Theor Chem Acc (2012) 131:1224
DOI 10.1007/s00214-012-1224-9

REGULAR ARTICLE

Insights into conformational changes of procarboxypeptidase A and B from simulations: a plausible explanation for different intrinsic activity

Jitrayut Jitonnom · Adrian J. Mulholland

Received: 21 January 2012 / Accepted: 17 April 2012 / Published online: 1 May 2012

Abstract Different forms of carboxypeptidase proenzymes (zymogens) are observed experimentally to show different behavior: procarboxypeptidase A (proCPA, forms proCPA1 and proCPA2) exhibit some activity against small substrates, but proCPB does not. In this work, these three zymogen forms (subtypes A1, A2 and B) are investigated by means of 15-ns molecular dynamics simulations and principal component analysis to shed light on their dynamic/conformational behaviors that may be relevant to those experimental observations. The simulations revealed that proCPA (both A1 and A2) shows different conformational behavior from proCPB: the former undergoes a major conformational change (opening and closing), and the latter exhibits only a minor conformational change (remaining closed throughout the simulation). Differences center on the interface between the globular moiety of the pro-segment and the catalytic domain. Analysis of the trajectories demonstrates the importance of hydrogen bonds and salt-bridges in stabilizing the zymogen structures and shows different hydrogen-bond patterns between proCPA and proCPB: the former shows fewer strong H-bonds formed between the globular domain and the catalytic domain. The observed difference in conformational behavior between proCPA and proCPB may explain why small substrates and inhibitors can access the active sites of proCPA1 and proCPA2 but not of proCPB.

Keywords Zymogen · Intrinsic activity · Procarboxypeptidase · Molecular dynamics · Principal component analysis · Conformational change

1 Introduction

Pancreatic and pancreatic-like carboxypeptidases (CPs) are a family of zinc-containing exopeptidases, involved in the digestion process, which catalyze hydrolysis of C-terminal amino acids from their substrates [1–3]. They were traditionally classified into types A and B form based on their different substrate specificity: carboxypeptidase A (CPA) has a stronger preference for aliphatic and aromatic C-terminal residues, and carboxypeptidase B (CPB) for basic C-terminal residues [4]. CPA was later split into subtypes CPA1 and CPA2, with the former preferring smaller aromatic and aliphatic C-terminal residues and the latter the bulkier aromatic C-terminal residues [5, 6].

Carboxypeptidases are secreted as inactive zymogens (proenzymes), known as procarboxypeptidases (proCPs) and are activated by the cleavage of another protease (e.g., enteropeptidase for CPA and trypsin for CPB), thereby releasing their pro-domains to produce the active forms [7, 8]. The common structure of proCPs (Fig. 1) is formed by a pro-segment of about 90–95 residues, which inhibits the mature enzyme by shielding the active site, connected to a CP moiety of about 300 residues via a pro-mature junction (or a connecting loop). The pro-segment consists of a globular region with an open-sandwich antiparallel-α antiparallel-β topology and a connecting segment which is a C-terminal alpha-helix that connects to the pro-mature

Electronic supplementary material The online version of this article (doi:10.1007/s00214-012-1224-9) contains supplementary material, which is available to authorized users.

J. Jitonnom (✉)
Department of Chemistry, School of Science,
University of Phayao, Phayao 56000, Thailand
e-mail: jitrayut.018@gmail.com

A. J. Mulholland
Centre for Computational Chemistry, School of Chemistry,
University of Bristol, Bristol BS8 1TS, UK

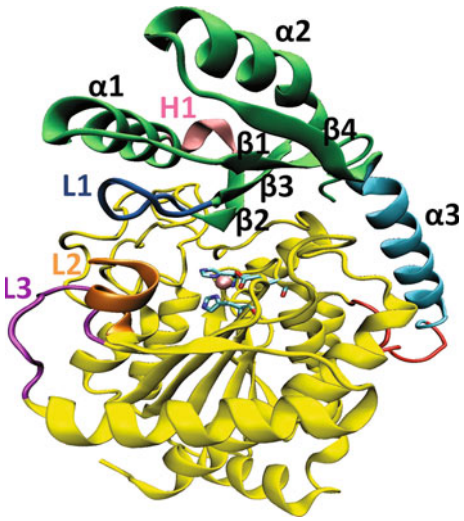

Fig. 1 Cartoon representation of the common structure of procarboxypeptidases. The catalytic domain is shown in *yellow*, which is linked by the pro-mature junction (in *red*) to the pro-segment (activation domain): the individual regions of the pro-segment, the globular domain and C-terminal alpha-helix of the connecting segment are shown in *green* and *cyan*, respectively. The Zn^{2+} is shown as a VDW sphere in *pink,* and the active site residues are shown in a stick representation, colored according to the element type (carbon in *cyan*, nitrogen in *blue* and oxygen in *red*). Alpha-helices and beta-strands on the pro-segment are named in sequence order. Other labels (H1 (helix 1), L1–L3 (loops 1–3)) used in this study are also indicated

junction, whereas the catalytic domain with α/β hydrolase topology has a preformed active site, which is shielded by the globular domain of the pro-segment [4, 6, 9], containing an active center Zn^{2+} atom essential for catalysis.

In addition to the inhibitory role of the pro-segment, it may also act as a postactivation mechanism of activity control [5], constitute important elements for proper folding of the enzyme [7]. It inhibits the proenzyme by preventing access of substrates to the active site cleft of the enzyme. However, a topic of particular interest is that some substrates (e.g., dipeptides and tripeptides) are small enough to penetrate inside the active site and undergo reaction. This has been observed in proCPA, but different intrinsic activity was shown in proCPB: proCPA1 and proCPA2 exhibited some activity against small substrates, but proCPB apparently did not [5]. Further evidence also supported this observation by structural comparison [6], showing the binding pocket of proCPA2 to be relatively large compared to that of proCPB. An inhibitor, benzylsuccinate, can diffuse into the active site of procarboxypeptidase A2 in crystals. However, analysis of crystal structures alone does not account for dynamic behavior of a protein that may be relevant to the different intrinsic activity [10]. The predominant conformation in solution may also be different from that observed crystallographically [11]. Important questions still remain and need to be addressed in order to fully understand the function of the proenzymes A1, A2 and B such as: What is the source of the different intrinsic activity between the proCP zymogens A and B, and do these relate to different dynamics and conformational behavior of these proteins? How do small substrates penetrate into the active site cleft of proCPA1 and proCPA2? Which residues play an important role in stabilizing the zymogen structures of the A and B forms of CP? These proenzymes have also been the subject of several structural studies [1, 5, 6, 9, 12, 13], focusing on their inhibition and activation. However, such processes are far more complex than the static picture provided by crystal structures [5]. Therefore, it is of great importance to gain insight into their dynamic behavior, which may be an important factor in the difference in intrinsic enzymatic activity between the proCP zymogens A and B. Understanding their dynamics may also perhaps be useful for studying understanding their activation processes.

Molecular dynamics (MD) simulations have been proved to be a very useful tool for understanding the structure and dynamics of biological systems at the atomic level [14–17]; among many applications of such biomolecular simulations are studying molecular interactions in a protein and its complex [18], helping in protein modeling and drug design [19, 20] and also providing initial structures for the study of enzymatic reaction [21–23]. They have previously been applied to study the structure, dynamic behavior and conformational change/transition associated with activation processes of other important zymogens: for example, matrix metalloprotease [24], human coagulation FVII and FX [25, 26], procaspase 3 [27], beta-secretase zymogen [28] and plasmepsin II [29, 30]. Despite a number of MD studies of the active CPs [31–35], their application to the investigation of dynamics and function of the whole proCPs has not been reported previously, which leads to the objective of this study.

Here, we performed three separate explicitly solvated MD simulations of proCPA1, proCPA2 and proCPB using available crystal structures. Collective motion along the MD trajectories of the three simulated systems was evaluated by principal component analysis to capture the importantly correlated motion that may be responsible for the experimental observation of different intrinsic activity. Important interactions including hydrogen bond (H-bond) and salt-bridges between the pro-segment and the catalytic domain for each system were comparatively evaluated. Analysis of the nature of the conformational change along the first eigenvector from PCA is also discussed in terms of H-bond interactions and conformational change. This study is, we believe, the first MD simulation study of these proenzymes and provides new insights into solution structures, dynamic behaviors and conformational changes of proCPA1, proCPA2 and proCPB.

2 Computational details

2.1 Molecular dynamics simulations

All molecular dynamics (MD) simulations were carried out using the MD software package AMBER 9 [36] with the ff03 force field [37]. The starting structures of the zymogens (proCPs) used in this study were taken from the crystallographic X-ray structures deposited in the Protein Databank (PDB) (proCPA1, PDB code 1PCA, 2.00 Å resolution, "pig" species [9]; proCPA2, PDB code 1AYE, 1.80 Å resolution, "human" species [6]; and proCPB, PDB code 1NSA, 2.30 Å resolution, "pig" species [4]). Small molecules such as citric acid and valine in proCPA1 and benzamidine in proCPB were removed, while the Zn^{2+} ions and X-ray waters were kept in all systems. Hydrogen atoms for each system were added to protein heavy atoms by the LEaP module in the AMBER package, whereas hydrogen atoms for crystal water molecules were added and optimized by the WHATIF program [38]. The protonation states of the N- and the C-termini were assigned to amino and carboxylated groups, respectively, following the default value of AMBER 9. The protonation states of the histidines and other titratable groups were determined with the WHATIF program [38]. To maintain the interaction of the zinc ion (Zn^{2+}) located at the active site of each proenzyme during the simulation, we employed the nonbonded parameters of Zn^{2+} atom, which were previously developed by Stote and Karplus [39]. The protein systems were solvated by a rectangular box of TIP3P waters [40], which extended at least 10 Å away from any protein atoms. Each system was neutralized by the addition of counter ions. The total numbers of atoms including protein, TIP3P waters and the counter ions are 51569, 42611 and 44804 for proCPA1, proCPA2 and proCPB, respectively. Each system was setup for simulation with the Particle Mesh Ewald (PME) method for treating long-range electrostatic interactions [41], a 8 Å cutoff for van der Waals nonbonded interactions and periodic boundary conditions. All covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm [42]. A time step of 2 fs was used to integrate the equations of motion. The solvated systems were minimized using the SANDER module of AMBER 9 with 2,500 steps of steepest descent followed by 2,500 steps of conjugated gradients to relax the systems and remove unfavorable contacts. The minimized solvated systems served as the initial structures for the subsequent MD simulations. Each system was gradually heated from 0 to 300 K over a 30 ps period with a weak restraint of 5 kcal mol^{-1} $Å^{-2}$ on all protein atoms, followed by another 80 ps of unrestrained MD for equilibration. A subsequent production MD run for each system was performed for 15 ns at constant temperature (300 K) and pressure (1 atm) using the Berendsen coupling algorithm [43] with a time constant for heat bath coupling of 0.2 ps. The *ptraj* module in Amber 9 was used to analyze trajectories. Prior to analysis, all the snapshots were superimposed onto a reference frame to remove the rotational degrees of freedom. Analysis of H-bonds was based on a definition of a H-bond existing between two residues if the distance between the donor (D) and acceptor (A) is shorter than 3.5 Å and the angle D–H…A is in the range from 120° to 180°, which is a default AMBER value. Generally speaking, hydrogen bonds (H-bonds) with distance between H and A of less than 2 Å are considered to be strong [44].

2.2 Principal component analysis

It is difficult to separate functionally important motions from complicated dynamics of protein extracted from MD sampling. Principal component analysis (PCA), or essential dynamics, is a very powerful analysis technique that can help to solve this issue by reducing the high dimensionality of such MD data into a few degrees of freedom, which represent importantly correlated motions of functional significance [45, 46]. PCA takes the trajectory from an MD simulation and extracts the dominant motions that are relevant to protein function (e.g., opening and closing and hinge bending motions) [46]. In principle, this analysis is based on the diagonalization of the covariance matrix built from the atomic fluctuations after removing the translational and rotational movement of protein by superimposing all MD trajectory snapshots onto the coordinate atoms of the reference structure. Upon diagonalization of the covariance matrix, a set of eigenvalues and eigenvectors is generated defining a new set of generalized coordinates. The eigenvectors correspond to directions in a 3 N dimensional space (where N is the number of atoms used for the analysis) along which collective fluctuations of atoms occur. The eigenvalues are the variance quantity represented the total mean square fluctuation along the corresponding eigenvector, and sum of all the eigenvalues is the total variance of the trajectory. Projection of the trajectory along the dominant eigenvectors defined by PCA enables us to examine the functionally interesting motions. More detailed information about PCA can be found elsewhere [45, 46] as well as some recent applications of this analysis for biological system [47, 48].

PCA was carried out on the Cα atoms of the proenzyme for each simulation, using the PCAZIP software developed by Meyer et al. [49] MD trajectories of corresponding atoms were extracted using the *ptraj* module in AMBER 9, and analysis was carried out for the entire simulation time using 1,500 frames.

3 Results and discussion

3.1 Structural and dynamic features

To gain insight into the structural and dynamic properties of proCPA1, proCPA2 and proCPB in aqueous solution, explicit solvent MD simulations were performed on each system at constant temperature and pressure, as described in Methods above. Analysis of total energy and root mean square deviation (RMSD) from the initial structure is important to determine the convergence of a MD simulation. Here, the total energy for each system gradually decreases for the first 5 ns of simulation and then becomes relatively stable after that (see Fig. S1 in Supplementary material), indicating that our simulations were long enough to capture the average properties of the system. Figure 2a shows the RMSD plot of the C_α atoms (with respect to the crystal structure) of the whole proenzyme and its different portions for the three systems. As can be seen, the overall structure of the proenzyme is apparently stable throughout the simulations with average RMSD values for the three systems of around 1.2–1.6 Å, showing a close resemblance of the structure to the initial crystal structure in each case. The region with the largest deviation is in the pro-segment (RMSD 1.2–1.7 Å) rather than in the catalytic segment (0.8–1.1 Å). The RMSD curves of only the pro-segment of proCPA1 and proCPA2 (Fig. 2a) are found to be similar, reaching apparent equilibrium after about 10 ns. This is in contrast to the proCPB pro-segment, whose RMSD plot reaches apparent equilibrium at about 7 ns, earlier than that of proCPA1 and proCPA2. The RMSD of other two portions, the catalytic/globular domain and the connecting segment, was also analyzed. As depicted in Fig. 2a, a relatively small change from the crystal structure is seen for the globular/catalytic domain, while large RMSDs are seen in the connecting segment of proCPA1 and proCPB, although a low RMSD is observed in proCPA2 due to its longer and more structured of α3 helix [6]. The latter observation was also found in the pro-mature junctions of proCPA1 and proCPB with the higher RMSD values of ~3 and ~4 Å, respectively (see Fig. S2). Together, after about 10 ns, these two portions of pro-segment reach full equilibration in all three systems, and therefore, the final 5 ns of trajectories for each system are collected for analysis of average structural properties below.

Root mean square fluctuations (RMSF) indicate some dynamic properties of the systems. The per-residue RMSF of the three systems is shown in Fig. 2b. The plot shows that all the systems undergo very similar fluctuations, although different magnitudes were observed in regions H1, L1, L2 and L3 (see also Fig. 1). The most flexible region among the three systems was in the pro-mature junction where the highest fluctuation was observed in proCPB (RMSF ~4 Å) followed by proCPA1 (RMSF ~3 Å) and proCPA2 (RMSF ~2 Å), respectively, also consistent with the lengths of the pro-mature junction from crystallographic structures (proCPB > proCPA1 > proCPA2; Fig. S2) [4, 6, 9]. These results further implied that the highly flexible pro-mature junctions allow more solvent

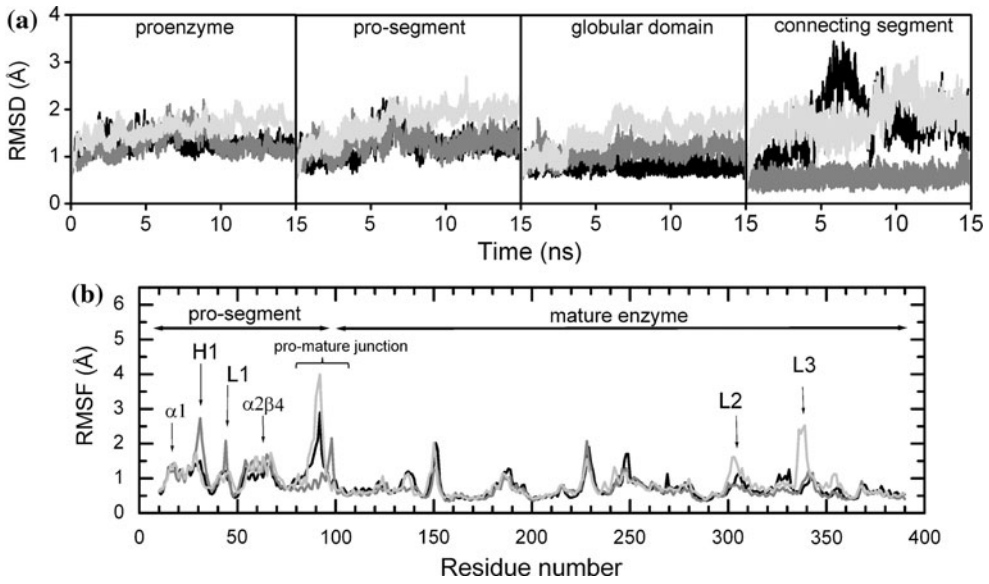

Fig. 2 **a** Root mean square displacement (RMSD) of the C_α atoms of the whole proenzyme and the pro-segment, globular domain and connecting segment portions, with respect to the initial crystal structure as a function of time and **b** per-residue root mean square fluctuation (RMSF) of the simulations of proCPA1 (*black*), proCPA2 (*dark gray*) and proCPB (*light gray*). 10 residues located at each of the N- and C-termini with very large fluctuations are excluded from (**b**) for clarity

exposure than other regions, thus helping facilitate the substrate recognition and cleavage of protease at the scissile peptide bond [5]. In addition, the RMSF also indicates that the pro-segment is more flexible than the mature enzyme (average RMSF values for each system are 1.06–1.20 Å and 0.64–0.75 Å for the pro-segment and the catalytic/globular domain, respectively), consistent with the greater structural change for this region from RMSD analysis. We also compared these calculated RMSF values to the experimental B-factors in the X-ray structure to judge the quality of our simulations [50]. The results show that there is a good correlation between these two quantities (Fig. S3).

3.2 Hydrogen bonding and salt-bridges

Previous crystallographic studies showed that H-bonds and salt-bridges play an important role in stabilizing the zymogen forms of metallo-procarboxypeptidases [2, 4, 6, 9] and related zymogens [7]. Here, the H-bonds and salt-bridges formed by residues between the pro-segment and the catalytic/globular domain for each system were analyzed over the equilibrated simulation (i.e., last 5 ns), and those residues having H-bond occupation >10 % during the simulation are summarized in Table 1. Our MD simulations reproduce most of the H-bonds present in the crystallographic structures. Some important H-bonds and salt-bridges, not observed by crystallography, are established in the simulations. For example, one additional salt-bridge formed by Glu88p and Arg277 was found in the connecting region of proCPA1. Note that the nomenclature "p" after each residue means the residue within pro-segment, while the residue without this symbol means the residue within catalytic domain. A weak H-bond formed by Glu33p and Ser162 was formed in proCPA2, and its role in the conformational change of the proenzyme is also discussed in this study (see below). Many H-bonds absent in the X-ray experiment were observed in the simulation of proCPB, such as Lys39p-Glu270, Gln45p-Tyr248 and Lys47p-Thr246. Most of these pro-segment residues are located in the globular domain of the pro-segment, thereby increasing the strength between the globular moiety of pro-segment and the mature enzyme. These new interactions arise as the structures relax away from their crystal lattice-induced conformations to ones that should be more representative of the solution-phase situation.

Figure 3 shows the H-bonds observed between the pro-segment and the catalytic domain, with different pattern of H-bonds among the three systems. For the connecting segment, most H-bonds are present during the simulation of proCPA2, followed by proCPA1 and proCPB, respectively. This is in reasonable agreement with the length of the α3 helix in the connecting segment (see Fig. 3a–c) in *right panel*; proCPA2 > proCPA1 > proCPB). In the case of the globular domain, a higher number of relatively strong H-bonds are clearly seen in proCPB (Fig. 3c), while the H-bond patterns found in proCPA1 and proCPA2 are generally similar (Fig. 3a, b). The greater number of H-bonds observed between the globular moiety and the catalytic domain of proCPB is probably due to the presence of a turn containing one additional 3_{10} helix in the globular moiety of the proCPB pro-segment (Fig. 4c). This 3_{10} helix, which is buried deeply inside the active site of proCPB, provides two additional salt-bridges: Lys39p-Glu270 and Asp41p-Arg145. These salt-bridges cause a stronger interaction between the globular moiety and the catalytic domain compared with proCPA1 and proCPA2. This is also shown the H-bond pattern. Altogether, this stronger interaction may make proCPB harder for small substrates to penetrate, explaining why null activity was observed in proCPB [4].

Important common residues of the pro-segment that strongly interact with the enzyme moiety are Asp36p, Asp53p and Glu89p for proCPA1and proCPA2; and Asp41p, Asp53p and Gln89p for proCPB. These show H-bonds for about 80–100 % of the time during the equilibrated simulations (Table 1; Fig. 3). Most of these residues form salt-bridges, that is, Asp36p-Arg71, Glu89p-Arg124, Asp41p-Arg145 (Table 1), indicating the importance of salt-bridging residues for structural stabilization of the zymogen. These salt-bridge residues, together with other H-bond residues, play a key role in the intramolecular binding of the pro-segment to the catalytic domain. In addition to H-bond interactions, hydrophobic interactions may also be important for this binding: two aromatic residues, Trp38p and Phe279, were observed near the binding interface of the pro-segment and the catalytic domain, making a π–π interaction 3 (Fig. 4).

3.3 Principal component analysis

In order to identify the essential dynamics of proCPA1, proCPA2 and proCPB, principal component analysis (PCA) was applied to those simulated zymogens using snapshots taken from MD trajectories at equal intervals, which represent all major conformational sampling. This technique allows us to identify the majority of protein motion in the main 3 N directions along which collective fluctuations of atoms occur [46]. The PCA data are summarized in Table 2, providing information on numbers of atoms, eigenvectors, total variance, explained variance, quality of compression and the 10 most significant eigenvectors with their corresponding eigenvalues, as well as the weight ratio in percent compared with the explained variance for the three simulated zymogens.

Table 1 Summary of the average distances between heavy atoms (Å) and percent occurrence data for important hydrogen bonds between the pro-segment and the catalytic domain of proCPA1, proCPA2 and proCPB

Hydrogen bond	Expt.[a]	MD-average	Occurrence (%)
proCPA1			
NH2(Arg14p) ··· O(Thr274)	2.9	2.97 ± 0.19	46.3
OD1(Asp36p) ··· NH1(Arg71)	2.7	2.94 ± 0.17	97.1
NH2(Arg39p) ··· O(Val246)	3.9	3.11 ± 0.21	11.1
NH1(Arg47p) ··· O(Ile244)	2.6	2.93 ± 0.18	47.0
OD1(Asp53p) ··· OH(Tyr198)	2.8	2.91 ± 0.21	77.4
OE1(Glu88p) ··· NH1(Arg277)	ND[b]	2.96 ± 0.22	36.8
OE1(Glu89p) ··· NE(Arg124)	2.8	2.88 ± 0.14	99.5
O(Ser96p) ··· ND2(Asn8)	4.2	3.16 ± 0.20	13.9
OD1(Asp53p) ··· NE(Arg14p)	2.8	2.83 ± 0.11	99.7
proCPA2			
OE1(Glu33p) ··· OG(Ser162)	ND	2.91 ± 0.30	53.8
OD1(Asp36p) ··· NH1(Arg71)	2.8	2.86 ± 0.15	91.8
O(Phe37p) ··· NH2(Arg71)	3.7	3.18 ± 0.20	19.7
OE1(Glu50p) ··· NZ(Lys239)	4.7	2.90 ± 0.20	19.0
ND1(His53p) ··· OH(Tyr198)	2.6	3.04 ± 0.18	95.9
O(Met78p) ··· NH1(Arg276)	ND	3.01 ± 0.20	10.5
OE1(Glu89p) ··· NE(Arg124)	2.8	2.81 ± 0.12	99.4
OE2(Glu92p) ··· NH2(Arg284)	3.4	2.98 ± 0.21	13.5
OD1(Asn96p) ··· ND2(Asn8)	3.0	3.00 ± 0.18	88.5
O(Arg97p) ··· N(Arg2)	3.5	3.13 ± 0.18	60.7
O(Arg98p) ··· N(Arg2)	3.0	3.23 ± 0.17	23.2
NE2(His53p) ··· OE1(Glu14p)	3.3	2.96 ± 0.18	62.2
proCPB			
NH2(Arg14p) ··· O(Lys274)	3.4	2.89 ± 0.15	97.9
OD1(Asp36p) ··· NH1(Arg71)	3.3	2.99 ± 0.20	<10
NZ(Lys39p) ··· OE1(Glu270)	ND	2.90 ± 0.18	22.7
OD2(Asp41p) ··· NH2(Arg145)	2.8	2.90 ± 0.18	85.9
O(Gln45p) ··· N(Tyr248)	5.3	2.96 ± 0.17	66.3
N(Lys47p) ··· O(Thr246)	ND	3.25 ± 0.16	28.8
NZ(Lys47p) ··· O(Ser244)	4.9	2.86 ± 0.15	18.0
OD1(Asp53p) ··· OH(Tyr198)	2.7	2.83 ± 0.17	99.4
O(Leu78p) ··· NH1(Arg276)	ND	2.91 ± 0.17	10.6
OE1(Gln89p) ··· NH2(Arg124)	3.2	2.86 ± 0.13	99.8
N(Asp91p) ··· OE1(Glu283)	3.1	3.08 ± 0.20	29.2
OD1(Asp53p) ··· NH2(Arg14p)	2.7	2.90 ± 0.20	99.6

Salt-bridges are indicated in italic

[a] H-bond distances were derived from crystal structures (PDB code 1PCA, 1AYE and 1NSA for proCPA1, proCPA2 and proCPB, respectively)

[b] ND, means no H-bond was detected within a distance <5.0 Å

As can be seen in Table 2, the first and second eigenvectors in all three systems account for ~40 % of the overall fluctuation of the protein. The respective eigenvalues are in the order proCPB > proCPA1 > proCPA2. Note that there is a good correlation between the number of eigenvectors and the variances in which small eigenvectors used to describe the mode of motion corresponding with large variances (eigenvalues) and vice versa. ProCPB has smaller value of normalized eigenvectors (0.072) compared with the other proenzymes (0.087 and 0.090 for proCPA1 and proCPA2, respectively). Accordingly, the lesser mode of motion of proCPB implies lesser complicated dynamics than proCPA1 and proCPA2.

Figure 5a shows the projection of the MD trajectory of each of the three systems on its two most significant eigenvectors, providing an indication of the conformational sampling of the protein in the essential subspace. It can be seen that proCPA1 and proCPA2 have similar sampling in subspace, although they have different numbers of conformational clusters (i.e., 3 clusters for proCPA1 and 4 clusters for proCPA2). This is in contrast to proCPB that sequentially visits with more random conformations and

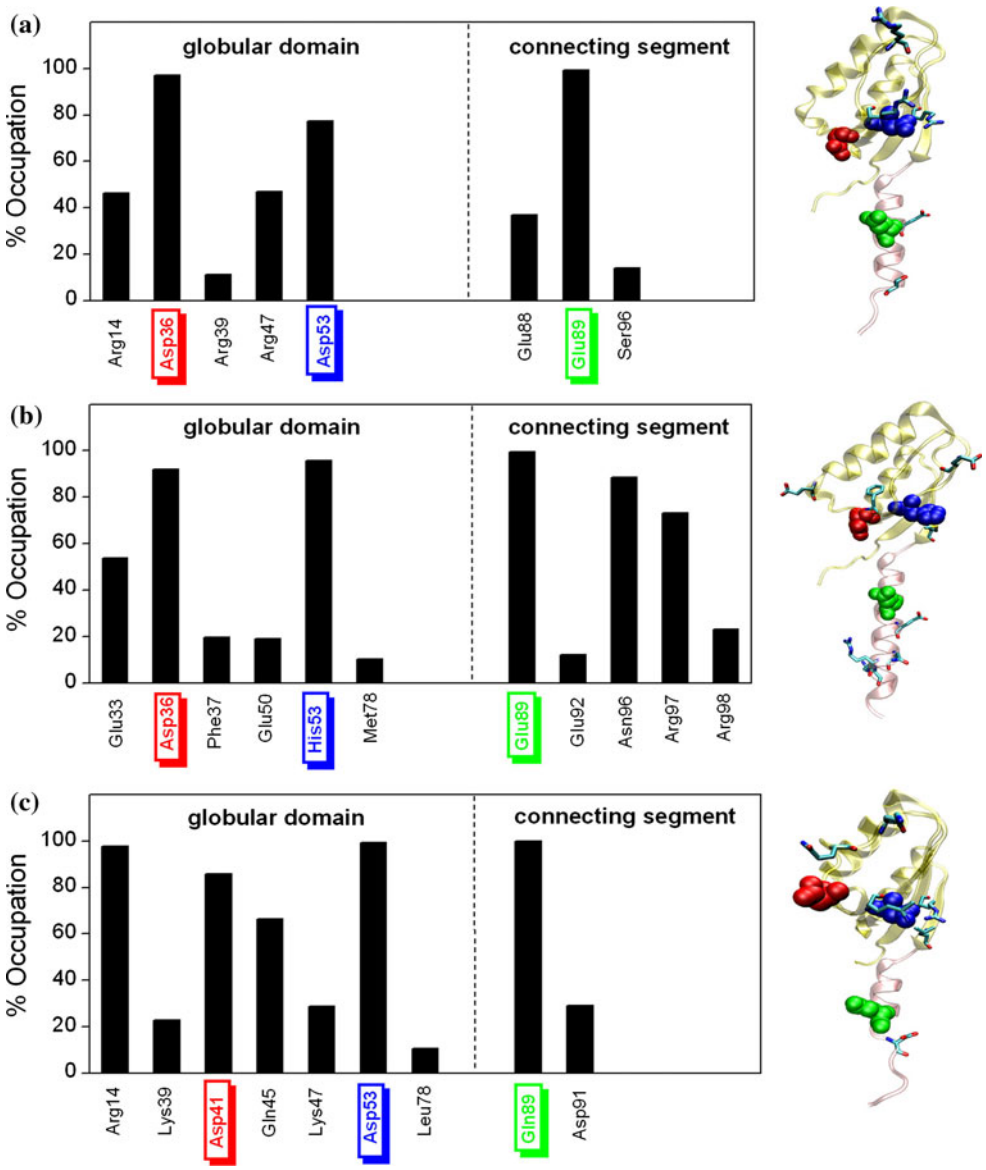

Fig. 3 Percentage occupations of hydrogen bonds between the catalytic domain and the pro-segment residues (>10 % occupied) during the last 5 ns of MD simulations. Only the pro-segment residues showing strong hydrogen bonds are indicated by *rectangular box* and their positions are displayed in space-filling model for the proCPA1 (**a**), proCPA2 (**b**) and proCPB (**c**)

clusters, and the conformational sampling is quite different to proCPA1 and proCPA2 in terms of shape, size and direction.

Further insight into the dominant motion among proCPA1, proCPA2 and proCPB can be seen by projecting the MD trajectories along the important eigenvectors. The projection over the trajectories along the first eigenvector revealed the difference in conformational change between the zymogen A and B, while the other modes (eigenvector 2 and 3) are also important, but show less difference between the zymogens (Fig. S4). Thus, here we will focus only on the first eigenvector motion among the three systems. Their animations are displayed in Fig. 5b.

Overall, for all systems, the most prominent motion is seen in the globular moiety of the pro-segment (Fig. 5b; see *orange* circle). ProCPA2 shows the largest amplitude of fluctuation in this region, which can be appreciated from the analysis of atomic displacement along the first eigenvector, as shown in Fig. 6. Furthermore, the plot clearly indicates three important regions (H1, L1 and $\alpha 2\beta 4$; see Fig. 1), which may characterize different movements between the three systems. A concerted motion between the globular moiety and the catalytic domain is dominant in proCPA1 and proCPA2, which is apparently an "opening" and "closing" motion. ProCPB seems to undergo a different mode of motion. The dominant motion of proCPB is located in loops L2 and L3 (Fig. 5b; see *green* circle), consistent with the large fluctuation of residues in these two loops found in the RMSF analysis (Fig. 2b; see L2 and L3). The high fluctuation between these two loops probably occurs because of the loss of a H-bond between Lys47p and the oxyanion L4 loop containing residues 244–247 (see Fig. S5). It is interesting to note that no major conformational change was observed in the globular moiety of

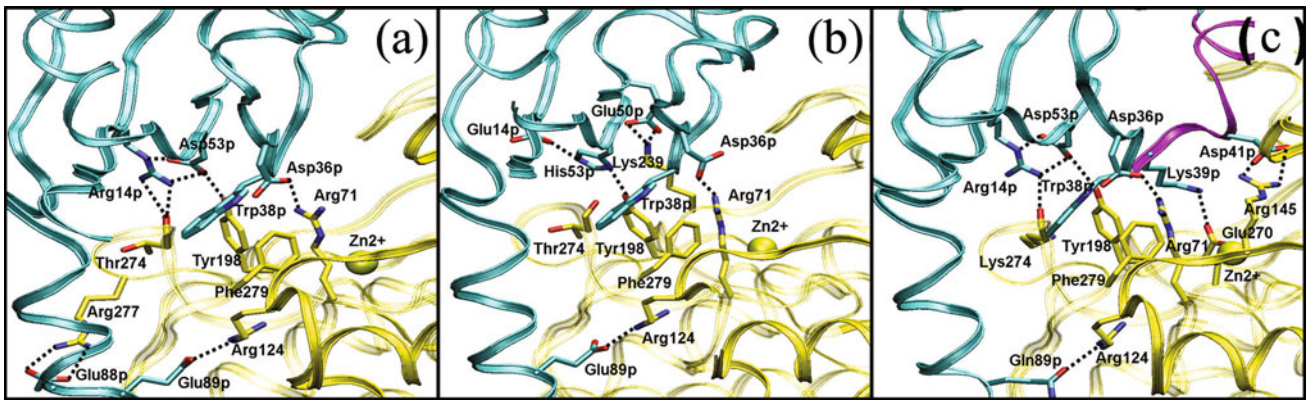

Fig. 4 Average structures from the last 5 ns of simulation of proCPA1 (**a**), proCPA2 (**b**) and proCPB (**c**), showing the important hydrogen bonds and salt-bridges at the interface between the pro-segment (shown in *cyan ribbon*) and the catalytic domain (shown in *yellow ribbon*). The turn region, which has an additional 3_{10} helix in the globular domain of the pro-segment of proCPB, is shown in *purple ribbon*. The side chains of binding residues of the pro-segment and the catalytic domain are shown in *stick* representation, in which carbon atoms are colored in *cyan* and *yellow*, respectively. Oxygen atoms are shown in *red* and nitrogen in *blue*. Hydrogen bonds are displayed as *black dashed lines*

Table 2 Important results showing numbers of atoms, eigenvectors, total variance, explained variance, quality of compression and the 10 most significant eigenvectors from principal component analysis of the simulation trajectories of proCPA1, proCPA2 and proCPB

	proCPA1		proCPA2		proCPB	
Number of atoms (C_α)	402		400		394	
Number of eigenvectors	104		108		85	
Normalized eigenvector[a]	0.087		0.090		0.072	
Total variance	336		295		403	
Explained variance	303		266		363	
Quality of the compression	90.2		90.2		90.1	
Eigenvectors	Eigenvalues/weight (%)					
1	92	30.3	80	30.4	112	30.8
2	31	10.2	26	9.8	42	11.5
3	25	8.3	20	7.5	35	9.6
4	20	6.7	17	6.4	21	5.8
5	11	3.5	10	3.7	18	4.9
6	8	2.8	9	3.4	11	3
7	8	2.6	8	2.9	10	2.8
8	7	2.3	6	2.3	9	2.5
9	6	1.9	5	2.0	8	2.2
10	5	1.8	5	1.8	6	1.6

[a] Obtained by using the equation; #eigenvectors/3N-6 where *N* is a number of carbon atom

proCPB, and specifically in the turn region with an additional 3_{10} helix. This is evidenced by the small value (<0.1 nm) atomic displacement along the first eigenvector of proCPB pro-segment (Fig. 6), further suggesting its high rigidity during the motion compared with proCPA1 and proCPA2. This structural rigidity is also shown by the strong H-bond interaction between the globular region and the enzyme moiety (Fig. 3). It is apparent from the MD simulations that proCPA1 and proCPA2 undergo an opening and closing motion, which should facilitate access to the active site, while proCPB does not, and is more rigid. This may explain why small molecules do not penetrate inside the active site and cause activity in the proCPB zymogen, but do in the other forms [4].

3.4 Nature of the conformational change

Figure 7a, b represent the conformational change observed during the simulations of proCPA1 and proCPA2, respectively. Hydrogen bonding interactions formed by Glu33p and Ser159/162, and Asp36p and Arg71 were found to be important for that change. Analysis of those residues showed the distances increased dramatically around 1.5–7.5 and 3–9 ns in simulations of proCPA1 and

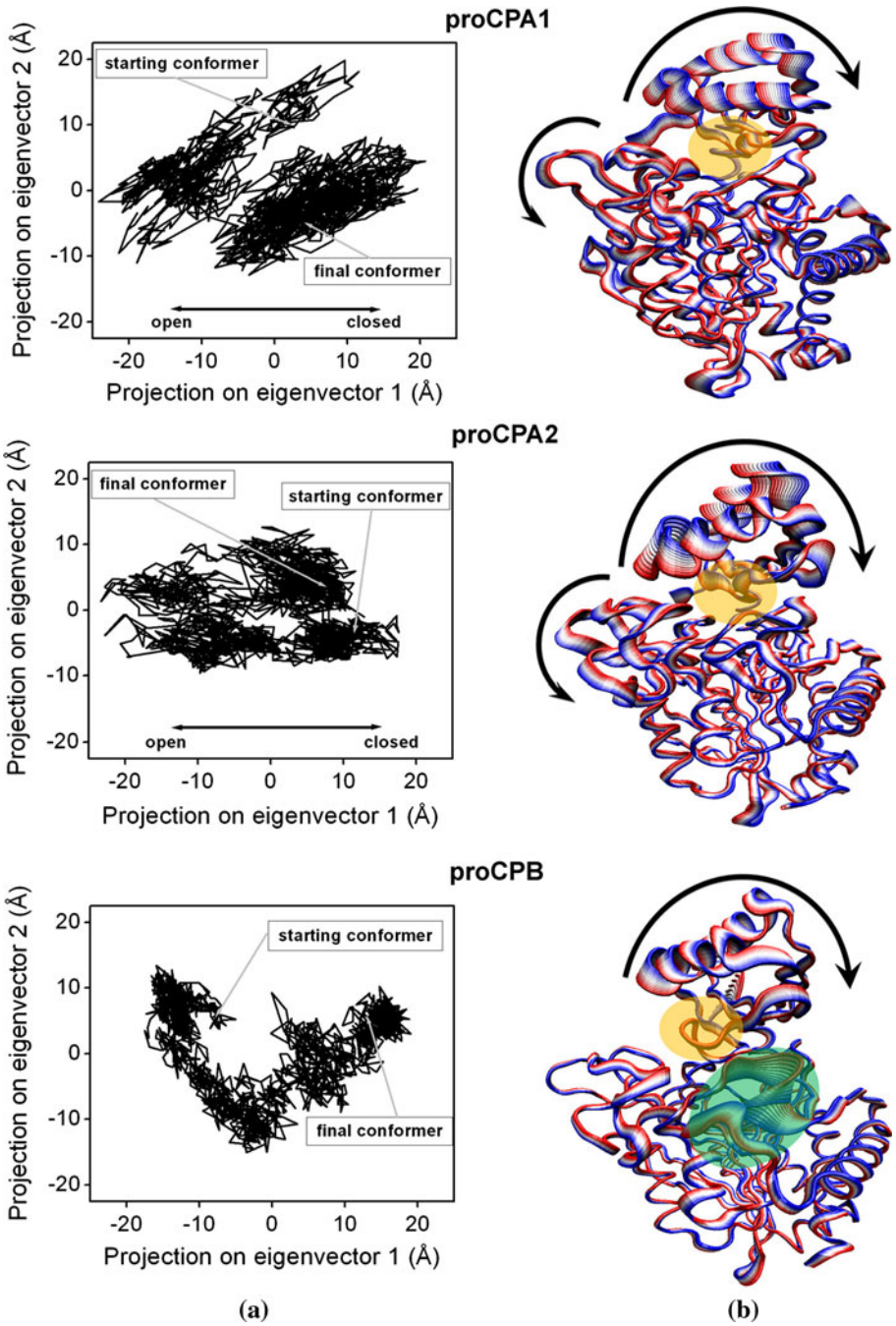

Fig. 5 Principal component analysis for proCPA1, proCPA2 and proCPB. **a** Projection of Cα trajectories along the first two eigenvectors. **b** Tube representation [48] of ten projections of the MD motions along the first eigenvector. The direction of the motion is indicated by the *arrows* whose size has no meaning and by the flanked tubes, the versus being defined from the *red* to the *blue color*. Regions with small-fluctuated motion located deeply in the binding pocket are also depicted in *orange* circle, whereas the *green circle* indicates the large fluctuation of L2 and L3 loops of proCPB

proCPA2, respectively (Fig. 7c, d), which are exactly the same time as the proteins undergo a large conformational change. The loss of those H-bond interactions destabilizes the interaction between the globular domain and the catalytic moiety, resulting in a major conformational change between the two domains. A different degree of conformational change is seen for proCPA1 and proCPA2, as clearly seen from the H-bond/distance plots (Fig. 7c, d). ProCPA1 shows a larger change of conformation than proCPA2, as evidenced by longer distances between the Glu33p OD1 atom and Ser159/162 OG atom (average distance from 1.5 to 7.5 ns and from 3 to 9 ns is 10.2 and 9.0 Å for proCPA1 and proCPA2, respectively). Furthermore, based on the H-bond analysis, we also found indications that the globular domain–catalytic domain interactions in proCPA2 may be stronger than the interactions in proCPA1: an additional salt-bridge formed by Glu50p and Lys239 was observed in proCPA2. This salt-bridge was also found to cooperate with the Glu33p-Ser162 and Asp36p-Arg71 during the simulations, as shown in Fig. 7d, which can be described into three periods following the conformational change (closed → open → closed) of

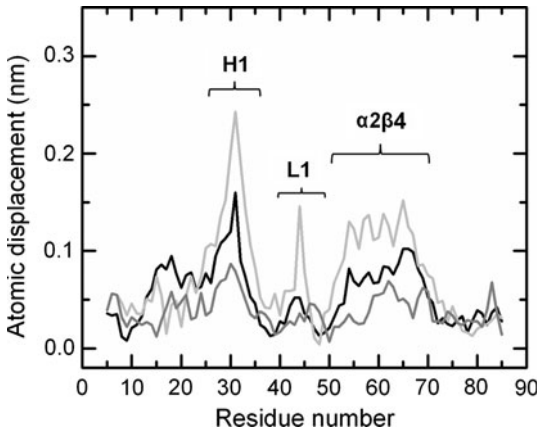

Fig. 6 Atomic displacements along the first eigenvector of the pro-segment of proCPA1 (*black*), proCPA2 (*light gray*) and proCPB (*dark gray*). Regions with large displacement are also indicated (H1, L1 and $\alpha 2\beta 4$, see Fig. 1)

proCPA2. During the first 3 ns, no salt-bridge appears between Glu50 and Lys239 residues (with distance of ~5.1 Å) while two H-bonds were observed between residues Glu33-Ser162 (~2.6 Å) and Asp36p-Arg71 (~2.0 Å). This period represents the major (closed) conformation. After that, the Glu50p-Lys239 salt-bridge is formed as shown by smaller distance (~2.9 Å), while the two H-bonds are now broken for ~6 ns (from 3 to 9 ns) with the Glu33-Ser162 distance being the highest fluctuation (~9.0 Å), and this period represents the minor (open) conformation. Finally, the two H-bonds are formed again with distances of ~2.7 Å for Glu33-Ser162 and ~2.0 Å for Asp36p-Arg71, while the Glu50-Lys239 salt-bridge is broken (~3.3 Å) with some slightly fluctuations observed during the rest of simulation. And this final period represents the major (closed) conformation again.

The different degree of conformational change between proCPA1 and proCPA2 may be due to the different pattern of a H-bond network formed by four residues (Arg/Glu14p, Asp/His53p, Tyr198 and Thr274) located at the binding interface between the globular moiety and the catalytic domain, as shown in Fig. 8. These residues are likely to act as a hinge during the conformational change of proCPA1 and proCPA2. Asp/His53p and Tyr198 may serve as the

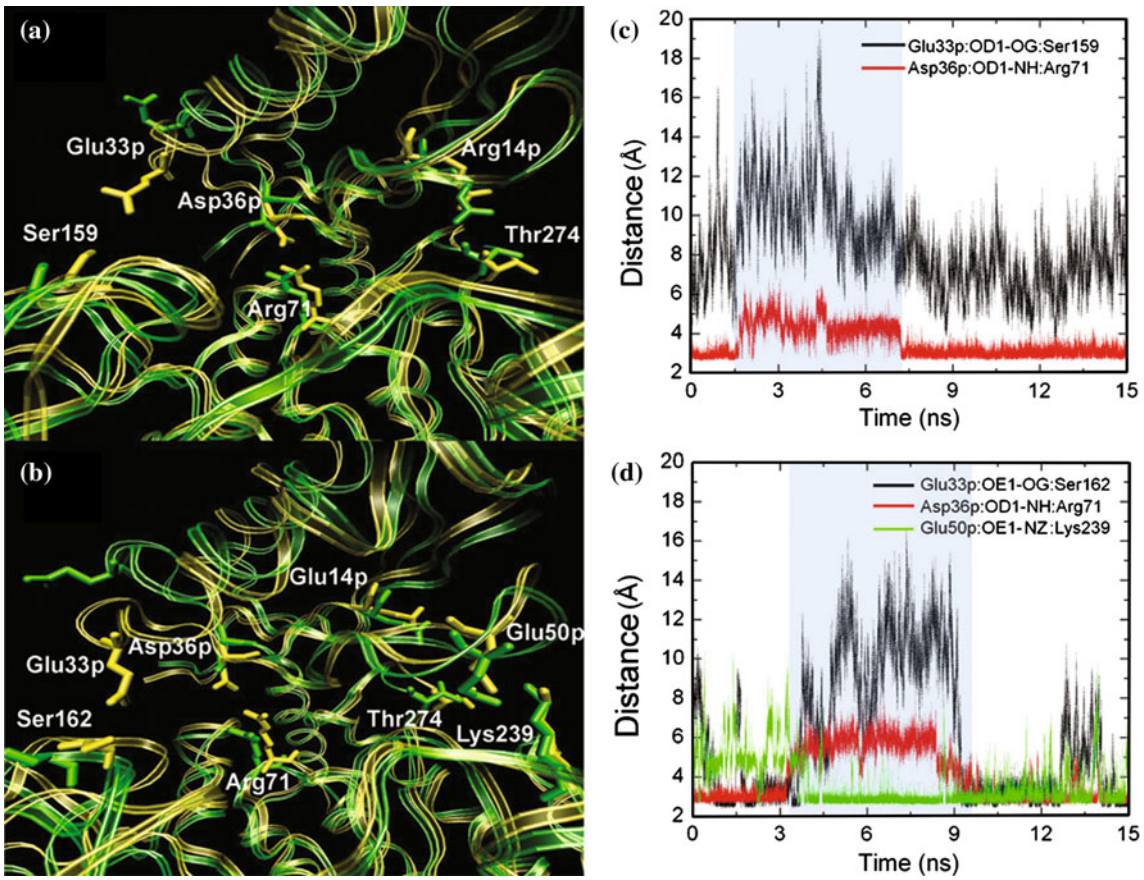

Fig. 7 Structural representation of the open (in *green ribbon*) and closed (in *yellow ribbon*) conformations of proCPA1 (**a**) and proCPA2 (**b**) from MD simulations. Some distances of important residues whose changing interactions are associated with the conformational change are shown for proCPA1 (**c**) and proCPA2 (**d**). Regions where a large conformational change was observed during the simulations are indicated in *blue*

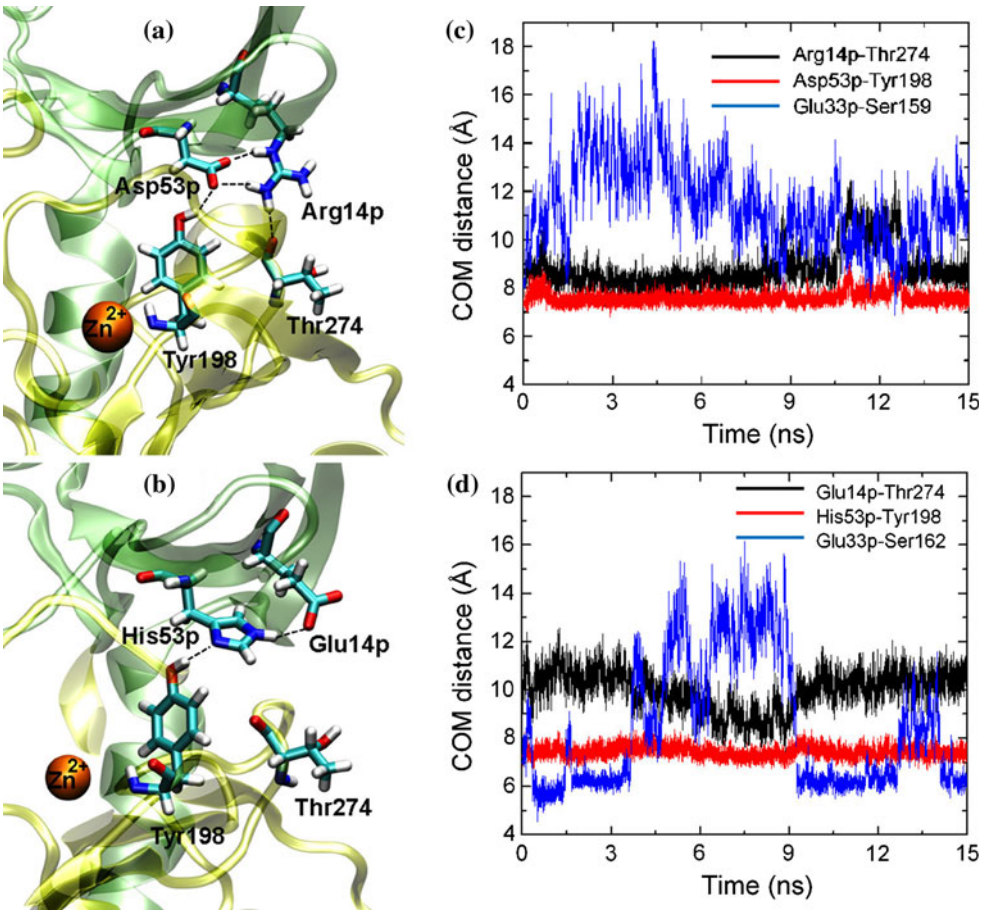

Fig. 8 Hydrogen-bond networks of the hinge residues at the interface between the globular moiety of pro-segment and the catalytic domain of proCPA1 (**a**) and proCPA2 (**b**). Some distances of hinge residues during the MD simulations of proCPA1 (**c**) and proCPA2 (**d**). For each distance, the center of mass (COM) between the two residues is calculated

center of the hinge motion, as indicated by the relatively stable center of mass (COM) distance between the two residues (see Fig. 8c, d) compared to the other hinge residues. Figure 8a, b clearly show that proCPA1 exhibits more of a H-bond network than proCPA2: three H-bonds formed by Asp53p-Tyr198 (77.4 %), Asp53p-Arg14p (99.7 %) and Arg14p-Thr274 (46.3 %) were observed in proCPA1 (Fig. 8a) while only two H-bonds formed by His53p-Tyr198 (95.9 %) and His53-Glu14p (62.2 %) were observed in proCPA2 (Fig. 8b). This indicates that proCPA1 has a stronger interaction between the hinge residues than proCPA2. Like proCPA1, proCPB also showed the same H-bond network of the hinge residues (see Fig. S6). The lack of H-bonds formed between the side chains of Glu14p and Thr274 in proCPA2 leads to a space between them that may allow Glu14p to move toward Thr274, as suggested by the highly decreased COM distance during the major conformational change (Fig. 8d). The equivalent movement of Arg14p toward Thr274 in proCPA1 seems difficult because less space is available for the movement of Arg14p, whose side chain is bigger than Glu14p. More importantly, Arg14p is constrained by hydrogen bonding with Asp53p and Thr274, which means there is only a slight decrease of Arg14p-Thr274 COM distance during the major (closed to open) conformational change (Fig. 8c) in proCPA1.

4 Conclusions

This work represents a case of how MD simulations can extract new information from X-ray structures, by investigating dynamic behavior. Here, we have shown differences in the conformational behavior of three variant proCPs (proCPA1, proCPA2 and proCPB). These differences are likely to be related to the molecular mechanisms responsible for the observed differences in enzymatic activity between proCPA and proCPB.

The most striking result from PCA analysis of the MD trajectories is the different conformational behavior observed between proCPA (both A1 and A2 forms) and proCPB. This difference particularly involves differences in the interface between the globular moiety of the pro-segment and the catalytic domain. The proCPA zymogens (both A1 and A2 forms) undergo a major conformational change (from closed to open), while proCPB exhibits only

very minor conformational changes. Conserved pro-segment residues that strongly interact with the catalytic domain (having >80 % of H-bond presence) are Asp36p, Asp53p and Glu89p for proCPA1; Asp36p, His53p and Glu89p for proCPA2; and Asp41p, Asp53p and Gln89p for proCPB. Most of these residues appear to associate via H-bonds and salt-bridges, supporting the role of these interactions in stabilizing the zymogens.

While the crystallographic structures [5, 6] give clear clues as to how the structure and interactions between the pro-domain and catalytic domain either preclude, or permit, residual activity, our simulations here provide important extra information. Firstly, the simulations confirm that the interactions in the crystal structure are also generally present in solution. Secondly and more importantly, the simulations predicted new interactions between these domains, such as an additional salt-bridge: Glu88p-Arg277 located in the connecting region of proCPA1; one weak new H-bond (Glu33p-Ser162) was observed in proCPA2; in proCPB, various additional H-bonds were observed, such as Lys39p-Glu270, Gln45p-Tyr248 and Lys47p-Thr246. By far, the most important observation from the MD simulations, however, is that the proCPA zymogens undergo a conformational change, from closed to open and back to closed, over the 15 ns timescale of the simulations, whereas the proCPB zymogen does not. This provides a plausible explanation for the experimental observation that small substrates can access the active site of the proCPA proteins, but not proCPB [5].

Further possible interesting investigations of proCPs would include, for example, modeling of proCPB mutants that might regain residual catalytic activity. Correlation analysis of interdomain motion could possibly be performed [51]. Understanding of the activation process of proCPs, which is quite complex and challenging, could also be performed, for example, using accelerated MD sampling techniques such as targeted [52] or steered MD simulations [28, 29] to gain insight into the conformational changes associated with the proCP activations, or using combined quantum mechanics/molecular mechanics (QM/MM) methods to model the chemical reaction of the cleavage process [53]. We are hopeful that the results presented here will not only provide important insight into the dynamics of this enzyme, but also may help guide future study of the activation processes of proCPA1, proCPA2 and proCPB.

Acknowledgments This work was financially supported by the National Research Council of Thailand. We would like to thank the Advanced Computing Research Centre, University of Bristol (see http://www.bris.ac.uk/acrc/) for providing the computational facilities of this work. AJM is an EPSRC Leadership Fellow (grant number EP/G007705/1).

References

1. Aviles FX, Vendrell J, Guasch A, Coll M, Huber R (1993) Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. Eur J Biochem 211(3):381–389
2. Gomis-Ruth FX (2008) Structure and mechanism of metallocarboxypeptidases. Crit Rev Biochem Mol Biol 43(5):319–345
3. Vendrell J, Aviles FX, Fricker LD (2004) Metallocarboxypeptidases. Handb Metalloproteins 3:176–189
4. Coll M, Guasch A, Aviles FX, Huber R (1991) Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. EMBO J 10(1):1–9
5. Vendrell J, Querol E, Aviles FX (2000) Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. Biochim Biophys Acta, Protein Struct Mol Enzymol 1477(1–2):284–298
6. Garcia-Saez I, Reverter D, Vendrell J, Aviles FX, Coll M (1997) The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. EMBO J 16(23):6906–6913
7. Khan AR, James MNG (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Sci 7(4):815–836
8. Lazure C (2002) The peptidase zymogen proregions: nature's way of preventing undesired activation and proteolysis. Curr Pharm Des 8(7):511–531
9. Guasch A, Coll M, Aviles FX, Huber R (1992) Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. J Mol Biol 224(1):141–157
10. Glowacki DR, Harvey JN, Mulholland AJ (2012) Taking Ockham's razor to enzyme dynamics and catalysis. Nat Chem 4(3):169–176
11. Pentikainen U, Pentikainen OT, Mulholland AJ (2008) Cooperative symmetric to asymmetric conformational transition of the apo-form of scavenger decapping enzyme revealed by simulations. Proteins 70(2):498–508
12. Vendrell J, Guasch A, Coll M, Villegas V, Billeter M, Wider G, Huber R, Wuthrich K, Aviles FX (1992) Pancreatic procarboxypeptidases: their activation processes related to the structural features of the zymogens and activation segments. Biol Chem Hoppe-Seyler 373(7):387–392
13. Villegas V, Vendrell J, Aviles X (1995) The activation pathway of procarboxypeptidase B from porcine pancreas: participation of the active enzyme in the proteolytic processing. Protein Sci 4(9):1792–1800
14. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652
15. Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. Proc Natl Acad Sci USA 102(19):6679–6685
16. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267(5612):585–590
17. van der Kamp MW, Shaw KE, Woods CJ, Mulholland AJ (2008) Biomolecular simulation and modelling: status, progress and prospects. J R Soc Interface 5(Suppl 3):S173–190
18. Lee VS, Tue-ngeun P, Nangola S, Kitidee K, Jitonnom J, Nimmanpipug P, Jiranusornkul S, Tayapiwatana C (2010) Pairwise decomposition of residue interaction energies of single chain Fv with HIV-1 p17 epitope variants. Mol Immunol 47(5):982–990
19. Jitonnom J, Lomthaisong K, Lee VS (2012) Computational design of peptide inhibitor based on modifications of proregion from *Plutella xylostella* midgut trypsin. Chem Biol Drug Des 79(4):583–593

20. Nimmanpipug P, Jitonnom J, Ngaojampa C, Hannongbua S, Lee VS (2007) A computational H5N1 neuraminidase model and its binding to commercial drugs. Mol Simul 33(6):487–493
21. Jitonnom J, Mulholland AJ, Nimmanpipug P, Lee VS (2011) Hybrid QM/MM study on the deglycosylation step of chitin hydrolysis catalysed by chitinase B from *Serratia marcescens*. Maejo Int J Sci Technol 5(1):47–57
22. Jitonnom J, Lee VS, Nimmanpipug P, Rowlands HA, Mulholland AJ (2011) Quantum mechanics/molecular mechanics modeling of substrate-assisted catalysis in family 18 chitinases: conformational changes and the role of Asp142 in catalysis in ChiB. Biochemistry 50(21):4697–4711
23. Lee VS, Kodchakorn K, Jitonnom J, Nimmanpipug P, Kongtawelert P, Premanode B (2010) Influence of metal cofactors and water on the catalytic mechanism of creatininase-creatinine in aqueous solution from molecular dynamics simulation and quantum study. J Comput Aided Mol Des 24(10):879–886
24. Kotra LP, Cross JB, Shimura Y, Fridman R, Schlegel HB, Mobashery S (2001) Insight into the complex and dynamic process of activation of matrix metalloproteinases. J Am Chem Soc 123(13):3108–3113
25. Perera L, Darden TA, Pedersen LG (2002) Predicted solution structure of zymogen human coagulation FVII. J Comput Chem 23(1):35–47
26. Venkateswarlu D, Perera L, Darden T, Pedersen LG (2002) Structure and dynamics of zymogen human blood coagulation factor X. Biophys J 82(3):1190–1206
27. Piana S, Rothlisberger U (2004) Molecular dynamics simulations of structural changes during procaspase 3 activation. Proteins Struct Funct Bioinf 55(4):932–941
28. Zuo Z, Chen G, Zou H, Mok PC, Zhu W, Chen K, Jiang H (2007) Why does beta -secretase zymogen possess catalytic activity? Molecular modeling and molecular dynamics simulation studies. Comput Biol Chem 31(3):186–195
29. Friedman R, Caflisch A (2008) Pepsinogen-like activation intermediate of plasmepsin II revealed by molecular dynamics analysis. Proteins Struct Funct Bioinf 73(4):814–827
30. Friedman R, Caflisch A (2007) The protonation state of the catalytic aspartates in plasmepsin II. FEBS Lett 581(21):4120–4124
31. Banci L, Bertini I, La Penna G (1994) The enzymatic mechanism of carboxypeptidase: a molecular dynamics study. Proteins Struct Funct Bioinf 18(2):186–197. doi:10.1002/prot.340180210
32. Banci L, Schröder S, Kollman PA (1992) Molecular dynamics characterization of the active cavity of carboxypeptidase A and some of its inhibitor adducts. Proteins Struct Funct Bioinf 13(4):288–305. doi:10.1002/prot.340130403
33. Makinen MW, Troyer JM, van der Werff H, Berendsen HJC, van Gunsteren WF (1989) Dynamical structure of carboxypeptidase A. J Mol Biol 207(1):201–216
34. Chiche L, Heitz A, Padilla A, Le-Nguyen D, Castro B (1993) Solution conformation of a synthetic bis-headed inhibitor of trypsin and carboxypeptidase A: new structural alignment between the squash inhibitors and the potato carboxypeptidase inhibitor. Protein Eng 6(7):675–682
35. Bayés A, Sonnenschein A, Daura X, Vendrell J, Aviles FX (2003) Procarboxypeptidase A from the insect pest *Helicoverpa armigera* and its derived enzyme. Eur J Biochem 270(14):3026–3035
36. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER version 9. University of California, San Francisco
37. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16):1999–2012
38. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graphics 8(1):52–56, 29
39. Stote RH, Karplus M (1995) Zinc binding in proteins and solution: a simple but accurate nonbonded representation. Proteins Struct Funct Genet 23(1):12–31
40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
41. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092
42. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341
43. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690
44. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 44(2):97–179
45. Daidone I, Amadei A (2012) Essential dynamics: foundation and applications. Wiley Interdiscip Rev Comput Mol Sci. doi:10.1002/wcms.1099
46. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. Proteins 17(4):412–425
47. Haider S, Parkinson GN, Neidle S (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. Biophys J 95(1):296–311
48. Falconi M, Biocca S, Novelli G, Desideri A (2007) Molecular dynamics simulation of human LOX-1 provides an explanation for the lack of OxLDL binding to the Trp150Ala mutant. BMC Struct Biol 7(1):73
49. Meyer T, Ferrer-Costa C, Perez A, Rueda M, Bidon-Chanal A, Luque FJ, Laughton CA, Orozco M (2006) Essential dynamics: a tool for efficient trajectory compression and management. J Chem Theory Comput 2(2):251–258
50. Hadfield AT, Mulholland AJ (1999) Active-site dynamics of ASADH—A bacterial biosynthetic enzyme. Int J Quantum Chem 73(2):137–146
51. Gargallo R, Hünenberger PH, Avilés FX, Oliva B (2003) Molecular dynamics simulation of highly charged proteins: Comparison of the particle–particle particle-mesh and reaction field methods for the calculation of electrostatic interactions. Protein Sci 12(10):2161–2172
52. Mátrai J, Jonckheer A, Joris E, Krüger P, Carpenter E, Tuszynski J, De Maeyer M, Engelborghs Y (2008) Exploration of the activation pathway of Δα-Chymotrypsin with molecular dynamics simulations and correlation with kinetic experiments. Eur Biophys J 38(1):13–23. doi:10.1007/s00249-008-0348-2
53. Mujika JI, Lopez X, Mulholland AJ (2012) Mechanism of C-terminal intein cleavage in protein splicing from QM/MM molecular dynamics simulations. Org Biomol Chem 10(6):1207–1218. doi:10.1039/c1ob06444d