J Mol Model (2010) 16:17–28
DOI 10.1007/s00894-009-0518-x

ORIGINAL PAPER

Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K

Shu-Qun Liu · Zhao-Hui Meng · Yun-Xin Fu · Ke-Qin Zhang

Received: 14 February 2009 / Accepted: 7 April 2009 / Published online: 23 May 2009

Abstract Serine protease proteinase K, a member of the subtilisin family of enzymes, is of significant industrial, agricultural and biotechnological importance. Despite the wealth of structural information about proteinase K provided by static X-ray structures, a full understanding of the enzymatic mechanism requires further insight into the dynamic properties of this enzyme. Molecular dynamics simulations and essential dynamics (ED) analysis were performed to investigate the molecular motions in proteinase K. The results indicate that the internal core of proteinase K is relatively rigid, whereas the surface-exposed loops, most notably the substrate-binding regions, exhibit considerable conformational fluctuations. Further ED analysis reveals that the large concerted motions in the substrate-binding regions cause opening/closing of the substrate-binding pockets, thus supporting the proposed induced-fit mechanism of substrate binding. The distinct electrostatic/hydrogen-bonding interactions between Asp39 and His69 and between His69 and Ser224 within the catalytic triad lead to different thermal motions and orientations of these three catalytic residues, which can be related to their different functional roles in the catalytic process. Statistical analyses of the geometrical/functional properties as well as evolutionary conservation of the glycines in proteinase K-like proteins reveal that glycines may play an important role in determining the folding architecture and structural flexibility of this class of enzymes. Our simulation study complements the biochemical and structural studies and provides new insights into the dynamic structural basis of the functional properties of this class of enzymes.

Keywords Subtilisin family · Molecular dynamics · Structural flexibility · Large concerted motions · Catalytic triad · Glycine · Structure-function relationship

Abbreviations

RMSD	Root mean square deviation
MD	Molecular dynamics
SPC	Single point charge
NHB	Number of hydrogen bonds
NNC	Number of native contacts
SSE	Number of residues in the secondary structure elements
Rg	Radius of gyration
SASA	Solvent accessible surface area
ED	Essential dynamics
PSL	Polar surface loop

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0518-x) contains supplementary material, which is available to authorized users.

S.-Q. Liu · Y.-X. Fu · K.-Q. Zhang (✉)
Laboratory for Conservation and Utilization of Bio-Resources,
Yunnan University,
Kunming,
650091 Yunnan, P. R. China
e-mail: kqzhang1@yahoo.com.cn

Z.-H. Meng
Department of Cardiology, No. 1 Affiliated Hospital,
Kunming Medical College,
Kunming,
650032 Yunnan, P. R. China

Y.-X. Fu (✉)
Human Genetics Center, School of Public Health,
University of Texas,
1400 Herman Pressler, E453,
Houston, TX 77025, USA
e-mail: shuqunliu@ynu.edu.cn

Introduction

Proteases are enzymes that catalyze the hydrolysis of peptide bonds in other proteins. The 279-residue serine protease proteinase K (EC 3.4.21.4) from the fungus *Tritirachium album* limber belongs to the subtilisin family of enzymes [1–3]. This family of enzymes has attracted intensive research interest from the academic, industrial, and agricultural communities. The academic interest is inspired by the ready amenability of subtilases to structural and functional investigation [4], and by applications of proteinase K in biotechnology research such as the removal of DNases and RNases when isolating DNA and RNA from tissues or cell lines [2, 5]. The industrial and agricultural applications of these enzymes include protein-degrading components in washing powders [6] and bio-control agents against parasites [7], respectively. Nearly all subtilases are synthesized as pre-proenzymes that are translocated over a membrane system and then activated by cleavage of the pro-sequence. Many of the properties of these enzymes involved in the catalysis, structure, substrate specificity, and stability to inactivation and pH profile have been probed in detail by biochemical, protein-engineering, and structural studies [4, 7–14]. Such studies have revealed that the subtilases follow a common mechanism of action that involves an identical stereochemistry of the catalytic triad Asp–His–Ser and the oxyanion hole. In the catalytic triad, Ser functions as the primary nucleophile and His plays a dual role as the proton acceptor and donor at different stages in the reaction. The proposed role of Asp in the triad is to bring His into the correct orientation to facilitate the nucleophilic attack by Ser. The role of the oxyanion hole is to stabilize the developing negative charge on the oxygen atom of the substrate upon formation of the tetrahedral intermediate [15].

The structure of proteinase K has previously been solved at 0.98–2.2 Å in a series of X-ray crystallographic investigations [11–13]. The backbone root mean square deviation (RMSD) values between the available structures range from 0.2 Å to 0.4 Å. Proteinase K (Fig. 1) has a well-defined global fold comprising 15 β strands, six α helices and one 3/10 helix. These secondary structure elements can be divided into an internal core that is composed of a parallel β sheet comprising nine β strands and two buried α helices, which is encapsulated by four amphiphilic α helices and three antiparallel two-stranded β sheets. The catalytic triad consists of Asp39, His69 and Ser224; the oxyanion hole is formed primarily by Asn161, and the substrate recognition site is formed primarily by two segments, Gly100-Tyr104 and Ser132-Gly136 that form a three-stranded antiparallel β sheet with the substrate [13]. The native proteinase K contains two Ca^{2+} cations, which are considered to enhance the thermal stability of the enzyme and increase its resistance to proteolysis [12, 16, 17]. The

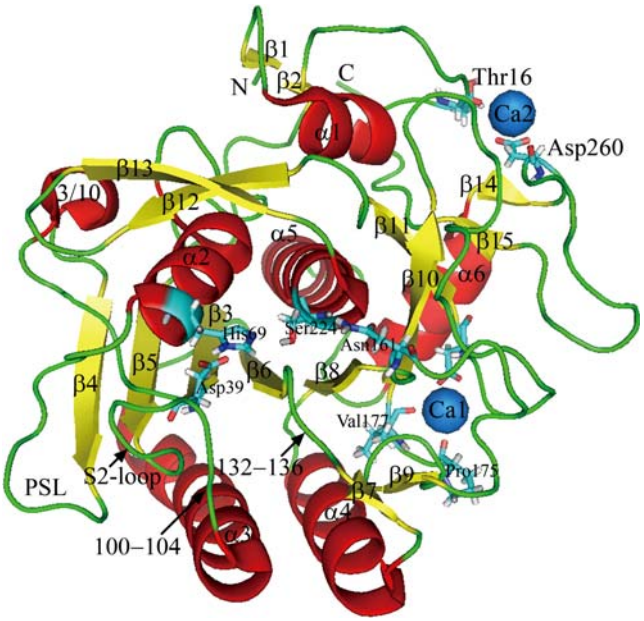

Fig. 1 Ribbon diagram of proteinase K (PDB code 1IC6). The α helices, β strands and loops/links are colored *red*, *yellow* and *green*, respectively. The residues of the catalytic triad (Asp39, His69 and Ser224), oxyanion hole (Asn161), and calcium binding sites (Ca1 site: Pro175, Val177 and Asp200; Ca2 site: Thr16 and Asp260) are shown as *stick models*. The two bound Ca^{2+} cations, Ca1 and Ca2, are shown as *blue spheres*

calcium cation Ca1 is bound tightly with high affinity by the $O_{\delta1}$ and $O_{\delta2}$ of Asp200 and the carbonyl oxygen atoms of Pro175 and Val177. The Ca2 is bound with weak affinity by the $O_{\delta1}$ and $O_{\delta2}$ of Asp260 and the carbonyl oxygen atom of Thr16.

Although crystallographic studies have yielded valuable structural information that provides insight into the function of proteinase K, detailed questions about dynamic aspects of the enzymatic mechanism, such as the dynamic behavior of the catalytic residues, substrate binding/product release, and how they are regulated, remain unanswered while only static structures are available. Proteases are dynamic entities, thus understanding their mechanism of action requires analysis of their dynamic behavior while they are performing their allotted function. Detailed information about the dynamics of proteinase K is therefore necessary for a full understanding of its function.

The aim of this paper is to report a detailed study on the dynamic properties of proteinase K using a molecular dynamics (MD) simulation approach. The flexibility behavior of proteinase K and the mechanism underlying its conformational fluctuations was investigated. The detailed interactions between residues within the catalytic triad were analyzed, and their functional implications are discussed. In addition, the role of glycines in determining the structural flexibility of proteinase K, as well as the dynamic behavior of the bound Ca^{2+} cations was investigated.

Materials and methods

Preparation of the starting structure

The high-resolution crystal structure (PDB code 1IC6 [11]) of the serine protease proteinase K from *Tritirachium album* limber at 0.98 Å resolution was used as starting model for the MD simulation. Before performing the MD simulation, all the hetero atoms such as NO3 and crystal waters were removed but the two Ca^{2+} cations, Ca1 and Ca2, were retained.

Molecular dynamics setup

The GROMACS molecular dynamics package [18, 19] was used with the GROMOS96 43a1 force field. Of the four histidines in proteinase K (His46, His69, His72 and His229), only His69 was protonated on the $N_{\delta 1}$ atom, whereas the assignment of the protons to other histidines was performed automatically using the program pdb2gmx within the GROMACS package. Protonation only on the $N_{\delta 1}$ atom of His69 guarantees the correct hydrogen bonds within the catalytic triad, i.e., the hydrogen bonds His69-$N_{\delta 1}$-H$\cdots$$O_{\delta 2}$-Asp39 and Ser224-$O_{\gamma}$-H$\cdots$$N_{\varepsilon 2}$-His69. The protonations of other histidines were based on the optimal hydrogen-bonding conformation. All atoms in the aromatic rings and the amino group in side chains were converted into virtual sites in order to eliminate fast improper dihedral fluctuations [20]. This modified model of proteinase K was solvated using the single point charge (SPC) water molecules [21] in a dodecahedron periodic box with a 1.4 nm solute-wall minimum distance. After a first steepest descent energy minimization with positional restraints on the solute, five chloride ions were introduced by replacing water molecules at the highest electrostatic potential to compensate for the net positive charge on the protein. This summed to a total of 36,947 atoms for the resulting system. A second energy minimization was performed until no significant energy change could be detected. Subsequently, the system was simulated by five successive 200 ps dynamics runs with decreasing harmonic positional restraint force constants on all protein atoms and the two Ca^{2+} cations (K_{posres}=10,000, 1,000, 100, 10 and 0 kJ mol^{-1} nm^{-2}), followed by the production MD run.

Protein, Ca^{2+} cations, solvent and counter-ions were coupled independently to a reference temperature bath at 300 K with a coupling constant τ_t of 0.1 ps [22]. The pressure was maintained by weakly coupling the system to an external pressure bath at 1 atmosphere with a coupling constant τ_p of 0.5 ps. The non-bonded pair was updated every 20 fs and the non-bonded interactions were calculated using twin range cutoffs of 8 Å and 14 Å. Long-range electrostatic interactions beyond the cutoff were treated with the generalized reaction field model using a dielectric constant of 54 [23]. The LINCS algorithm [24] was used to constrain the bond lengths to their equilibrium positions, in conjunction with the virtual atom for aromatic rings and amino group in side chains, allowing a time step of 2 fs. The production simulation was performed for 20 ns, and coordinates were saved every 8 ps.

Analysis techniques

Conventional structural and geometrical analyses such as the number of hydrogen bonds (NHB), number of native contacts (NNC), number of residues in the secondary structure elements (SSE), radius of gyration (Rg), solvent accessible surface area (SASA) and RMSD were performed using the programs g_hbond, g_mindist, do_dssp [25], g_gyrate, g_sas and g_rms within the GROMACS software package, respectively. The non-bonded energies (Coulomb's electrostatic energy and van der Waals energy) were calculated with the GROMOS96 43a1 force field using a twin range cutoffs of 8 Å and 14 Å with a reaction field correction. The essential dynamics (ED) technique [26, 27] was utilized to investigate large concerted motions in proteinase K. This method is based on the diagonalization of the covariance matrix of atomic fluctuations, which yields a set of eigenvectors and eigenvalues. The eigenvectors indicate directions in a $3N$ (where N=number of atoms) configurational space and describe concerted motions of the atoms. The eigenvalues define the mean square fluctuation of the motion along the eigenvectors. The central hypothesis of ED is that only a few eigenvectors with large corresponding eigenvalues are important for describing the overall motion of a protein. ED analyses were performed on the MD trajectory using the g_covar and g_anaeig programs within GROMACS. Only the main chain atoms N, C_{α}, C and O were included in the analyses.

Identification of the hinge-bending properties of glycines was conducted using the DYNDOM program [28]. This program analyzes conformational changes in terms of the rotational properties of residues. Substructures are identified by clustering each residue's rotation vector between the two extreme structures extracted from the projections of the first three eigenvectors. The glycines are considered to serve as hinge points if they fall within the hinge-bending regions between the identified substructures. The following steps were performed to identify the conservation of glycine: (1) the PSIBLAST program [29] was used to search for related sequences with respect to proteinase K against the non-redundant protein database (SDSC) [30]; (2) the CLUSTALW program [31] was used to obtain a multiple sequence alignment; and (3) the position of each glycine in the multiple sequence alignment was examined. The glycine was classified arbitrarily as absolutely con-

served, strongly conserved, conserved, and non-conserved given a percentage of glycine occupation at a given position of 100%, > 80%, > 50%, and < 50%, respectively.

Results

Geometrical property analyses

To evaluate the structural stability of proteinase K during the MD simulation, geometrical properties such as NHB, SASA, NNC, Rg, RMSD, and SSE are calculated. A comparison of these geometrical properties between the starting structure and the simulated structures is shown in Table 1. The results indicate that the simulated structures show a stable trajectory as the differences in NHB, SASA, NNC, Rg, and SSE are minute. The RMSD values were calculated with respect to the starting structure after superposition on the secondary structure elements as defined by DSSP [25]. Of the three RMSD values shown in Table 1, the "all atom" RMSD has the largest average value (1.96 Å) and standard deviation (SD; 0.25 Å), whereas the RMSD of the secondary structure backbone ("SS bb") has the lowest average value (0.97 Å) and SD (0.11 Å). The average value and standard deviation of the backbone RMSD ("All bb") are intermediate, i.e., 1.34 Å and 0.20 Å, respectively. These results indicate that the backbone atoms, particularly the secondary structure backbone, are relatively stable when compared to the entire molecule during simulation. Stability and equilibration of the system was further examined by plotting the three RMSD values as a function of simulation time (Fig. 2). The largest, medium and lowest fluctuations were also observed in RMSDs of the all atom, the backbone, and the secondary structure backbone, respectively, indicating that the structural deviation originates primarily from fluctuations in the loops/links located between the secondary structure elements. This can be readily seen by superimposing the snapshots extracted from the MD trajectory (Fig. 3). The superimposed snapshots also demonstrate that the two loop regions of residues 100–104 and 160–168 undergo the largest fluctuations. In addition, Fig. 2 also shows a slight

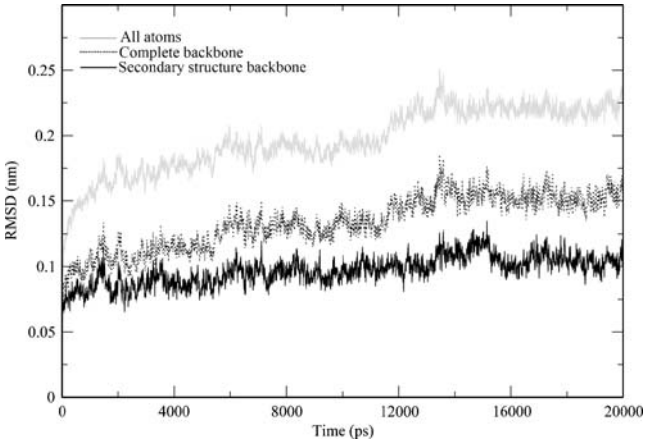

Fig. 2 Time evolution of the root mean square deviations (RMSDs) with respect to the starting structure during molecular dynamics (MD) simulation of proteinase K. RMSDs were calculated every 8 ps after superimposing on the backbone of the secondary structure elements (SSE) as defined by DSSP [25] in the crystal structure 1IC6. The backbone RMSD of the SSE defined in the crystal structure are indicated by the *black line*. The RMSDs of the complete backbone and all atoms are shown by the *dotted* and *grey lines*, respectively

Table 1 Comparison of the geometrical properties of proteinase K between the starting structure 1IC6 and the structural ensemble generated by molecular dynamics (MD) simulation. *NHB* Number of hydrogen bonds, *NNC* number of native contacts, *SASA* total solvent accessible surface area, *Rg* radius of gyration, *RMSD* root mean square deviation, *SSE* secondary structure elements

	NHB[a]	SASA (Å^2)	NNC[b]	Rg (Å)	RMSD[c] (Å)			SSE[g]		
					All bb[d]	SS bb[e]	All atom[f]	α helix	β sheet	Turn
1IC6	223	10992.0	138794	166.9	0	0	0	69	66	43
Ensemble[h]	216 (7)	11225.9 (196.9)	137353 (1003)	167.2 (0.5)	1.34 (0.2)	0.97 (0.11)	1.96 (0.25)	69 (2)	59 (2)	32 (4)

[a] A hydrogen bond is considered to exist when the donor-hydrogen-acceptor angle is larger than 120° and the donor-acceptor distance is smaller than 3.5 Å

[b] A native contact is considered to exist if the distance between two atoms is less than 6 Å

[c] RMSD with respect to 1IC6. RMSD values were calculated with respect to the starting structure 1IC6 after superposition on the SSE as defined by DSSP [25]

[d] Backbone RMSD values of all residues

[e] Backbone RMSDs of the SSE

[f] All atom (including hydrogen atoms) RMSDs

[g] Number of residues in the corresponding SSE

[h] Average geometrical properties calculated over MD trajectory. Standard deviations are shown in parentheses

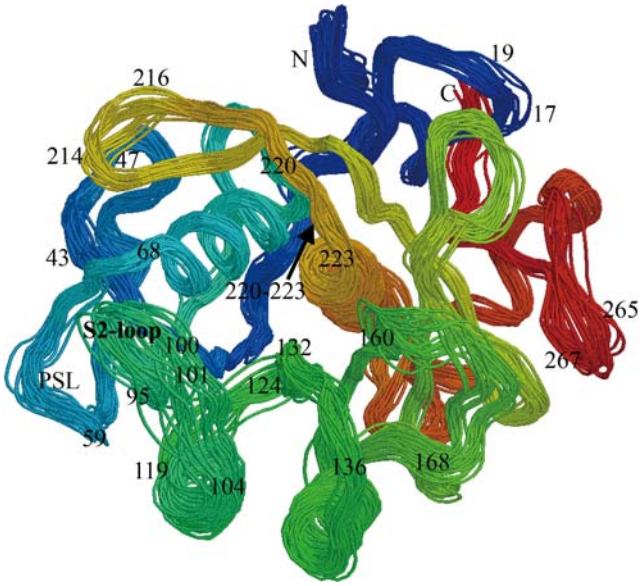

Fig. 3 Snapshots of the structures of the proteinase K during MD simulation. The backbone traces were taken from the MD trajectory every 800 ps. The structures were superimposed on the secondary structure backbone as defined by DSSP [25] in the crystal structure 1IC6. The snapshots are colored from *blue* at the N-terminus to *red* at the C-terminus. Loops involved in significant conformational difference and/or structural variability are labeled

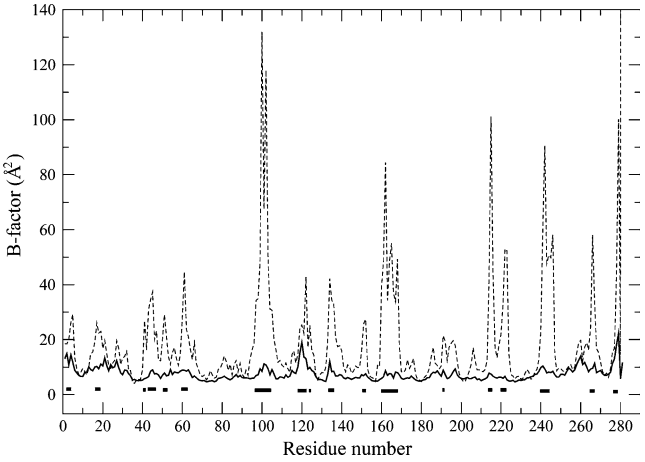

Fig. 4 Comparison of experimental C_α B-factors (*solid lines*) taken from PDB entry 1IC6 with the C_α B-factors (*dashed lines*) calculated from MD simulation of proteinase K. The residue regions with simulated B-factors greater than 20 Å^2 are indicated with *black bars* along the horizontal axis

increase in all three RMSDs from 0 to 2 ns, after which the system appeared to reach equilibration, and therefore the remaining18 ns of the trajectory can be used for subsequent ED analysis.

B-factors and structural flexibility

Crystallographic temperature factors (B-factors) provide one of the few measures of protein flexibility so that simulation studies often attempt to reproduce them in order to test the verisimilitude of the dynamics observed. A comparison of the crystallographic B-factors of 1IC6 with the B-factors calculated from the MD trajectory is shown in Fig. 4. Simulated B-factors are generally larger than those determined crystallographically. There are two possible reasons for this: (1) simulations sample far more conformations of the protein than crystallography; (2) experimental and simulation conditions differ. However, the calculated B-factors do not detract from qualitative agreement with the experimental values as the two curves fluctuate in a very similar way. Specifically, those regions of the molecule that present increased B-factors in the crystal structure also show increased B-factors in the ensemble of simulated structures, and this relationship also holds true for the regions with decreased B-factors, suggesting that the simulation predicts the identity of the mobile regions better than it predicts the degree of their mobility. In short, we consider that the simulated B-factors are a scaled version of the experimental B-factors. The analysis of the protein flexibility, therefore, is focused mainly on the B-factors calculated from the MD trajectory.

The most significant mobile regions of the molecule, arbitrarily defined as those segments with a B-factor greater than 40 Å^2, comprise residues 61, 99–103, 134, 161–162, 164–165, 168, 214–216, 222–223, 241–246, 266, and 278–279 (Fig. 4). Many of them are located within, or close to, the substrate-binding sites (e.g., 99–103, 134, 161–162, 164–165, 168, and 222–223). The remaining residues are located either in the surface-exposed loops/turns or at the C-terminus of the protein. For instance, Tyr61 is located in the polar surface loop (PSL) comprising residues 59–68 [7]; Gly214, Gly215 and S216 are located in a turn between β12 and β13, which form a two-stranded antiparallel β sheet at the protein surface; region 241–246 is located in a loop between α5 and α6; Phe266 is located in a well exposed loop between β14 and β15; and Gln278 and Ala279 are the C-terminal residues of the protein.

The B-factors of the catalytic triad residues, Asp39, His69 and Ser224, are 5.5, 6.9 and 15.0 Å^2, respectively. The relatively smaller fluctuations of Asp39 and His69 are probably caused by a strong electrostatic interaction between them (for details, see the section on "Flexibility of the catalytic triad"), whereas the relatively larger fluctuation of Ser224 can be attributed to the high flexibility of its neighboring residues 220–223. The B-factor values for the strong Ca^{2+}-binding site (Ca1 site) residues, Pro175, Val177 and Asp200, are 10.9, 7.7 and 5.7 Å^2, respectively, and 19.2 and 19.7 Å^2, respectively, for the weak Ca^{2+}-binding site (Ca2 site) residues, Thr16 and Asp260, indicating that the Ca1 site is not involved in significant motion during simulation and that it is more

stable than the Ca2 site. The differences in flexibility between the Ca1 and Ca2 sites are discussed below (see section on "Ca^{2+} motions").

Flexibility of the catalytic triad

The high resolution crystal structure of proteinase K [11] has provided a clear picture of the arrangements of Asp39, His69 and Ser224 in the catalytic triad, in which the two hydrogen bonds, His69-$N_{\delta1}$-H$\cdots$$O_{\delta2}$-Asp39 and Ser224-$O_{\gamma}$-H$\cdots$$N_{\varepsilon2}$-His69 can be clearly observed. In order to further investigate the dynamic nature of the catalytic triad, the distances, hydrogen bond life and interaction energies between Asp39 and His69 and between His69 and Ser224 were calculated over the MD trajectory. Figure 5a shows the distances between the centers of mass of Asp39 and His69, and of His69 and Ser224 as a function of time. It can be seen that the Asp39-His69 distance, which fluctuates around a constant value, is smaller and more stable than the His69–Ser224 distance. This leads to the speculation that the hydrogen bond formed between Asp39 and His69 should be more stable than that between His69 and Ser224. This prediction was confirmed by monitoring the hydrogen bond life during the 20 ns MD trajectory. We found that the $N_{\delta1}$ of the imidazole group in His69 (which acts as the hydrogen donor) can be hydrogen bonded to either $O_{\delta2}$ or $O_{\delta1}$ of the carboxyl group in Asp39. The life of the former and latter hydrogen bonds is 63.9% and 45.2% of the total simulation time, respectively. When hydrogen bonds are considered to exist between the $N_{\delta1}$ of His69 and the carboxyl group ($O_{\delta2}$ or $O_{\delta1}$) of Ser224, they have a life of 80.3%. In contrast, the hydrogen bond Ser224-O_{γ}-H$\cdots$$N_{\varepsilon2}$-His69 occupies only 32.5% of the simulation time. This shorter hydrogen bond life is probably caused by the higher mobility (i.e. a large B-factor) of Ser224 as described above.

The electrostatic and van der Waals interaction energies between Asp39 and His69, and between His69 and Ser224 as a function of time are shown in Fig. 5b and c, respectively. In the case of Asp39–His69 (Fig. 5b), both the electrostatic and van der Waals interaction energies are stable and fluctuate around respective average values of −88.3 and −5.3 kJ mol^{-1}, indicating that the electrostatic force makes a substantial contribution to maintenance of the Asp39–His69 association. In the case of His69–Ser224 (Fig. 5c), although the electrostatic interaction energy is generally lower than the van der Waals energy, both curves show large fluctuations, with the calculated average values (standard deviation) being −15.4 (15.2) and −2.0 (4.3) kJ mol^{-1}, respectively. The weaker and unstable electrostatic interaction of His69–Ser224, when compared to that of Asp39–His69, can explain why His69–Ser224 exhibits a weaker association than His39–Ser69, and may also explain why Ser224 is more flexible than Asp39 and His69.

Ca^{2+} motions

The presence of Ca^{2+} is a common feature of members of the subtilisin family. The stability of Ca1 and Ca2 during simulation was examined by monitoring the distances of the calcium ions from their binding sites (Fig. 6a). The distance between the Ca1 and the center of mass of Ca1 site is rather stable, fluctuating around 0.27 nm (standard deviation= 0.03) with the exception of only a slight increase before 800 ps, whereas the Ca2–Ca2 site distance changes drastically during MD simulation, indicating different

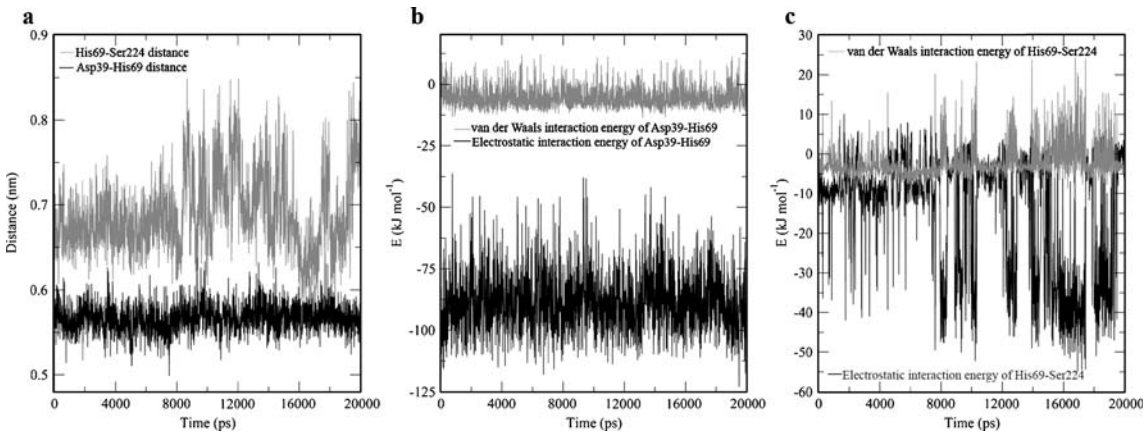

Fig. 5a–c Distances and interaction energies within the catalytic triad. **a** Distances between the centers of mass of Asp39 and His69 (*black line*) and of His69 and Ser224 (*grey line*) as a function of time. **b** Interaction energies between Asp39 and His69 as a function of time. The coulomb's electrostatic and van der Waals (Lennard-Jones) interaction energies of Asp39–His69 are shown in *black* and *grey lines*, respectively. **c** Interaction energies between His69 and Ser224 as a function of time. The coulomb's electrostatic and van der Waals (Lennard-Jones) interaction energies of His69–Ser224 are shown in *black* and *grey lines*, respectively. The energies are the sum of the short (SR) and long range (LR) terms; the 1–4 terms are not included

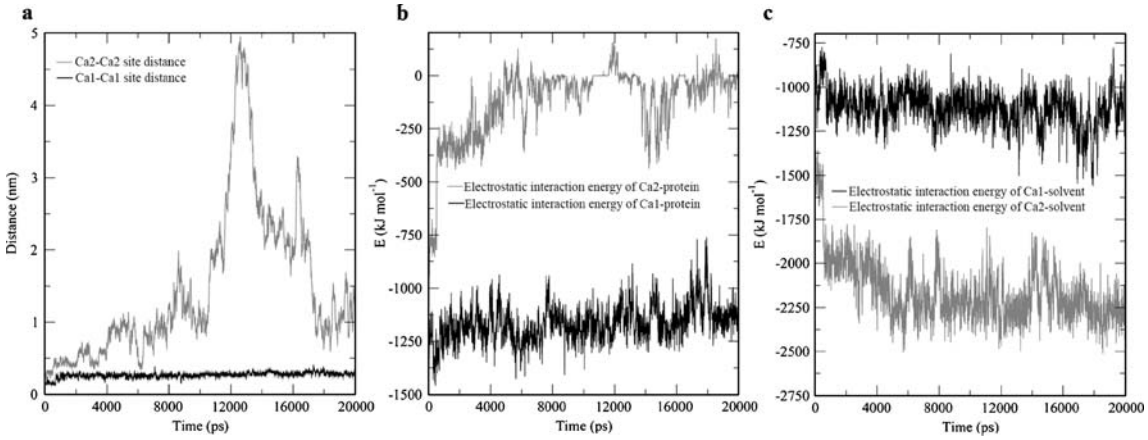

Fig. 6a-c Distances and interaction energies between the calcium cations and the protein and between the calciums and the solvent. **a** The distances between the two calcium cations and the centers of mass of their corresponding binding sites as a function of time. The distances of the Ca1–Ca1 site and Ca2–Ca2 site are indicated by *black* and *grey lines*, respectively. **b** Electrostatic interaction energies between the two calcium cations and the protein. The interaction energies of the Ca1–protein and the Ca2–protein are indicated by *black* and *grey lines*, respectively. **c** Electrostatic interaction energies between the two calcium cations and the solvent. The interaction energies of the Ca1–solvent and the Ca2–solvent are indicated in *black* and *grey lines*, respectively. The energies are the sum of the short (SR) and long range (LR) terms; the 1–4 terms are not included

protein-binding affinity of the two calcium cations. In order to ascertain the factors responsible for such a difference, the electrostatic interaction energies between the Ca^{2+} cations and the protein and between the Ca^{2+} cations and the solvent were calculated. As shown in Fig. 6b, the Ca1-protein interaction energy is relatively stable, fluctuating around an average value of −1,163.9 (standard deviation= 171.1) kJ mol^{-1}, whereas the Ca2-protein interaction energy is rather unstable, increasing from −948.0 to 120.5 kJ mol^{-1} at 0–5.6 ns and subsequently fluctuating drastically around 0 kJ mol^{-1} (average value and standard deviation are −120.0 and 190.5 kJ mol^{-1}, respectively). These results indicate that the Ca1–protein interaction is stronger and more stable than the Ca2–protein interaction. The electrostatic interaction energies between the two Ca^{2+} cations and the solvent are shown in Fig. 6c, which indicates that the Ca2–solvent interaction is stronger than the Ca1–solvent interaction, with the average values (standard deviations) being −2,160.9 (171.1) and −1,118.0 (107.5) kJ mol^{-1}, respectively. Accordingly, we can conclude that it is the competition between the Ca2–protein and Ca2–water molecule interactions that gives rise to fluctuation of Ca2. The stronger electrostatic interaction of Ca2 with water, compared to that with protein, drives this Ca^{2+} to diffuse away from the protein, whereas the interaction between Ca1 and the water molecules cannot overcome the interaction between Ca1 and the protein, thus keeping Ca1 at its binding site. The differences in flexibility between the Ca1 and Ca2 sites, i.e., the higher thermal motions (B-factors) for the Ca2 site than for the Ca1 site, may arise from the distinct dynamic behaviors of these two Ca^{2+} cations. These results are in agreement with published structural and biochemical investigations [12, 14], suggesting that the weak Ca^{2+}-binding site Ca2 is only partially occupied by calcium in solution structure.

It should be pointed out that the van der Waals interactions of Ca^{2+}–protein and Ca^{2+}–solvent are minute (about 35 and 90 kJ mol^{-1}) in comparison to the electrostatic interactions. Therefore, we consider that the van der Waals interaction has a negligible effect on Ca^{2+} binding affinity.

Large concerted protein motion

Essential dynamics was used to investigate the large concerted motions in proteinase K. The covariance matrix built from atomic fluctuations in the last 18 ns MD trajectory was diagonalized and the eigenvectors and their corresponding eigenvalues were obtained. Here, we focused mainly on the first three eigenvectors, as it has been shown that the overall internal motion of the protein can be described adequately by using only a few degrees of freedom [26, 27].

Figure 7 shows the projection extremes of the first three eigenvectors. The large concerted motions described by the first three eigenvectors are shown in Animations 1–3, respectively, in the electronic supplementary material (ESM). For eigenvector 1, the large concerted motions originate mainly from displacements of the segments 100–105, 160–169, 214–216, 221–223 and 240–246 (Fig. 7a, and Animation 1). Of these, segment 100–105 forms part of the substrate-binding pocket S4; segments 160–169 and 221–223 are parts of the substrate-binding pocket S1; and segment 240–246 is located between α6 and α5, at whose

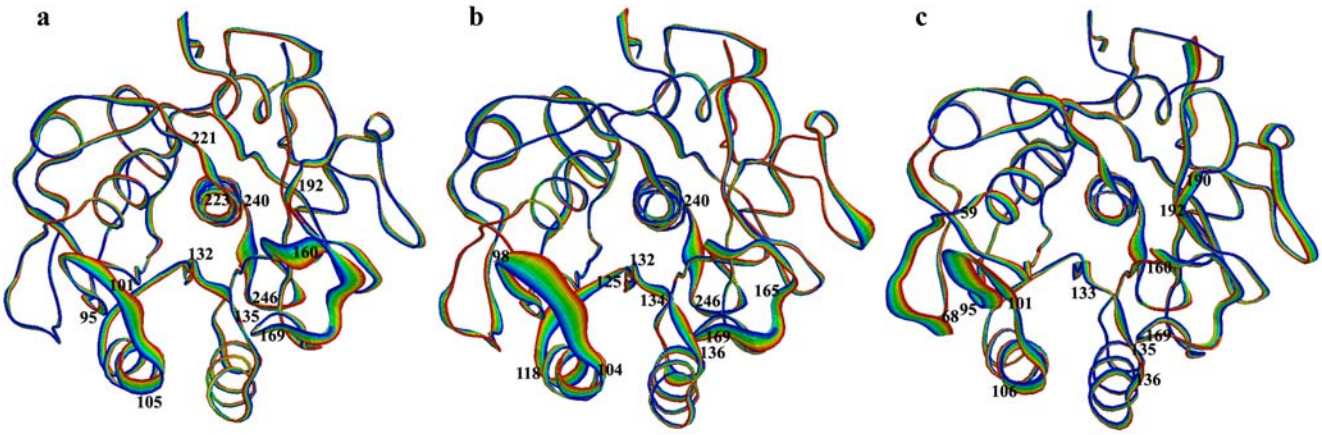

Fig. 7 Large concerted motions of proteinase K described by **a** eigenvector 1, **b** eigenvector 2 and **c** eigenvector 3. The linear interpolations between the two extremes extracted from projection of the trajectory onto the eigenvectors are colored from *blue* to *red* to highlight the structural differences between the two states, but do not represent the transition pathway

end the catalytic residue Ser224 is located. Similarly, the large concerted motions described by eigenvectors 2 and 3 also arise from displacements of the residues that are within (or closed to) the substrate-binding pockets, or are located in regions opposite the substrate-binding sites. The most significant effect of these motions is that they can influence dynamic variations of the substrate-binding regions, leading to the opening or closing of the S1 and S4 pockets. Another interesting effect is that large concerted motions in regions opposite the substrate-binding sites can mediate or modulate conformational changes in substrate-binding regions. For example, for eigenvector 1, both the segments 221–223 and 240–246 exhibit large conformational displacements (Fig. 7a). As described above, the segment 240–246 is linked to the segment 221–223 through helix α5 and this helix shows only a small fluctuation. We therefore consider that the large displacement of segment 221–223 is mediated by structural changes in loop 240–246, indicating that concerted motion opposite the active Ser224 and S1 pocket can affect the dynamics of the active site.

In the case of the collective motions described by eigenvector 2 (Fig. 7b), a large displacement is observed in the loop region 118–125, which is located between the α3 that follows segment 100–104 and the β4 that precedes segment 132–136. Both the α-helix and the β-strand show only small fluctuations, indicating that displacement of segments 100–104 and 132–136 may correlate with conformation changes in the loop region 118–125. This segment, like segment 240–246, may also play a role in modulating the dynamics of the substrate-binding site opposite its location. For concerted motions described by the third eigenvector (Fig. 7c), the region of residues 59–68 (termed PSL because it is a surface-exposed loop comprising several polar residues [7]) exhibits the largest fluctuation in concert with displacement of loop S2 (composed of residues 95–101). Such concerted motions between PSL and the S2-loop could, on the one hand, avoid steric hindrance between them as they are spaced by only ~3 Å at the narrowest part of the two loops; on the other hand, they could modulate the flexibility and orientation of the S2-loop, which plays a critical role in recognition and binding of the P2 and P3 residues of the substrate [7].

Dynamic properties of glycines

It has been suggested that internal protein motions are often correlated with the locations of glycines in the protein structure as their hinge-bending properties allow substructures in protein to fluctuate with respect to each other [32–34]. There are 33 glycines in proteinase K. For these glycines, the RMSD values calculated from the minimum and maximum extremes of motions along the first three eigenvectors, their locations in the tertiary structure of proteinase K, their conservation and the results for identification of their hinge-bending properties are shown in Table 2. Absolutely conserved glycines are observed at positions 30, 41, 68, 75, 83, 100, 102, 134, 160, 196, 205, 222 and 232. Most of them are identified to have hinge-bending properties when the conformation changes between the two extremes of some of the first three eigenvectors, except for the glycines at positions 100, 102 and 134. Interestingly, most of the absolutely conserved glycines are located in the loop and/or link regions between secondary structure elements with the exception of Gly75 and Gly232, which are located in the middle parts of two α helices, α2 and α5. These two helices are spaced by ~ 9 Å in the crystal structure of proteinase K, and have at their beginnings the two catalytic residues, His69 and Ser224, respectively. Therefore, the hinge-bending properties of the two glycines may impart a certain structural flexibility to the two helices, allowing them to adjust the positions of

Table 2 Statistical analyses of the geometrical/functional properties of the glycines in proteinase K and their conservation in related sequences

Glycine[a]	9	19	30	32	41	51	66	68	70	75	78	83	92	100	102	110	126	134	135	136
Eigenvector 1[b] (Å)	0.62	0.79	0.04	0.64	1.37	0.42	0.63	0.45	0.29	0.52	0.49	0.57	0.46	1.74	2.94	0.58	0.24	1.52	0.75	0.18
Eigenvector 2[b] (Å)	0.42	0.36	0.42	0.24	0.93	1.11	1.07	0.24	0.36	0.27	0.56	0.41	0.22	2.27	2.60	0.41	0.50	1.36	1.48	0.87
Eigenvector 3[b] (Å)	0.36	0.23	0.50	0.48	0.69	0.35	0.95	0.72	0.33	0.50	0.74	0.72	0.21	4.34	2.34	0.19	0.34	1.10	0.65	0.97
Conservation[c]	SC	NC	AC	SC	AC	SC	SC	AC	SC	AC	C	AC	NC	AC	AC	SC	SC	AC	SC	SC
Location[d]	H	L	L	L	L	L	L	L	H	H	H	L	S	L	L	H	L	L	S	S
Hinge-point[e]	+	+	+	+	+	+	−	+	+	+	+	+	+	−	−	+	+	−	+	+
Glycine[a]	152	160	181	196	203	205	214	215	222	232	241	259	267							
Eigenvector 1[b] (Å)	0.73	1.81	0.39	1.29	0.55	0.87	1.45	2.26	2.16	0.60	1.70	0.60	1.23							
Eigenvector 2[b] (Å)	1.27	0.40	0.13	0.40	0.21	0.43	0.56	1.28	0.48	0.37	1.18	0.48	0.36							
Eigenvector 3[b] (Å)	0.34	0.87	0.36	0.43	0.27	0.28	0.85	0.84	1.34	0.27	1.04	0.47	1.31							
Conservation[c]	SC	AC	SC	AC	NC	AC	C	NC	AC	AC	NC	NC	NC							
Location[d]	L	L	S	L	S	L	L	L	L	H	L	L	L							
Hinge-point[e]	+	+	+	+	+	+	−	−	+	+	−	−	−							

[a] Positions of glycines in the amino acid sequence of proteinase K

[b] RMSD values of glycines calculated from minimum and maximum extremes of the motions along the corresponding eigenvectors

[c] Conservation of glycines among proteinase K-related sequences obtained from PSIBLAST search. AC, SC, C and NC indicate that the glycine is absolutely conserved, strongly conserved, conserved and non-conserved at a position within the multiple sequence alignment, respectively

[d] Locations of glycines in the tertiary structure of proteinase K. H, S and L indicate that the glycine is located in the α helix, β strand and loop, respectively

[e] '+' indicates that the glycine acts as a hinge point when conformation changes between the two extremes along any of the three motional modes. '−' indicates that the glycine does not serve as the hinge point when conformation changes between the two extremes in any mode described by the first three eigenvectors

His69 and Ser224 concertedly to facilitate the formation of a hydrogen bond between these two residues.

We also note that most of these absolutely conserved glycines are either neighbored by residues in the catalytic triad or located at, or close to, substrate-binding sites. For instance, Gly41, Gly68 and Gly222 are found in the vicinity of the catalytic triad residues Asp39, His69 and Ser224, respectively, probably providing a certain flexibility for the active triad residues to guarantee the correct orientation between them. A strongly conserved glycine, Gly70 (conservation at this position is 99%), is also found adjacent to His69. Since His69 needs to form hydrogen bonds with both Asp39 and Ser224, the presence of glycines before and after His69 is likely to provide more space for its imidazole group to favor these hydrogen-bonding conformations. The absolutely conserved Gly100 and Gly102 are located in the substrate-binding segment 100–104; Gly134, together with two strongly conserved glycines, G1y135 (conservation 98%) and Gly136 (conservation 86%), are located in the substrate-binding segment 132–136; and Gly160 is located close to the substrate-binding pocket S1. These glycines have relatively high RMSD values (Table 2), reflecting large conformational displacements in motional modes described by the first three eigenvectors. On the other hand, these glycines also make a substantial contribution to the flexibility of the substrate-binding sites. Also worth noting is the absolutely conserved glycine at position 160, neighboring the oxyanion hole residue, Asn161; Gly160 may provide more space for the side chain of Asn161 to search for the oxyanion in the tetrahedral intermediate.

Two strongly conserved glycines, Gly126 (conservation 94%) and Gly152 (conservation 99%) are found in the loop regions opposite the substrate-binding segment 132–136. Gly126 is followed by strand β6 that precedes segment 132–136; Gly152 is linked to segment 132–136 through helix α4. Both β6 and the α4 show only small fluctuations (Fig. 4), suggesting that the hinge-bending properties of Gly126 and Gly152 might mediate structural changes in the substrate-binding segment 132–136. The N-terminus of 132–136 is linked to the 9-stranded β sheet through β6, and its C-terminal residues 135–136 form an antiparallel β sheet with residues 169–170. This may explain why this segment has a lower conformational freedom than the other well-exposed substrate-binding segment 100–104 (Fig. 4). However, the three consecutive glycines, Gly134, Gly135 and Gly136, in conjunction with the two hinge-bending glycines, Gly126 and Gly152, could aid in enhancing the flexibility of this segment. Such flexibility is necessary for proteinase K to accommodate the large P1 and P4 residues of the substrate as segment 132–136 participates in formation of both the S1 and S4 pockets.

Another notable glycine in proteinase K is Gly241. Despite its weak conservation (~ 5%), Gly241 has a markedly large B-factor during MD simulation. In addition, this glycine is located in the loop region 241–246 that is linked to the catalytically active residue Ser224 through the rigid helix α5. This implies that Gly241 may enhance the flexibility of segment 241–246, which, in turn, as suggested above, increases the conformational freedom of Ser224 that is required for the catalytic process.

Discussion

Like those of other members of the subtilisin-like protease family, the three-dimensional structure of the serine protease proteinase K presents a well-defined global fold and is often considered as a “rigid body” protein [35]. However, distortions in the substrate-binding region of subtilisin crystal structures have been observed that have been attributed traditionally to crystal contacts [36, 37]. An NMR structural study of serine protease PB92 revealed large local conformational differences in substrate-binding regions between solution conformers—an indication that structural flexibility exists within the substrate-binding region of this group of enzymes [14]. As a result, we believe detailed information on the dynamics of proteinase K would greatly facilitate understanding of the structure–function relationship of this important class of enzymes. Thus, we performed a 20 ns MD simulation on proteinase K with two bound Ca^{2+} cations using explicit SPC water.

Geometrical property and B-factor analyses for proteinase K during simulation revealed a rigid structural core with the exception of a limited number of flexible loops/links. Among the limited number of loops/links involved in the largest conformational fluctuations are those of regions located in the substrate-binding sites. The presence of flexibility in the substrate-binding region, which has emerged from studies of many enzymes, supports the proposed induced-fit mechanism of substrate binding [38]. We surmise that the highly flexible substrate-binding sites may play a role both in modulating the thermodynamics and kinetics of the enzyme-substrate interaction so that the substrates can be effectively recognized and bound, and in allowing interactions with substrates with a variety of sequence motifs, thus broadening the substrate specificity of proteinase K.

Although the thermal motions of the side chains of the three catalytic triad residues may be identical [11], our MD simulation reveals a stronger electrostatic interaction between His69 and Asp39 than between His69 and Ser224, resulting in a more stable and tighter association of Asp39–His69 than His69–Ser224. This implies that proton transfer between His69 and Asp39 should be more effective than

that between His69 and Ser224. We therefore speculate that proton transfer from the $N_{\delta 1}$ of His69 to the carboxyl group of Asp39 may be a prerequisite for proton transfer from the Ser224 O_{γ} to the His69 $N_{\varepsilon 2}$. This speculation is supported by results obtained from ab initio molecular orbital calculations of the catalytic triad [39], which indicate that the hydrogen transfer energy between the serine hydroxyl and the $N_{\varepsilon 2}$ of the histidine imidazole is higher than that between the $N_{\delta 1}$ of the histidine imidazole and the carboxyl group of aspartate. In addition, the stronger electrostatic/hydrogen-bonding interaction between His69 and Asp39 is also important for the correct orientation of the imidazole group in the triad. Upon proton abstraction from $N_{\delta 1}$ of the correctly orientated imidazole by the aspartate carboxyl group, the pK_a and the alkalinity of the imidazole group increase somewhat [40], thus making it ready for extraction of a proton from the serine O_{γ}. The relatively weaker interaction between serine and histidine residues leaves the serine hydroxyl with a relatively large conformational freedom. This observation is supported by the unfavorable geometry at the O_{γ} and/or the $N_{\varepsilon 2}$ in some crystal structures of both free and inhibited enzymes [41, 42]. The high flexibility of the serine hydroxyl is necessary as it needs to assume different orientations appropriate for either proton transfer or nucleophilic attack and, subsequently, for release of the cleaved peptide product.

Our MD simulation revealed that competition of electrostatic interactions between Ca2–water and Ca2–protein leads to diffusion of Ca2 away from the protein. It is possible that the occupation of Ca2 in the crystal structure might be stabilized by the cryocooling of the crystallization condition, whereas the relatively higher simulation temperature raises the possibility of Ca2 diffusion. Although diffusion of Ca2 from the protein gives rise to thermal fluctuation of its binding residues, the overall thermal stability of the simulated structure appears not to be affected. Several stable hydrogen bond and salt bridge networks, which were observed to be centered around the Ca2-binding site during simulation (data not shown), contribute to the stability of the Ca2 site. On the other hand, such bond networks can also help explain why the presence of EDTA has only a minor effect on the catalytic activity of proteinase K-like proteases.

Motions along the first few eigenvectors are mainly large concerted fluctuations and can be generally linked to the functional properties of proteins [43–47]. Large concerted motions within the substrate-binding regions can result in opening or closing of the substrate-binding pockets, which may facilitate binding or release of the peptide substrate/product. Interestingly, the concerted motions located opposite the substrate-binding pockets may also exert an effect on the dynamics of the substrate-binding pockets. This is further reflected by the hinge-bending motions (mediated mainly by glycines) between correlated segments, in agreement with previously published NMR [14] and simulation data [32].

Glycines play an important role in determining protein flexibility. It has been observed that single-site mutations of glycines, or mutation of non-glycine residues at other sites to glycines, can have remarkable effects on enzymatic action, such as changes in catalytic activity and substrate specificity [48, 49]. Such observations suggest that the absence of conserved glycines in different structures can cause altered protein flexibility and hence different modes of substrate recognition. In the case of proteinase K-like serine proteases, most of the conserved glycines are located either adjacent to the catalytic triad residues or within/opposite the substrate-binding regions. This has led to the predictions that these glycines could (1) provide a certain flexibility for the catalytic triad residues so that they can be shifted to the correct orientation relative to each other and to the substrate for the catalytic reaction to take place; (2) serve as hinge points to allow catalytically crucial regions to fluctuate relative to each other; (3) guarantee structural flexibility of the substrate-binding segments for induced fit. In addition, we also note that there are 40 absolutely conserved positions in the multiple sequence alignment of sequences homologous to proteinase K. Of these, 13 positions (32.5%) are occupied by glycines. This observation implies that glycine residues play important roles in determining not only the structural flexibility but also the folding architecture of this class of enzymes. However, a full understanding of the mechanism of how glycines affect the flexibility and the folding architecture of proteinase K requires further investigation using both experimental (such as site-directed mutagenesis) and theoretical methods—a huge workload that is beyond the scope of the current presentation.

Conclusions

In summary, during simulation proteinase K presents a well-defined rigid structural core with the exception of a limited number of surface-exposed loops/residues. It is possible that the high rigidity of the internal core of the molecule has evolved as a protective measure against autolysis. Among the limited number of loops/residues exhibiting significant mobility, many are directly involved in substrate binding, indicating that the process of enzyme–substrate binding is an induced fit. In particular, the large concerted motions in certain regions close to or opposite substrate-binding regions could also modulate the dynamic behavior of the substrate-binding pockets, especially regarding the two pockets S1 and S4. The high flexibility of these two pockets, together with their large size, could

explain the broad P1 and P4 selectivity of proteinase K towards substrates. The relatively strong electrostatic interaction between Asp39 and His69 contributes substantially to the maintenance of the Asp39–His69 association and to the orientation of the imidazole group of His69. The stable orientation of the imidazole group, which is ready for extraction of a proton from the hydroxyl group of Ser224, is necessary as Ser224 exhibits a higher conformational freedom than His69. Furthermore, the relatively higher conformational freedom of Ser224 can be related to its multifunctional roles in proton transfer, nucleophilic attack, and product release. Finally, analyses of the dynamic properties and conservation of glycines imply that glycines are not only essential in determining structural flexibility of substrate-binding sites, but are also important for the folding architecture of this class of enzymes.

Acknowledgments We thank the High Performance Computer Center in Yunnan University for computational support. This work was funded by the National Natural Science Foundation of China (approved numbers 30630003 and 30860011) and partially supported by grants from Yunnan Province (2006C008M, 2007C163M, 07Z10756, 2007PY-22 and 08Y0026) and Innovation Group Project from Yunnan University (KL070002).

References

1. Ebeling W, Hennrich N, Klockow M, Metz H, Orth HD, Lang H (1974) Eur J Biochem 47:91–97
2. Wiegers U, Hilz H (1971) Biochem Biophys Res Commun 44:513–519
3. Pahler A, Banerjee A, Dattagupta JK, Fujiwara T, Lindner K, Pal GP, Suck D, Saenger W (1984) EMBO J 3:1311–1314
4. Wells JA, Estell DA (1988) Trends Biochem Sci 13:291–297
5. Hilz H, Wiegers U, Adamietz P (1975) Eur J Biochem 56:103–108
6. Shaw WV (1987) Biochem J 246:1–17
7. Liu SQ, Meng ZH, Yang JK, Fu YX, Zhang KQ (2007) BMC Struct Biol 7:33
8. Pantoliano MW, Ladner RC, Bryan PN, Rollence ML, Wood JF, Poulos TL (1987) Biochemistry 26:2077–2083
9. Siezen RJ, de Vos WM, Leunissen JA, Dijkstra BW (1991) Protein Eng 4:719–737
10. Siezen RJ, Leunissen JAM (1997) Protein Sci 6:501–523
11. Betzel C, Gourinath S, Kumar P, Kaur P, Perbandt M, Eschenburg S, Singh TP (2001) Biochemistry 40:3080–3088
12. Müller A, Hinrichs W, Wolf WM, Saenger W (1994) J Biol Chem 269:23108–23111
13. Wolf WM, Bajorath J, Müller A, Raghunathan S, Singh TP, Hinrichs W, Saenger W (1991) J Biol Chem 266:17695–17699
14. Martin JR, Mulder FAA, Karimi-Nejad Y, van der Zwan J, Mariani M, Schipper D, Boelens R (1997) Structure 5:521–532
15. Dodson G, Wlodawer A (1998) Trends Biochem Sci 23:347–352
16. Betzel C, Pal GP, Saenger W (1988) Eur J Biochem 178:155–171
17. Betzel C, Teplyakov AV, Harutyunyan EH, Saenger W, Wilson KS (1990) Protein Eng 3:161–172
18. Kutzner C, van der Spoel D, Fechner M, Lindahl E, Schmitt UW, de Groot BL, Grubmuller H (2007) J Comp Chem 28:2075–2084
19. Lindahl E, Hess B, van der Spoel D (2001) J Mol Model 7:306–317
20. Feenstra KA, Hess B, Berendsen HJC (1999) J Comp Chem 20:786–798
21. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342
22. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR (1984) J Chem Phys 81:3684–3690
23. Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) J Chem Phys 102:5451–5459
24. Hess B, Bekker H, Berendsen HJC, Fraaije J (1997) J Comp Chem 18:1463–1472
25. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
26. Amadei A, Linssen ABM, Berendsen HJC (1993) Proteins 17:412–425
27. Balsera MA, Wriggers W, Oono Y, Schulten K (1996) J Phys Chem 100:2567–2572
28. Hayward S, Berendsen HJC (1998) Proteins 30:144–154
29. Altschul SF, Madden TL, Schaffer AA (1997) Nucleic Acids Res 25:3389–3402
30. Subramaniam S (1998) Proteins 32:1–2
31. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673–4680
32. Peters GH, Frimurer TM, Andersen JN, Olsen H (1999) Biophys J 77:505–515
33. Vreede J, van der Horst MA, Hellingwerf KJ, Grielaard W, van Aalten DMF (2003) J Biol Chem 278:18434–18439
34. van Aalten DMF, Haker A, Hendriks J, Hellingwerf KJ, Joshua-Tor L, Crielaard W (2002) J Biol Chem 277:6463–6468
35. Horn JR, Ramaswamy S, Murphy KP (2003) J Mol Biol 331:497–508
36. Sobek H, Hecht HJ, Aehle W, Schomburg D (1992) J Mol Biol 228:108–117
37. Dauberman JL, Ganshaw G, Simpson C, Graycar TP, McGinnis S, Bott R (1994) Acta Cryst D 50:650–656
38. Koshland DEJ (1958) Proc Natl Acad Sci USA 44:98–104
39. Nishihira J, Tachikawa H (1996) J Theor Biol 196:513–519
40. Moult J, Sussman F, James MNG (1985) J Mol Biol 182:555–566
41. Dauter Z, Betzel C, Genov N, Pipon N, Wilson KS (1991) Acta Cryst B 47:707–730
42. Rypniewski WR, Dambmann C, von der Osten C, Dauter M, Wilson KS (1995) Acta Cryst D 51:73–84
43. van Aalten DMF, Findlay JBC, Amadei A, Berendsen HJC (1996) Protein Eng 8:1129–1135
44. Mello LV, de Groot BL, Li S, Jedrzejas MJ (2002) J Biol Chem 277:36678–36688
45. Barrett CP, Noble ME (2005) J Biol Chem 280:13993–14005
46. Liu SQ, Liu CQ, Fu YX (2007) J Mol Graphics Model 26:306–318
47. Liu SQ, Liu SX, Fu YX (2008) J Mol Model 14:857–870
48. Jancso A, Szent-Györgyi AG (1994) Proc Natl Acad Sci USA 91:8762–8766
49. Vermersch PS, Tesmer JJG, Lemon DD, Quiocho FA (1990) J Biol Chem 265:16592–16630