



Original article

3D-QSAR studies of boron-containing dipeptides as proteasome inhibitors with CoMFA and CoMSIA methods

Yong-Qiang Zhu^{a,*}, Meng Lei^b, Ai-Jun Lu^a, Xin Zhao^a, Xiao-Jin Yin^a, Qing-Zhi Gao^{c,**}^a Jiangsu Simcere Pharmaceutical Research Institute, No. 699-18 Xuan Wu Avenue, Xuan Wu District, Nanjing 210042, PR China^b School of Sciences, Nanjing Forestry University, Nanjing 210037, PR China^c Chemistry Department, XenoPort Inc., 3410 Central Expressway, Santa Clara, CA 95051, United States

ARTICLE INFO

Article history:

Received 13 June 2008

Received in revised form 9 July 2008

Accepted 11 July 2008

Available online 24 July 2008

Keywords:

3D-QSAR

CoMFA

CoMSIA

Proteasome inhibitors

Dipeptide boronate

Drug design

ABSTRACT

Three-dimensional quantitative structure–activity relationship (3D-QSAR) studies were performed for a series of dipeptide boronate proteasome inhibitors using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) techniques. A training set containing 46 molecules served to establish the models. The optimum CoMFA and CoMSIA models obtained for the training set were all statistically significant with cross-validated coefficients (q^2) of 0.676 and 0.630 and conventional coefficients (r^2) of 0.989 and 0.956, respectively. The predictive capacities of both models were successfully validated by calculating a test set of 13 molecules that were not included in the training set. The predicted correlation coefficients (r^2_{pred}) of CoMFA and CoMSIA are 0.963 and 0.919, 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 constructed in this paper can be used to guide the development of novel dipeptide boronate inhibitors of 20S proteasome.

© 2008 Elsevier Masson SAS. All rights reserved.

1. Introduction

The ubiquitin–proteasome system (UPS) was firstly identified more than 20 years ago [1,2], and it was found to degrade most long- and short-lived normal and abnormal intracellular proteins [3–5]. In fact, the bulk of proteins in mammalian cells (up to 80–90% of all intracellular proteins) is degraded via the UPS, which is hence considered to be the major pathway of intracellular protein degradation. And growing evidence demonstrated that it plays an essential role in the regulation of many physiological processes [6–11] and in the development of a number of major human diseases [12–18], such as cancer, Alzheimer, Parkinson, diabetes, cardiac diseases, etc. So, further emphasis on this pathway could provide new therapeutic strategies to attack proliferative disorders.

Within the UPS, the 26S proteasome is involved, which is formed from a cylinder-shaped multicatalytic proteinase complex (MPC) [19,20], referred to as the 20S proteasome core particle (CP), capped at each end by a regulatory component termed the 19S complex (regulatory particle, RP or 19S proteasome). X-ray

crystallography of 26S proteasome [21,22] revealed that the 20S proteasome is composed of four stack rings, with each ring consisting of seven α - and β -type subunits, following an $\alpha_1\text{--}\beta_1\text{--}\beta_2\text{--}\alpha_2\text{--}\beta_3\text{--}\beta_4$ stoichiometry. It belongs to the N-terminal nucleophile (Ntn) hydrolase superfamily [23] and is responsible for catalyzing the cleavage of peptide bonds on the carboxyl side in both natural peptides and synthetic substrates (including the falsely folded proteins) [24] with the Thr10 residues located at the $\beta 1$, $\beta 2$ and $\beta 5$ subunits. The 19S proteasome is made up of 18 subunits, which controls the recognition and unfolding of the protein substrates prior to its translocation into the 20S proteasome [25,26].

As stated above, a vast array of biological pathways are controlled by the proteasomes, which are responsible for regulating degradation of many critical proteins. Thus, proteasome inhibitors are promising candidates as anti-tumor and anti-inflammatory drugs. Furthermore, these inhibitors would act as useful research tools for determining the essential roles of the proteasome and UPS in various physiological processes. Due to the promising therapeutic potential of proteasome inhibition, much attention has been paid to design, synthesize and biologically evaluate various proteasome inhibitors [27]. Of all the proteasome inhibitors, it is noteworthy to mention a dipeptide boronic acid analog, Velcade® (also named bortezomib, MG341 and PS341), which was approved by the FDA in May 2003, the EMEA in April 2004 and Japan in October 2006, respectively, for the treatment of multiple myeloma (MM) patients who have received one prior

* Corresponding author. Tel.: +86 025 85566666x1726; fax: +86 025 85566666x1835.

** Corresponding author. Tel.: +1 408 616 7356; fax: +1 408 616 7210.

E-mail addresses: zhuyongqiang@simcere.com (Y.-Q. Zhu), qgao@xenoport.com (Q.-Z. Gao).

therapy but failed [28]. On December 8, 2006, the FDA granted full approval of Velcade® for the treatment of patients with relapsed mantle cell lymphoma (MCL). And now it is also under investigation in Phase I, Phase II and Phase III studies for a wide variety of haematological malignancies and solid tumors, including non-Hodgkin's lymphoma, prostate, breast and non-small-cell lung cancers [29–31]. This drug sets the milestone of the development of proteasome inhibitors: it is the first marketed drug of the proteasome inhibitors and also the first marketed peptide compound containing B atom. Such kind of compounds covalently inhibit the proteasome by reacting with the hydroxyl group of the N-terminal threonines of the β subunits to form a reversible tetrahedral adduct, which has been confirmed by X-ray diffraction of complex of MG341–20S proteasome [32]. Compared with other proteasome inhibitors, such as peptide aldehyde, peptide boronates are much more selective due to their weaker interactions with thiol proteases and serine proteases [33]. However, recent clinical data showed some toxic side-effects of Velcade®, including fatigue, nausea, sensory neuropathy, etc. [34]. So how to decrease or eliminate the toxicities and design more potent and specific boronate inhibitors is becoming an urgent task. Though many different peptide boronate inhibitors have been synthesized and experimentally assessed, to the best of our knowledge, no QSAR or 3D-QSAR (three-dimensional quantitative structure–activity relationship) investigations for such kind of compounds has been reported to date. As a continuation of our program on proteasome inhibitors [35], this study aims to build predictive 3D-QSAR models by using comparative molecule field analysis (CoMFA) and comparative molecule similarity indices analysis (CoMSIA) methods, and using them to find the correlation between the structure of dipeptide boronates and activities to 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 statistical techniques [36]. The CoMFA methodology was first introduced by Cramer et al. [37] and CoMSIA was an extension of the CoMFA method reported by Klebe et al. [38]. 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.

So far, carborane-based drug design utilizing computer-aided molecular design (CAMD) has been applied only in relatively few cases [39–43]. This is mainly due to the complex structures of carboranes with sixfold coordinated carbon and boron atoms. In addition, CAMD approaches containing boron atoms are not supported by most commercially available software packages because they do not provide the required empirical potential energy functions for the boron atom in default settings [44]. This paper presents the first study using a ligand-based 3D-QSAR computational approach dealing with peptide boronate analogs as proteasome inhibitors.

2. Materials and methods

2.1. Data set for analysis

Though a lot of peptide boronate inhibitors were synthesized, their biological activities were assessed by quite different methods. As a result, 3D-QSAR models built with mixing biological data are considered risky and unreliable. To avoid this experimental bias, we selected 59 dipeptide boronates from the work of Adams et al. [45], since the biological activities of these compounds were measured under identical conditions. And the remaining molecules that do not share the common scaffold with dipeptide boronates were discarded. The K_i values of the 59

molecules were converted to pK_i ($-\log K_i$), and as a customary for CoMFA and CoMSIA, the data set was divided into a training set (46 molecules) for final model development and a test set (13 molecules) for model validation. Selection of the test-set compounds was conducted randomly, except that some bias was applied to ensure adequate coverage in terms of binding affinity and structural diversity. Structural variations of the parent structure, present in all molecules, were allowed at P1, P2 and P3 positions (Tables 1 and 2).

2.2. Molecular conformation

All molecular modeling including CoMFA and CoMSIA was performed using SYBYL software package (Tripos Inc., St. Louis, USA, version 7.3) running on Red Hat Enterprise Linux workstation.

The conformation of a template molecule is one of the most important parameters in CoMFA and CoMSIA methods. Generally speaking, a crystal structure is preferred to be selected as template molecule especially for flexible molecules. In this paper, all the dipeptide boronates are highly flexible, therefore, compared with rigid molecules, it is more difficult to obtain meaningful CoMFA and CoMSIA models. To our pleasure, Groll et al. [32] published the crystal structure of MG341–20S proteasome complex (PDB 2F16) in 2006. So we extracted MG341 crystal structure from the complex and took it as the starting conformation (Fig. 1a). For the absence of the empirical potential energy function of B atom in SYBYL 7.3, so in this paper, we adopted the same strategy reported in the literatures [46,47] and changed the boron atom 'B' in MG341 to carbon atom 'C.3' using the Built/Edit option in SYBYL and colored it with 'Magenta'. (For interpretation of the references to colour in this text, the reader is referred to the web version of this article.) 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 moving to false regions. Tripos force field [48–50] 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 Gasteiger–Hückel (GH) method [51,52].

2.3. Alignment procedure

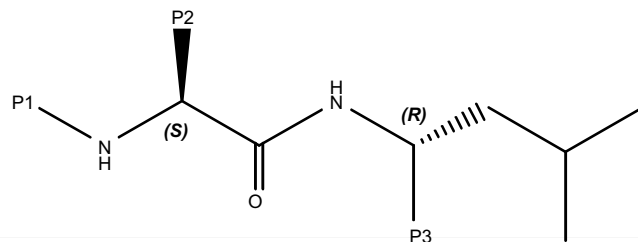
The most crucial input for CoMFA is the alignment of the molecules. The template molecule MG341 was taken and the common 12 atoms labeled with numbers on MG341 in Fig. 1b were used as the reference atoms to align all the molecules using the 'DATABASE ALIGNMENT' option in SYBYL. The aligned molecules are shown in Fig. 2.

2.4. CoMFA studies

All the molecules were placed in a 3D regular lattice (2 Å spacing) extending at least 2 Å beyond the volumes of all investigated molecules on all axes. 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. The cutoff values for both steric and electrostatic fields were set to 30.00 kcal/mol with a distance-dependent dielectric constant.


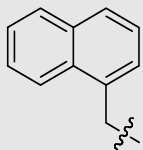

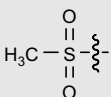
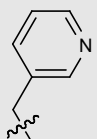


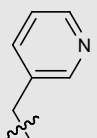

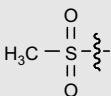
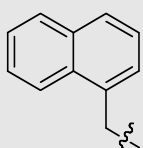
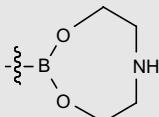
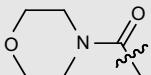
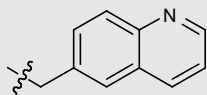

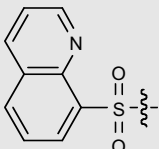
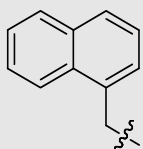

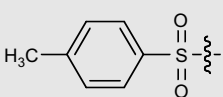
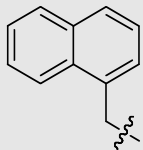

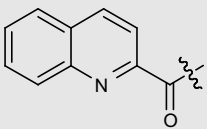
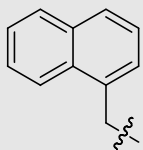

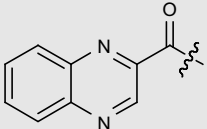
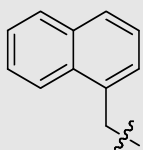

Table 1

The structure of 46 inhibitors in the training set and their actual activities



Compound	P1	P2	P3	p <i>K</i> _i
MG267				10.00
MG268				6.00
MG270				10.08
MG273				9.74
MG274				8.52
MG285				8.52
MG286				9.82

Table 1 (continued)

Compound	P1	P2	P3	pK _i
MG287				9.77
MG289				8.20
MG290				8.27
MG291				9.55
MG292				8.22
MG296				8.77
MG297				9.77
MG298				10.12
MG299				9.85

(continued on next page)

Table 1 (continued)

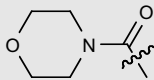
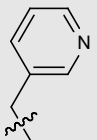

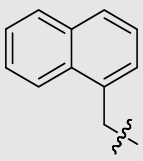

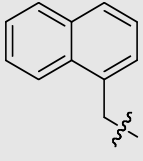

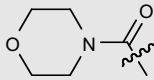
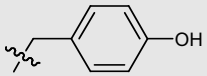

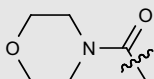
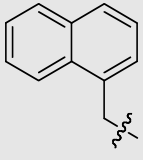

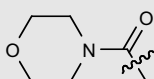
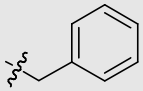

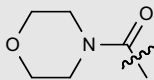
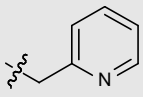

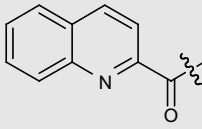
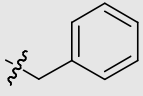

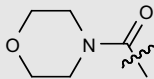
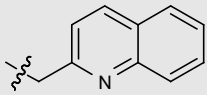

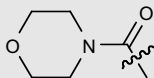
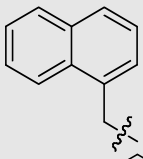

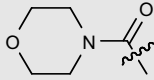
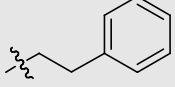

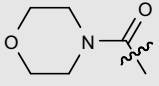
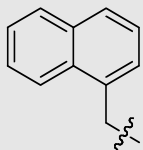

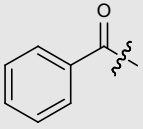
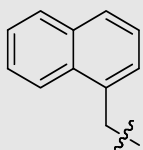

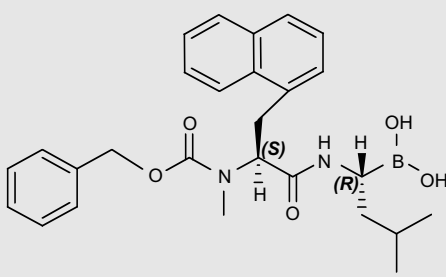
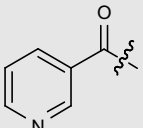
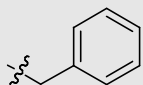

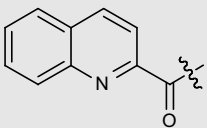
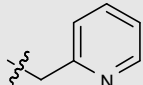

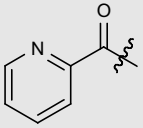
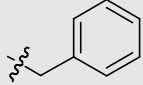

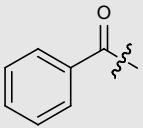
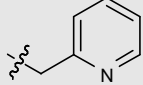

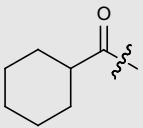
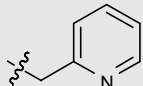

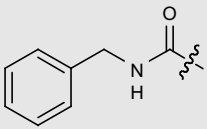
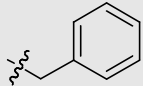

Compound	P1	P2	P3	p <i>K_i</i>
MG300				8.89
MG302	H			8.12
MG303	H·HCl			8.41
MG307				9.29
MG308				9.14
MG309				9.09
MG312				8.20
MG314				9.72
MG315				8.66
MG318				6.19
MG322				8.66

Table 1 (continued)

Compound	P1	P2	P3	pK _i
MG324				8.70
MG328				10.06
MG332				9.02
MG336				9.60
MG338				8.85
MG343				9.38
MG345				9.12
MG346				8.96
MG350				9.85

(continued on next page)

Table 1 (continued)

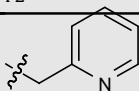

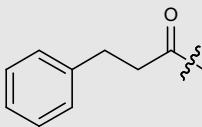
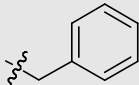

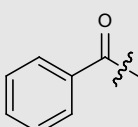
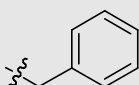

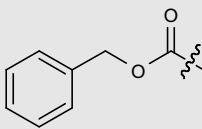
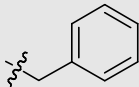

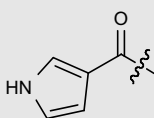
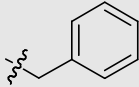

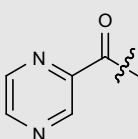
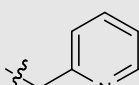

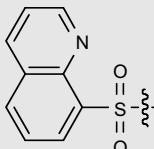
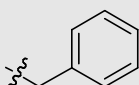

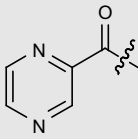
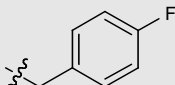

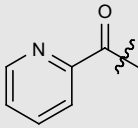
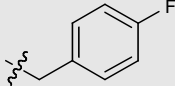

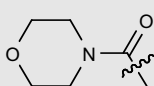
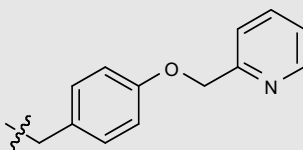

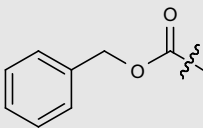


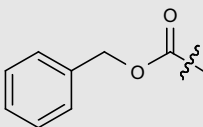
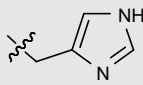

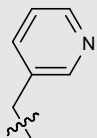

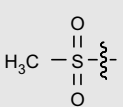
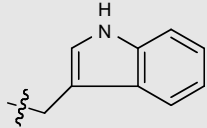

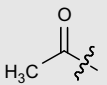
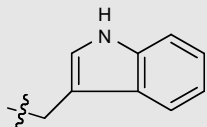

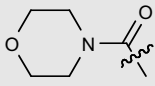
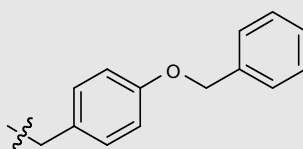

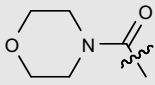
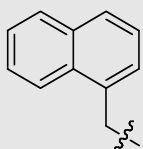

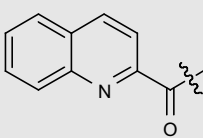
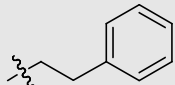

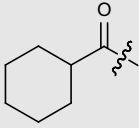
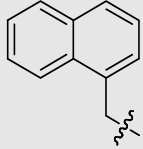

Compound	P1	P2	P3	pK _i
MG351	H			7.50
MG352				9.82
MG353				9.82
MG356				9.89
MG361				9.85
MG368				8.15
MG380				8.36
MG382				9.02
MG383				9.08
MG387				9.70

Table 2

The structure of 13 inhibitors in the test set and their actual activities

Compound	P1	P2	P3	pK _i
MG264				6.92
MG282				7.60
MG294	H			6.82
MG295				8.24
MG301				8.89
MG306				9.64
MG317				7.00
MG325				8.55
MG329				10.52

(continued on next page)

Table 2 (continued)

Compound	P1	P2	P3	pK _i
MG334				8.96
MG341				9.22
MG347				7.54
MG358				9.77

2.5. CoMSIA studies

The aligned molecules were placed in a 3D lattice with regular grid points separated by 2 Å similar to those of CoMFA studies. CoMSIA calculates also steric and electrostatic fields, and in addition hydrophobic, hydrogen-bond donor and acceptor fields and it uses Gaussian equations for field calculation that do not require cutoff values. The analysis was carried out 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 [53].

2.6. Partial Least Square (PLS) analysis and validations

CoMFA and CoMSIA descriptors were used as independent variables and the corresponding pK_i values as the dependent variables in the PLS regression analyses for the development of 3D-QSAR models. SAMPLS (SAMple distance PLS) analysis was performed using the cross-validated 'Leave-One-Out' (LOO) option

to determine the optimum number of components to be used in the final non-cross-validated analysis. The optimum number of components produces the smallest root mean predictive sum of squared errors, which correspond to the highest cross-validated coefficient (q^2). Then the non-cross-validated analysis was carried out using the optimum number of components to give the conventional correlation coefficients (r^2), with the column filtration σ set to 2.0.

2.7. Predictive r^2 values

The predictive ability of each analysis was determined from the test set of 13 compounds that were not included in the training set. These molecules were aligned in the same way as those in the training set and their activities were predicted by each PLS analysis. The predictive correlation coefficient (r^2_{pred}) is defined as

$$r^2_{\text{pred}} = (\text{SD} - \text{PRESS})/\text{SD}$$

where SD is the sum of squared deviations between the biological

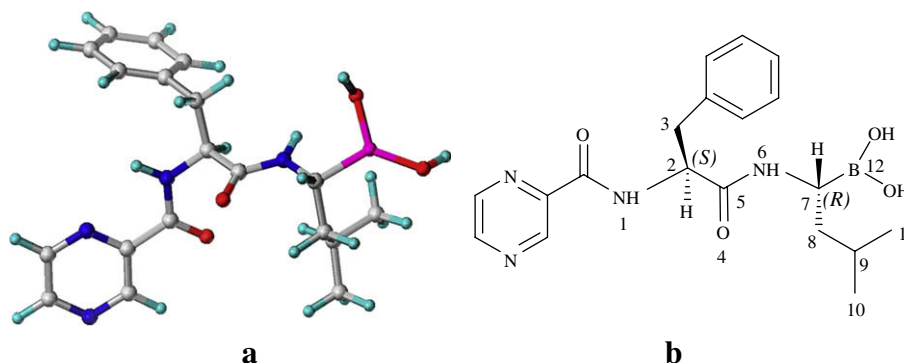


Fig. 1. Extracted crystal structure of MG341.

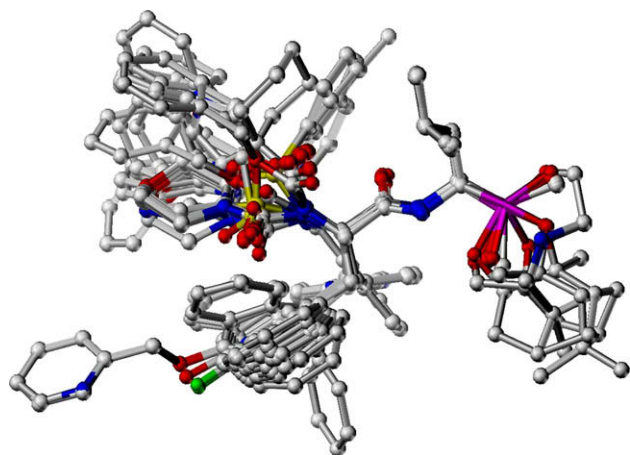


Fig. 2. Alignment of dipeptide boronate proteasome inhibitors in the training set. Hydrogens are not shown.

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.

3. Results and discussion

3.1. Results of the CoMFA

The statistical results obtained from standard CoMFA models constructed with steric and electrostatic fields are summarized in Table 3. The q^2 , S_{PRESS} , r^2 , F and s values were computed as defined in SYBYL. The optimum number of components (13) was determined by SAMPLS analysis implemented in SYBYL 7.3 with an LOO cross-validated q^2 of 0.676, which indicated a good predictive capacity of the model ($q^2 > 0.5$). A high correlation coefficient (r^2) of 0.989 for the non-cross-validated final model showed the self-consistency of the model ($r^2 > 0.9$). In both steric and electrostatic field contributions, the former accounts for 0.491, while the latter contributes 0.509, indicating that these two factors nearly contribute the same to the binding affinities.

3.2. Results of the CoMSIA

As stated above, CoMSIA not only offered steric and electrostatic field information as CoMFA but also gave hydrophobic, hydrogen-bond donor and acceptor field information, whereas these three factors are always important to the binding affinities. Thus different

combination of the three fields with steric and electrostatic fields may result in different models. So, in this paper eight models were built by varying these five fields (Table 3). The optimum component of each model was determined by SAMPLS analysis implemented in SYBYL 7.3 with maximum q^2 value. The model developed by using only steric and electrostatic fields gave a poor cross-validated q^2 of 0.332 (< 0.5) with five variables. Addition of a third field such as hydrophobic, hydrogen-bond donor or acceptor to this model did not obtain a statistical significant model. But among these three 3-field models, the model including hydrophobic field (SEH model) gave the best result (q^2 of 0.431 versus 0.338 and 0.386). However, both the SEHD and SEHA models with the combination of four fields produced significant models (q^2 of 0.524 and 0.511, respectively; r^2 of 0.935 and 0.930, respectively) except SEDA model (no hydrophobic field, $q^2 = 0.364$). So a conclusion can be drawn that hydrophobic field plays a key role in building the models, which is also explained by the greatest field contributions of hydrophobic field in each model. The best 10-component model with the maximum q^2 value consisted of all the five fields as the descriptors, which gave the highest cross-validated coefficient q^2 of 0.630, thereby indicating the most powerful predictive capacity. The highest conventional correlation coefficient r^2 of 0.956 was also obtained for this model, showing the strong internal consistency and therefore, this model was chosen for the final analysis. The standard error of estimate and the F -test values of this model were 0.053 and 752.3, respectively.

3.3. 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 13 inhibitors (Table 2) that were not included in the training set were used as a test set. Table 4 summarizes the predicted results obtained from the CoMFA and CoMSIA models. The plots of predicted versus actual binding affinities for the test set inhibitors are shown in Fig. 3a and b, which represent models based on CoMFA and CoMSIA, respectively. By comparison of the experimentally observed and theoretically predicted pK_i values of dipeptide boronate inhibitors, it can be seen that both CoMFA and CoMSIA models performed well in the prediction of the activities of the test inhibitors. In almost all the cases, the predicted values were close to the observed pK_i values, deviating by less than 1 logarithm unit in binding affinities to the 20S proteasome. Especially for marketed MG341, both CoMFA and CoMSIA models gave ideally predictive values, both of which are less than 0.2 logarithm units. The predicted correlation coefficients (r^2_{pred}) are 0.963 for the CoMFA model and 0.919 for the CoMSIA model. All results including statistical data from this predictive validation analysis further confirmed the robustness of the

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

PLS statistics	CoMFA	CoMSIA (SE)	CoMSIA (SEH)	CoMSIA (SEA)	CoMSIA (SED)	CoMSIA (SEHD)	CoMSIA (SEDA)	CoMSIA (SEHA)	CoMSIA (ALL)
Components	13	5	3	6	6	8	6	11	10
q^2	0.676	0.332	0.431	0.338	0.386	0.524	0.364	0.511	0.630
S_{PRESS}	0.663	0.838	0.758	0.842	0.826	0.73	0.826	0.763	0.736
r^2	0.989	0.800	0.818	0.828	0.874	0.935	0.865	0.947	0.956
F	187.8	40.7	79.503	40.0	57.9	321.1	53.2	384.4	552.3
s	0.157	0.459	0.429	0.430	0.367	0.143	0.381	0.112	0.153
Field contributions (%)									
S	0.491	0.456	0.236	0.338	0.318	0.216	0.263	0.207	0.160
E	0.509	0.544	0.232	0.376	0.382	0.211	0.274	0.215	0.183
H			0.532		0.300	0.404		0.397	0.321
D						0.170	0.242		0.158
A				0.286			0.220	0.181	0.178

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, A and ALL denote the steric, electrostatic, hydrophobic, hydrogen-bond donor and acceptor and the combination of all the five fields, respectively.

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

Compound	EA (pK _i)	PA (pK _i)		Residuals	
		CoMFA	CoMSIA	CoMFA	CoMSIA
MG264	6.92	6.80	7.23	0.12	−0.31
MG282	7.60	7.81	8.18	−0.21	−0.58
MG294	6.82	6.92	7.17	−0.10	−0.35
MG295	8.24	7.98	7.85	0.26	0.39
MG301	8.89	8.77	8.60	0.12	0.29
MG306	9.64	9.42	9.84	0.22	−0.20
MG317	7.00	6.77	7.38	0.23	−0.38
MG325	8.55	9.03	7.87	−0.48	0.68
MG329	10.52	10.22	10.06	0.30	0.46
MG334	8.96	9.01	8.38	−0.05	0.58
MG341	9.22	9.15	9.10	0.07	0.12
MG347	7.54	8.06	7.75	−0.52	−0.01
MG358	9.77	9.51	9.66	0.26	0.11

EA: Experimental Activity; PA: Predicted Activity.

obtained models in terms of their predictive abilities. There was no significant difference between the CoMFA and CoMSIA models for this predictive validation test.

3.4. CoMFA contour maps analysis

Table 3 shows that the steric fields and electrostatic fields nearly gave the same contribution, accounting for 0.491 and 0.509, respectively, which suggests that both fields are critical 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 and electrostatic interactions. Fig. 4a shows the steric contour map for the CoMFA models with the highly active inhibitor MG314 (pK_i = 9.72) as a reference. Three regions at P1 and P2 positions of MG314 have been identified with green polyhedra, which indicate that bulky substituents at these positions may improve the activities. The quinoline moiety at P1 position of MG314 was surrounded by two green isopleths lying above and below the aromatic ring, respectively. So addition of a bulky group at this position is favorable to the binding affinities. This is indeed the case for MG287 and MG329. Sterically favored cyclohexyl group at P1 position of MG329 increased the pK_i value to 10.52; however, while it was substituted by a comparatively smaller methyl group at P1 position (compound MG287), the pK_i value was decreased to 9.77. Compounds MG302 and MG334 also account for such results. MG302, with the amino group at P1 position substituted by a hydrogen atom, gave a pK_i value of 8.12, but when the same amino group was methylated (compound MG334), the pK_i value was increased to 8.96. Beside the benzyl group at P2 position of the

reference molecule MG314, a green polyhedron appears, which means that sterically favored substituents will improve the biological activities of proteasome inhibitors. The presence of a bulky group 2-naphthyl methylene at P2 position of MG270 showed a higher binding affinity (pK_i = 10.08), while the smaller group benzyl at P2 position of MG356 decreased the pK_i value to 9.89. For compounds MG306 and MG309, the former bearing an additional bulky *p*-benzyloxy on benzyl group at P2 position exhibited more potent activity than the latter (pK_i: 9.64 versus 9.09). However, the CoMFA models did not give any isopleth at P3 position in the contour map.

Electrostatic fields based on the PLS analysis of the CoMFA models are shown in Fig. 4b with MG314 as a reference. A small red isopleth above the quinoline ring of P1 position represents an area where a negative GH charge is favored. Methyl group at P1 position of MG287, bearing positive GH charges on the C atom, decreases the activity (pK_i = 9.77), while benzyloxy group at P1 position of MG270, having negative charges on the O atom and benzyl ring, increases the activity (pK_i = 10.08). At the same time, negatively charged nitrogen atom on the quinoline is necessary for a red favorable isopleth in proximity to this area. Nitrogen atoms in such position usually can form H-bonds with amino acid residues of proteasome. Three small blue polyhedra encompassing the quinoline ring indicate positively charged groups, such as hydrogen atoms, which are beneficial to the activity. A large region of red contour surrounding two oxygen atoms of carbonyl suggests that negatively charged oxygen atoms are necessary to increase activity. At P2 position, a red polyhedron lying below benzyl group indicates substituents with GH negative charges may improve the binding affinities. Positively charged *tert*-butyl group at P2 position of compound MG264 leads to its inactivity (pK_i = 6.92), while negatively charged 2-naphthyl methylene at the same position of MG267 greatly contributes to the activity (pK_i = 10.00). It is noteworthy to point out that at P3 position, positively charged groups linked with two hydroxyls of boronic acid, such as pinane, are beneficial to high activity. Compound MG264 having positive GH charge at P3 position shows a higher activity (pK_i = 6.92), whereas the 'naked' boronic acid at P3 position of compound MG268 decreased the binding affinity (pK_i = 6.00).

3.5. 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.

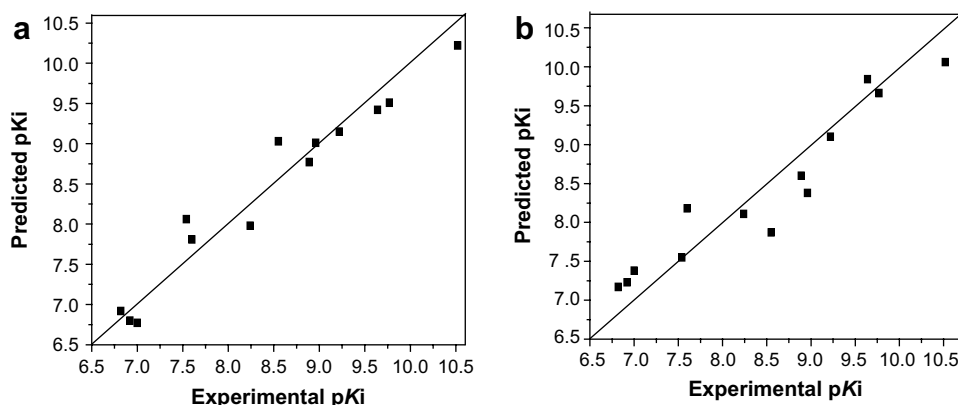


Fig. 3. Predicted versus actual binding affinities for the 13 inhibitors in the test set for (a) CoMFA and (b) CoMSIA models. The predicted correlation coefficients (r^2_{pred}) are 0.963 for the CoMFA model and 0.919 for the CoMSIA model.

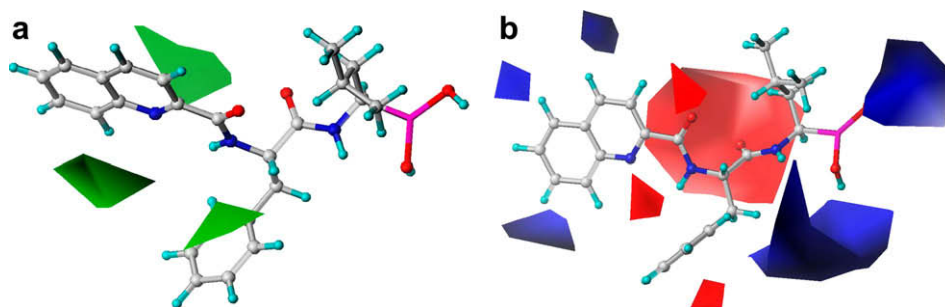


Fig. 4. Std *coeff contour maps of final CoMFA analysis with 2 Å grid spacing in combination with MG314. (a) Steric contour map. Green contours refer to sterically favored regions. (b) Electrostatic contour map. Blue contours (80% contribution) refer to regions where positively charged substituents are favored; red contours (20% contribution) indicate regions where negatively charged substituents are favored. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5 provides graphical representations of the CoMSIA models. To aid in visualization, the highly active inhibitor MG314 was overlaid in the maps once again.

Fig. 5a shows the CoMSIA steric field contour map. Green isopleths indicate regions where more steric substituents will enhance the activities. Just as discussed in CoMFA contour maps, benzyl group of P2 position is surrounded by a large green polyhedron, which suggests bulky substituents, such as 2-naphthyl methylene and 4-(*p*-benzoxyl)-benzyl groups (for example MG270, MG298, MG306 and MG314), are favored. A small green polyhedron embedded in methylene of *tert*-butyl group indicates bulky groups instead of small groups attaching on the methylene are preferred to this region. As in CoMFA field information, the CoMSIA models also did not give any isopleths at P3 position in the contour map.

In the electrostatic CoMSIA plots (Fig. 5b), it nearly gave the same results as in CoMFA. At P1 position, a red isopleth lying above the nitrogen atom of quinoline ring indicates the important role of such a negative atom. Large red polyhedra embedded in phenyl ring of P2 position and oxygen atoms of the two carbonyls denote regions of preferential negatively charged substituents. Blue isopleths around the boronic acid indicate that groups with positive charges are favored, which is consistent with the results of CoMFA contour map.

The hydrophobic field is presented in Fig. 5c. 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 affinities. This is consistent with the hydrophobic properties of the amino acid residues in the $\beta 5$ subunit of 20S proteasome [32]. Within 5 Å around MG341, ALA20, MET45, ALA46, ALA49 and ALA50 are all hydrophobic amino acid residues, which can effectively interact with inhibitors through hydrophobic interaction [35]. Three large yellow polyhedra surrounding P1, P2 positions and *tert*-butyl group indicate that hydrophobic groups are beneficial to enhance the activity. The white isopleths around the hydrogen atoms of quinoline ring reveal the necessity of the hydrophilic hydrogen atoms on the aromatic ring to the activities.

The graphical interpretation of the hydrogen-bond donor interaction in the CoMSIA model is represented in Fig. 5d. It highlights areas beyond the molecules where putative hydrogen-bond acceptor groups in the enzyme can form H-bonds with 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 amide between P1 and P2 positions, which can form hydrogen bond with residues of proteasome, the other is beside the

hydrogen atom of NH group of the amide between P2 and P3 positions, which indicates the necessity of the hydrogen atoms at these positions for high activities. In the MG341–20S proteasome crystal complex, two 2.7 Å and 3.1 Å bond-length H-bonds were formed between the two NH groups of MG341 and Thr21 and Gly47 residues of $\beta 5$ subunit, respectively [32]. However, due to the methylation of the N atoms in MG268 and MG332, the binding affinities (pK_i value) reduced greatly to 6.00 and 9.02, compared with non-methylated MG274 ($pK_i = 8.52$) and MG270 ($pK_i = 10.08$), respectively.

Fig. 5e indicates areas where hydrogen-bond acceptors in the ligand promote or decrease binding affinities. Proton acceptors in the ligand directing to red regions increase the binding affinities. A red polyhedron below the nitrogen atom of the quinoline ring at P1 position of the reference molecule suggests that the subunit bearing hydrogen-bond donors will increase activity, which is consistent with the results discussed in the electrostatic fields of CoMFA and CoMSIA contour maps. Furthermore, large red isopleths encompassing two oxygen atoms of the carbonyls indicate hydrogen-bond acceptors in the ligands promote activity. Groll et al. [32] reported that the two oxygen atoms of MG341 formed two 3.2 Å and 3.1 Å bond-length H-bonds with NH groups of Ala49 and Thr21 residues of 20S proteasome.

The CoMFA and CoMSIA contour maps offered enough information for us to understand the binding mode between the inhibitors and the proteasome. Based on the analysis, several structurally novel compounds were designed and synthesized. The biologically assessed experiments indicated that the constructed models were reliable enough to be applied both in the rational design and library screening. The related application results will be published later.

4. Conclusions

A 3D-QSAR study using CoMFA and CoMSIA methods had been applied for the first time to a series of boron-containing dipeptide boronate proteasome inhibitors. As a result, the 3D-QSAR models for the boron-containing clusters with the accessible software package SYBYL were successfully constructed. Both models from CoMFA and CoMSIA are satisfactory according to the statistical validation results as well as the contour map analysis. The CoMFA model was obtained with LOO cross-validation q^2 and non-cross-validated r^2 values of 0.676 and 0.989, respectively. The best CoMSIA model with the most statistically significant correlation result has been generated with the cross-validated coefficient q^2 of 0.630, and the conventional correlation coefficient r^2 of 0.956. The standard error of estimate and the *F*-test value of this model were 0.053 and 752.3, respectively. All of the constructed models possessed good internal and external consistency and showed statistical significance and predictive abilities. Both the predictive

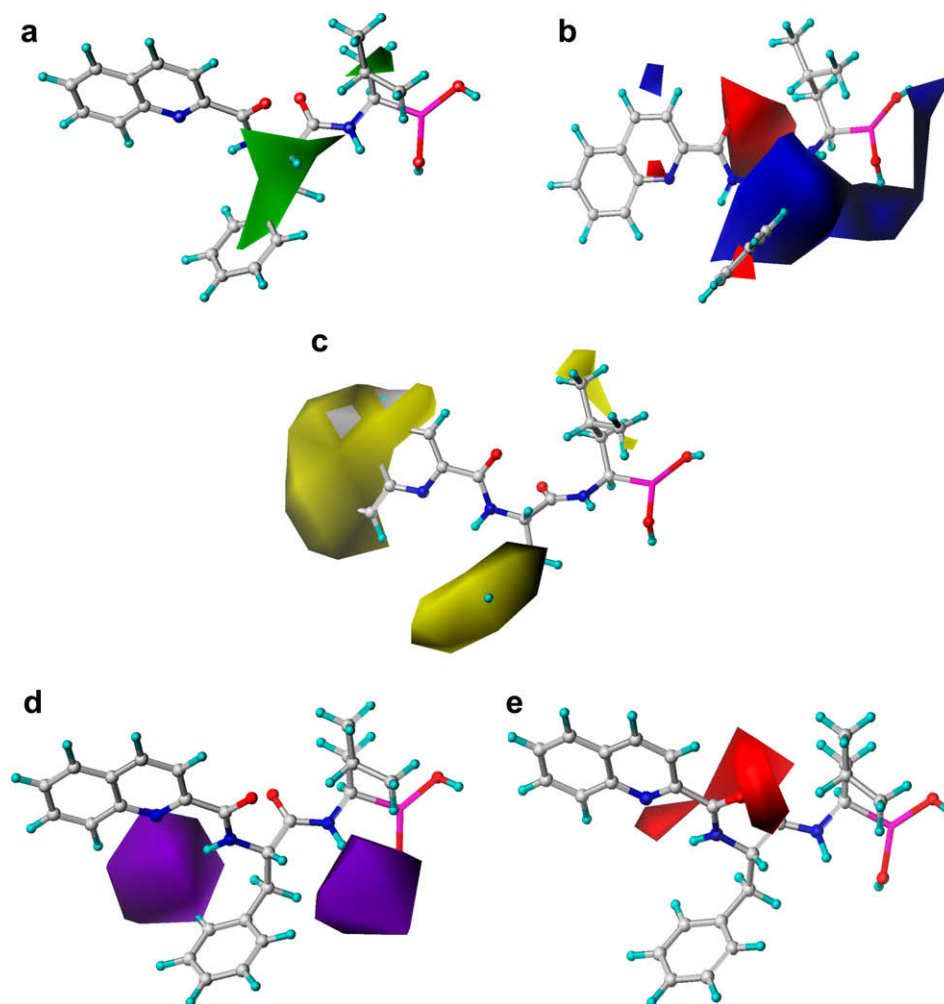


Fig. 5. Std *coeff contour maps of final CoMSIA analysis with 2 Å grid spacing in combination with MG314. (a) Steric contour map. Green contours refer to sterically favored regions. (b) Electrostatic contour map. Blue contours (75% contribution) refer to regions where negatively charged substituents are disfavored; red contours (25% contribution) indicate regions where negatively charged substituents are favored. (c) Hydrophobic contour map. White contours (10% contribution) refer to regions where hydrophilic substituents are favored; yellow contours (90% contribution) indicate regions where hydrophobic substituents are favored. (d) Hydrogen-bond donor contour map. Cyan contours indicate regions where hydrogen-bond acceptor groups on the receptor increase activity. (e) Hydrogen-bond acceptor contour map. Red contours refer to areas where hydrogen-bond donors on the receptor promote the affinity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

evaluation and the contour map analysis accorded well with the experimental interaction mode of the first marketed proteasome inhibitor MG341 with 20S proteasome. A combined application of the obtained CoMFA and CoMSIA models was further employed for the design of new dipeptide boronate inhibitors. Several structurally novel compounds were synthesized and evaluated. The experimental results from our biological assay indicated that the constructed models were reliable enough to be applied in both the rational design and library screening. The related application results will be published later.

Acknowledgements

We thank Dr. Xuedong Dai at XenoPort Inc. and Ms. Wei Wang at NEOTRIDENT for their helpful discussions and dedicated support.

References

- [1] A. Hershko, Trends Biochem. Sci. 21 (1996) 445–449.
- [2] J.D. Etlinger, A.L. Goldberg, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 54–58.
- [3] A.L. Goldberg, Nature (London) 426 (2003) 895–899.
- [4] K.L. Rock, C.F. Gramm, L. Rothstein, K. Clark, R. Stein, L. Dick, D. Hwang, A.L. Goldberg, Cell 78 (1994) 761–771.
- [5] A. Craiu, M. Gaczynska, T. Akopian, C.F. Gramm, G. Fenteany, A.L. Goldberg, K.L. Rock, J. Biol. Chem. 272 (1997) 13437–13445.
- [6] K.L. Rock, A.L. Goldberg, Annu. Rev. Immunol. 17 (1999) 739–779.
- [7] W.E. Mitch, A.L. Goldberg, N. Engl. J. Med. 335 (1996) 1897–1905.
- [8] Y. Murakami, S. Matsufuji, T. Kameiji, S. Hayashi, K. Igarashi, T. Tamura, K. Tanaka, A. Ichihara, Nature (London) 360 (1992) 597–599.
- [9] G.C. Turner, F. Du, A. Varshavsky, Nature (London) 405 (2000) 579–583.
- [10] N. Naidoo, W. Song, M. Hunter-Ensor, A. Sehgal, Science 285 (1999) 1737–1741.
- [11] D.G. Chain, J.H. Schwartz, A.N. Hegde, Mol. Neurobiol. 20 (1999) 125–142.
- [12] A. Alves-Rodrigues, L. Gregori, M.E. Figueiredo-Pereira, Trends Neurosci. 21 (1998) 516–520.
- [13] K.M. Sakamoto, Mol. Genet. Metab. 77 (2002) 44–56.
- [14] J.Z. Sun, S. Nam, C.S. Lee, B.Y. Li, D. Coppola, A.D. Hamilton, Q.P. Dou, S.M. Sebt, Cancer Res. 61 (2001) 1280–1284.
- [15] S. Keck, R. Nitsch, T. Grune, O. Ullrich, J. Neurochem. 85 (2003) 115–122.
- [16] K.J. Rodgers, R.T. Dean, Int. J. Biochem. Cell Biol. 35 (2003) 716–727.
- [17] W.S. Willis, C. Patterson, J. Mol. Cell. Cardiol. 41 (2006) 567–579.
- [18] O. Zolk, C. Schenke, A. Sarikas, Cardiovasc. Res. 70 (2006) 410–421.
- [19] A.P. Arrigo, K. Tanaka, A.L. Goldberg, W.J. Welch, Nature (London) 331 (1988) 192–194.
- [20] O. Cux, K. Tanaka, A.L. Goldberg, Annu. Rev. Biochem. 65 (1996) 801–847.
- [21] J. Löwe, D. Stock, B. Jap, P. Zwickl, W. Baumeister, R. Huber, Science 268 (1995) 533–539.
- [22] M. Groll, L. Ditzel, J. Löwe, D. Stock, M.B. Bochtler, R. Huber, Nature (London) 386 (1997) 463–471.
- [23] J.A. Brannigan, G. Dodson, H.J. Duggleby, P.C. Moody, J.L. Smith, D.R. Tomchick, A.G. Murzin, Nature (London) 378 (1995) 416–419.

- [24] U. Schubert, L.C. Anton, J. Gibbs, C.C. Norbury, J.W. Yewdell, J.R. Bennink, *Nature (London)* 404 (2000) 770–774.
- [25] C.N. Larsen, D. Finley, *Cell* 91 (1997) 431–434.
- [26] W. Baumeister, J. Walz, F. Zühl, E. Seemüller, *Cell* 92 (1998) 367–380.
- [27] J.G. Delcros, M. Baudy Floc'h, C. Prigent, Y. Arlot-Bonnemains, *Curr. Med. Chem.* 10 (2003) 479–503.
- [28] P.G. Richardson, P. Sonneveld, M.W. Schuster, D. Irwin, E.A. Stadtmauer, T. Facon, J.L. Harousseau, D. Ben-Yehu, S. Lonial, H. Goldschmidt, *N. Engl. J. Med.* 352 (2005) 2487–2498.
- [29] J. Adams, *Trends Mol. Med.* 8 (Suppl.) (2002) S49–S54.
- [30] J. Adams, V.J. Palombella, E.A. Sausville, J. Johnson, A. Destree, D.D. Lazarus, J. Maas, C.S. Pien, S. Prakash, P.J. Elliott, *Cancer Res.* 59 (1999) 2615–2622.
- [31] L. Pham, A. Tamayo, P. Lo, L. Yoshimura, R.J. Ford, *Blood* 98 (2001) 465a.
- [32] M. Groll, C.R. Berkens, H.L. Ploegh, H. Ova, *Structure* 14 (2006) 451–456.
- [33] J. Adams, M. Behnke, S.W. Chen, A.A. Cruickshank, L.R. Dick, L. Grenie, J.M. Klunder, Y.T. Ma, L. Plamondon, R.L. Stein, *Bioorg. Med. Chem. Lett.* 8 (1998) 333–338.
- [34] J. Adams, *Drug Discov. Today* 8 (2003) 307–315.
- [35] Y.Q. Zhu, J.F. Pei, Z.M. Liu, L.H. Lai, J.R. Cui, R.T. Li, *Bioorg. Med. Chem.* 14 (2006) 1483–1496.
- [36] H. Kubinyi, *Drug Discov. Today* 2 (1997) 457–467.
- [37] R.D. Cramer III, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* 110 (1988) 5959–5967.
- [38] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* 37 (1994) 4130–4146.
- [39] Y. Endo, T. Iijima, K. Yaguchi, E. Kawachi, N. Inoue, H. Kagechika, A. Kubo, A. Itai, *Bioorg. Med. Chem. Lett.* 11 (2001) 1307–1311.
- [40] Y. Endo, T. Iijima, Y. Yamakoshi, H. Fukasawa, C. Miyaura, M. Inada, A. Kubo, A. Itai, *Chem. Biol.* 8 (2001) 341–355.
- [41] Y. Endo, K. Yaguchi, M. Tsuji, K. Yamaguchi, K. Shudo, *Chem. Pharm. Bull.* 47 (1999) 699–701.
- [42] Y. Endo, T. Yoshimi, K. Kimura, A. Itai, *Bioorg. Med. Chem. Lett.* 9 (1999) 2561–2564.
- [43] M. Tsuji, Y. Koiso, H. Takahashi, Y. Hashimoto, Y. Endo, *Biol. Pharm. Bull.* 23 (2000) 513–516.
- [44] P. Tao, L.H. Lai, *J. Comput.-Aided Mol. Des.* 15 (2001) 429–446.
- [45] J. Adams, Y.T. Ma, R. Stein, M. Baevsky, L. Grenier, L. Plamonda, *PCT Int. Appl.*, WO 1996013266, 1995.
- [46] J. Johnsamuel, Y. Byun, T.P. Jones, Y. Endo, W. Tjarks, *J. Organomet. Chem.* 680 (2003) 223–231.
- [47] J. Johnsamuel, Y. Byun, T.P. Jones, Y. Endo, W. Tjarks, *Bioorg. Med. Chem. Lett.* 13 (2003) 3213–3216.
- [48] *Sybyl Theory Manual*, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO, 1988.
- [49] *Sybyl Molecular Spreadsheet Manual*, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO, 1988.
- [50] *Sybyl Tutorial Manual*, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO, 1988.
- [51] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.* 107 (1985) 3902–3909.
- [52] J. Gasteiger, M. Marsili, *Tetrahedron* 36 (1980) 3219–3228.
- [53] M. Böhm, J. Stürzebecher, G. Klebe, *J. Med. Chem.* 42 (1999) 458–477.