Theor Chem Acc (2012) 131:1203
DOI 10.1007/s00214-012-1203-1

REGULAR ARTICLE

Structural features and binding free energies for non-covalent inhibitors interacting with immunoproteasome by molecular modeling and dynamics simulations

Beilei Lei · Hamza Adel · Chang-Guo Zhan

Received: 23 January 2012 / Accepted: 5 March 2012 / Published online: 29 March 2012

Abstract Immunoproteasome subunit LMP7 is an important target for development of various novel therapeutic agents. In the present study, we examined the detailed binding structures and free energies for a promising series of non-covalent inhibitors interacting with LMP7 by carrying out homology modeling, molecular docking, molecular dynamics (MD) simulations, binding free energy calculations, and binding energy decompositions. The obtained protein-inhibitor binding structures and energetic results have revealed some interesting structural features of the non-covalent inhibitors binding with the LMP7 subunit and valuable insights into the factors affecting the activity of these non-covalent inhibitors. Based on the MD-simulated protein–ligand binding structures, the calculated binding free energies are in good agreement with the experimental activity data for all of the inhibitors examined, which suggests that the computational protocol and the obtained structural insights are reasonable. The obtained computational insights, along with the binding free energy calculation protocol tested in this study, are expected to be valuable for rational design of new, more potent non-covalent inhibitors of LMP7.

Keywords Molecular dynamics simulation · Protein–ligand interaction · Binding free energy · Immunoproteasome · Non-covalent inhibitor

1 Introduction

The 20S proteasome contains the catalytic subunits $\beta5$, $\beta1$, and $\beta2$ accounting for chymotrypsin-like, caspase-like, and trypsin-like activities. This protein plays a major role in the regulation of essential cellular processes such as transcription, cell cycle progression, differentiation, and inflammatory responses [1–3]. The 20S proteasome is composed of 28 subunits arranged in four homoheptameric rings (two outer α rings and two inner β rings) that form a cylindrical structure with two-fold axial symmetry [4]. There are seven different subunits in each ring and it is organized as $\alpha7\beta7\beta7\alpha7$. Each of the two β rings contains the three catalytic subunits ($\beta5$, $\beta1$, and $\beta2$) with distinct substrate specificities and activities. Figure 1 shows the relative positions of the 28 subunits in the mammalian 20S proteasome (PDB code 1IRU) from both the top and side views.

Another form of the 20S proteasome is immunoproteasome expressed in cells of hematopoietic origin, particularly in lymphocytes and monocytes. It is an alternative form of the proteasome present in all eukaryotic cells. In the immunoproteasome, the three catalytic subunits are encoded by homologous genes that code for the low-molecular mass polypeptide-7 (LMP7/$\beta5$i), LMP2/$\beta1$i, and multicatalytic endopeptidase complex subunit-1 (MECL-1/$\beta2$i) that replace the constitutive proteasome $\beta5$, $\beta1$, and $\beta2$ subunits, respectively [5]. The immunoproteasome catalytic subunits

Electronic supplementary material The online version of this article (doi:10.1007/s00214-012-1203-1) contains supplementary material, which is available to authorized users.

B. Lei
Center of Bioinformatics, Northwest A&F University, Yangling 712100, Shaanxi, China

B. Lei
College of Life Sciences, Northwest A&F University, Yangling 712100, Shaanxi, China

B. Lei · H. Adel · C.-G. Zhan (✉)
Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 789 South Limestone Street, Lexington, KY 40536, USA
e-mail: zhan@uky.edu

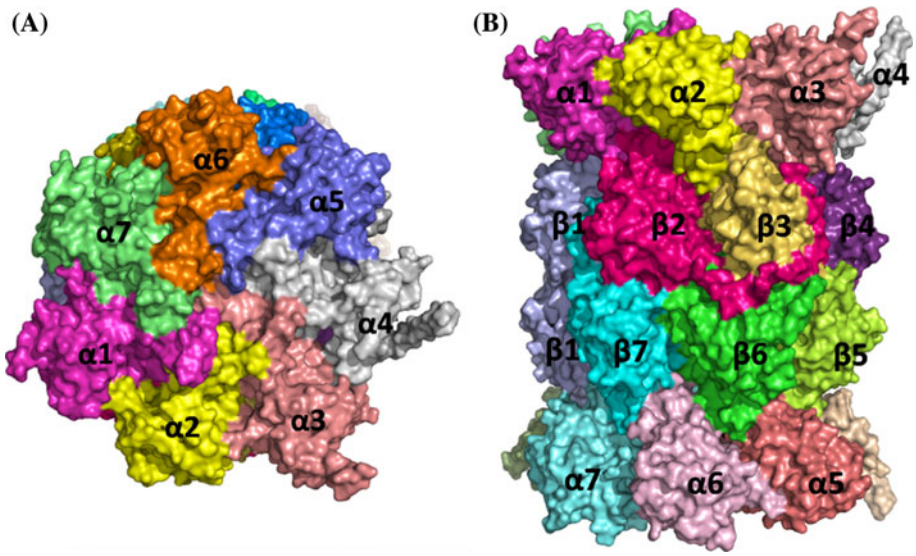


Fig. 1 Surface representation of the crystal structure of the mammalian 20S proteasome (PDB code 1IRU) from *top* **a** and side **b** views of the particle. The figure was prepared using PyMOL

can also be induced in non-hematopoietic cells following exposure to inflammatory cytokines such as interferon-γ (IFN-γ) and tumor necrosis factor-α (TNF-α) [6].

It has been reported that immunoproteasome is involved in the generation of antigenic peptides for class I major histocompatibility complexes (MHC-I) [7] and the function for the immunoproteasome in cytokine production has also been described by a selective inhibitor of LMP7, i.e., PR-957, the first LMP7-selective inhibitor which forms two covalent bonds with the protein. It was revealed that selective inhibition of LMP7 subunit blocked production of cytokine (such as interleukin-23, interferon-γ, and interleukin-2) and provides a therapeutic rationale in autoimmune disorders such as rheumatoid arthritis [8].

So far, a number of small-molecule inhibitors of the proteasomes have been developed for use as molecular probes of proteasome function and potential therapeutics [9, 10]. Most of the known inhibitors form pseudo-covalent or irreversible covalent bonds with N-terminal residue Thr1 of the active site and have been found to induce apoptosis and cell death [11–13], such as peptide aldehydes, peptide boronates, lactacystin and derivatives, peptide vinyl sulfones, and peptide expoxyketones, etc. [10]. Moreover, most of the known proteasome inhibitors, including bortezomib/PS-341, inhibit both proteasome chymotrypsin-like activities (β5 and LMP7) [14–16] and have considerable toxicities that probably limit their clinical use in chronic inflammatory disease [17, 18]. Bortezomib/PS-341 is the dipeptide boronic acid bortezomib in clinical use for the treatment of multiple myeloma [19] and refractory mantle cell lymphoma [20] and is being evaluated for the treatment of other malignancies [21]. Recently, a new series of non-covalently capped peptide inhibitors were characterized, showing that these compounds can potently inhibit LMP7 [22]. It is interesting for rational design of more potent LMP7 inhibitors to understand the more detailed protein-inhibitor interactions and explore the structure–activity correction for this unique series of inhibitors for their binding with LMP7.

Currently, X-ray crystal structures are available for constitutive (regular) proteasome, but not for immunoproteasome. The X-ray structure of constitutive proteasome shows that the binding cavity in catalytic subunits is usually formed between two proteasome subunits [23]. For example, the binding site for chymotrypsin-like activity is formed by association of β5 and β6 subunits [24]. Another example is that the epoxide group of epoxomicin, a well-known inhibitor of proteasome, binds to the β5 active site by through covalent bonds and residues from the β6 subunit form a part of the binding cavity and interact with the other end (N-terminus) of epoxomicin [25].

In the present study, we aimed to investigate the structure–activity correlation for all of the 16 capped peptides in the above-mentioned series as they are non-covalent inhibitors of LMP7 catalytic subunit. For this purpose, homology modeling and molecular docking were employed to obtain the binding mode for each inhibitor. Molecular docking was followed by molecular dynamics (MD) simulations and binding free energy calculations to refine the binding structures and to understand the structure–activity correlation of the LMP7 inhibitors. The insights obtained in this study are expected to be valuable for future rational design of novel, more potent LMP7 inhibitors.

2 Methods

2.1 Dataset

The inhibitors examined in the present study include a capped tri-peptide and 15 capped di-peptides [22]. The

structures of the 16 inhibitors and their IC_{50} values (that range from 1.3 μM to 1.1 nM) are summarized in Table 1. For convenience of comparison with computational data, the experimental IC_{50} value against LMP7 was converted to the corresponding experimental binding free energy (ΔG_{bind}^{Exp}).

2.2 Homology modeling

To study the binding interactions between the inhibitors and LMP7 subunit, we first need to build the 3D structure of the LMP7 subunit by homology modeling. SWISS-MODEL server [26] was used to generate the homology model of the LMP7 subunit based on a suitable template and LMP7 sequence. The FASTA sequence ID for LMP7 is P28062 (residues 73–276). The template for modeling LMP7 subunit was identified through a BLAST search [27]. Once the template and query sequence are given, the SWISS-MODEL server generates the homology model for the query sequence automatically. The final model quality was evaluated using the Ramachandran plot (from PROCHECK) [28, 29] in SWISS-MODEL workspace [26].

2.3 Molecular docking

After the LMP7 homology model was built, molecular docking was carried out by using GOLD program [30, 31]. All of the 16 inhibitors were constructed in SYBYL 6.9 molecular modeling package [32]. Energy minimizations were performed using the Tripos force field with a distance-dependent dielectric and Powell method with a convergence criterion of 0.1 kcal/mol [33]. Partial atomic charges were calculated using Gasteiger-Hückel method. The active site of the LMP7 catalytic subunit is located at the interface of LMP7 and β6 subunits. Thus, both of the two subunits were used in the subsequent docking and MD simulations. To facilitate the docking binding-site definition, the complex $\beta5\beta6$ binding with UK101 in our previous work [32] was aligned to $\beta5i\beta6$ and ligand UK101 was transferred into the active site of $\beta5i\beta6$. Before docking, the protein was prepared by adding hydrogen atoms, assigning protonation states, and carrying out energy minimization with a small number of steps to relax amino-acid residue side chains and UK101 to relative appropriate positions. The protonation state of histidine residues was assigned according to the structure of residues surrounding the histidine side chain, i.e., at ε-position for His117, His165, and His178 in LMP7 and His36 and His 8 in β6, and δ-position for His10 and His197 in LMP7 and His77 and His163 in β6.

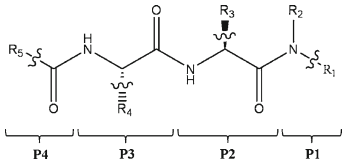
The binding site was defined as all atoms of the protein within 10 Å of the ligand UK101 in LMP7. Subsequently, the docking was performed with GOLD software using the genetic algorithm (GA) search strategy. The GA parameters included 200,000 genetic operations on an initial population of 100 divided into five subpopulations. The number of generated poses was set to 50 for each compound and early termination was turned off. Atom types for ligands and the protein were set automatically by the GOLD. GoldScore was selected as the scoring function. For each ligand, the protein–ligand structure with the best-scoring pose was selected as the initial structure for the energy minimization and then MD simulation on the solvated system as described below.

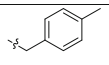
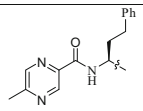
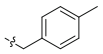
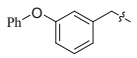
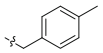
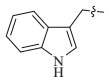
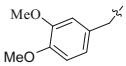
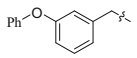
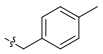
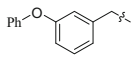
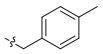
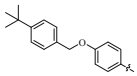
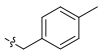
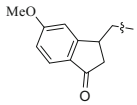
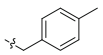
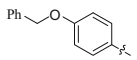
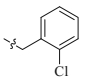
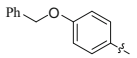
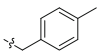
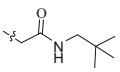
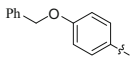
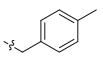
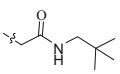
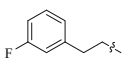
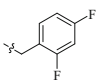
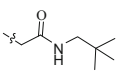
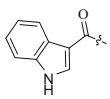
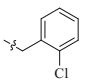
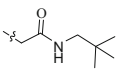
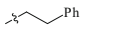
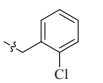
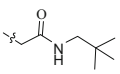
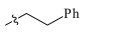
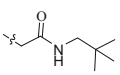
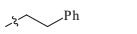
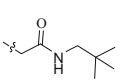
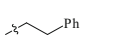
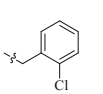
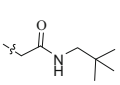
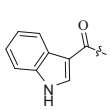
2.4 Molecular dynamics simulation

The general procedure for carrying out the MD simulations in water is similar to that used in our previously reported computational studies [32, 34–39]. Briefly, the MD simulations were performed using the Sander module of Amber (version 9) with Amber ff03 force field [40]. The partial atomic charges for the ligand atoms were calculated using the RESP protocol [41] after the electrostatic potential calculations at the HF/6-31G* level using Gaussian (03 version) [42]. Each protein–ligand binding complex was neutralized by adding suitable counter-ions and was solvated in a truncated octahedron box of TIP3P water molecules [43] with a minimum solute-wall distance of 10 Å. The solvated systems were carefully equilibrated and fully energy-minimized. These systems were gradually heated from $T = 10$ K to $T = 298.15$ K in 60 ps before a production MD simulation run for 1 ns, making sure that we obtained a stable MD trajectory for each of the simulated systems. The time step used for the MD simulations was 2 fs. Periodic boundary conditions in the NPT ensemble at $T = 298.15$ K with Berendsen temperature coupling [44] and $P = 1$ atm with isotropic molecule-based scaling [41] were applied. The SHAKE algorithm [45] was used to fix all covalent bonds containing hydrogen atoms. The non-bonded pair list was updated every 10 steps. The particle mesh Ewald (PME) method [46] was used to treat long-range electrostatic interactions. A residue-based cutoff of 12 Å was utilized for the non-covalent interactions.

For each complex, the whole system was first fully energy-minimized before heating and, after that, the backbone C, Cα, and N atoms of the protein were constrained in order to avoid possible artifacts of the MD simulations on the model systems. The time-dependence of the geometric parameters was carefully examined to make sure that we obtained a stable MD trajectory for each simulated protein–ligand binding system. The atomic coordinates of the simulated system were collected every 1 ps during the simulation. One-hundred snapshots of the simulated structure within the stable MD trajectory (with

Table 1 Molecular structures and inhibitory activity of the 16 non-covalent inhibitors examined



No.	R1	R2	R3	R4	R5	IC_{50} (nM)
1		H				8.9
2		H				230
3		H				37
4		H				41
5		H				27
6		H				10
7		CH3				1300
8		H				4.1
9		H				6.7
10		H				3.2
11		H	CH3			1.1
12		H	H			3.5
13		H				3.4
14		H				200
15		H				44
16		H				1.1

equal time-intervals between the neighboring snapshots) were selected to perform the MM-PBSA calculations (see below).

2.5 Binding free energy calculation and energy decomposition

The binding free energies were calculated by using the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method [47]. In the MM-PBSA method, the free energy of the protein-inhibitor binding, ΔG_{bind}, is obtained from the difference between the free energies of the protein–ligand complex (G_{cpx}) and the unbound receptor/protein (G_{rec}) and ligand (G_{lig}) as following:

$$\Delta G_{\text{bind}} = G_{\text{cpx}} - G_{\text{rec}} - G_{\text{lig}}. \quad (1)$$

The binding free energy (ΔG_{bind}) was evaluated as a sum of the changes in the molecular mechanical (MM) gas-phase binding energy (ΔE_{MM}), solvation free energy (ΔG_{solv}), and entropic contribution ($-T\Delta S$):

$$\Delta G_{\text{bind}} = \Delta E_{\text{bind}} - \text{T}\Delta S \quad (2)$$

$$\Delta E_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} \quad (3)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{ele}} + \Delta E_{\text{vdW}} \quad (4)$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{np}} \quad (5)$$

$$\Delta G_{\text{np}} = \gamma \cdot \Delta\text{SASA} + \beta. \quad (6)$$

The MM binding energies were calculated with the Sander module of the Amber9. Electrostatic solvation free energy was calculated by the PBSA module of the Amber9. The dielectric constant used in PB calculations was 1 for the solute and 80 for the surrounding solvent. The MSMS program [48] was used to calculate the solvent accessible surface area (ΔSASA) for the estimation of the non-polar solvation free energy (ΔG_{np}) using Eq. (6) with $\gamma = 0.00542$ kcal/(mol Å^2) and $\beta = 0.92$ kcal/mol. The entropy contribution to the binding free energy ($-\text{T}\Delta S$) was obtained by using a local program developed in our own laboratory. In this method, the entropy contribution is attributed to two contributions: solvation free entropy and conformational free entropy. The detail of the method has been described previously [49]. The final binding free energy calculated for each protein–ligand binding mode was taken as the average of the ΔG_{bind} values calculated for the 100 snapshots.

The binding energy decomposition was calculated by using the molecular mechanics-Generalized Born surface area (MM-GBSA) method implemented in Amber9 [47]. Linear combinations of pairwise overlaps (LCPO) method was used to calculate the solvent accessible surface area (ΔSASA) and evaluate the ΔG_{np} value with $\gamma = 0.0072$ kcal/(mol Å^{-2}) and $\beta = 0.00$ kcal/mol [50]. In addition, five representative compounds (**5**, **7**, **8**, **14**, and **16**) were selected for the binding energy decomposition calculations according to the structural differences and the range of the ligand activity.

3 Results and discussion

Through the BLAST search, it was determined that the constitutive proteasome β5 subunit is the most suitable template for modeling LMP7 with a sequence identity of 70 %. Then, SWISS-MODEL server was used to generate the homology model of the immunoproteasome LMP7 subunit [26]. The Ramachandran plot for the LMP7 model (Figure S1 in Supplementary Data) indicates that 87.2 % of residues are in the most favored regions and 12.2 % of residues are in additional allowed regions, and 0.6 % of residues are in generously allowed regions. There are no residues in disallowed regions in the model, which indicates that the homology model is adequate.

The binding structures of the 16 non-covalent inhibitors have been studied by performing molecular docking. The docking results revealed that all of these inhibitors bind with LMP7 catalytic subunit in a similar mode; the superposition of the docked binding structures for the 16 inhibitors is shown in Fig. 2. Provided in Supplementary Material (Table S3) are the obtained docking score (GoldScore) values for all inhibitors. The subsequent MD simulations indicate that the docked binding structures were dynamically stable. The MD-simulated binding structures along with the stable MD trajectories are shown in Fig. 3 for the most active inhibitor **16**, Fig. 4 for the least active inhibitor **7**, and Fig. 5 for a moderate inhibitor **14**. The RMSDs for the atomic positions of the protein-inhibitor complexes from the MD trajectories obtained for other compounds are provided in Supplementary Material (Figure S2). In addition, the plots (see Figure S3 of Supplementary Material) of the positional RMSDs of the inhibitor atoms in the MD-simulated protein-inhibitor structures from those in the docked structures *versus* the simulation time also suggest that the MD-simulated inhibitor conformations were dynamically stable and close to those in the docked structures. Below, we will first discuss the binding structures for several representative inhibitors. Then, we will discuss the calculated binding free energies and insights for future rational design of more potent inhibitors.

3.1 Binding of compound 16 with LMP7

The calculated binding free energy based on the MD-simulated binding structure of the most active compound **16** in the active site of LMP7 is the highest one and

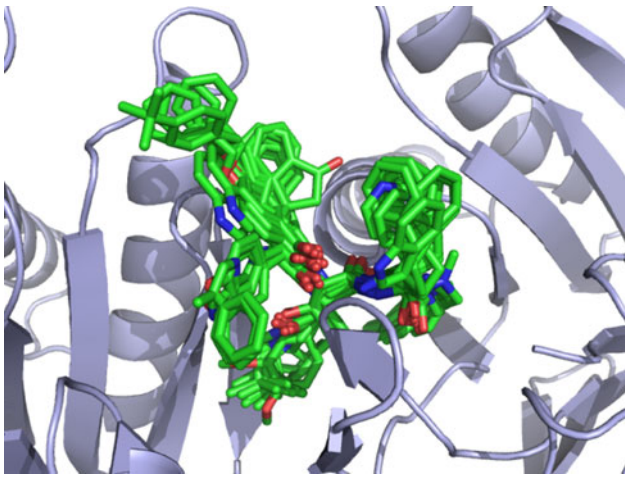
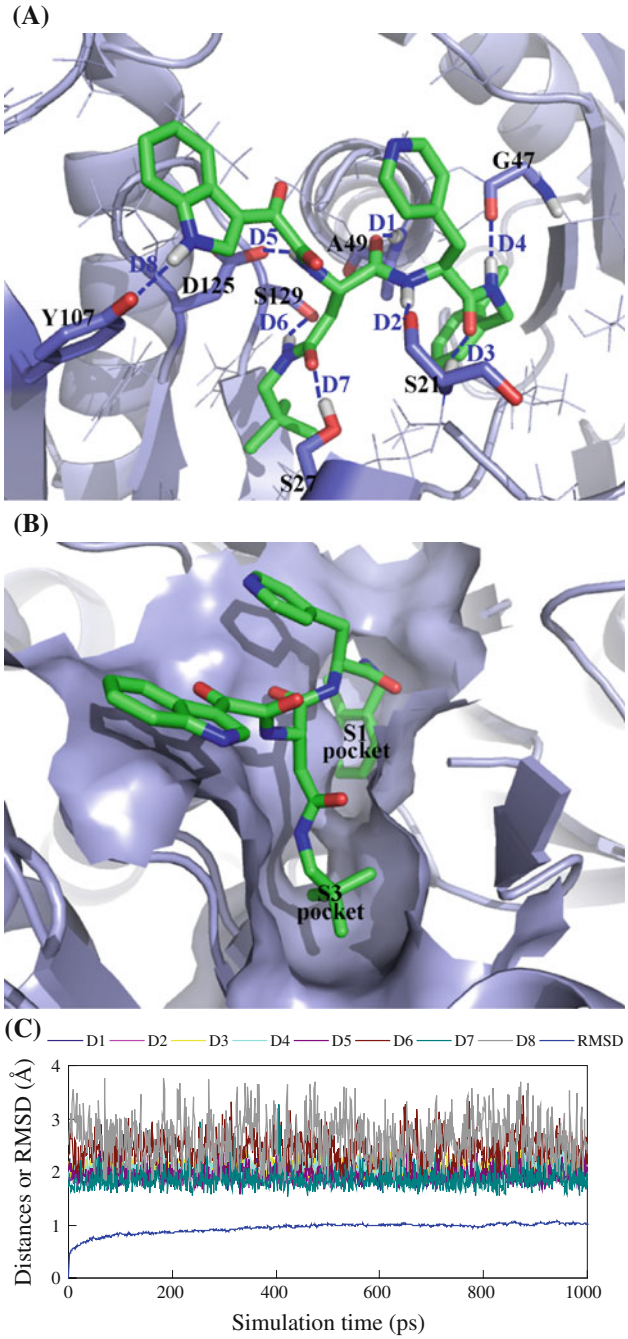


Fig. 2 Superposition of the docked binding structures for the 16 inhibitors in the active site of LMP7 catalytic subunit

consistent with the experimental data. The binding interactions are shown in Fig. 3a and b; the fluctuations of the RMSD and key hydrogen-bond distances *versus* the simulation time are shown in Fig. 3c. As seen in the Figures, the peptide backbone of compound **16** is stabilized on the interface between LMP7 and β6 by five key hydrogen bonds (associated with D1-D5 in Fig. 3): (1) the carbonyl O in the P3 segment of compound **16** and the amino H of Ala49 backbone; (2) the amino H in the P2 segment of compound **16** and the carbonyl O of Ser21 backbone; (3) the carbonyl O in the P2 segment of compound **16** and the backbone amino H of Ser21; (4) the amino H in the P1 segment of compound **16** and the carbonyl O of Gly47 backbone; (5) the amino H in the P3 segment of compound **16** and the carboxyl O of Asp125 side chain (from β6 subunit). The distances associated with the five hydrogen bonds are all stabilized at 1.9–2.0 Å, showing favorable hydrogen-bond interactions, especially the one with Asp125 because it has a negative charge. The five hydrogen bonds allow the compound located at the cleft of the active site by stabilizing the backbone of peptide inhibitors. The side chains of the inhibitor fit in the pockets around the cleft to make the interactions more potent and stable.

In addition to the five key hydrogen bonds, the side chains of each segment in compound **16** have other interactions in the active site to stabilize the hydrogen bonds and strengthen the interactions. The surface representation of the interactions between compound **16** and the active site with the S1 and S3 pockets highlighted is shown in Fig. 3b. The side chain in the P1 segment of **16** is located in the S1 pocket, which is composed of Val31, Lys33, Met45, and Ala49. The 2-Cl-benzyl substituent for R1 fits well in the binding pocket. The side chain in the P3 segment fits into S3 pocket formed by Ser27 (from LMP7), Ser123, Phe124, Ser129, Gln131, and Lys136 (from β6



subunit). The *tert*-butyl fits in the pocket perfectly for steric interactions. Two hydrogen bonds associated with D6 and D7 in Fig. 3 formed between the amino H and the carbonyl O in the P3 segment side chain of compound **16** and between the hydroxyl O of Ser129 (from β6 subunit) and the hydroxyl H of Ser27, respectively. The interactions between the substituent group at R5 and the β6 subunit include two parts: one is a strong van der Waals (vdW) interaction with Pro126 (Fig. 6); the other is a hydrogen bond with Tyr107 (Fig. 3a). The hydrogen bond and the vdW interaction with Pro126 help to stabilize the R5

◀ **Fig. 3** **a** Binding mode of the MD-simulated structure of compound **16** in the active site of LMP7. The intermolecular hydrogen bonds are highlighted in *dashed blue line*. **b** Surface representation of S1 and S3 pockets in the active site with compound **16**. **c** Plots of MD-simulated internuclear distances and RMSD *versus* simulation time for compound **16** with LMP7. D1 refers to the distance between the carbonyl O in the P3 segment of compound **16** and the amino H of Ala49 backbone; D2 refers to the distance between the amino H in the P2 segment of compound **16** and the carbonyl O of Ser21 backbone; D3 refers to the distance between the carbonyl O in the P2 segment of compound **16** and the backbone amino H of Ser21; D4 refers to the distances between the amino H in the P1 segment of compound **16** and the carbonyl O of Gly47 backbone; D5 refers to the distance between the amino H in the P3 segment of compound **16** and the carboxyl O of Asp125 side chain (from β6 subunit); D6 and D7 refer to the distances between the amino H and the carbonyl O in the P3 segment side chain of compound **16** and between the hydroxyl O of Ser129 (from β6 subunit) and the hydroxyl H of Ser27, respectively; D8 refers to the hydrogen-bond distance between Y107 (from β6 subunit) and R5 group

substituent on the protein surface and, thus, decrease the overall free energy.

3.2 Binding of compound 7 with LMP7

Compound **7** is a relatively less active inhibitor among these 16 molecules, with $IC_{50} = 1.3$ μM. Based on the MD simulation, the low affinity of the ligand is mainly due to the substituent CH_3 on the amide H in the P1 segment, which causes the loss of the hydrogen bond with Gly47 (see Fig. 4). Due to the loss of this favorable interaction, another key hydrogen bond in the P3 segment with Ala49 is also broken. Compared to the most active compound (**16**), two of the five key hydrogen bonds are missing in the binding with compound **7**, which makes it go slightly outside the active-site cleft and the interactions with S1 pocket (mainly Lys33 and Met45) decreased (see the decomposition energies in Fig. 7). In addition, the side chain in the P3 segment is small and does not fit the S3 pocket well. Moreover, an additional hydrogen bond between the R4 substituent of the P3 segment and Asp125 (D6 in Fig. 4) does not increase the activity. During the MD simulation, compound **7** slightly moved out of the cleft and the binding mode became slightly different compared to other compounds.

3.3 Binding of compound 14 with LMP7

Compounds **12**–**14** have similar chemical structures, but **14** shows a lower activity compared to compounds **12** and **13**. Figure 5 shows the binding of compound **14** with LMP7 and it can be seen that all of the five key hydrogen bonds are stable except the one close to the R1 substituent. R1 substituent in **14** is smaller than that in **12** and **13**. The surface representation of compound **14** in the active site

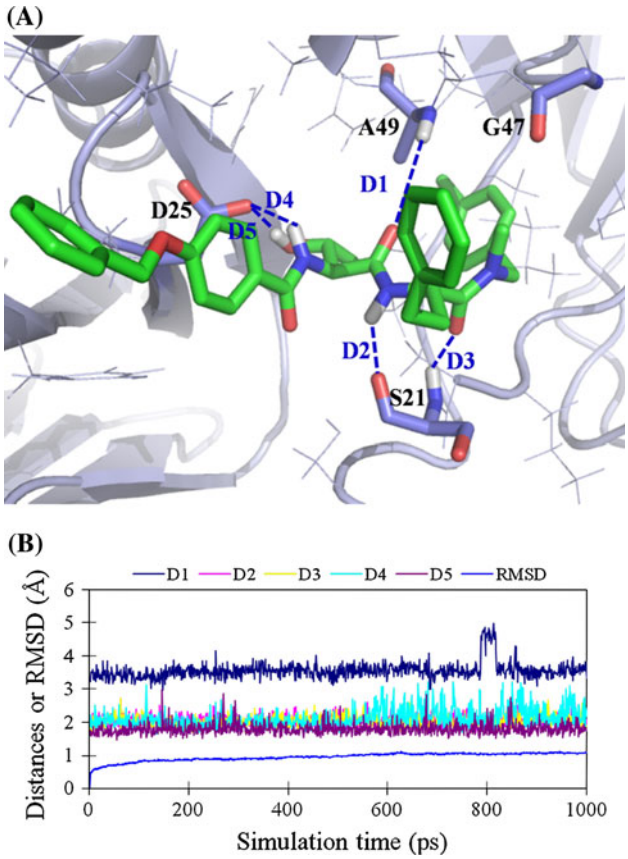
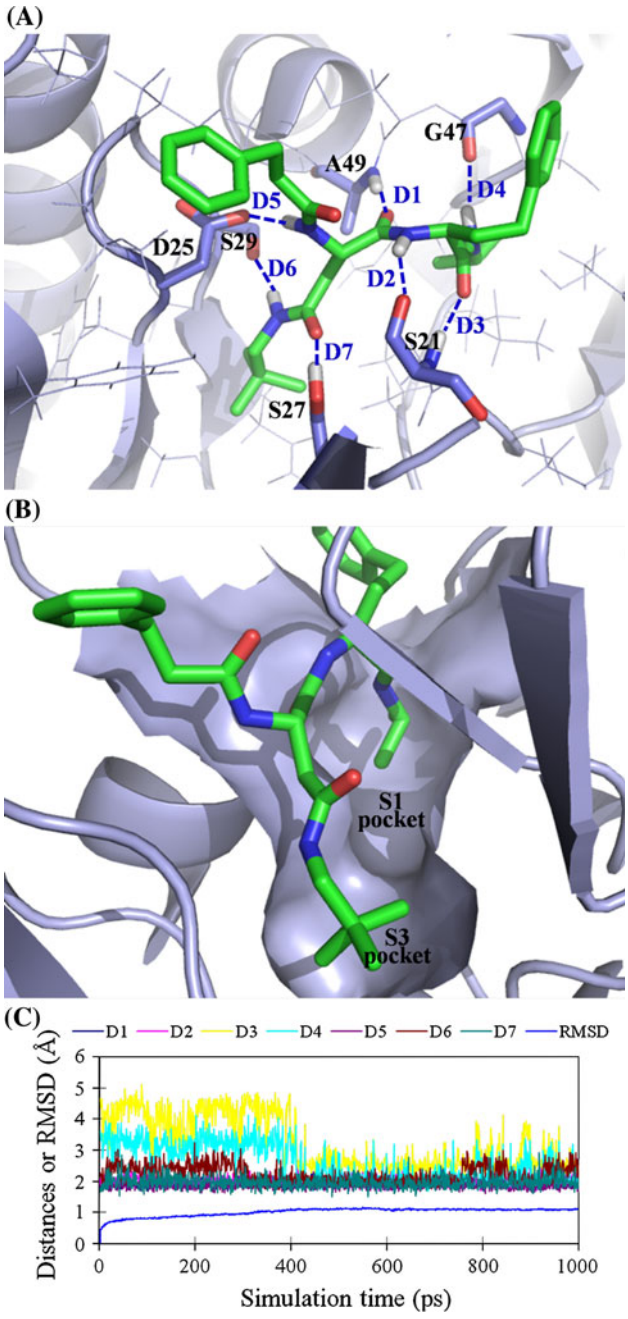


Fig. 4 **a** Binding mode of the MD-simulated structure of compound **7** in the active site of LMP7. The intermolecular hydrogen bonds are highlighted in *dashed blue line*. **b** Plots of MD-simulated internuclear distances and RMSD *versus* simulation time for compound **7** with LMP7. D1 refers to the distance between the carbonyl O in the P3 segment of compound **7** and the amino H of Ala49 backbone; D2 refers to the distance between the amino H in the P2 segment of compound **7** and the carbonyl O of Ser21 backbone; D3 refers to the distance between the carbonyl O in the P2 segment of compound **7** and the backbone amino H of Ser21; D4 and D5 refer to the distances between the carboxyl O of Asp125 (from β6 subunit) and the amino H and hydroxyl H in the P3 segment of compound **7**, respectively

(Fig. 5b) indicates that the isopropyl substituent for R1 cannot fit well into the S1 pocket compared to that of **16** (Fig. 3b). The instability of the hydrogen bond with Gly47 can be seen in Fig. 5c. The R5 group for compound **14** has no significant interaction with the protein.

3.4 Main factors affecting the protein-inhibitor binding

As seen in the above illustrations, the five key hydrogen bonds play a central role in the binding of these capped peptide inhibitors. Once one or more of them is/are broken, the inhibitor will be unstable in the cleft. Specially, for the hydrogen bonding with Asp125, multiple (weaker) hydrogen bonds with Asp125 are not necessarily more favorable



◀ **Fig. 5** **a** Binding mode of the MD-simulated structure of compound **14** in the active site of LMP7. The intermolecular hydrogen bonds are highlighted in *dashed blue line*. **b** Surface representation of S1 and S3 pockets in the active site with compound **14**. **c** Plots of MD-simulated internuclear distances and RMSD *versus* simulation time for compound **14** with LMP7. D1 refers to the distance between the carbonyl O in the P3 segment of compound **14** and the amino H of Ala49 backbone; D2 refers to the distance between the amino H in the P2 segment of compound **14** and the carbonyl O of Ser21 backbone; D3 refers to the distance between the carbonyl O in the P2 segment of compound **14** and the backbone amino H of Ser21; D4 refers to the distances between the amino H in the P1 segment of compound **14** and the carbonyl O of Gly47 backbone; D5 refers to the distance between the amino H in the P3 segment of compound **14** and the carboxyl O of Asp125 side chain (from β6 subunit); D6 and D7 refer to the distances between the amino H/carbonyl C in the P3 segment side chain of compound **14** and hydroxyl O of Ser129 (from β6 subunit)/hydroxyl H of Ser27, respectively

than a single (stronger) hydrogen bond for a couple of reasons. First of all, the overall strength of the multiple hydrogen bonds with a substituent is not necessarily stronger than that of the single one. In addition, with the substituent change, one should also account for the possible change in the vdW interaction and solvation free energy associated with the change in the hydrogen bonding. As shown in Fig. 7, the decomposition energy of Asp125 for compounds **14** and **16** are lower than compounds **5** and **8**, which means that the interactions of Asp125 with compounds **14** and **16** are stronger. For compounds **5** and **8**, an additional hydrogen bond slightly strengthens the electrostatic interaction, but decreases the vdW interaction and increases the solvation free energy (see Table S3) and, thus, the overall contribution of Asp125 to binding free energy decreases (Fig. 7). Besides, the side chains in P1, P3, and P5 segments contribute to the binding affinity significantly, according to the binding energy decomposition.

For R4 substituent in P3 segment, hydrogen bonds with Ser129 (from β6 subunit) can be beneficial (compound **16** in Fig. 7). The substituent should fit into the S3 pocket well. The main contributions from Ser123 and Gln131 are shown in Fig. 7, indicating that compounds **14** and **16** with a *tert*-butyl group have more favorable contributions from these residues to the binding energies. The *tert*-butyl group fits in the pocket perfectly from the steric aspect, but most of the residues forming S3 pocket are polar and not hydrophobic. For example, the electrostatic interaction of *tert*-butyl group with Lys136 is unfavorable according to the results of the energy decomposition (see data in Fig. 7 and Table S3 in Supplementary Data).

As for R5 substituent in P4 segment, the substituent interacts with the residues on the surface of β6 subunit. The R5 substituent of compound **16** is the best one out of these 16 inhibitors, which has a combination of a perfect interaction with Pro126 and a hydrogen-bond interaction with Tyr107.

Finally, solvation free energy cannot be ignored in accounting for the relative binding free energies. For example, compound **2** has favorable electrostatic and vdW interactions with the target, but it also has a lower activity because it has the highest solvation free energy change within all of the 16 inhibitors (Table 2). The highest solvation free energy is due to the hydrophobic properties of all the four side chains.

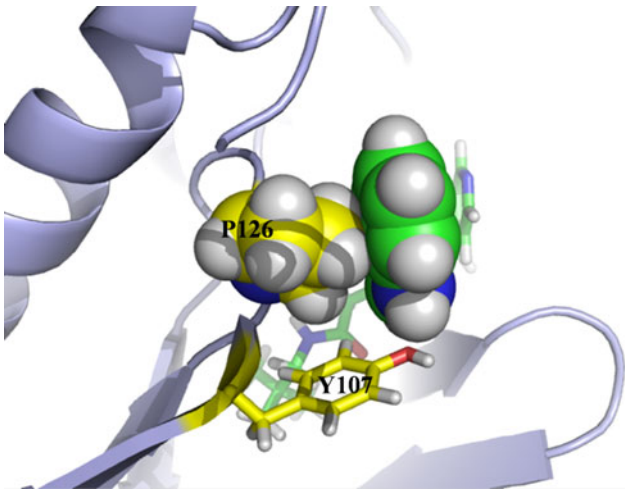


Fig. 6 The interactions of R5 substituent in compound **16** with P126 and Y107 (from β6 subunit)

3.5 Binding free energies

The binding free energies calculated for the 16 compounds binding with LMP7 catalytic subunit are summarized in Table 2 in comparison with the corresponding binding free energies derived from the experimental data. As seen in Table 2, the $\Delta G_{\text{bind}}^{\text{Cal}}$ values are all qualitatively consistent with the corresponding $\Delta G_{\text{bind}}^{\text{Exp}}$ values in terms of relative binding free energies for different inhibitors. The correlation between the $\Delta G_{\text{bind}}^{\text{Exp}}$ and $\Delta G_{\text{bind}}^{\text{Cal}}$ values is excellent, with the squared correlation coefficient (R^2) being 0.973. As seen in Table 2, for the directly calculated binding free energies, the MM-PBSA calculations systematically overestimated the binding affinity for all ligands. Nevertheless, reasonable energetic results can be obtained from the use of a linear correlation relationship:

$$\Delta G_{\text{bind}}^{\text{Cal}}(\text{corrected}) = 0.2214 \times \Delta G_{\text{bind}}^{\text{Cal}} - 5.6638\ (\text{kcal/mol}) \quad (7)$$

The corrected $\Delta G_{\text{bind}}^{\text{Cal}}$ values are also listed in Table 2 (in parentheses) and Fig. 8 shows the correlation between $\Delta G_{\text{bind}}^{\text{Exp}}$ and the corrected $\Delta G_{\text{bind}}^{\text{Cal}}$.

In addition, we wanted to know where any component (*i.e.*, the electrostatic energy, vdW energy, or solvation free energy) of the calculated binding free energy dominates the correlation between the $\Delta G_{\text{bind}}^{\text{Exp}}$ and $\Delta G_{\text{bind}}^{\text{Cal}}$ values. So, we also tried to explore the possible correlation of the experimental affinity with each component of the calculated binding free energy, but we did not find a satisfactory correlation relationship with any component. The linear correlation between the $\Delta G_{\text{bind}}^{\text{Cal}}$ (based on the MM-PBSA calculation) and GoldScore (based on the docking) values is also not good ($R^2 = 0.157$, as shown in Figure S4 of Supplementary Material).

3.6 Insights for future design of more potent inhibitors

The computational results discussed above have demonstrated some interesting insights into the factors affecting the activity of LMP7 inhibitors, such as the five key protein–ligand hydrogen bonds involving the backbone of the capped peptide inhibitors, a suitable R1 group with a moderate size, the hydrogen bond with Ser129, the vdW and electrostatic interactions with pocket S3 considering Lys136 (from β6 subunit), and the solvation effects of the side chain in inhibitors. It is essential for rational design of new, more potent inhibitors of LMP7 catalytic subunit to carefully account for all of the favorable intermolecular interactions. For example, a potent inhibitor should at least maintain the essential five hydrogen bonds with the backbone of protein in the active-site cleft and have suitable R1 and R4 groups in pockets S1 and S3, respectively, considering both the steric and electrostatic interactions. In addition, the solvation effects should be taken into account while maintaining the favorable interactions during the rational design of a virtual molecular library. Once a virtual molecular library of various possible compounds has been built, one may perform molecular docking, MD simulations, and MM/PBSA binding free energy calculations on all compounds in the library, along with the empirical corrections using Eq. (7), to quantitatively estimate/predict their binding free energies. Based on the estimated/predicted binding free energies, the

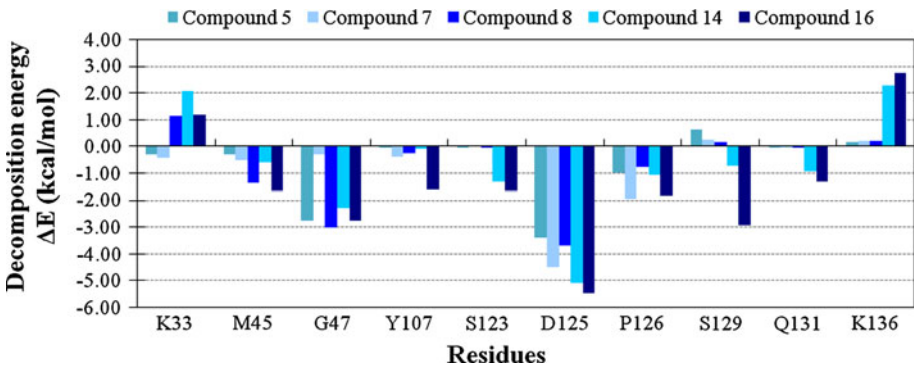


Fig. 7 Decomposition energy contributions from key residues in the active site (K33, M45, and G47 from LMP7; Y107, S123, D125, P126, S129, Q131, and K136 from β6) for compounds **5**, **7**, **8**, **14,** and **16**. The detailed values of ΔE_{ele}, ΔE_{vdw}, ΔE_{MM}, ΔG_{solv}, and ΔE_{bind} can be seen in Supplementary Data (Table S3)

Table 2 Calculated binding free energies obtained from MM-PBSA calculations in comparison with available experimental data for the 16 compounds binding with the LMP7 catalytic subunit

Ligand in the complex	Cal.							Exp.[a]
	ΔE_{ele}	ΔE_{vdw}	ΔE_{MM}	ΔG_{solv}	ΔE_{bind}	$-T\Delta S$[b]	ΔG_{bind}^{Cal} [c]	ΔG_{bind}^{Exp}
1	−49.43	−78.43	−127.86	74.77	−53.08	28.23	−24.85 (−11.17)	−10.99
2	−48.20	−76.01	−124.21	85.23	−38.97	24.34	−14.63 (−8.90)	−9.06
3	−54.76	−75.77	−130.53	82.26	−48.27	27.18	−21.09 (−10.33)	−10.15
4	−57.69	−55.22	−112.91	69.64	−43.28	22.31	−20.97 (−10.31)	−10.09
5	−47.30	−59.45	−106.76	63.76	−43.00	22.30	−20.82 (−10.27)	−10.33
6	−68.94	−57.79	−126.73	77.46	−49.27	24.76	−24.51 (−11.09)	−10.92
7	−42.51	−52.16	−94.67	63.94	−30.73	19.14	−11.59 (−8.23)	−8.04
8	−50.15	−53.98	−104.13	57.51	−46.62	20.65	−25.97 (−11.41)	−11.45
9	−52.02	−73.04	−125.05	76.78	−48.28	22.66	−25.62 (−11.34)	−11.16
10	−57.36	−70.16	−127.52	73.75	−53.77	28.09	−25.68 (−11.35)	−11.60
11	−56.52	−69.78	−126.29	75.63	−50.66	21.49	−29.17 (−12.12)	−12.23
12	−64.67	−61.53	−126.21	71.88	−54.33	27.48	−26.85 (−11.61)	−11.55
13	−56.47	−74.72	−131.18	81.72	−49.46	24.55	−24.91 (−11.18)	−11.56
14	−56.42	−62.68	−119.10	77.04	−42.05	27.70	−14.35 (−8.84)	−9.15
15	−60.25	−56.70	−116.94	70.20	−46.74	26.42	−20.32 (−10.16)	−10.04
16	−51.19	−74.22	−125.41	69.59	−55.83	26.08	−29.75 (−12.25)	−12.23

[a] Binding free energies derived from the experimental IC_{50} values in ref [22]

[b] The conformational free entropy (ΔS_{conf}) contribution was calculated from atoms within a distance of 4 Å from any of the ligand atoms

[c] The values in parentheses are the corrected binding free energies calculated by using Eq. (7)

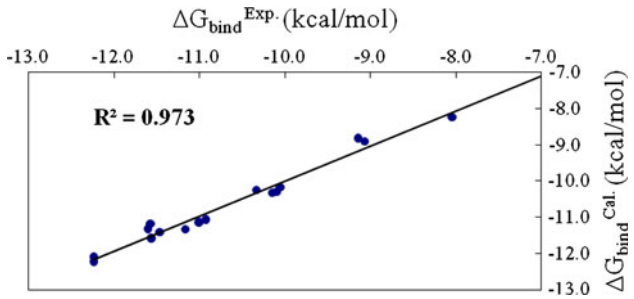


Fig. 8 Correlation between the experimental binding free energy (ΔG_{bind}^{Exp}) and the calculated binding free energy (ΔG_{bind}^{Cal}, corrected using Eq. 7) for the 16 inhibitors

compounds corresponding to the higher binding affinity with the LMP7 subunit will be recommended for chemical synthesis and bioassays.

4 Conclusion

The combined homology modeling, molecular docking, molecular dynamics (MD) simulations, binding free energy calculations, and binding energy decompositions have revealed some detailed structural features of non-covalent inhibitors binding with LMP7 catalytic subunit of immunoproteasome and interesting insights into the factors affecting the activity of these non-covalent inhibitors. Based on the MD-simulated protein–ligand binding structures, the calculated binding free energies are in good agreement with the experimental data for all of the non-covalent inhibitors examined, suggesting that the computational protocol and the obtained structural insights are reasonable. The obtained computational insights into the favorable protein–ligand binding and the binding free energy calculation protocol tested in this study provide a solid base for rational design of new, more potent non-covalent inhibitors of LMP7.

Acknowledgments This work was supported in part by the NIH (grants RC1MH088480) and NSF (grant CHE-1111761) to Zhan. The authors also acknowledge the Computer Center at University of Kentucky for supercomputing time on IBM X-series Cluster with 340 nodes or 1,360 processors.

References

1. Hershko A, Ciechanover A (1998) Annu Rev Biochem 67:425–479
2. Myung J, Kim KB, Crews CM (2001) Med Res Rev 21(4):245–273
3. Groll M, Heinemeyer W, Jager S, Ullrich T, Bochtler M, Wolf DH, Huber R (1999) Proc Nat Acad Sci USA 96(20):10976–10983

4. Unno M, Mizushima T, Morimoto Y, Tomisugi Y, Tanaka K, Yasuoka N, Tsukihara T (2002) Structure 10(5):609–618
5. Kloetzel PM (2001) Nat Rev Mol Cell Biol 2(3):179–187
6. Griffin TA, Nandi D, Cruz M, Fehling HJ, Van Kaer L, Monaco JJ, Colbert RA (1998) J Exp Med 187(1):97–104
7. Basler M, Moebius J, Elenich L, Groettrup M, Monaco JJ (2006) J Immunol 176(11):6665–6672
8. Muchamuel T, Basler M, Aujay MA, Suzuki E, Kalim KW, Lauer C, Sylvain C, Ring ER, Shields J, Jiang J, Shwonek P, Parlati F, Demo SD, Bennett MK, Kirk CJ, Groettrup M (2009) Nat Med 15(7):781–787
9. Genin E, Reboud-Ravaux M, Vidal J (2010) Curr Top Med Chem 10(3):232–256
10. Borissenko L, Groll M (2007) Chem Rev 107(3):687–717
11. ImajohOhmi S, Kawaguchi T, Sugiyama S, Tanaka K, Omura S, Kikuchi H (1995) Biochem Biophys Res Commun 217(3):1070–1077
12. Shinohara K, Tomioka M, Nakano H, Tone S, Ito H, Kawashima S (1996) Biochemical J 317:385–388
13. Hideshima T, Richardson P, Chauhan D, Palombella VJ, Elliott PJ, Adams J, Anderson KC (2001) Cancer Res 61(7):3071–3076
14. Berkers CR, Verdoes M, Lichtman E, Fiebiger E, Kessler BM, Anderson KC, Ploegh HL, Ovaa H, Galardy PJ (2005) Nat Methods 2(5):357–362
15. Chauhan D, Catley L, Li GL, Podar K, Hideshima T, Velankar M, Mitsiades C, Mitsiades N, Yasui H, Letai A, Ovaa H, Berkers C, Nicholson B, Chao TH, Neuteboom STC, Richardson P, Palladino MA, Anderson KC (2005) Cancer Cell 8(5):407–419
16. Demo SD, Kirk CJ, Aujay MA, Buchholz TJ, Dajee M, Ho MN, Jiang J, Laidig GJ, Lewis ER, Parlati F, Shenk KD, Smyth MS, Sun CM, Vallone MK, Woo TM, Molineaux CJ, Bennett MK (2007) Cancer Res 67(13):6383–6391
17. Bross PF, Kane R, Farrell AT, Abraham S, Benson K, Brower ME, Bradley S, Gobburu JV, Goheer A, Lee SL, Leighton J, Liang CY, Lostritto RT, McGuinn WD, Morse DE, Rahman A, Rosario LA, Verbois SL, Williams G, Wang YC, Pazdur R (2004) Clin Cancer Res 10(12):3954–3964
18. Groll M, Huber R, Moroder L (2009) J Pept Sci 15(2):58–66
19. Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T, Harousseau JL, Ben-Yehuda D, Lonial S, Goldschmidt H, Reece D, San-Miguel JF, Blade J, Boccadoro M, Cavenagh J, Dalton WS, Boral AL, Esseltine DL, Porter JB, Schenkein D, Anderson KC, Investigators A (2005) N Engl J Med 352(24):2487–2498
20. Fisher RI, Bernstein SH, Kahl BS, Djulbegovic B, Robertson MJ, de Vos S, Epner E, Krishnan A, Leonard JP, Lonial S, Stadtmauer EA, O'Connor OA, Shi HL, Boral AL, Goy A (2006) J Clin Oncol 24(30):4867–4874
21. Orlowski RZ, Kuhn DJ (2008) Clin Cancer Res 14(6):1649–1657
22. Blackburn C, Gigstad KM, Hales P, Garcia K, Jones M, Bruzzese FJ, Barrett C, Liu JX, Soucy TA, Sappal DS, Bump N, Olhava EJ, Fleming P, Dick LR, Tsu C, Sintchak MD, Blank JL (2010) Biochem J 430:461–476
23. Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R (1997) Nature 386(6624):463–471
24. Momose I, Umezawa Y, Hirosawa S, Iinuma H, Ikeda D (2005) Bioorg Med Chem Lett 15(7):1867–1871
25. Groll M, Kim KB, Kairies N, Huber R, Crews CM (2000) J Am Chem Soc 122(6):1237–1238
26. Arnold K, Bordoli L, Kopp J, Schwede T (2006) Bioinformatics 22(2):195–201
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) J Mol Biol 215(3):403–410
28. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
29. Melo F, Feytmans E (1998) J Mol Biol 277(5):1141–1152
30. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267(3):727–748
31. Jones G, Willett P, Glen RC (1995) J Mol Biol 245(1):43–53
32. Lei BL, Hameed M, Hamza A, Wehenkel M, Muzyka JL, Yao XJ, Kim KB, Zhan C-G (2010) J Phys Chem B 114(38):12333–12339
33. Clark M, Cramer RD III, van Opdenhosch N (1989) J Comp Chem 10:982–1012
34. AbdulHameed MDM, Hamza A, Zhan C-G (2006) J Phys Chem B 110(51):26365–26374
35. Hamza A, Cho H, Tai HH, Zhan C-G (2005) J Phys Chem B 109(10):4776–4782
36. Hamza A, Tong M, AbdulHameed MDM, Liu JJ, Goren AC, Tai HH, Zhan C-G (2010) J Phys Chem B 114(16):5605–5616
37. Huang XQ, Gu HH, Zhan C-G (2009) J Phys Chem B 113(45):15057–15066
38. Liu JJ, Hamza A, Zhan C-G (2009) J Am Chem Soc 131(33):11964–11975
39. Liu JJ, Zhang YK, Zhan C-G (2009) J Phys Chem B 113(50):16226–16236
40. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan, J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2006) Amber 9. University of California, San Francisco
41. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) J Phys Chem 97(40):10269–10280
42. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JJA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford
43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79(2):926–935
44. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) J Chem Phys 81(8):3684–3690
45. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23(3):327–341
46. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103(19):8577–8593
47. Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Acc Chem Res 33(12):889–897
48. Sanner MF, Olson AJ, Spehner JC (1996) Biopolymers 38(3):305–320
49. Pan YM, Gao DQ, Zhan C-G (2008) J Am Chem Soc 130(15):5140–5149
50. Harvey SC (1989) Proteins-Struct Func Genet 5(1):78–92