

3D-QSAR studies on tripeptide aldehyde inhibitors of proteasome using CoMFA and CoMSIA methods

Yong-Qiang Zhu,^a Jian-Feng Pei,^b Zhen-Ming Liu,^b Lu-Hua Lai,^{b,*}
Jing-Rong Cui^a and Run-Tao Li^{a,*}

^a*School of Pharmaceutical Sciences, Peking University Health Science Center, Beijing 100083, China*

^b*College of Chemistry and Molecular Engineering, State Key Laboratory for Structural Chemistry of Stable and Unstable Species, Peking University, Beijing 100871, China*

Received 5 August 2005; revised 30 September 2005; accepted 1 October 2005

Available online 26 October 2005

Abstract—The ubiquitin–proteasome pathway plays a crucial role in the regulation of many physiological processes and in the development of a number of major human diseases, such as cancer, Alzheimer’s, Parkinson’s, diabetes, etc. As a new target, the study on the proteasome inhibitors has received much attention recently. Three-dimensional quantitative structure–activity relationship (3D-QSAR) studies using comparative molecule field analysis (CoMFA) and comparative molecule similarity indices analysis (CoMSIA) techniques were applied to analyze the binding affinity of a set of tripeptide aldehyde inhibitors of 20S proteasome. The optimal CoMFA and CoMSIA models obtained for the training set were all statistically significant with cross-validated coefficients (q^2) of 0.615, 0.591 and conventional coefficients (r^2) of 0.901, 0.894, respectively. These models were validated by a test set of eight molecules that were not included in the training set. The predicted correlation coefficients (r^2) of CoMFA and CoMSIA are 0.944 and 0.861, respectively. The CoMFA and CoMSIA field contour maps agree well with the structural characteristics of the binding pocket of $\beta 5$ subunit of 20S proteasome, which suggests that the 3D-QSAR models built in this paper can be used to guide the development of novel inhibitors of 20S proteasome.

© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

The ubiquitin–proteasome ATP-dependent pathway, which was identified more than 20 years ago,^{1,2} plays a pivotal role in the regulation of many physiological processes^{3–10} and in the development of a number of major human diseases,^{11–17} such as cancer, Alzheimer’s, Parkinson’s, diabetes, etc. However, the development of the studies on its physiological roles in mammalian cells was rather slow until the introduction of the cell-permeable proteasome inhibitors that greatly simplified such studies.^{18,19} The use of these inhibitors demonstrated that most of the degradation of the short-lived and long-lived proteins was catalyzed by the proteasome, while the long-lived proteins comprise the bulk of the proteins in mammalian cells.^{20,21} Studies using these inhibitors have also

demonstrated that the ubiquitin–proteasome pathway is essential for many cellular regulatory mechanisms.²²

The proteasome is a highly conserved intracellular multicatalytic proteinase complex (MPC),^{23,24} which can catalyze cleavage of peptide bonds on the carboxyl side of basic, acidic, and hydrophobic amino acid residues in both natural peptides and synthetic substrates.^{25–29} Within the ubiquitin–proteasome system, the 26S proteasome is involved, which is composed of a 19S capped-shape regulatory particle (also termed RP or 19S proteasome) and a 20S catalytic core particle (CP or 20S proteasome).³⁰ X-ray crystallography of 26S proteasome^{25,31} further revealed the two following facts: (1) the 19S proteasome is made up of 18 subunits, which control the recognition, deubiquitinylation, and unfolding of the protein substrates prior to its translocation into the 20S proteasome^{32,33}; (2) the cylindrically shaped 20S proteasome is composed of two copies of 14 different subunits, arranged as $(\alpha 1-\alpha 7, \beta 1-\beta 7)_2$ with four stacked rings. The 20S proteasome is a member of the N-terminal nucleophile (Ntn)-hydrolase superfamily.³⁴ Among the seven different β subunits, $\beta 1$, $\beta 2$, and $\beta 5$,

Keywords: 3D-QSAR; CoMFA; CoMSIA; Proteasome inhibitors; Tripeptide aldehyde; Drug design.

* Corresponding authors. Tel.: +86 10 62757520; fax: +86 10 62751725 (L.-H.L.); tel.: +86 10 82801504; fax: +86 10 62346154 (R.-T.L.); e-mail addresses: lhilai@pku.edu.cn; lirt@bjmu.edu.cn

which were originally characterized as post-glutamyl peptide hydrolytic (PGPH), trypsin-like (T-L), and chymotrypsin-like (CT-L) activity,^{35,36} respectively, are catalytically active portions with the N-terminal residues of Thr1. They are responsible for cleaving small chromogenic substrates and cutting polypeptides at their active sites, meantime generating peptides with 3–22 residues.^{31,37,38}

The fact that the proteasome has been implicated in many disease states suggests that proteasome inhibitors have therapeutic potentials. Furthermore, these inhibitors would provide outstanding tools for determining the role of the proteasome in various physiological functions. So in recent years, much attention has been drawn to design, synthesize, and biologically evaluate various proteasome inhibitors.³⁹ Nowadays, a dipeptide boronic acid analogue Velcade (also named bortezomib) was approved by both USA and EU in 2003 and 2004, respectively, for the treatment of multiple myeloma patients, and now it is also under investigation in Phase I and II studies for a wide variety of solid cancers.^{40–42}

Among all the proteasome inhibitors ever studied, peptide aldehydes were the first to be developed and are still the most widely used inhibitors in *in vitro* and *in vivo* studies for their rapid reversibility and low cost.^{39,43} They inhibit the proteasome by reacting with the hydroxyl group of the N-terminal threonines of the β subunits to form a reversible hemiacetal, which has been confirmed by X-ray diffraction of complex of tripeptide aldehyde (Ac-Leu-Leu-nLeu-al, commonly termed ALLN or MG101) and 20S proteasome.²⁵ Aldehyde inhibitors have fast dissociation rates, and are rapidly oxidized into inactive acids by cells and are transported out of cells by the multi-drug resistance (MDR) system carrier.⁴⁴ Consequently, in experiments involving cultured mammalian cells and yeast, effects of these inhibitors can be rapidly reversed by removal of the inhibitors.⁴⁵ However, due to their poor selectivity and specificity, their usage was greatly limited. So designing more selective and specific aldehyde inhibitors is an urgent task. Though many different aldehyde inhibitors have been synthesized and experimentally assessed,³⁹ up to now, no QSAR or three-dimensional quantitative structure–activity relationship (3D-QSAR) studies on them were reported. The purpose of the present study was to build predictive 3D-QSAR models by using comparative molecule field analysis (CoMFA)⁴⁶ and comparative molecule similarity indices analysis (CoMSIA)⁴⁷ methods, and using them to find the correlation between the structure and activity of the inhibitors and design more potent and selective inhibitors.

CoMFA and CoMSIA are computer programs that are particularly effective in correlating the 3D structures of the molecules and their bioactivities based on statistic techniques.⁴⁸ These two methods sample the potential fields surrounding a set of ligands and construct 3D-QSAR models by correlating these 3D fields with the corresponding experimental activities of ligands interacting with a common target receptor.

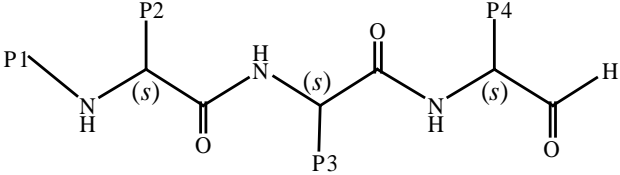
2. Results and discussion

2.1. 3D-QSAR models

The statistical results obtained from standard CoMFA and CoMSIA models are summarized in Table 3. The CoMFA models were constructed from steric and electrostatic descriptor fields, and the CoMSIA models were built by varying the steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor descriptor fields. The q^2 , S_{PRESS} , r^2 , F , and s values were computed as defined in SYBYL. The training set was initially checked for outliers. Generally, if the residual of an inhibitor between observed $\text{p}K_i$ and predicted $\text{p}K_i$ values is greater than 1 logarithm unit, the inhibitor is considered as outlier. According to this rule, inhibitors MG216 and MG217 are regarded as outliers. We examined these two molecules and found that a hydroxyl was inserted between P4 position and aldehyde group for MG216, and between P3 and P4 position for MG217, respectively. Therefore, these two compounds did not share the same scaffold as the other compounds, and the potential field distributions surrounding the molecules are different from the others. The hydroxyl group may form hydrogen bonds with the amino acid residues of the β subunit. The variation of hydrogen-bond donor contributions from 0.165 to 0.118 may account for this. Thus, both CoMFA and CoMSIA models based on the 47 compounds led to a lower cross-validated correlation coefficient q^2 of 0.516 and 0.463, respectively. However, after dropping these two compounds from the training set, the 3D-QSAR models re-performed by the remaining 45 compounds produced a much higher cross-validated q^2 of 0.615 for CoMFA and 0.591 for CoMSIA (see Table 3). Therefore, the models based on the 45 compounds are much better than those based on 47 compounds. So the following results and discussion are based on the 45-compound models.

2.2. Predictive power of CoMFA and CoMSIA models

In order to test the predictive power of the CoMFA and CoMSIA models obtained by the training set, other eight inhibitors (Table 2) that were not included in the training set were used as a test set. Table 4 summarizes the results obtained from the CoMFA and CoMSIA predictions. The plots of predicted versus actual binding affinities for the test set inhibitors are shown in Figures 4A and B, which represent models based on CoMFA and CoMSIA, respectively. Both CoMFA and CoMSIA models performed well in the prediction of the activities of the test inhibitors. The predicted correlation coefficients (r^2) are 0.944 for the CoMFA model and 0.861 for the CoMSIA model. In almost all the cases, the predicted values were close to the observed $\text{p}K_i$ values, deviating by less than 1 logarithm unit in binding affinity to the 20S proteasome. Especially for inhibitors MG101, MG119, and MG142, both CoMFA and CoMSIA models gave ideal predictive values, all of which are less than 0.3 logarithm units. Since all the inhibitors were built with the template of crystal conformation of MG101, the difference between predicted versus actual values may be explained by comparing different flexible

Table 1. Structures of 47 inhibitors in the training set and their actual activities


Compound	P1	P2	P3	P4	p <i>K</i> _i
MG110					7.32
MG111				CH ₃	6.70
MG114					7.33
MG115					7.68
MG118				H	5.42
MG120					7.03
MG121					7.60
MG129					7.13
MG131					8.16
MG132					8.40
MG133					6.33
MG134					6.98
MG135					6.54
MG139					7.70

(continued on next page)

Table 1 (continued)

Compound	P1	P2	P3	P4	p <i>K</i> _i
MG140					7.55
MG141					7.30
MG150					9.40
MG151					8.00
MG153					6.55
MG154					7.18
MG155					7.33
MG157					7.85 (racemate)
MG158					8.94
MG160					7.29
MG161					7.19
MG165					9.62
MG166					10.46
MG167					10.82
MG168			CH ₃		9.31

Table 1 (continued)

Compound	P1	P2	P3	P4	p <i>K</i> _i
MG170					10.00
MG171					10.10
MG173					7.11
MG176					7.15
MG178					9.72
MG184					8.10
MG191					9.42
MG192					7.02
MG208					6.57
MG210					8.26
MG216					6.81

(continued on next page)

Table 1 (continued)

Compound	P1	P2	P3	P4	p <i>K_i</i>
MG217					7.24
MG224					7.70
MG229					7.47
MG232				CH ₃ , CH ₃	5.31
MG234					7.96
MG235					8.26
MG271		H			6.26

All the chiral carbon atoms are in *S* form unless otherwise indicated.

substituents in P1, P2, P3, and P4 positions. MG119 share common *t*-butyl substituents in P2 and P3 positions with MG101, but a less flexible ethyl instead of *n*-butyl substituent in P4 position, so the local minimization conformation is close to crystal one and a good prediction was obtained. As for the substituents in P1 position, they affect the binding affinities much less than those in P4 position, which is clearly shown in Figure 5. Figure 5 shows *n*-butyl group in P4 position of MG101 penetrates deeply into the active site of $\beta 5$ subunit and can effectively interact with MET45 and ALA49 residues by hydrophobic interaction, which was proved by the hydrophobic field contribution 0.411 of CoMSIA results listed in Table 3 and the hydrophobic contour map of CoMSIA results shown in Figure 7B. But acetyl group in P1 position of MG101 lies outside of the active site and interacts with adjacent subunits. The same explanation can account for the case of MG142. However, the predictive results of the other five inhibitors are not as perfect as those of other three inhibitors. It may

be caused by the more flexible substituents in P1, P2, P3, and P4 positions.

2.3. The CoMFA contour map analysis

The steric field contribution accounts for 0.655, while the electrostatic field contribution only accounts for 0.345, 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. Figure 6A shows the steric contour maps for the CoMFA model with the most active inhibitor MG167 ($K_i = 0.015$ nM) as a reference. Green and yellow polyhedra indicate regions where bulky groups will improve or decrease the activity, respectively. Sterically favored regions (green polyhedra) around P2 (2-naphthyl methyl) and P4 (*t*-butyl) positions suggest an increase in the activity for the compounds having bulky substituents at P2 and P4

Table 2. Structures of eight inhibitors in the test set and their actual activities

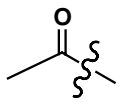
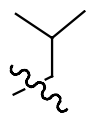
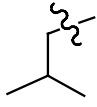
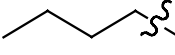
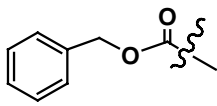
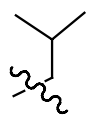
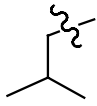
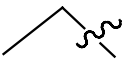
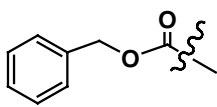
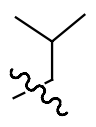
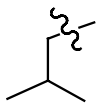
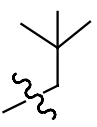
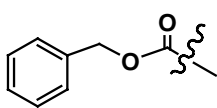
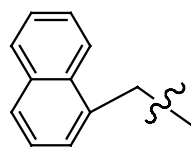
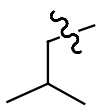
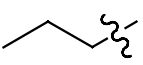
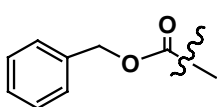
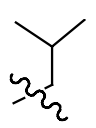
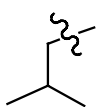
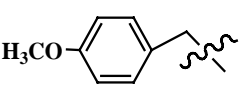
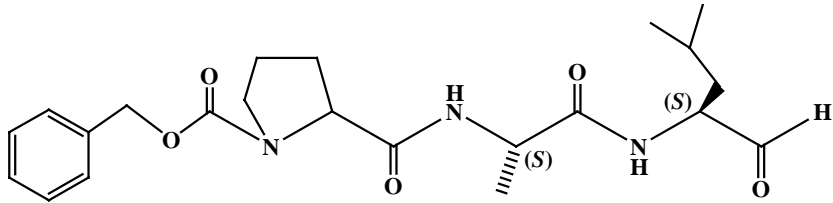
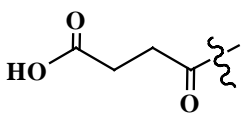
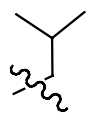
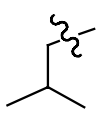
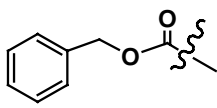
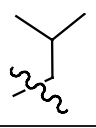
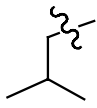
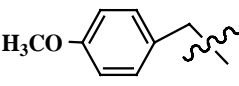
Compound	P1	P2	P3	P4	pK_i
MG101					6.85
MG119					7.30
MG136					6.00
MG142					9.70
MG152					8.22
MG169					6.16
MG183				H, H	7.41
MG209					8.04

Table 3. Summary of 3D-QSAR analysis results obtained using CoMFA and CoMSIA

PLS statistics	47-compound models		45-compound models	
	CoMFA	CoMSIA	CoMFA	CoMSIA
q^2	0.516	0.463	0.615	0.591
S_{PRESS}	0.887	0.944	0.818	0.835
r^2	0.819	0.802	0.901	0.894
F	76.701	50.545	85.311	79.296
s	0.503	0.574	0.416	0.430
PLS components	3	4	5	4
<i>Field contributions (%)</i>				
S	0.565	0.174	0.655	0.197
E	0.435	0.246	0.345	0.205
H		0.243		0.411
D		0.165		0.118
A		0.172		0.068

q^2 , squared cross-validated coefficient; S_{PRESS} , root mean predictive error sum of squares; r^2 , squared conventional coefficient; F , F test value; s , standard error of estimate; S, E, H, D, and A denote the steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor, respectively.

Table 4. Residuals of the predictions of the test set by CoMFA and CoMSIA models

Compound	EA	PA		Residuals	
		CoMFA	CoMSIA	CoMFA	CoMSIA
MG101	6.85	6.66	6.83	0.19	0.02
MG119	7.30	7.01	7.27	0.29	0.03
MG136	6.00	6.82	6.75	-0.82	-0.75
MG142	9.70	9.71	9.62	-0.01	0.08
MG152	8.22	7.81	7.54	0.41	0.68
MG169	6.15	6.06	6.95	0.09	-0.80
MG183	7.41	7.78	6.54	-0.37	0.87
MG209	8.04	7.98	7.50	0.06	0.54

EA, experimental activity; PA, predictive activity; Residuals = EA - PA.

positions. This is indeed the case for MG142, MG150, MG165, MG166, MG167, MG168, MG170, MG171, MG178, and MG191, all of which showed less than 1 nM binding affinity and have bulky groups at P2 and P4 positions. For example, sterically favored 1-naphthyl methyl group occupies P2 positions of MG142, MG150, MG166, MG168, and MG170, respectively, and bulky 2-naphthyl methyl occupies those of MG165 and MG167, while P2 positions of MG171, MG178, and MG191 are substituted by bulky indole-3-methyl and benzothiofuran-3-methyl, respectively. As for P4 positions, these compounds all have a sterically favored *t*-butyl groups, except for MG142 with a comparatively smaller *n*-propyl substituent. But for inhibitors MG111 ($K_i = 210$ nM), MG118 ($K_i = 3800$ nM), and MG232 ($K_i = 4900$ nM), because they have sterically disfavored methyl, hydrogens, and dimethyl substituents at P4 positions, respectively, they were less potent. However, it is an exception for MG153 with a much bulky substituent, *p*-benzyloxy benzyl, at P4 position, which shows a poor binding affinity of 285 nM. Perhaps, this is due to the collision between the too long arm and the residues in the 20S proteasome in addition to the small group at P2 position. Detailed analysis indicates that the structural features of other inhibitors also concord with the contour maps. The presence of a large yellow isopleth around the aromatic moiety of P1 position indicates that less bulky substituents are favored at this position.

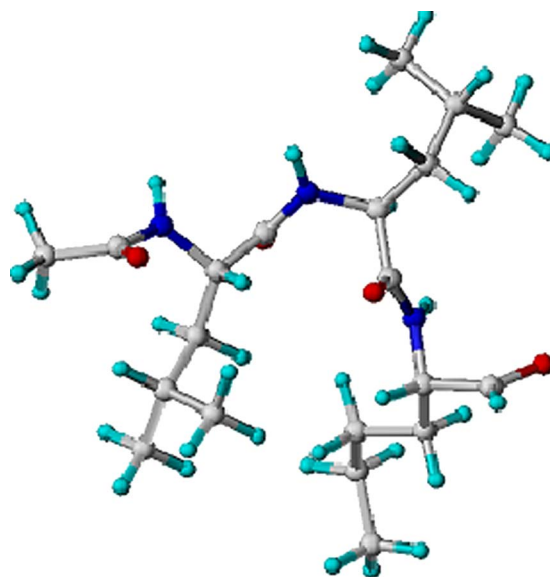
Electrostatic fields based on the PLS analysis of the CoMFA models are shown in Figure 6B with MG167 as a reference. A large red isopleth above the phenyl ring of P1 position represents an area where a negative Gasteiger–Hückel (GH) charge is favored. Methyl group at P1 position of MG101, bearing positive GH charges on the C atom, decreases the activity ($K_i = 140$ nM), while benzyloxy group at P1 position of MG114, having negative charges on the O atom and benzyl ring, increases the activity greatly ($K_i = 47$ nM). Negatively charged aldehyde group is essential for high activity, which is represented by the red favorable isopleth near this area. A small red polyhedron beside the P2 positions indicates electron-rich groups are beneficial to the activity. Such as, MG168 with a negatively charged bulky 1-naphthyl group at P2 position showed a potent activity of K_i 0.49 nM;

while the C atom at P2 position of MG169 was cyclized with neighboring N atom by a positively charged *n*-propyl greatly decreased the activity ($K_i = 700$ nM). A large region of blue contour near the P4 position suggests that positively charged groups such as *t*-butyl (MG132) instead of negatively groups such as benzyl substituted by halogen atoms (MG157, MG158, MG160, and MG161), hydroxyl (MG131), *p*-nitryl (MG135), *p*-methoxy (MG152), and *p*-benzyloxy (MG153), respectively, allyl (MG139), propargyl (MG140), and indole-3-methyl (MG151) increase activity. Furthermore, positively charged hydrogens surrounded by a blue polyhedron at P1 positions are beneficial to activity.

2.4. CoMSIA contour maps analysis

CoMSIA calculates both steric and electrostatic fields, as in CoMFA, but additionally uses hydrophobic, hydrogen-bond donor and acceptor fields. Since a Gaussian function is used to determine the distance dependence, therefore, the similarity indices can also be calculated at grid points inside the molecule, not just outside, as with CoMFA.⁴⁷

Figure 7 provides graphical representation of the CoMSIA models. To aid in visualization, the most potent inhibitor MG167 was overlaid in the maps once again. Figure 7A shows the combination of CoMSIA steric and electrostatic field contour maps. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk will enhance the activity, respectively. Red and blue polyhedra represent regions where negatively charged or positively charged groups will increase the activity. Just as discussed in Section 2.3, P2 and P4 positions are surrounded by two large green polyhedra, which suggests bulky substituents, such as naphthyl methyl and *t*-butyl groups (for example, MG165, MG166, MG167, MG168, and MG170), are favored. A small yellow polyhedron beside the methyl of P2 position indicates small groups instead of bulky groups

**Figure 1.** The starting conformation of MG101.

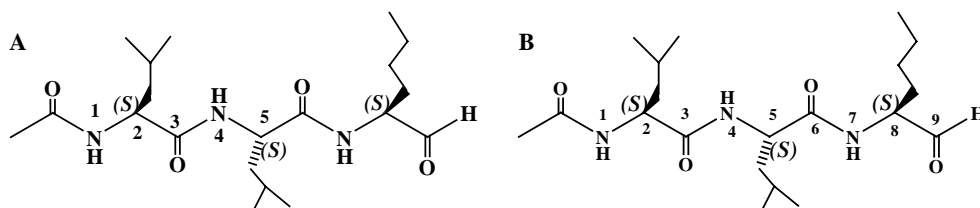


Figure 2. Templates of two models with 55 molecules and 53 molecules. (A) Template of 55 molecules. (B) Template of 53 molecules.

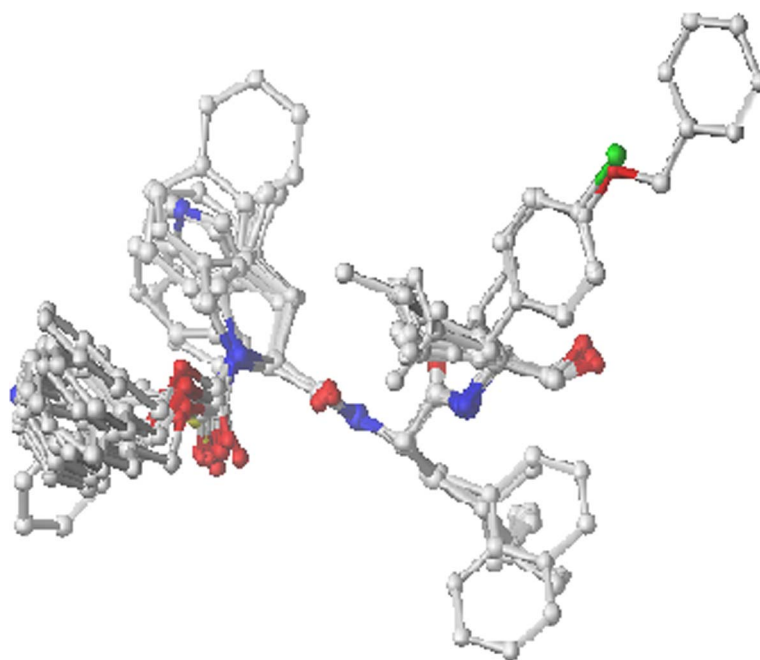


Figure 3. Alignment of inhibitory compounds in the training set. Hydrogens are not shown.

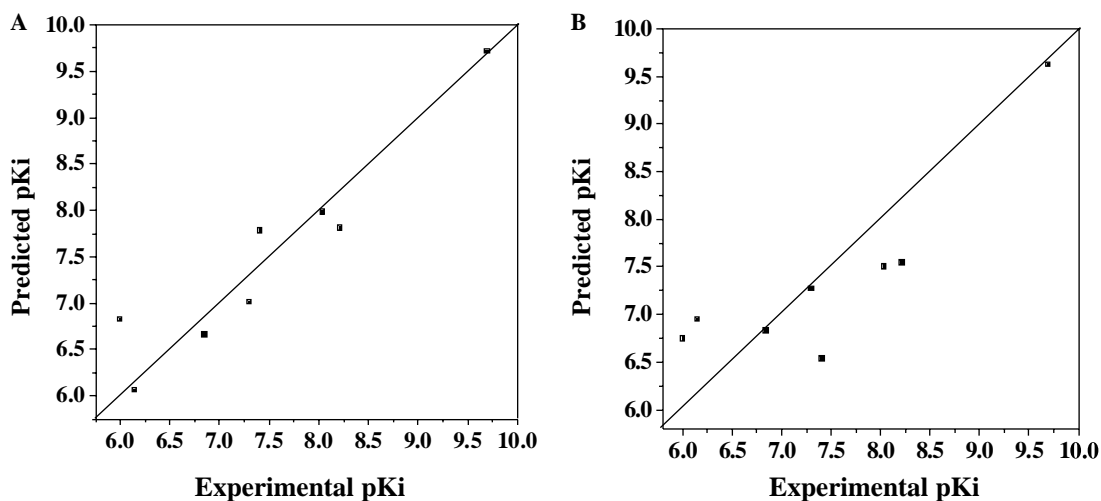


Figure 4. Predicted versus actual binding affinities for the eight inhibitors in the test set for (A) CoMFA and (B) CoMSIA models. The correlation coefficients (r^2) are 0.944 for the CoMFA model and 0.861 for the CoMSIA model.

attaching on the methyl are referred to this region. Also two large red polyhedra covering P1 and P2 positions indicate regions where negatively charged groups will increase the activity, as discussed in detail in CoMFA. The

presence of a blue region beside the methyl of P2 position again proves it is preferred to be positively charged atoms at this region in consideration of their less steric bulk. A red isopleth, near the O atom of the carbonyl

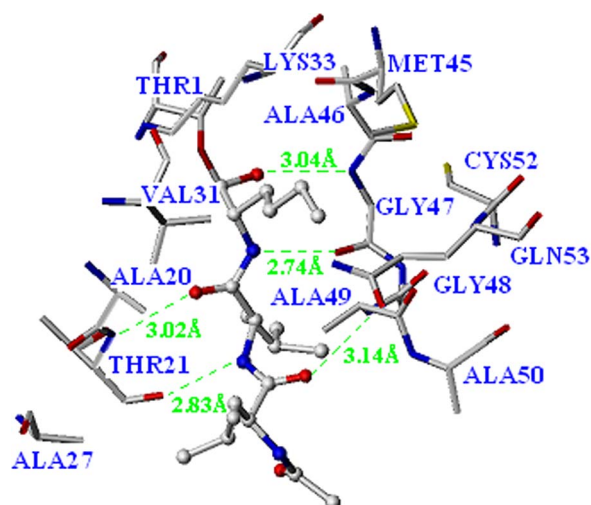


Figure 5. The binding of MG101 with $\beta 5$ subunit in 20S proteasome (crystal structure coordinates kindly provided by R. Huber). Only the residues within 5 Å around MG101 are shown for clarity. MG101 is displayed in ball–stick style and amino acid residues are displayed in stick style. Green dotted lines: hydrogen bonds between MG101 and the amino acid residues.

attaching to P2 position, correlates with hydrogen-bond donor on the receptor. The results agree well with what has been discussed in Section 2.3.

The hydrophobic field is presented in Figure 7B. White and yellow contours highlight areas where hydrophilic and hydrophobic properties are preferred, respectively. Table 3 shows that hydrophobic field made the largest contribution to the CoMSIA QSAR models, which suggests that among the descriptors considered, the hydrophobicity of inhibitors is the most important factor

affecting the binding affinity. This is consistent with the hydrophobic properties of the amino acid residues in the $\beta 5$ subunit of 20S proteasome.²⁵ Within 3.5 Å around MG101, ALA20, MET45, ALA46, ALA49, and ALA50 are all hydrophobic amino acid residues, which can effectively interact with inhibitors through hydrophobic interaction (see Fig. 5). Two large yellow polyhedra surrounding P2 and P4 positions indicate hydrophobic groups are beneficial to enhance the activity, for these two positions can interact with ALA20 and MET45, ALA46, respectively, which is consistent with the result of Adams et al.⁴⁹ Two white isopleths around the P1 position and P4 position of MG167 reveal that hydrophilic groups at these areas are beneficial to the activities. These groups are close to the hydrophilic residues LYS33 and GLN53 of $\beta 5$ subunit. The hydrophobic and hydrophilic characteristics of the amino acid residues in the binding pocket of $\beta 5$ subunit of 20S proteasome concord well with those of the P1, P2, and P4 positions in the inhibitors.

The graphical interpretation of the hydrogen-bond donor interaction in the CoMSIA model is represented in Figure 7C. It highlights areas beyond the molecules where putative hydrogen acceptor groups in the enzyme can form H-bonds with the molecule thereby influencing binding affinities. The cyan contours represent locations where a hydrogen-bond acceptor on the receptor will improve activity, that is, hydrogen-bond donors in the ligand directing to these regions are favorable. For example, there are two cyan areas: one surrounds the hydrogen atoms of NH group attached to the carbobenzyloxy, which can form a hydrogen bond with the residues in the neighboring subunit, the other is beside the hydrogen atom of the NH group between P2 and P3 position forming a 2.83 Å strong hydrogen bond with the

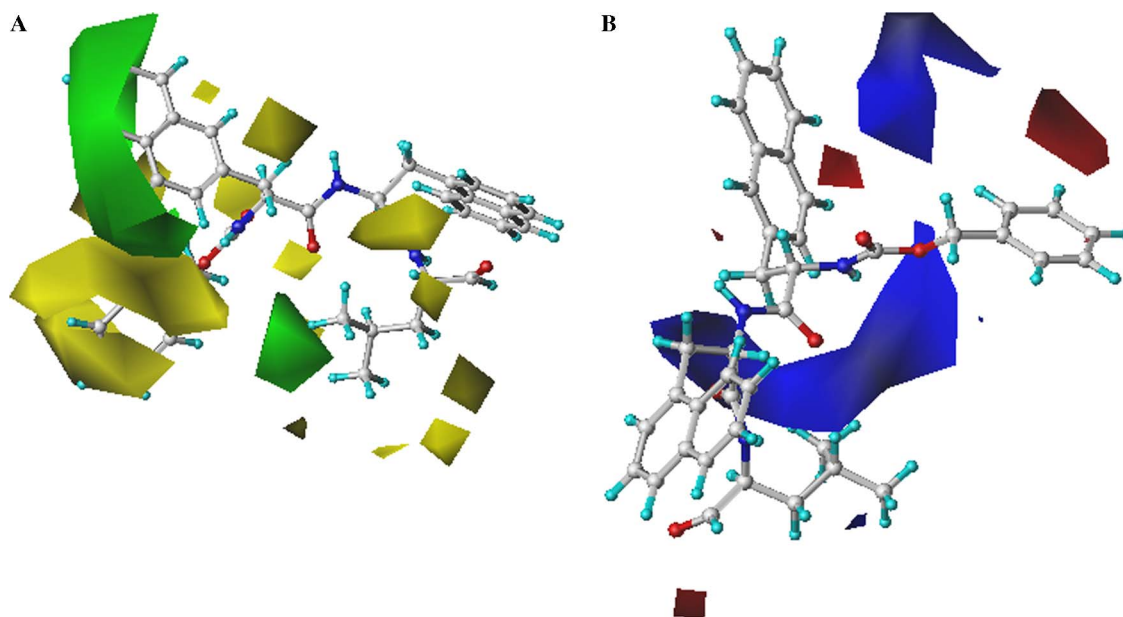


Figure 6. Std * coeff contour maps of final CoMFA analysis with 2 Å grid spacing in combination with MG167. (A) Steric contour map. Green contours (80% contribution) refer to sterically favored regions; yellow contours (20% contribution) indicate sterically disfavored areas. (B) Electrostatic contour map. Blue contours (80% contribution) refer to regions where positively charged substituents are favored; red contours (20% contribution) indicate regions where negatively charged substituents are favored.

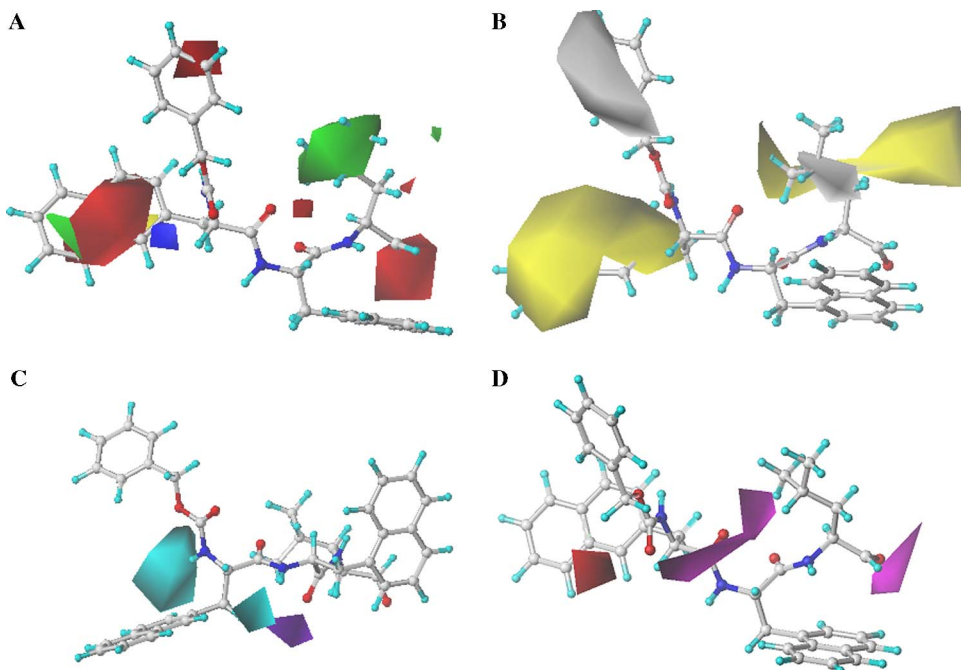


Figure 7. Std * coeff contour maps of final CoMSIA analysis with 2 Å grid spacing in combination with MG167. (A) Steric and electrostatic contour map. Green contours (0.017 level) refer to sterically favored regions; yellow contours (−0.01 level) indicate disfavored areas. Blue contours (0.017 level) refer to regions where negatively charged substituents are disfavored; red contours (−0.01 level) indicate regions where negatively charged substituents are favored. (B) Hydrophobic contour map. White contours (0.017 level) refer to regions where hydrophilic substituents are favored; yellow contours (−0.01 level) indicate regions where hydrophobic substituents are favored. (C) Hydrogen-bond donor contour map. Cyan contours (0.017 level) indicate regions where hydrogen-bond acceptor groups on the receptor increase activity. Purple contours (−0.01 level) refer to regions where hydrogen-bond acceptor groups on the receptor decrease activity. (D) Hydrogen-bond acceptor contour map. Magenta contours (85% contribution) encompass regions where hydrogen-bond donors on the receptor are expected. Red contours (15% contribution) refer to areas where hydrogen-bond donors on the receptor decrease the affinity.

carbonyl group of THR21 (see Fig. 5), which indicates the necessity of the hydrogen atom at this position for activity. Due to the cyclization of the N atom with adjacent C atom in MG169 and MG176, the binding affinities reduce greatly to 700 and 71 nM. Conversely, a purple polyhedron below the carbonyl oxygen between P3 and P4 position is indicative of a disfavored hydrogen-bond donor region, which can be proved by a 3.02 Å H-bond between its carbonyl and NH group of ALA20 from another profile.

Figure 7D indicates areas where hydrogen-bond acceptors on the ligand promote or decrease binding affinities. 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 red polyhedron beside the carbonyl oxygen of P2 position suggests that the subunit bearing hydrogen donors will increase the activity. A large magenta isopleth between two carbonyls from P2 to P4 position indicates that hydrogen-bond donors on the receptor promote activity. From the crystal structure, it is clear that two hydrogen bonds formed between these two carbonyls and NH groups of THR21 and GLY47 with the length of 3.02 and 2.74 Å (Fig. 5). Another large polyhedron surrounding the oxygen atom of the aldehyde group reveals that the hydrogen-bond donor on the receptor stabilizes the hemiacetal and pro-

notes the activity. The aldehyde group formed a 3.04 Å hydrogen bond with amide NH of GLY47 (Fig. 5).

In general, the CoMFA and CoMSIA contour maps agree well with the structural characteristics of the binding pocket of β_5 subunit of 20S proteasome in P1, P2, and P4 positions. However, both in CoMFA and CoMSIA contour maps, no information was given at P3 position, which suggests that this position is not in contact with any amino acid residue in the active part of β_5 subunit but perhaps only contributes to the ADME process. It was also proved by Groll et al.²⁵

3. Conclusions

Though tripeptide aldehydes are highly flexible, we obtained 3D-QSAR models with statistical significance and predictive abilities by using CoMFA and CoMSIA methods. And the results can quite well accord with the actual interaction mode of inhibitor MG101 with β_5 subunit of 20S proteasome. Overall, the predictive power of the CoMFA models appeared to be better than that of the CoMSIA models. However, a combined application of the CoMFA and the CoMSIA models may be most suitable to predict the binding affinity of new tripeptide aldehyde inhibitors. The CoMFA and CoMSIA steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor

potential fields provide insights into important properties at given positions that are crucial for activities.

4. Materials and methods

4.1. Data set

Biological data are prone to variation between laboratories and even between studies within the same laboratory. As a result, 3D-QSAR models built with mixing biological data are considered risky and unreliable. Though a lot of aldehyde inhibitors were synthesized, their biological activities were assessed by quite different methods. To avoid this experimental bias, we used the 55 tripeptide aldehydes from the work of Stain,⁵⁰ since the biological activities of these compounds were measured under identical conditions. The K_i values of the 55 molecules were transformed to pK_i , and the molecules were randomly divided into two groups containing 47 and 8 molecules as the training set and the test set, respectively. The training set of the 47 molecules (Table 1) was represented by a combination of highly active (K_i 0.015–50 nM), moderately active (K_i 50–500 nM), and less active (K_i 500–4900 nM) molecules, and the K_i values of the test set molecules (Table 2) were within the same distribution range as those of the training set molecules. Structural variations of the parent structure, present in all molecules, were allowed at P1, P2, P3, and P4 positions (Tables 1 and 2).

4.2. Molecular conformation

All molecular modeling including CoMFA, CoMSIA analyses was performed on a SGI O₂ workstation using SYBYL version 6.91 software (Tripos Inc., St. Louis, MO).

The conformation of a template molecule is one of the important parameters in CoMFA and CoMSIA methods. In this paper, all the tripeptide aldehydes are highly flexible, therefore, compared with rigid molecules, it is more difficult to obtain meaningful CoMFA and CoMSIA models. In 1997, Huber et al.²⁵ published the crystal structure of 20S proteasome (PDB 1RYP), however, it is not a complex structure with any inhibitors. Due to the lack of structural data for the aldehyde inhibitor–proteasome complex, a systematic search of the most active compound MG167 was carried out in SYBYL. The search was performed by using a rotation increment of 30° for all the torsional angles. Since the number of conformations generated was quite large, only those conformers within an energy threshold value of 5 kcal/mol of the local minimized structure were saved. The resulted global minimum conformer of MG167 was used as a template to build the remaining compounds. Fortunately, soon we acquired the latest crystal structural data of 20S proteasome combined with MG101 that was kindly provided by Prof. Huber on our request.⁵¹ The root-mean-square distance (RMSD) between the crystal structure and the conformer of MG101 generated based on the lowest-energy conformer of MG167 was 0.242 Å.

These two conformations are considered highly similar. So we took the crystal structure of MG101 as the starting conformation (Fig. 1) and the rest of the molecules were created from it using the ‘sketch molecule’ function in SYBYL. A constrained minimization followed by full minimization was carried out on these molecules in order to prevent the conformations from moving to false regions. Tripos force field^{52–54} with a distance-dependent dielectric and the Powell conjugate-gradient algorithm with a convergence criterion of 0.05 kcal/mol were used. Partial atomic charges were calculated using the GH method.^{55,56}

4.3. Alignment

Initially we attempted to use all the 55 molecules to build the 3D-QSAR models. The common five atoms of all the 55 molecules labeled on MG101 in Figure 2A were used as the reference atoms to align all the molecules using the ‘align database’ option in SYBYL. But the partial least-squares (PLS) results showed it produced poorly predictive models, with cross-validated q^2 of 0.516 and 0.463 for CoMFA and CoMSIA, respectively. In fact, MG216 and MG217 are quite different from the other molecules, as their peptide backbones are longer than other molecules. So we removed the two molecules from the dataset. Since the remaining 53 molecules share the common tripeptide aldehyde scaffold and have 9 common atoms (labeled in Fig. 2B), we aligned all the 53 molecules using these 9 reference atoms and the ‘align database’ option in SYBYL and obtained statistically significant models with cross-validated q^2 of 0.615 and 0.591 for CoMFA and CoMSIA, respectively. The final models were built with the 53 molecules, among which 45 molecules (listed in Table 1, excluding MG216 and MG217) were taken as the training set and the other eight molecules (listed in Table 2) were selected as the test set. The structural diversity of the aligned molecules is shown in Figure 3.

4.4. CoMFA studies

All the molecules were placed in a 3D lattice with regular grid points separated by 2 Å. The van der Waals potentials and Coulombic term representing the steric and electrostatic fields were calculated using standard Tripos force field for CoMFA. A C_{sp3} atom with a formal charge of +1 and a van der Waals radius of 1.52 Å served as probe atom to generate steric (Lennard-Jones 6-12 potential) and electrostatic (Coulombic potential) field energies, which were obtained by summing the individual interaction energies between each atom of the molecule and the probe atom at every grid point. A distance-dependent dielectric constant was used. The steric and electrostatic fields were truncated at ±30.00 kcal/mol.

4.5. CoMSIA studies

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, hydro-

gen-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 0.3.⁵⁷

4.6. PLS analysis and validations

The PLS technique was employed to generate a linear relationship that correlates changes in the various computed potential fields with changes in the corresponding experimental values of the binding affinities (pK_i) for the data set of the inhibitors. Employing the CoMFA and CoMSIA potential energy fields for each molecule as the independent variable and the corresponding pK_i value as the dependent variable, PLS converts these potential energy field descriptors to the so-called latent variables or principal components (PCs) that consist of linear combinations of the original independent variables.

To assess the internal predictive ability of the CoMFA and CoMSIA models, we employ the 'leave-one-out' (LOO) cross-validation procedures. In this procedure, each data set or compound is excluded one at a time, and then its activity is predicted by the model constructed from the remaining compounds in the data set. Cross-validation determines the optimum number of PCs, corresponding to the smallest error of prediction and the highest cross-validated q^2 . PLS analysis was repeated without validation using the optimum number of PCs to generate final CoMFA and CoMSIA models from which the conventional correlation coefficient r^2 , a measure of the internal consistency of the models, was derived. To improve efficiency and reduce 'noise,' a minimum filter value σ of 1.00 kcal/mol was used.

4.7. Predictive correlation coefficient

The predictive abilities of the 3D-QSAR models were determined by predicting the test set of eight compounds that were not included in the training set. These molecules were aligned in the same way as those in the training set. The predictive correlation coefficient (r_{pred}^2), based on the molecules of the test set, is defined as

$$r_{\text{pred}}^2 = (\text{SD} - \text{PRESS})/\text{SD},$$

where SD is the sum of squared deviations between the biological activities of the test set and mean activity of the training set molecules, and PRESS is the sum of squared deviation between predicted and actual activity values for every molecule in the test set.

Acknowledgments

We thank Prof. Huber and Mr. Groll kindly offered the coordinates of MG101–20S proteasome complex.

References and notes

- Hershko, A. *Trends Biochem. Sci.* **1996**, *21*, 445.
- Etlinger, J. D.; Goldberg, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 54.
- Rock, K. L.; Goldberg, A. L. *Annu. Rev. Immunol.* **1999**, *17*, 739.
- Prösch, S.; Priemer, C.; Höflich, C.; Liebenthal, C.; Babel, N.; Krüger, D. H.; Volk, H.-D. *Antiviral Ther.* **2003**, *8*, 555.
- Mitch, W. E.; Goldberg, A. L. *N. Eng. J. Med.* **1996**, *335*, 1897.
- Hampton, R. Y.; Gardner, R. G.; Rine, J. *Mol. Biol. Cell* **1996**, *7*, 2029.
- Murakami, Y.; Matsufuji, S.; Kameji, T.; Hayashi, S.; Igarashi, K.; Tamura, T.; Tanaka, K.; Ichihara, A. *Nature* **1992**, *360*, 597.
- Turner, G. C.; Du, F.; Varshavsky, A. *Nature* **2000**, *405*, 579.
- Naidoo, N.; Song, W.; Hunter-Ensor, M.; Sehgal, A. *Science* **1999**, *285*, 1737.
- Chain, D. G.; Schwartz, J. H.; Hegde, A. N. *Mol. Neurobiol.* **1999**, *20*, 125.
- Alves-Rodrigues, A.; Gregori, L.; Figueiredo-Pereira, M. E. *Trends Neurosci.* **1998**, *21*, 516.
- Sakamoto, K. M. *Mol. Genet. Metab.* **2002**, *77*, 44.
- Hideshima, T.; Richardson, P.; Chauhan, D.; Palombella, V. J.; Elliott, P. J.; Adams, J.; Anderson, K. C. *Cancer Res.* **2001**, *61*, 3071.
- Sun, J. Z.; Nam, S.; Lee, C. S.; Li, B. Y.; Coppola, D.; Hamilton, A. D.; Dou, Q. P.; Sebt, S. M. *Cancer Res.* **2001**, *61*, 1280.
- Keck, S.; Nitsch, R.; Grune, T.; Ullrich, O. *J. Neurochem.* **2003**, *85*, 115.
- Rodgers, K. J.; Dean, R. T. *Int. J. Biochem. Cell Biol.* **2003**, *35*, 716.
- Merforth, S.; Kuehn, L.; Osmer, A.; Dahlmann, B. *Int. J. Biochem. Cell Biol.* **2003**, *35*, 740.
- Palombella, V. J.; Rando, O. J.; Goldberg, A. L.; Maniatis, T. *Cell* **1994**, *78*, 773.
- Figueiredo-Pereira, M. E.; Berg, K. A.; Wilk, S. J. *Neurochem.* **1994**, *63*, 1578.
- Rock, K. L.; Gramm, C. F.; Rothstein, L.; Clark, K.; Stein, R.; Dick, L.; Hwang, D.; Goldberg, A. L. *Cell* **1994**, *78*, 761.
- Craiu, A.; Gaczynska, M.; Akopian, T.; Gramm, C. F.; Fenteany, G.; Goldberg, A. L.; Rock, K. L. *J. Biol. Chem.* **1997**, *272*, 13437.
- Lee, D. H.; Goldberg, A. L. *Trends Cell Biol.* **1998**, *8*, 397.
- Arrigo, A. P.; Tanaka, K.; Goldberg, A. L.; Welch, W. J. *Nature* **1988**, *331*, 192.
- Coux, O.; Tanaka, K.; Goldberg, A. L. *Annu. Rev. Biochem.* **1996**, *65*, 801.
- Groll, M.; Ditzel, L.; Löwe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. *Nature* **1997**, *386*, 463.
- Cardozo, C.; Michaud, C.; Orłowski, M. *Biochemistry* **1999**, *38*, 9768.
- Dick, T. P.; Nussbaum, A. K.; Deeg, M.; Heinemeyeri, W.; Groll, M.; Schirle, M.; Keilholz, W.; Stevanović, S.; Wolfi, D. H.; Huber, R.; Rammensee, H. G.; Schild, H. J. *Biol. Chem.* **1998**, *273*, 25637.
- Orłowski, M. *Biochemistry* **1990**, *29*, 10289.
- Rivett, A. *Biochem. J.* **1993**, *291*, 1.
- Voges, D.; Zwickl, P.; Baumeister, W. *Annu. Rev. Biochem.* **1999**, *68*, 1015.
- Löwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. *Science* **1995**, *268*, 533.
- Larsen, C. N.; Finley, D. *Cell* **1997**, *91*, 431.
- Baumeister, W.; Walz, J.; Zühl, F.; Seemüller, E. *Cell* **1998**, *92*, 367.
- Brannigan, J. A.; Dodson, G.; Duggleby, H. J.; Moody, P. C.; Smith, J. L.; Tomchick, D. R.; Murzin, A. G. *Nature* **1995**, *378*, 416.
- Wilk, S.; Orłowski, M. *J. Neurochem.* **1980**, *35*, 1172.

36. Orłowski, M.; Wilk, S. *Biochem. Biophys. Res. Commun.* **1981**, *101*, 814.
37. Kisselev, A. F.; Akopian, T. N.; Woo, K. M.; Goldberg, A. L. *J. Biol. Chem.* **1999**, *274*, 3363.
38. Nussbaum, A. K.; Dick, T. P.; Keilholz, W.; Schirle, M.; Stevanovic, S.; Dietz, K.; Heinemeyer, W.; Groll, M.; Wolf, D. H.; Huber, R.; Rammensee, H.-G.; Schild, H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12504.
39. Delcros, J. G.; Baudy Floc'h, M.; Prigent, C.; Arlot-Bonnemains, Y. *Curr. Med. Chem.* **2003**, *10*, 479.
40. Adams, J. *Trends Mol. Med.* **2002**, *8*(Suppl.), S49.
41. Adams, J.; Palombella, V. J.; Sausville, E. A.; Johnson, J.; Destree, A.; Lazarus, D. D.; Maas, J.; Pien, C. S.; Prakash, S.; Elliott, P. *J. Cancer Res.* **1999**, *59*, 2615.
42. Pham, L.; Tamayo, A.; Lo, P.; Yoshimura, L.; Ford, R. J. *Blood* **2001**, *98*, 465a.
43. Vinitsky, A.; Michaud, C.; Powers, J. C.; Orłowski, M. *Biochemistry* **1992**, *31*, 9421.
44. Kisselev, A. F.; Goldberg, A. L. *Chem. Biol.* **2001**, *8*, 739.
45. Lee, D. H.; Goldberg, A. L. *J. Biol. Chem.* **1996**, *271*, 27280.
46. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
47. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
48. Kubinyi, H. *Drug Discovery Today* **1997**, *2*, 457.
49. Adams, J.; Behnke, M.; Chen, S. W.; Cruickshank, A. A.; Dick, L. R.; Grenie, L.; Klunder, J. M.; Ma, Y. T.; Plamondon, L.; Stein, R. L. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 333.
50. Stain, R. L. U.S. Patent 5,693,617, 1997; *Chem. Abstr.* **1997**, *128*, 48498.
51. Personal communication with Prof. Huber. The crystal structure is not publicly published at the time this paper was written.
52. Sybyl Theory Manual, 1988, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO 63144.
53. Sybyl Molecular Spreadsheet Manual, 1988, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO 63144.
54. Sybyl Tutorial Manual, 1988, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO 63144.
55. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
56. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
57. Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458.