a**, **B3b**, and **B3c**). CoMSIA hydrogen-bond donor contour maps contain two cyan regions and one purple region. The two cyan regions correspond to the two amide linkage positions of compound **A8a**, indicating that hydrogen-bond donor groups in these positions will be conducive to greater potency. The purple region of the CoMSIA hydrogen-bond donor contour maps occupied the same location as the zinc ion (Fig. 7), while the CoMSIA hydrogen-bond acceptor contour map showed a magenta region located at the same position as the purple region of the donor contour map, indicating that this region prefers hydrogen acceptor to donor. Another magenta region of the CoMSIA hydrogen-bond acceptor contour maps near the amide group on the R4 group of compound **A8a** indicates preference for hydrogen-bond acceptor groups (Fig. 8). This observation can be confirmed by the docking result, which showed that the amino group of Leu83

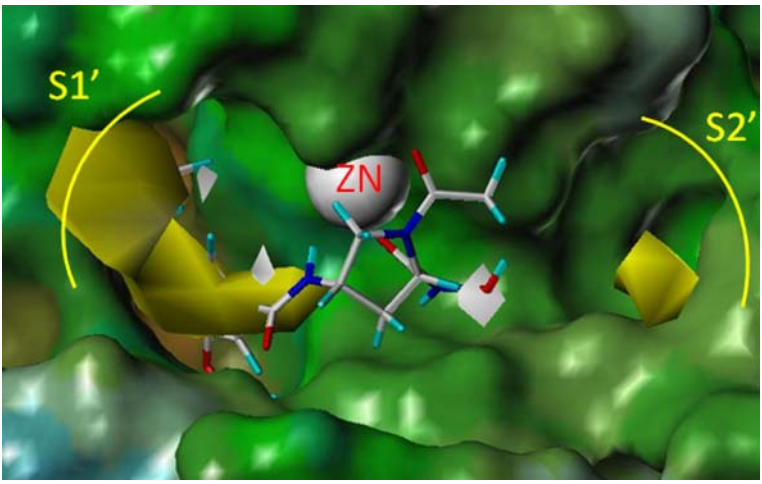


Fig. 6 CoMSIA hydrophobic contour map from model B plus compound **A8a** on the lipophilic potential surface (brown, hydrophobic; blue, hydrophilic) of the MMP-2 active site [refer color figure in online version]

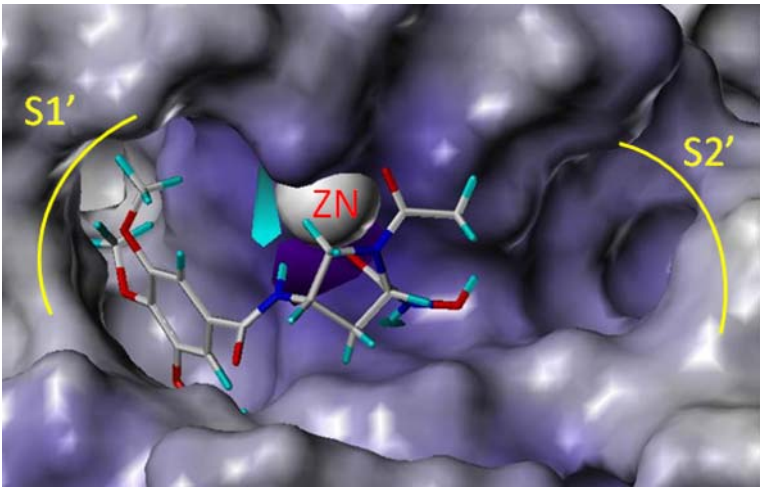


Fig. 7 CoMSIA hydrogen-bond donor contour map from model B plus compound **A8a** on the hydrogen-bond acceptor surface (blue: high acceptor density; white: low acceptor density) of the MMP-2 active site [refer color figure in online version]

might form a hydrogen bond with the carbonyl oxygen on the R4 group of compound **A8a** (Fig. 1).

Model validation

To evaluate the predictive ability of the models, a training set of 50 molecules and a test set of 17 molecules were used. Table 3 showed the actual pIC_{50} of the

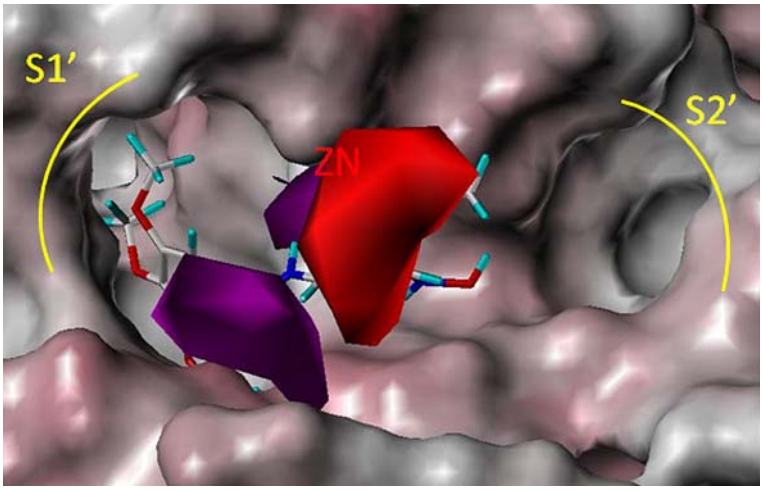


Fig. 8 CoMSIA hydrogen-bond acceptor contour map from model B plus compound **A8a** onto the hydrogen-bond donor surface (red: high acceptor density; white: low acceptor density) of the MMP-2 active site [refer color figure in online version]

compounds and the activity data predicted by the 3D-QSAR models. It should be noted that the predicted activities of the compounds are close to the actual values (Fig. 9).

Early study indicated that a high value of q^2 alone is an insufficient criterion for a QSAR model to be highly predictive. Tropsha and coworkers have introduced a new set of validation criteria for QSAR or QSPR models (Golbraikh and Tropsha, 2002; Golbraikh *et al*., 2003), and those criteria were accepted and have been applied by other researchers (Liu and Gramatica, 2007; Li *et al*., 2008). Tropsha holds that a predictive QSAR model should satisfy the following conditions:

$$q^2 > 0.5$$

$$R^2 > 0.6$$

$$\frac{\left(R^2 - R_0^2\right)}{R^2} < 0.1 \text{ and } 0.85 \leq K \leq 1.15$$

Here R^2 is the correlation coefficient between the predicted and observed activities of the test set and R_0^2 is a quantity characterizing linear regression of the test set with the Y-intercept set to zero (i.e., described by $Y = kX$, where Y and X are the actual and predictive activity, respectively), which is different from the conventional R for the best-fit linear regression (i.e., $Y = aX + b$) (Liu and Gramatica, 2007; Li *et al*., 2008).

The QSAR models established in the present study match the Tropsha criterion very well. The statistical data are shown in Table 2.

Table 3 Actual activity versus predicted activity of the compounds in 3D-QSAR models

No.	Actual	CoMFA predicted pIC_{50}		CoMSIA predicted pIC_{50}	
	pIC_{50}	Model A	Model B	Model A	Model B
A3a	5.7	5.768	5.759	5.762	5.813
A3b	5.54	5.513	5.474	5.749	5.456
A4a[a]	6.52	6.131	5.992	5.648	6.081
A4b	5.96	6.018	5.935	5.615	5.911
A5a	5.59	5.693	5.569	5.751	5.51
A5b	5.47	5.359	5.503	5.738	5.536
A6a	5.82	5.781	5.784	5.553	5.827
A6c[a]	5.68	5.562	5.477	5.665	5.268
A7a	5.57	5.554	5.546	5.722	5.481
A7b	5.38	5.358	5.336	5.773	5.352
A7c	5.21	5.309	5.311	5.787	5.328
A7d	5.28	5.095	5.15	5.554	5.278
A7e[a]	7	7.131	7.081	6.379	6.881
A7f	6	5.971	6.111	5.712	6.11
A7g	5.66	5.7	5.647	5.673	5.636
A7h	5.68	5.758	5.764	5.72	5.703
A7i	7	6.746	6.799	6.079	6.705
A7[a]	6.1	6.844	6.972	6.218	6.654
A7k	5.52	5.651	5.625	5.559	5.761
A7l	5.66	5.687	5.562	5.685	5.61
A7m	5.8	5.854	5.831	6.027	5.868
A8a	9	8.987	9.035	8.548	9.045
A8b[a]	8.3	7.795	8.091	7.881	8.248
A8c	8.4	8.563	8.445	8.243	8.561
A8d	7.92	7.739	7.817	7.851	7.724
A8e	7.96	8.012	8.08	8.207	7.958
A9[a]	7.8	9.07	7.483	6.441	7.106
A10	6.96	7	6.954	6.869	6.899
B3a	7	6.829	6.858	6.353	6.905
B3b	6.1	6.17	6.25	6.227	6.266
B3c[a]	5.15	5.465	5.49	5.826	5.518
B4a	6.4	6.38	6.42	6.092	6.358
B4b[a]	5.7	5.754	5.839	5.966	5.746
B4c	5.02	5.041	5.06	5.565	5.007
B5a	7	6.913	6.884	6.402	6.864
B5b	6.7	6.683	6.626	6.305	6.649
B5c	6.52	6.53	6.53	6.219	6.512
B5d[a]	6.7	6.251	6.312	6.276	6.258
B5e	5.96	6.028	6.047	6.179	6.043
B5f	5.82	5.903	5.945	6.101	5.906

Table 3 continued

No.	Actual	CoMFA predicted pIC_{50}		CoMSIA predicted pIC_{50}	
	pIC_{50}	Model A	Model B	Model A	Model B
B6a	8.52	8.513	8.606	8.574	8.665
B6b[a]	8.22	8.34	8.35	8.479	8.449
B6c	8.1	8.362	8.254	8.359	8.275
B6d	8.1	8.04	8.042	8.445	8.054
B6e[a]	7.96	7.825	7.781	8.349	7.841
B6f	7.92	7.733	7.679	8.241	7.668
C4a	5.49	5.853	5.959	6.54	5.847
C4b[a]	5.66	6.524	6.324	6.732	6.691
C4c	8.15	7.782	7.991	7.818	7.999
C4d	7.3	7.455	7.489	7.491	7.51
C4e[a]	6.7	6.186	6.067	6.862	6.354
C4f	7.4	7.596	7.501	6.629	7.391
C4g	5.46	5.626	5.941	6.173	5.847
C4h[a]	5.57	6.298	6.324	6.366	6.691
C4i	7.4	7.578	7.542	7.451	7.611
C4j	8	7.25	7.039	7.124	7.122
C4k	5.89	5.964	5.614	6.466	5.875
C4l[a]	7.05	7.366	7.055	6.263	7.003
C4m	5.28	4.931	4.953	5.134	4.765
C4n	5.82	5.601	5.334	5.326	5.609
C4o	6.7	6.836	6.998	6.43	6.916
C4p	6.15	6.512	6.485	6.102	6.427
C4q	5.05	5.201	5.072	5.445	5.175
C4r[a]	6.3	6.668	6.504	5.223	6.307
C5a	8.7	8.661	8.601	8.612	8.778
C5b[a]	8.4	8.458	8.147	8.243	8.389
C5c	7.7	7.701	7.705	7.23	7.707

[a] Molecules in test set

Conclusion

In the present study, we examined the 3D-QSAR models of a set of pyrrolidine derivatives as MMP-2 inhibitors using CoMFA and CoMSIA methods. Both molecular-dynamics simulated annealing and molecular docking techniques were used in the identification of the template conformation. Tropsha's criteria were introduced to validate the predictive power of these 3D-QSAR models, and all models matched these criteria well. Contour maps of the best models matched well with the surface characteristics of the MMP-2 activity site. The 3D-QSAR models were helpful in understanding the structure requirement of the MMP-2 active site,

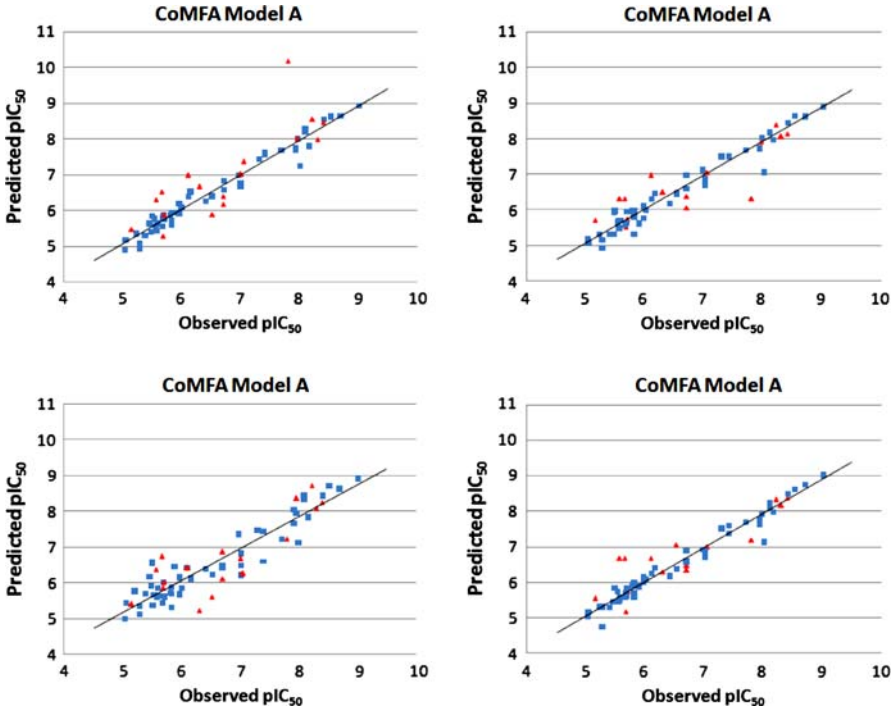


Fig. 9 Scatterplots of the CoMFA/CoMSIA predicted and experimental values of MMP-2 inhibitory activity for molecules in models A and B. ■ Training set. ▲ Test set

and will give reasonable suggestions for modification of pyrrolidine derivatives as MMP-2 inhibitors.

Acknowledgements This work was supported by the National High Technology Research and Development Program of China (863 project; Grant No. 2007AA02Z314), the National Natural Foundation Great Research Program (No. 90713041), and the National Natural Foundation Research Grant (Grant No. 30772654).

References

Awale M, Mohan CG (2008) Molecular docking guided 3D-QSAR CoMFA analysis of N-4-Pyrimidinyl-1H-indazol-4-amine inhibitors of leukocyte-specific protein tyrosine kinase. J Mol Model 14:937–947. doi:10.1007/s00894-008-0334-8

Björklund M, Koivunen E (2005) Gelatinase-mediated migration and invasion of cancer cells. Biochim Biophys Acta 1755:37–69

Cheng XC, Wang Q, Fang H, Tang W, Xu WF (2008a) Design, synthesis and evaluation of novel sulfonyl pyrrolidine derivatives as matrix metalloproteinase inhibitors. Bioorg Med Chem 16:5398–5404. doi:10.1016/j.bmc.2008.04.027

Cheng XC, Wang Q, Fang H, Tang W, Xu WF (2008b) Design, synthesis and preliminary evaluation of novel pyrrolidine derivatives as matrix metalloproteinase inhibitors. Eur J Med Chem 43:2130–2139. doi:10.1016/j.ejmech.2007.12.020

Cheng XC, Wang Q, Fang H, Tang W, Xu WF (2008c) Synthesis of new sulfonyl pyrrolidine derivatives as matrix metalloproteinase inhibitors. Bioorg Med Chem 16:7932–7938. doi:10.1016/j.bmc.2008.07.073

Clark M, Cramer RD, Opdenbosch NV (1989) Validation of the general purpose tripos 5.2 force field. J Comput Chem 10:982–1012. doi:10.1002/jcc.540100804

Cramer RD, Bunce JD, Patterson DE, Frank IE (1988a) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant Struct Act Relat 7:18–25. doi:10.1002/qsar.19880070105

Cramer RD, Patterson DE, Bunce JD (1988b) Comparative molecular field analysis (CoMFA) 1 Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967. doi:10.1021/ja00226a005

Exner T, Keil M, Moeckel G, Brickmann J (1998) Identification of substrate channels and protein cavities. J Mol Model 4:340–343. doi:10.1007/s008940050091

Feng Y, Likos JJ, Zhu L, Woodward H, Munie G, McDonald JJ, Stevens AM, Howard CP, De Crescenzo GA, Welsch D, Shieh HS, Stallings WC (2002) Solution structure and backbone dynamics of the catalytic domain of matrix metalloproteinase-2 complexed with a hydroxamic acid inhibitor. Biochim Biophys Acta 1598:10–23

Fisher JF, Mobashery S (2006) Recent advances in MMP inhibitor design. Cancer Metastasis Rev 25:115–136. doi:10.1007/s10555-006-7894-9

Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. Tetrahedron 36:3219–3228. doi:10.1016/0040-4020(80)80168-2

Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A (2003) Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 17:241–253. doi:10.1023/A:1025386326946

Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20:269–276. doi:10.1016/S1093-3263(01)00123-1

Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146. doi:10.1021/jm00050a010

Kontogiorgis CA, Papaioannou P, Hadjipavlou-Litina DJ (2005) Matrix metalloproteinase inhibitors: a review on pharmacophore mapping and (Q)SARs results. Curr Med Chem 12:339–355

Li M, Ni N, Wang B, Zhang Y (2008) Modeling the excitation wavelengths (lambda(ex)) of boronic acids. J Mol Model 14:441–449. doi:10.1007/s00894-008-0293-0

Liu H, Gramatica P (2007) QSAR study of selective ligands for the thyroid hormone receptor beta. Bioorg Med Chem 15:5251–5261. doi:10.1016/j.bmc.2007.05.016

Murphy GJ, Murphy G, Reynolds JJ (1991) The origin of matrix metalloproteinases and their familial relationships. FEBS Lett 289:4–7. doi:10.1016/0014-5793(91)80895-A

Nagase H, Visse R, Murphy G (2006) Structure and function of matrix metalloproteinases and TIMPs. Cardiovasc Res 69:562–573. doi:10.1016/j.cardiores.2005.12.002

Nurbo J, Peterson SD, Dahl G, Helena Danielson U, Karlen A, Sandstrom A (2008) Beta-amino acid substitutions and structure-based CoMFA modeling of hepatitis C virus NS3 protease inhibitors. Bioorg Med Chem 16:5590–5605. doi:10.1016/j.bmc.2008.04.005

Polette M, Nawrocki-Raby B, Gilles C, Clavel C, Birembaut P (2004) Tumour invasion and matrix metalloproteinases. Crit Rev Oncol Hematol 49:179–186. doi:10.1016/j.critrevonc.2003.10.008

Powell MJD (1977) Restart procedures for the conjugate gradient method. Math Program 12:241–254. doi:10.1007/BF01593790

Verma RP, Hansch C (2007) Matrix metalloproteinases (MMPs): chemical-biological functions and (Q)SARs. Bioorg Med Chem 15:2223–2268. doi:10.1016/j.bmc.2007.01.011

Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8:127–134. doi:10.1093/protein/8.2.127