J Mol Model (2012) 18:4941–4954
DOI 10.1007/s00894-012-1541-x

ORIGINAL PAPER

Conformational dynamics of threonine 195 and the S1 subsite in functional trypsin variants

Trevor Gokey · Teaster T. Baird Jr. · Anton B. Guliaev

Received: 31 May 2012 / Accepted: 16 July 2012 / Published online: 8 August 2012

Abstract Replacing the catalytic serine in trypsin with threonine (S195T variant) leads to a nearly complete loss of catalytic activity, which can be partially restored by eliminating the C42-C58 disulfide bond. The 0.69 μs of combined explicit solvent molecular dynamics (MD) simulations revealed continuous rearrangement of T195 with different conformational preferences between five trypsin variants tested. Among three conformational families observed for the T195 residue, one showed the T195 hydroxyl in a conformation analogous to that of the serine residue in wild-type trypsin, positioning the hydroxyl oxygen atom for attack on the carbonyl carbon of the peptide substrate. MD simulations demonstrated that this conformation was more populated for the C42A/C58V/S195T and C42A/C58A/S195T triple variants than for the catalytically inactive S195T variant and correlated with restored enzymatic activities for triple variants. In addition, observation of the increased motion of the S214-G219 segment in the S195T substituted variants suggested an existence of open and closed conformations for the substrate binding pocket. The closed conformation precludes access to the S1 binding site and could further reduce enzymatic activities for triple variants. Double variants with intact serine residues (C42A/C58A/S195 and C42A/C58V/S195) also showed interchange between closed and open conformations for the S214-G219 segment, but to a lesser extent than the triple variants. The increased conformational flexibility of the S1 subsite, which was not observed for the wild-type, correlated with reduced enzymatic activities and suggested a possible mode of substrate regulation for the trypsin variants tested.

T. Gokey · T. T. Baird Jr. · A. B. Guliaev (✉)
Department of Chemistry and Biochemistry, San Francisco State University,
San Francisco, CA 94132, USA
e-mail: Guliaev@sfsu.edu

Keywords Catalytic triad · Conformational rearrangement · Molecular dynamics · S1 binding site · Trypsin variants

Abbreviations

MD	Molecular dynamics
WT	Wild-type trypsin
S195T	C42/C58/S195T variant
AA	C42A/C58A/S195 variant
AV	C42A/C58V/S195 variant
AAT	C42A/C58A/S195T variant
AVT	C42A/C58V/S195T variant
PDB	Protein data bank
RMSD	Root mean square deviations

Introduction

The serine proteases comprise one of the most studied and well-understood families of enzymes. Indeed, investigations involving members of this family have been instrumental to our current understanding of catalysis and substrate recognition in enzymes. However, some aspects of their structure-function relationships are still unclear. Biologically, serine proteases are involved in the regulation of a number of processes including digestion, immune responses, hemostasis and protein synthesis and turnover. Additionally, many serine proteases have been implicated as playing roles in disease and are attractive targets in medicinal chemistry [1–4], making a more thorough understanding of structure-function relationships in serine proteases essential.

A large number of biochemical and structural studies have been devoted to the catalytic mechanism of the serine proteases, including the canonical serine protease, trypsin. Indeed, crystallographic studies have shed light on the structural features of serine protease complexes with their substrates and various inhibitors [5–12]. Trypsin is a digestive

protease that cleaves peptide bonds at the carboxyl side of lysine or arginine residues via a conserved catalytic triad (D102, S195 and H57), sometimes referred to as the charge relay system. In the currently accepted catalytic mechanism, the hydrolysis of the peptide bond occurs in two major steps [1, 13]. The first step, acylation, involves the imidazole group of H57, which acts as a general base in accepting a proton from the hydroxyl group of serine S195, increasing its nucleophilicity. The activated hydroxyl group then attacks the carbonyl carbon of the substrate scissile bond, leading to formation of a tetrahedral intermediate and subsequent release of the C-terminal fragment. The oxyanionic tetrahedral intermediate is stabilized by the backbone amide hydrogens of residues S195 and G193 (the oxyanion hole) of the S1 binding site. In the second step, deacylation occurs when an incoming solvent water molecule attacks the acyl-enzyme, leading to the release of the N-terminal fragment via the second tetrahedral intermediate [5, 9, 14]. The strong electrostatic effect of D102 stabilizes a protonated H57 which accepts a hydrogen bond from D102. The details of the D102 – H57 interactions, however, are still under discussion [15–17]. Density functional theory calculations together with the ultrahigh-resolution x-ray structure of trypsin complexed to an inhibitor suggested that D102 could be protonated prior to the acylation step. However, when H57 receives the proton from the catalytic S195, the aspartate becomes simultaneously negatively charged and provides additional stabilization of the H57 [17]. Stabilization of the catalytic triad is mediated through a network of additional H-bonds provided by several highly conserved amino acid residues surrounding the catalytic triad, including S214. Peptide binding, in addition to interactions with the catalytic triad, is supported by a hydrogen bond network between the substrate amino acid residues and trypsin residues such as D189 at the bottom of S1 binding pocket, and the residues lining the S3-S4' subsites at both sides of the S1 pocket [18–20].

For trypsin and most serine proteases, the catalytic reaction proceeds with minor structural changes, and the conformational position of the catalytic triad with most of the active site residues, is conserved. Superimposition of the catalytic structures of bovine pancreatic trypsin inhibitor (BPTI), one of the most extensively studied protease inhibitors, shows close geometries to all functional groups at different stages of the catalytic mechanism [9, 21–23]. The correct orientation of the S195 hydroxyl group in the enzyme substrate complex is essential for efficient nucleophilic attack on the carbonyl of the peptide bond [18, 24]. D189, at the bottom of the S1 pocket, is the primary determinant of arginine and lysine specificity. Molecular dynamics (MD) simulations have shown that the internal core of serine proteases is usually rigid, whereas the surface-exposed loops around the substrate-binding region are rather flexible [25]. The stability of the catalytic core is supported by a highly conserved disulfide bridge (C42–C58) near the catalytic triad [26].

Previously, biochemical and mutagenic studies were carried out to address the question as to why *threonine* proteases are infrequently observed in nature. Enzymatic assays demonstrated that replacement of serine with threonine (S195T variant) virtually eliminated the enzymatic activity toward a number of substrates, including the para-nitroanilide (pNA) and 7-amino-4 methyl coumarin (AMC) derivatives of benzyloxycarbonyl-glycylprolylarginine (Z-GPR). Removing the conserved disulfide bridge near position 195 in the S195T background (C42A/C58A/S195T (AAT) and C42A/C58V/S195T (AVT) triple variants) restored the activity to observable rates as measured by k_{cat}/K_M [27]. The AVT variant was 20-40 fold less active than AAT (or about half of the AAT activity based on log k_{cat}/K_M). In the same work, two double variants, C42A/C58V (AV) and C42A/C58A (AA), were tested to assess possible destabilizing effects of removing the C42-C58 disulfide bridge with retention of serine at position 195. Both double variants exhibited higher activities toward the Z-GPR-pNA (pNA) and Z-GPR-AMC (AMC) substrates, with exception of the AAT variant, which had higher activity than the AA variant toward the pNA substrate [27]. It was suggested that removal of the C42-C58 disulfide bridge provides structural flexibility of H57 in T195 variants, leading to the repositioning of H57 and T195 that is essential for a functional enzyme [27].

Despite the large amount of structural information available for serine proteases provided by X-ray crystallography, a complete understanding of the structural constraints on and requirements for catalytic activity requires understanding the dynamic properties of the enzymes. MD simulations can provide valuable insights into the structural dynamics of proteins that lead to conformations and features not observable in static structures, thereby uniquely complementing experimental work [28, 29]. In this work, we used MD simulations to investigate conformational properties of trypsin and S195-substituted trypsin variants with and without the C42-C58 disulfide bridge. This work provided a structural rationale for the observed enzymatic differences and identified structural features that should modulate the activity of S195T variants.

Methods

The MD package AMBER 11 was used for simulations utilizing the ff99SB force field [30]. Initial coordinates for the protein were acquired from the Protein Data Bank (PDB code 1ANE) [31]. Trypsin variants were generated by appropriate residue substitution in the 1ANE coordinates

leading to five threonine variants (C42/C58/S195T (S195T), C42A/C58A/S195T (AAT), C42A/C58V/S195T (AVT), C42A/C58A/S195 (AA) and C42A/C58V/S195 (AV). To guarantee correct hydrogen bonding with the catalytic triad, H57 was protonated on the Nδ1; protonations of other histidines were based on the optimal hydrogen-bonding conformations. The proteins were solvated with TIP3P-explicit water in a 10 Å periodic box and seven counterions were added to neutralize the overall charge of the system [32]. Sodium ions were introduced by replacing water molecules at the highest electrostatic potential to compensate for the net negative charge on the protein. A cutoff distance of 12 Å was used for the coulombic interactions using the Particle-Mesh-Ewald (PME) method using 1 Å charge grid spacing with B-spline interpolation and a sum tolerance of 10^{-6} Å [33]. The simulations were performed using a 2 fs time step with SHAKE algorithm for X-H bonds using periodic boundary condition. Van der Waals interactions were modeled using the Lennard-Jones 6–12 potential [34]. The water box was minimized first, holding the solute fixed with a positional restraint of 100 kcal mol^{-1} $Å^{-2}$. Then the entire system was minimized using 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization using the Hingerty distance-dependent dielectric function. After minimization, the system was subjected to a heating phase from 100 K to 310 K with constant volume for 30 ps using the Berendsen thermostat with a 2 ps coupling constant [35]. The heating phase was followed by a second equilibration phase for another 20 ps to obtain the correct density and volume with a 20 kcal mol^{-1} $Å^{-2}$ constraint on the solute. The entire system was then equilibrated without any constraints on the system for a total of 500 ps. This run was done to ensure that the entire system was stable before the production run. The additional details on MD procedures and theory can be found elsewhere [36–41]. A total of six proteins (five variants and wild-type, WT) were subjected to 115 ns MD production runs (0.69 μs combined) to assess the conformational dynamics of the proteins. One 115 ns simulation took about 12 days on a 16 core Intel system.

The xLeap module was used to prepare the initial coordinates for simulation, PMEMD as the simulation engine, and the ptraj module of the AMBER package was utilized to process generated trajectories [30] and VMD was the primary visualization tool [42]. Structural parameters and computation of energy contributions were calculated for 11,500 snapshots at 10 ps intervals along the 115 ns trajectory for each variant and the wild-type. Each trajectory was post-processed to assess the energetics of the systems. Free energy calculations were made with the molecular mechanics AMBER force field and the solvation energy term and the solvent-accessible surface (GB/SA) were based on continuum-generalized Born equation for implicit solvation (MM-GBSA) [43, 44]. Using an implicit solvation model, the energy of the variants was decomposed in a pairwise fashion allowing evaluation of the interaction energies between individual amino acids in catalytic triad (between H57 and T195 for example). B-factors were calculated from atomic positional fluctuations by multiplying squared fluctuations by 8/3 π^2.

Results and discussion

In this work we used explicit solvent MD simulation to describe conformational differences between five trypsin variants (S195T, AAT, AVT, AV, and AA) in comparison to a wild-type trypsin (WT). Overall, MD simulation of WT trypsin showed good correlation between the simulation data and values calculated from the crystal structure. MD trajectories initiated from the crystal structure of the WT enzyme (PDB ID: 1ANE) rapidly reach equilibrium, and the backbone WT root mean square deviation (RMSD) values calculated against the crystal coordinates fell in the range of 1.61 Å±0.15 Å (Table 1). The secondary structure motifs RMSD values for the WT simulations with respect to the crystal coordinates had an average value of 1.01 Å with a standard deviation of 0.10 Å. The secondary structure motifs were comprised of α helixes and β-strands of protein and were assigned based on the STRIDE algorithm implemented in VMD [45]. The integrity of the secondary motifs supports the structural stability of our simulations, indicating the conformational fluctuations originate primarily from the motions in the loops and links between the secondary structure elements. The all-atom RMSD values for the WT simulations were measured at 2.20 Å±0.15 Å.

The initial analysis of MD trajectories was focused on the three residues forming the catalytic triad in trypsin. In the WT simulations, D102, H57, and S195 showed stable conformational geometry, supporting the hydrogen bond network needed for the catalytic mechanism (Fig. 1). The H57 Nε2 was clearly positioned for deprotonation of Oγ of S195. However, the average distance between H57 Nε2 and S195 Oγ during 115 ns MD simulation of WT were higher (3.8±0.5 Å) than expected for a strong hydrogen bond interaction (2.8–3.0 Å between heavy atoms). Nevertheless, the distance measured during MD simulations agrees well with the corresponding distance in a crystal structure (3.6 Å) of the WT trypsin complexed to benzamidine [31]. It has been suggested that substrate binding can reduce the distance between the hydroxyl oxygen of S195 and Nε2 of H57 [9, 14, 46]. The position of H57 in the WT simulations was stabilized by electrostatic interactions with nearby D102. The Nσ1 of the imidazole ring of H57 was hydrogen-bonded to the side chain of D102, and the average distance between H57 Nσ1 and Asp 102 Cγ was measured

Table 1 Average root mean square deviations (RMSD) and corresponding standard deviations for the WT trypsin and five variants: C42/C58/S195T (S195T), C42A/C58A/S195 (AA), C42A/C58V/S195 (AV), C42A/C58A/S195T (AAT) and C42A/C58V/S195T (AVT). For each structure the RMSD was calculated from the initial crystal coordinates (PDB: 1ANE) over 115 ns trajectory

RMSD (*Å*)	WT	S195T	AA	AV	AAT	AVT
All atoms	2.20±0.15	1.96±0.14	2.37±0.19	2.50±0.26	2.35±0.19	2.06±0.14
Backbone	1.61±0.14	1.37±0.15	1.65±0.18	2.02±0.28	1.76±0.16	1.46±0.15
Backbone Cα	1.62±0.14	1.35±0.15	1.65±0.18	2.01±0.29	1.74±0.16	1.45±0.15
2^{ry} structure	1.00±0.10	0.97±0.09	1.00±0.10	1.10±0.11	1.25±0.13	1.03±0.12

at 2.47±0.16 Å during the course of MD calculations. This rotational flexibility allows both Oσ1 and Oσ2 to from hydrogen bond interactions with H57 Nσ1. The 2.47±0.16 Å value compares favorably with distance observed in the crystal structure of the WT protein (2.8 Å). The loss of the strong electrostatic stabilization of H57 by nearby D102 has been recognized as an anti-catalytic factor [47]. In addition, our initial analysis monitored the distance between S214 Oγ and Asp102 Cγ. S214 forms hydrogen bond interaction to the D102. The distance between S214 Oγ and Asp102 Cγ was measured at 3.64±0.30 Å and agrees well with the value from the crystal structure (3.68 Å).

Analysis of the RMSD values was extended to the five trypsin variants: S195T, AAT, AVT, AV, and AA (Table 1). All six structures tend to rapidly reach a steady equilibrium based on the time evolution plots of the RMSDs. Despite the fact that all atoms and backbone RMSD values are very broad measurement of the specific dynamics features, one observation can be made for the WT and trypsin variants based on all atoms RMSD. Larger deviations from the crystal coordinates were measured for the AA and AV variants (2.37±0.19 Å and 2.50±0.26 Å respectively, Table 1). In contrast, S195T shows smaller RMSD values (average at 1.96±0.14 Å) than WT. These differences can be related to the nature of the amino acid substitution. The loss of the C42-C58 disulfide bridge in the AA and AV variants should increase overall structural flexibility. For the S195T variant, the replacement on the S195 residue by threonine adds an additional 20 $Å^3$ of the van der Waals (vdW) volume into the active site. This should impose steric constrains on the enzyme motions, and this is reflected in the lower average RMSDs for the S195T variant. In the AVT triple variant, the increase in vdW volume due to replacement of the alanine (A58) by valine (V58) was also reflected in RMSD calculations (Table 1). The AVT variant has lower positional fluctuations than the AAT variant (2.06±0.14 vs. 2.35±0.19 Å respectively). In order to confirm statistical significance of the reported RMSD differences we performed t-test analysis. The statistical analysis showed that p values for RMSD comparison of means were less that 0.0001 indicating that described differences are statistically significant. For example, for the AA/AV comparison the t= 128 with $p<0.0001$. The RMSD data agree with the previously reported higher thermostability of trypsin with

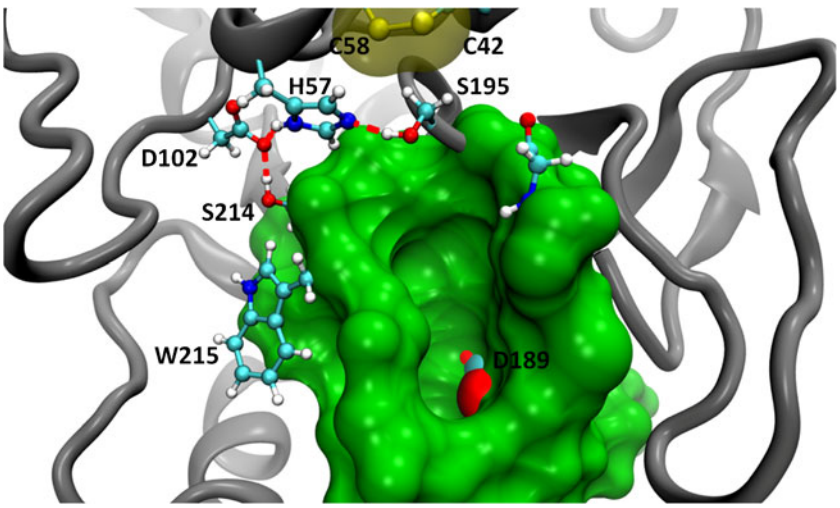

Fig. 1 Representative structure generated based on the 115 ns MD simulation of WT trypsin showing the mutual orientation of the catalytic triad (D102, H57 and S195). The H57 Nε2 forms a hydrogen bond interaction with the hydroxyl oxygen of S195. D102 interacts with H57 and S214, facilitating proton abstraction from S195 for the first step of the catalytic mechanism. The enzyme S1 pocket is shown in *green*. D189 at the bottom of the S1 binding site is shown in *red*. The C42-C58 disulfide bridge is shown by vdW surface representation

substituted T195 relative to its S195 counterpart [27]. Initial analysis of the hydrogen bond network for the catalytic triad showed that all variants, including S195T, had stable interactions between D102 and H57. However, distances between S/T195 and H57 and between D102 and S214 had a very large standard deviation for variants without the C42-C58 bridge. The observation of the large deviations in distance measurements suggested increases in local flexibility for the catalytic triad for these variants.

Conformational flexibility of threonine 195

Visual inspection of the MD trajectories of all variants (except S195T) revealed significant conformational motions of the catalytic residue at the 195 position. The structural analysis focused on the mutual orientation of residues comprising the catalytic triad, particularly the orientation of the residue at the 195 position (S or T) relative to H57. It has been suggested that, in order to display catalytic activity, the hydroxyl group of threonine in the S195T trypsin variant should be placed in a catalytically viable orientation similar to that of the hydroxyl group of S195 in WT. The hydroxyl group of the substituted threonine can then participate in the critical hydrogen bond network needed for the catalytical reaction [27].

In our work, MD simulations of the S195T variant with an intact C42-C58 bridge demonstrated that the preferred spatial orientation of the T195 residue, particularly the position of the hydroxyl group, was opposite to that observed for the S195 residue in the WT. As expected, the steric clash between the methyl group of T195 and the C42-C58 bridge resulted in rotation of T195 about the Cα-Cβ bond. In this orientation, the methyl group of T195 occupied the position of the hydroxyl group of S195 observed in the MD simulations of the WT trypsin. The conformation of S195 in the WT active site produced in our calculations agrees well with reported crystallographic data [14, 31]. In this conformation, the -OH group of S195 is positioned toward H57 and the S1 binding pocket to facilitate its interaction with the imidazole ring of H57 and prepare Oγ of S195 for nucleophilic attack on the peptide carbonyl bound at the enzyme pocket. Based on MD simulations for S195T, the -OH group of T195 was positioned away from H57, losing its essential catalytical location observed with S195 in WT. The orientation of the T195 or S195 -OH group relative to H57 and the substrate S1 binding pocket can be characterized by the rotation along the -Cα-Cβ- bond (Fig. 2, top panel). The measurements for the S195 N-Cα-Cβ-Oγ angle from the WT simulations showed values in the −50°±20° range for almost 97 % of the simulations (over 11,100 structures out of 11,500). This S195 N-Cα-Cβ-Oγ angle measures at −52° in the crystal structure. The analysis of the 115 ns MD trajectory for S195T showed that the T195 N-Cα-Cβ-Oγ angle lies in the 50°±15° range for almost 93 % of the simulation time. This value indicates over a 100° rotation of the T195 side chain from the position characteristic of the S195 side chain in WT. Based on the dihedral angle values, we can clearly distinguish between the orientation of T195 in the S195T variant and S195 in WT trypsin, thereby identifying at least two conformational families for the residue in position 195. Inspection of the MD trajectories of the trypsin variants without the cysteine bridge (AVT and AAT variants) revealed increased conformational flexibility of the threonine residue, with continuous rotation around the Cα-Cβ bond. Figure 2 shows time-dependent changes of the N-Cα-Cβ-Oγ dihedral angle for the T195 residue in S195T and two triple variants (AVT and AAT), and the S195 residue in WT and two double variants (AA and AV). Three distinct conformational families were identified when the N-Cα-Cβ-Oγ dihedral angle was plotted as a function of simulation time (Fig. 2). In addition to two conformational families identified based on WT and S195T simulations, there is a third conformational family with values of around 150° (150°±20°). This conformation was populated the least and was noticeably observed for the AAT, AA, and S195T variants. Based on the dihedral angle values, the identified conformational families were assigned as A, B, and C (Fig. 2).

To provide better visual characterization of the observed conformational families, Fig. 3 shows representative conformations of the catalytic triad for WT trypsin, assigned to conformation A (Fig. 3, panel 1), the catalytic triad in S195T assigned to conformation B (Fig. 3, panel 2), and the catalytic triad for the AAT variant showing all three conformational families A, B, and C (Fig. 3, panels 3, 4, and 5 respectively). In conformation A, the orientation of the hydroxyl moiety of the catalytic residue in position 195 (S195 or T195 in our work) agrees well with the currently accepted initial step in the catalytic mechanism. This conformation exhibits key features expected for the catalytically active enzyme in which mutual orientation of hydroxyl moiety of S/T195 and H57 positions the hydroxyl proton close to Nε2 of H57. In addition, the Oγ atom of the hydroxyl group is oriented toward the enzyme active site pocket (shown in green, Fig. 3) to facilitate nucleophilic attack on the carbonyl carbon of the incoming substrate. Structures in conformation A (WT with an intact C42-C58 bridge and AAT variant with the cysteine bridge removed (Fig. 3, panel 1 and 3 respectively)) showed similar orientation of the -OH group, characterized by the same dihedral angle (N-Cα-Cβ-Oγ). In agreement with the initial step for the catalytic mechanism proposed for trypsin, conformation A could be assigned as the catalytically active conformation.

The conformation of T195 in the S195T structure (panel 2, conformation B) was clearly different from conformation

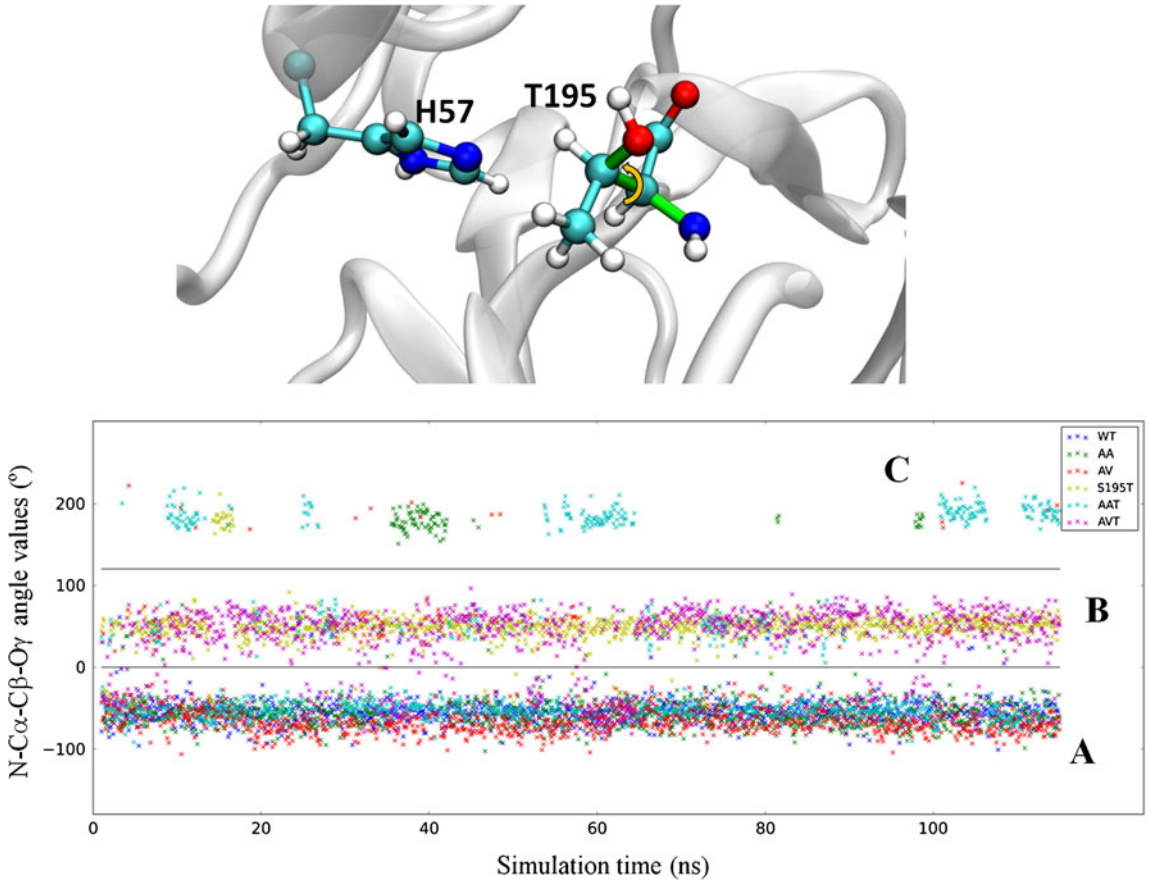

Fig. 2 Time evolution of the N-Cα-Cβ-Oγ dihedral angle for the residue in the 195 position of all proteins tested (S195 N-Cα-Cβ-Oγ for WT, AA and AV variants and T195 N-Cα-Cβ-Oγ for the S195T, AAT and AVT variants). The distribution of the dihedral angle values shows three distinct conformational families. All values between 0 and -100° were assigned to conformation A, 0 to 100° to conformation B, and all others to conformation C. Conformation A was assigned as catalytically active, and the corresponding value of the S195 N-Cα-Cβ-Oγ dihedral angle in the crystal structure (PDB: 1ANE) was measured at -52°. The top figure shows T195 in conformation B, proposed to be catalytically inactive. The rotation around the Cα-Cβ bond is indicated by the *yellow arrow*

A. In this conformation, the bulky methyl group of T195 occupies the location of the hydroxyl group observed in the catalytically active conformation. In this configuration, the methyl group of T195, and not the Oγ, would make contact with the carbonyl group of an incoming substrate. Moreover, the location of the -OH is too far away to make hydrogen-bond contact with H57 and would prevent formation of the essential hydrogen network that is characteristic of the catalytic triad. This conformation, conformation B, was also observed for the AAT variant (Fig. 3, panel 4). Owing to the absence of the essential interaction between H57 Nε2 and T195 Oγ and the fact that the methyl group of T195 would occupy the position of the -OH needed for catalysis, conformation B was designated catalytically inactive.

The representative conformation for the third conformational family is shown on panel 5 of Fig. 3 (conformation C). In this conformation, the orientation of the T195 side chain places the -OH group away from the protein active site, making it less accessible for the incoming substrate. However, the short distance between H57 and the hydroxyl group of T195 makes this conformation more suitable for the H57-T195 hydrogen bond interaction in comparison with catalytically inactive conformation B.

Based on MD simulations, all three conformational families (A, B and C) were observed for the trypsin variants, but with significantly different populations (percent occupancy) among different structures (Table 2). Based on analysis of MD trajectories, conformation A was undoubtedly dominant for WT trypsin. It was present during 96.6 % of the simulation time (96.6 % percent occupancy; Table 2). In contrast, the catalytically inactive S195T variant had a small percent occupancy for conformation A (3.95 %), and the conformational space covered by MD simulation was predominantly occupied by conformation B (93.83 % occupied). For the AAT and AVT variants, conformation A was 65.62 % and 19.48 % percent occupied, respectively. The higher occupancy for the catalytically active conformation A was also observed for the AA and AV variants (87.42 and 91.70 %, respectively). The ability of T195 in trypsin variants to obtain similar configuration to S195 in WT correlates with

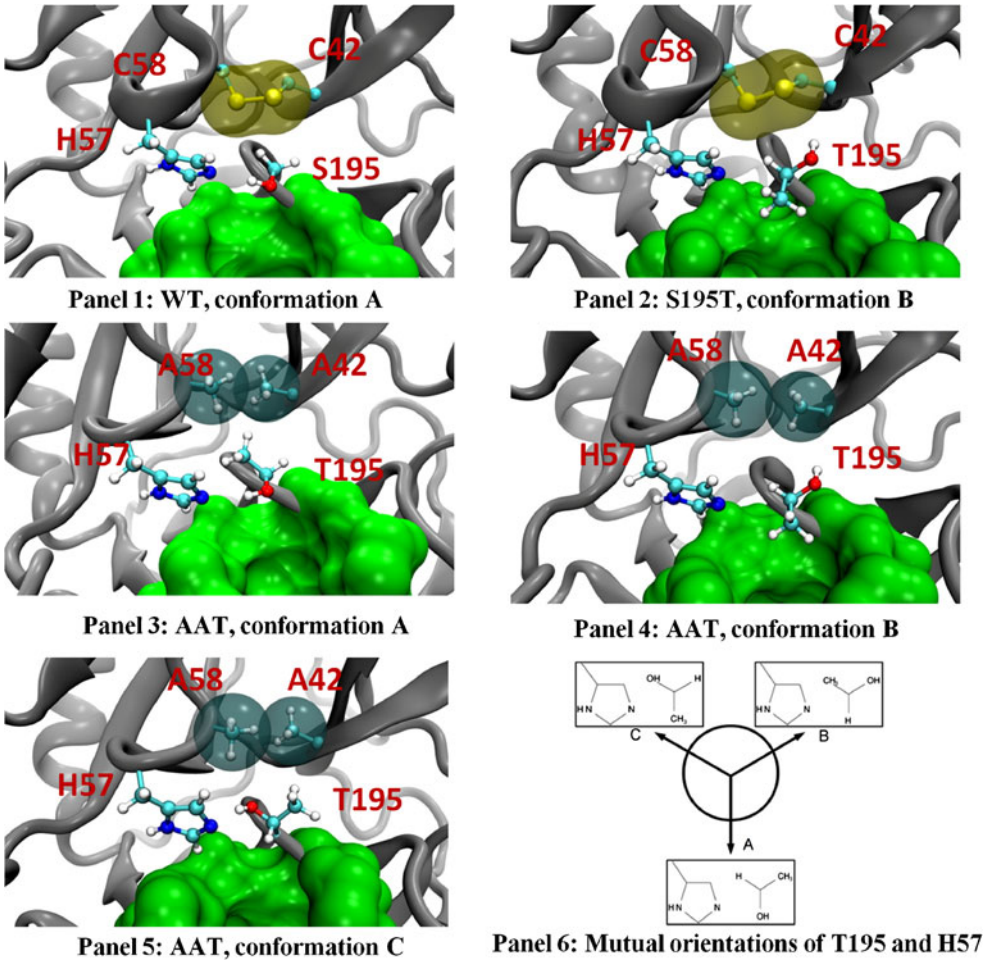

Fig. 3 Representative conformations for the catalytic triad for WT, S195T, and AAT variants based on 115 ns MD simulations. The part of the S1 binding pocket is shown in *green*. Panel 1: S195 in the WT positions the OH group toward the S1 binding pocket, making it accessible to incoming substrate. In this conformation (A), the OH group of S195 can attack the carbonyl carbon of the substrate's scissile bond. This conformation was highly populated during MD simulation, with almost 97 % occupancy. VdW surface of the nearby disulfide bridge is shown in *yellow*. Panel 2: S195T variant with an intact C42-C58 bridge. Steric constraints imposed by the disulfide bridge on the methyl group pushed the OH moiety away from the binding pocket, leading to an inactive S195T variant. In this conformation (94 % occupied in S195T), H57 cannot accept a hydrogen bond from T195, and the methyl group of T195 positions itself toward the S1 binding pocket. The C42A and C58A substitutions (AAT variant) provide the extra space needed for conformational flexibility of T195, leading to the observation of the three conformational families during MD simulation: conformations A (panel 3), B (panel 4) and C (panel 5). Panel 3: (conformation A) threonine in an orientation similar to S195 in the WT protein. This conformation has 65 % occupancy for the catalytically active AAT variant. In conformations B and C, the hydroxyl group of T195 positions away from S1 binding pocket, making these conformations much less suitable for catalysis. Panel 6: schematic representation of the mutual orientation between T195 and H57 for A, B, and C conformational families

previously observed enzymatic activity. Variants with higher enzymatic activity had a higher percentage of occupancy for conformation A during MD simulation (Table 2). The last two columns in Table 2 show enzymatic activities based on *log* k_{cat}/K_M toward different substrates [27]. The increase in the catalytically inactive conformation B population correlates with decreased enzymatic activity.

The conformational preferences for the N-Cα-Cβ-Oγ dihedral angle observed during the MD simulations correlated with the interaction energies calculated between S195 or T195 and the surrounding residues (F41-G43, H57, C58, D102, D189, S190-G197 &V213-C220) using the MM-GBSA model. The preference toward A, between A and B conformations, was supported by about 2.7 kJ mol^{-1} lower electrostatic interaction energy for the WT, AA and AV variants. The largest contribution to the electrostatic term came from S195 - H57 interaction. Conformation C was also favored by electrostatic energy (2.0 kJ mol^{-1} lower than A and B), but has about 4.0 kJ mol^{-1} higher vdW energy term. For the serine substituted variants, such as S195T and AVT, conformation B was favored by about 3.2 kJ mol^{-1} and 4.2 kJ mol^{-1} lower electrostatic and vdW energies, respectively. The AAT variant demonstrated less correlation between the calculated energies and observed conformational families.

Table 2 The conformational preferences for the WT and 5 variants based on the S/T195 N-Cα-Cβ-Oγ angle. Conformation A is almost exclusively favored by WT, AA, and AV while S195T, a largely inactive variant, avoids this conformation. In the AAT variant, conformation A was present in 66 % of the simulation in comparison with WT for which conformation A was 97 % occupied. The last 2 columns show *log* k_{cat}/K_M for WT and trypsin variants activities toward Z-GPR-pNA and Z-GPR-AMC substrates [27]. The activity of the wild type was assigned to 100 %

	Conf. A (%)	Conf. B (%)	Conf. C (%)	Activity toward Z-GPR-AMC	Activity toward Z-GPR-pNA
WT	96.62	3.37	0.02	100 %	100 %
S195T	3.95	93.83	2.22	<1 %	<1 %
AAT	65.62	14.51	19.87	24 %	50 %
AVT	19.48	80.51	0.01	10 %	22 %
AA	87.42	5.43	7.16	92 %	45 %
AV	91.70	7.25	1.04	65 %	82 %

The preference toward conformation A was favored with an electrostatic energy of about 2.1 kJ mol^{-1} lower than in conformation B. However, the vdW energy was about 1.8 kJ mol^{-1} lower for the B conformation.

Table 3 compares distances associated with the hydrogen bond network characteristic for the catalytic triad in conformation A for all structures used in our calculations. For variants with S195 substituted by T195 (S195T, AAT, and AVT), the distance between T195 Oγ and H57 Nε was higher than for the WT. The double variants (with a substituted C42-C58 bridge, but with S195 in the active site - AA and AV) showed distances similar or lower than the WT. The largest distance between catalytic serine or threonine and H57 was observed for the S195T variant, indicating movement of the H57 from the position observed in the WT. For all structures, MD simulations showed stable interactions between D102 and H57, and the average distance remained close to the value observed in the crystal structure (2.4 Å).

Figure 4 summarizes the conformational characteristics of the catalytic triad in WT trypsin and the five trypsin variants analyzed in our computational work. Each point in Fig. 4 represents a conformational snapshot for which the distance between H57 and S/T159 was plotted against the N-Cα-Cβ-Oγ dihedral angle. In addition, each point was colored according to the interaction energy between H57 and the catalytic residue in the 195 position. Two distinct conformational families corresponding to conformations A and B can be found on the WT plot (Fig. 4, WT panel). Conformation A is much more populated than B, and is favored by low interaction energies between H57 and S195. The area corresponding to conformation A in WT (highlighted by a blue rectangle) was projected to the other plots to emphasize the conformational differences between WT and the variants. Deviations from conformation A for the trypsin variants correlated with reduced enzymatic activities, with S195T being virtually inactive [27]. The majority of the conformational snapshots for the S195T variant fell outside the area characteristic of the WT. In comparison, AA and AV variants had much larger populations of structures in agreement with the conformational footprint for the WT. These double variants had higher enzymatic activity than S195T variant and triple variants, with the exception of AAT activity toward Z-GPR-pNA (Table 2). The lower activities of the triple variants correlated with the observation of larger conformational populations outside the ranges defined by the WT (Fig. 4).

Induced flexibility of the S214-G219 segment in trypsin variants

The conformational flexibility of T195 may only be a contributing factor to the restored but still relatively low enzymatic activities of the S195T variants with the removed C42-C58 cysteine bridge. To evaluate the motions of various structural motifs in the WT and five variants, we compared B-factors for individual residues. Figure 5 shows the B-factors calculated from 115 ns MD trajectories. Whereas relatively small values were observed for the catalytic triad residues H57 and D102, slightly larger values were

Table 3 Averaged distances and corresponding standard deviations for the catalytic triad in conformation A for all proteins tested in this work

Distance (*Å*)	WT	S195T	AA	AV	AAT	AVT
S/T195 Oγ -H57 Nε	3.83±0.48	4.67±0.52	3.68±0.61	3.24±0.60	4.15±0.76	3.91±0.87
H57 Nε - A102 Cγ	2.47±0.16	2.49±0.19	2.50±0.16	2.40±0.14	2.36±0.13	2.39±0.18
S214 Oγ - A102 Cγ	3.64±0.30	3.59±0.17	3.70±0.58	4.1±0.70	4.85±1.23	5.30±0.80

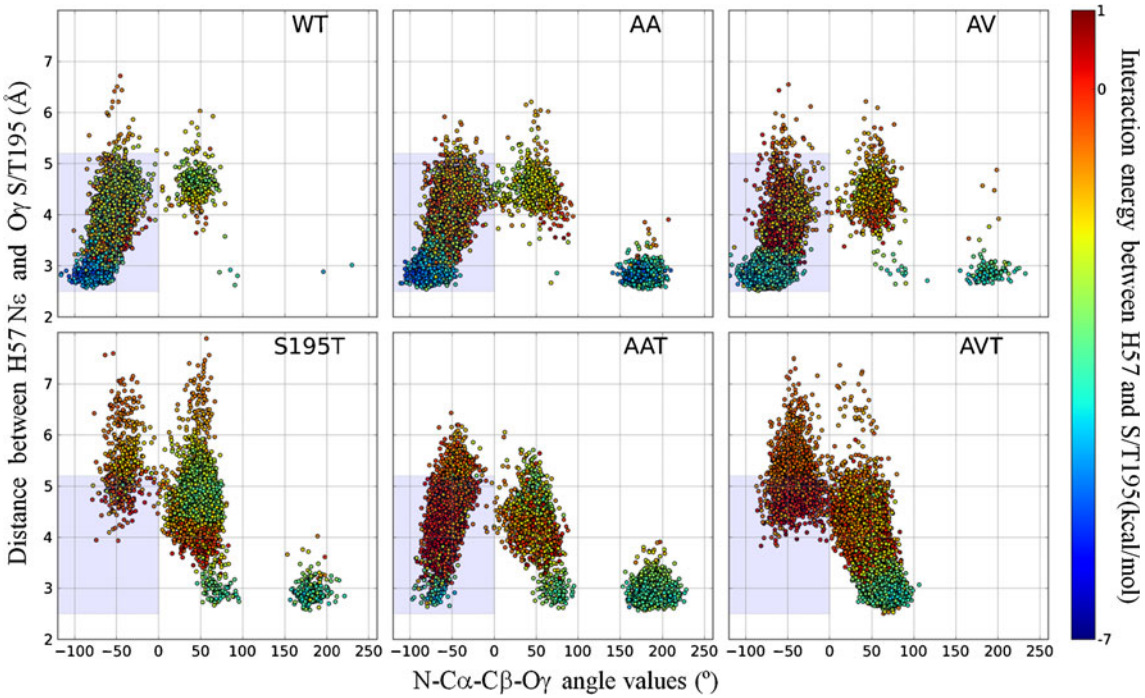

Fig. 4 Two-dimensional diagrams showing the interaction between the catalytic S/T195 residue and H57 computed for six trajectories. The x-axis refers to the distribution of the N-Cα-Cβ-Oγ dihedral angle of S/T195, and the y-axis refers to the distance between the Oγ of S/T195 and the Nε of H57. Each point represents a single snapshot and is colored according to the interaction energy value between H57 and S/T195. There are 11,500 snapshots per each trajectory. The color map for the energy values (sum of the vdW and electrostatic terms) obtained from GBSA calculation is shown on the right. The plot for the WT shows two conformational families (A and B based on dihedral angle values), with A being clearly dominant. This conformation is favored by relatively small distances between H57 and S195, low interaction energies (*dots* are mostly *blue* in color), and an orientation of the hydroxyl group appropriate for catalytic action. The area corresponding to conformation A in the WT (highlighted by a *blue rectangle*) was projected to the other plots to emphasize the conformational differences between the WT and variants. The deviation from conformation A for the mutants agrees with reduction in enzymatic activities, with S195T being virtually inactive. Most points fall outside the blue area. In comparison, AA and AV variants have much larger population of structures in agreement with conformation A. These structure shows much higher activity based on biochemical data

observed for the T195 residue, indicating observed conformational rearrangements of this residue. The large concerted motions comprise residues 18–26, 36–40, 60–64, 72–82, 95–100, 112–118, 144–154, 184–180, and 214–219 (Fig. 5). Most of these conformational motions are associated with surface-exposed loops and correlate with the position of the eight surface loops identified for the trypsin family. Based on the convention introduced by Perona and Craik [18, 24], five surface loops designated A–E correspond to residues 34–41, 56–64, 97–103, 143–149, and 74–80, respectively. The other two loops, loop 1 (residues 185–188) and loop 2 (residues 217–225), connect three β-strands that, together with the C191-C220 disulfide bond, form the S1 binding pocket (residues 184–195 and 213–228). Loop

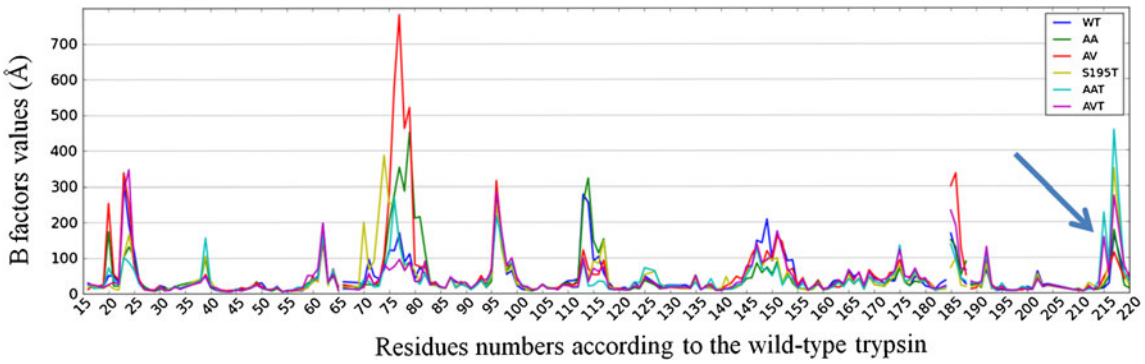

Fig. 5 B-factors calculated for the WT and five trypsin variants (S195T, AA, AV, AAT, and AVT) based on MD simulation. Rises in simulated B-factors are associated with the eight surface loops identified for the trypsin family. The region for the S214-G219 segment is indicated by an *arrow*

Table 4 Average root mean square deviations (RMSD) and corresponding standard deviations for the S214-G219 segment for WT trypsin and five variants. Second column shows average number of waters molecules found within 5 Å of S214 during the MD simulations

	RMSD value for the S214-G219 backbone (Å)	Waters within 5 Å of S214
WT	1.34±0.25	1.20±0.81
S195T	2.09±0.48	0.88±0.71
AA	2.00±0.27	0.51±0.63
AV	3.18±0.71	0.29±0.61
AAT	3.46±0.76	4.00±1.62
AVT	3.18±0.54	2.02±0.95

3, residues 169–174, can be also characterized as a surface loop associated with the S1 subsites.

The most interesting observation based on the calculated B-factors was significant conformational movement of the S214–G219 segment. This segment is a part of loop 2 and belongs to the S1 subsites (S2–S3) (Fig. 5). The S214–G219 motif includes a bulky tryptophan residue (W215) that is located near the β-strand that forms the enzyme active site. Conformational motions were clearly observed for all variants with a substituted C42-C58 disulfide bridge and can be initially monitored by RMSDs for the S214–G219 residues (Table 4). The WT structure showed minimum deviation from the starting coordinates, and the position of this segment was similar to that observed in the crystal coordinates. The largest RMSD values were observed for the triple variants and indicated substantial conformational dynamics in this part of the protein. Examination of the MD trajectories revealed that the S214–G219 motif undergoes a reversible conformational change between the position similar to the one observed in WT and a new position, in which S214-G219 collapses onto the S1 binding pocket. In the collapsed conformation, S214-G219 covers (or closes) the S1 substrate binding pocket. Figure 6 shows the S214-G219 segment (light green) for the AAT variant in two distinct conformations: the top figure shows the open conformation with clear access to the S1 binding pocket (this conformation of the S1 binding pocket is almost identical to the conformation observed in WT, as shown in Fig. 1) and the bottom figure shows the closed conformation, with the S214-G219 segment collapsing onto the S1 pocket. The conformational changes in the S1 subsite correlate with the movement of the W215 side chain, which in closed conformation blocks access to D189 at the bottom of the S1 subsite (Fig. 6, bottom). For all variants tested, the conformational rearrangement of the S214-G219 segment was observed during the first 20 ns of the production runs. As expected, this type of motion is much slower than the observed conformational motions of T195 which occur almost instantly at the start of the production runs.

The conformational flexibility of S214-G219 in the triple variants correlated with solvation of S214. In the WT structure, this residue mainly interacts with D102 and several water molecules can be found close to D102 near the substrate-binding site. Removal of the conserved C42-C58 bridge, as described above, allows rotation of the T195 side chain, leading to a catalytically active orientation of the

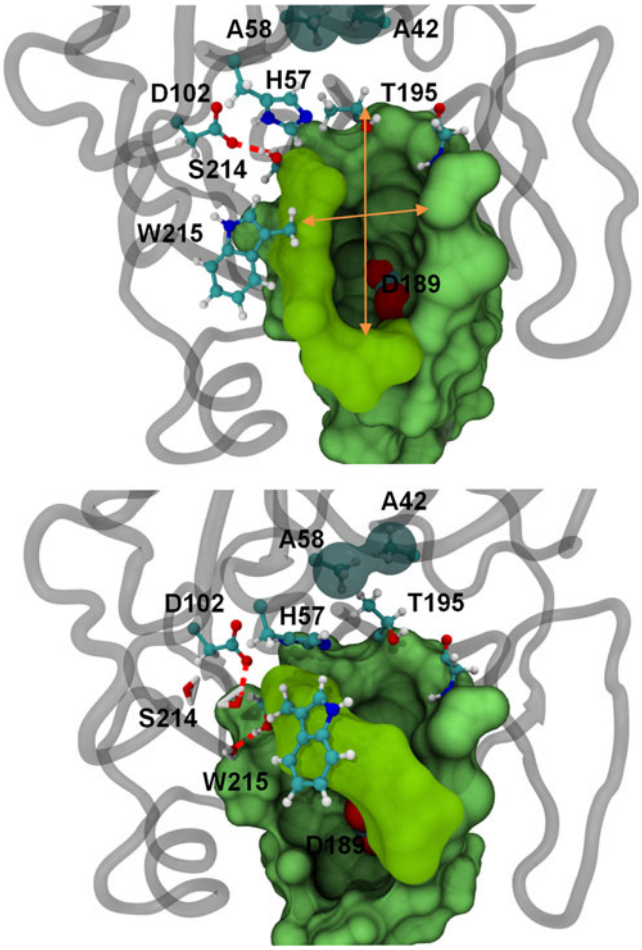

Fig. 6 Conformational rearrangement of the S214-G219 fragment (*light green*) for the AAT variant. *Top*: the position of the loop is identical to that observed in WT (open conformation). The hydrogen bond between S214 and A102 remained intact. *Bottom*: displacement of the S214-G219 fragment toward to the S1 binding site, precluding access to the primary binding site for putative substrates (closed conformation). In the closed conformation for the AAT variant, water molecules can gain access near W215 and S214. *Vertical* (S195 Cβ - Y217 Cγ) and *horizontal* (E192 Cα -W215 Cα) measurements for the opening of the S1 binding pockets are shown by *orange arrows* in the top picture. These measurements were used to monitor the motions of the S214-G219 fragment

T195 hydroxyl group. Conformational flexibility of T195 is supported by the minor displacement of H57, which also protrudes to D102 and S214. While the interaction between H57 and D102 remains intact, the interaction between D102 and S214 is disrupted, leading to additional water molecules entering the protein cavity near S214. Table 4 shows the average number of water molecules found near S214 during MD simulation. Based on our calculations, additional solvation around S214 for triple variants induces conformational rearrangement of S214 and an increase in hydrophilicty around W215, leading to destabilization of mostly hydrophobic indole side chain. The presence of additional water molecules near S214 and W215 was not observed for the double variants (AA and AV) indicating that removal of the C42-C58 bridge alone has smaller conformational impact on trypsin structure.

To characterize the movement of the S214-G219 segment, in addition to RMSD values, the distance between S195 or T195 Cβ and Y217 Cγ and between E192 Cα and W215 Cα were calculated from MD trajectories. These distances represent vertical and horizontal measurements across the S1 binding pocket (Fig. 6, orange arrows). A closed conformation can be described by smaller distances relative to the open conformation. The values for the WT simulations were 13.7±1.0 Å and 9.2±0.5 Å for S195 Cβ - Y217 Cγ and E192 Cα -W215 Cα distances, respectively, and correspond to the open conformation. The S195 Cβ - Y217 Cγ and E192 Cα -W215 Cα distances for the closed conformation in case of AAT variant for example, were measured at 10.7±0.4 Å and 6.0±0.5 Å respectively.

Figure 7 summarizes the dynamic changes for the S214-G219 segment for WT and all variants during MD simulations. Conformational clustering for this figure was achieved by plotting the distance between S195 or T195 Cβ and Y217 Cγ against RMSDs of all atoms for the S214-G219 segment. In addition, each point on the plot was colored by the E192 Cα -W215 Cα distance value. Similar to Fig. 4, the area highlighted by the blue rectangle corresponds to the conformational footprint of WT trypsin. Based on the calculated values and examination of the MD

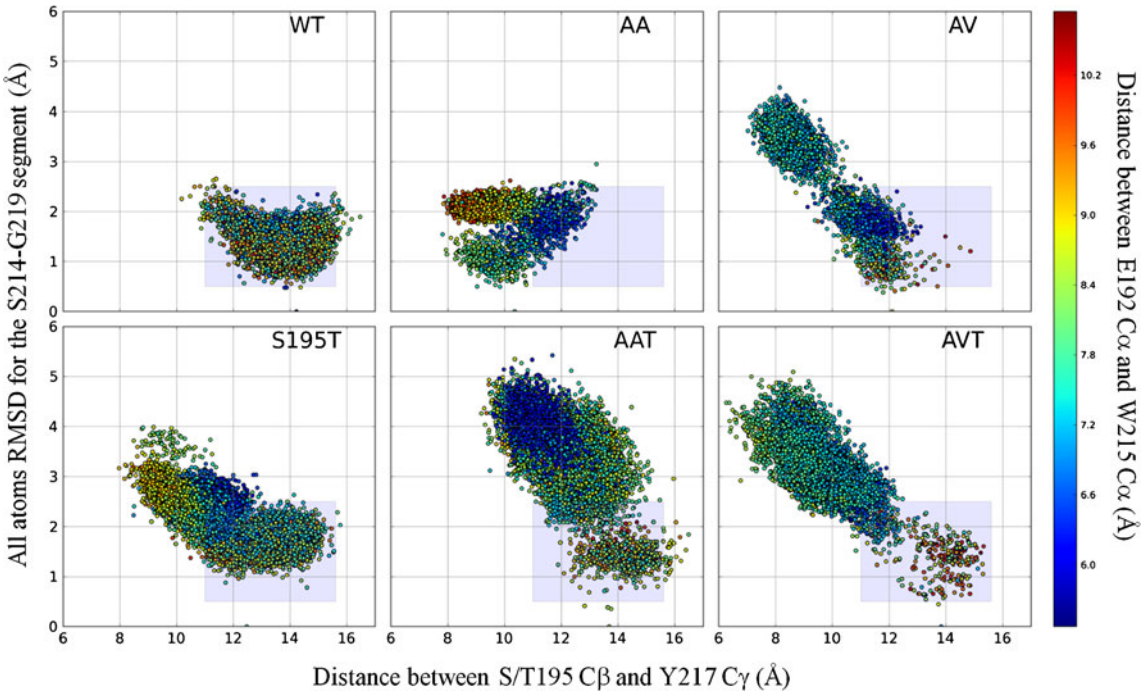

Fig. 7 Two-dimensional diagrams showing the conformational motions of the S214-G219 segment. The x-axis refers to the distribution of the distances between S195 or T195 Cβ and Y217 Cγ, and the y-axis refers to the all atoms RMSD values for the S214-G219 segment. Similar to Fig. 4, each point represents a single snapshot and is colored according to the distance between E192 Cα and W215 Cα. The color map for the distances is shown on the right. *Red* indicates the longest and blue the shortest distance ranges (Å). There were 11,500 snapshots per trajectory. The plot for the WT shows that the S214-G219 fragment is in the open conformation for almost all trajectory snapshots. The area corresponding to the WT conformational family (highlighted by a *blue rectangle*) was projected to the other plots to emphasize differences between WT and variants. Values calculated for S195T with an intact C42-C58 disulfide bridge also support an open conformation. Note that for the S195T simulation, the structures outside the WT conformational family are less than 10 %. MD trajectories for the variants with a removed C42-C58 disulfide bond revealed considerable deviations from the geometrical properties observed for the WT. For AAT, AVT, and AV structures, the RMSD values and distance measurements fall outside the ranges observed for the WT and correlate with a closed conformation of the S214-G219 motif. For the AA variant, displacement is clearly measurable by the S195 Cβ - Y217 Cγ distance, but not by RMSD values

trajectories, WT clearly favors an open conformation. The S195T variant was also mostly in the open conformation, with a smaller (less than 10 %) population outside the WT range. For other variants examined, a large number of structures fell outside the WT footprint, indicating larger populations of the closed conformation. All variants showed multiple conversions between open and closed conformations during the 115 ns simulations. Larger populations of the closed conformation correlated with reduced enzymatic activities for the triple in comparison to double variants. In addition, high K_M values for the AAT and AVT variants (214 and 217.7 μM, respectively [27]) for the Z-GPR-pNA substrate suggest that existence of the larger closed conformation populations could reduce affinity of this substrate toward triple variants. The larger populations of open conformations for the AA and AV variants agree with the biochemically observed higher affinity toward the same substrate. AA and AV K_M values are 44.6 and 29.5 μM, respectively; however, for the Z-GPR-AMC substrate, the K_M values were similar [27].

The conformational flexibility at the S1 subsite was previously observed for a number of proteases. The rotation of the W215 side chain toward the enzyme binding pocket has been reported for the S214K and S214E variants of trypsin, and this displacement resulted in reduced substrate recognition [11]. However, the activity of the enzyme was unchanged in the S214A variant, suggesting size of the substituted side chain could also play a role in substrate recognition [11, 48]. Large conformational differences for the W215-G219 segment leading to a closed conformation of the binding pocket was also reported for prostasin complexed to an inhibitor [49] and α1-typtase, which only been shown to open by D216G mutation [50, 51]. Structures of thrombin and thrombin variants showed the active site exists in a conformation that is either collapsed or accessible (open) to the substrate. Opening and closing of the primary specificity pocket was achieved by repositioning of the W215 side chain together with the W215-Y217 segment, which in the closed form prevented access to the primary specificity pocket [52, 53]. The existence of both closed and open conformations for thrombin suggested the closed form can be converted to the open one by specific co-factor or allosteric activators [54]. To our knowledge, there is no structural data so far suggesting displacement of the entire S214-G219 segment leading to a closed conformation of the S1 binding site in the WT trypsin. Our MD simulation showed this type of closed conformation can exist in the trypsin variants.

Conclusions

We employed MD simulations to investigate the conformational features of five trypsin variants. Previously reported biochemical studies showed that replacement of S195 by T195 at the catalytic triad leads to almost a complete loss in catalytic activity, but removal of the C42-C58 disulfide bond partially restored the activity of the S195T variants, albeit with lower efficiency as compared to WT. Removal of the C42-C58 disulfide bond also reduced the activity up to three orders of magnitude (in k_{cat}/K_M) of the variants with serine at position 195 [27]. Based on our conformational studies, we propose that one component leading to the observed restored enzymatic activity of the S195T variants with the substituted cysteine bridge is the conformational flexibility of the threonine residue (T195). In order to form the essential tetrahedral intermediate in the acylation step, the hydroxyl group of the catalytic residue should be positioned properly for an attack on the carbonyl group of the substrate-scissile carbonyl. Steric constraints imposed by the C42-C58 disulfide bridge in the S195T variant resulted in a conformation wherein the methyl group of T195 occupies the essential position of the hydroxyl group observed in the WT (a catalytically inactive conformation). Removal of the C42-C58 disulfide bond allows the rotation along the Cα-Cβ bond that places the hydroxyl moiety of T195 in the catalytically active orientation. Based on MD simulations, variants with higher enzymatic activities have a larger population of this catalytically active conformation. The catalytically inactive conformation was clearly favored by the S195T variant with an intact C42-C58 bridge. The observation of the conformational flexibility of the S214-G219 fragment at the S1 subsite, particularly for the variants with the removed C42-C58 disulfide bond, suggested this fragment could also modulate enzymatic activity. The dynamic motions of the S214-G219 motif in double and triple variants resulted in an exchange between open and closed conformations for the S1 binding site. The closed conformation for the S214-G219 motif precludes access to the primary specificity pocket and could contribute to enzyme activity. This was reflected in the lower K_M values for the AA and AV variants relative to the AAT and AVT variants, which demonstrated a higher population of the closed conformation during the MD simulations. Similar closed conformations for the S1 pocket were previously reported for serine proteases with various amino acids substitutions [49, 54–57].

The chemical nature of the substrate used in biochemical studies undoubtedly plays a vital role in substrate recognition and cleavage. Besides the primary binding site at the S1 pocket, the trypsin active site possesses multiple interaction sites with putative substrates that control catalytic efficiencies toward these substrates. Detailed interpretation of the biochemical data will require further structural studies, including conformational studies of the substrate-enzyme complexes. The present work provides possible insights on the substrate recognition for a number of trypsin variants. The dynamic motions of the S214-G219 segment, which

lead to the conformational population with the closed S1 binding pocket observed in this study, should modulate substrate recognition. At least two scenarios are possible: the closed conformation obstructs binding of substrate, or the chemical and/or structural nature of substrate facilitate opening of the pocket leading to a formation of a substrate enzyme complex. The observation of the open and closed conformations for the S1 binding pocket could be a new mode of the substrate regulation for the functional trypsin variants tested in this work.

Acknowledgments This work was supported by the National Institutes of Health grant 5SC2GM095448 (to Anton Guliaev), Center for Computing for Life Sciences (http://cs.sfsu.edu/ccls/) Mini-grant, San Francisco State University (to Anton Guliaev) and National Science Foundation Award MCB0643988 (to Teaster T. Baird Jr.).

References

1. Page MJ, Di Cera E (2008) Serine peptidases: classification, structure and function. Cell Mol Life Sci 65(7–8):1220–1236. doi:10.1007/s00018-008-7565-9
2. Ross J, Jiang H, Kanost MR, Wang Y (2003) Serine proteases and their homologs in the Drosophila melanogaster genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene 304:117–131
3. Puente XS, Sanchez LM, Overall CM, Lopez-Otin C (2003) Human and mouse proteases: a comparative genomic approach. Nat Rev Genet 4(7):544–558. doi:10.1038/nrg1111
4. Puente XS, Sanchez LM, Gutierrez-Fernandez A, Velasco G, Lopez-Otin C (2005) A genomic view of the complexity of mammalian proteolytic systems. Biochem Soc Trans 33(Pt 2):331–334
5. Ding X, Rasmussen BF, Petsko GA, Ringe D (1994) Direct structural observation of an acyl-enzyme intermediate in the hydrolysis of an ester substrate by elastase. Biochemistry 33(31):9285–9293
6. Blanchard H, James MN (1994) A crystallographic re-investigation into the structure of Streptomyces griseus proteinase A reveals an acyl-enzyme intermediate. J Mol Biol 241(4):574–587
7. Katona G, Wilmouth RC, Wright PA, Berglund GI, Hajdu J, Neutze R, Schofield CJ (2002) X-ray structure of a serine protease acyl-enzyme complex at 0.95-A resolution. J Biol Chem 277 (24):21962–21970. doi:10.1074/jbc.M200676200
8. Katona G, Berglund GI, Hajdu J, Graf L, Szilagyi L (2002) Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. J Mol Biol 315(5):1209–1218. doi:10.1006/jmbi.2001.5305
9. Radisky ES, Lee JM, Lu CJ, Koshland DE Jr (2006) Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. Proc Natl Acad Sci USA 103 (18):6835–6840
10. Brady K, Wei AZ, Ringe D, Abeles RH (1990) Structure of chymotrypsin-trifluoromethyl ketone inhibitor complexes: comparison of slowly and rapidly equilibrating inhibitors. Biochemistry 29(33):7600–7607
11. McGrath ME, Vasquez JR, Craik CS, Yang AS, Honig B, Fletterick RJ (1992) Perturbing the polar environment of Asp102 in trypsin: consequences of replacing conserved Ser214. Biochemistry 31 (12):3059–3064
12. Neidhart D, Wei Y, Cassidy C, Lin J, Cleland WW, Frey PA (2001) Correlation of low-barrier hydrogen bonding and oxyanion binding in transition state analogue complexes of chymotrypsin. Biochemistry 40(8):2439–2447
13. Hedstrom L (2002) Serine protease mechanism and specificity. Chem Rev 102(12):4501–4523. doi:10.1021/Cr000033x
14. Perona JJ, Tsu CA, Craik CS, Fletterick RJ (1993) Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. J Mol Biol 230(3):919–933
15. Frey PA, Whitt SA, Tobin JB (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases. Science 264(5167):1927–1930
16. Warshel A, Papazyan A (1996) Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. Proc Natl Acad Sci USA 93(24):13665–13670
17. Wahlgren WY, Pal G, Kardos J, Porrogi P, Szenthe B, Patthy A, Graf L, Katona G (2011) The catalytic aspartate is protonated in the Michaelis complex formed between trypsin and an in vitro evolved substrate-like inhibitor: a refined mechanism of serine protease action. J Biol Chem 286(5):3587–3596
18. Perona JJ, Craik CS (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. J Biol Chem 272(48):29987–29990
19. Helland R, Otlewski J, Sundheim O, Dadlez M, Smalas AO (1999) The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI. J Mol Biol 287(5):923–942
20. Helland R, Berglund GI, Otlewski J, Apostoluk W, Andersen OA, Willassen NP, Smalas AO (1999) High-resolution structures of three new trypsin-squash-inhibitor complexes: a detailed comparison with other trypsins and their complexes. Acta Crystallogr D: Biol Crystallogr 55(Pt 1):139–148. doi:10.1107/S090744499801052X
21. Zakharova E, Horvath MP, Goldenberg DP (2009) Structure of a serine protease poised to resynthesize a peptide bond. Proc Natl Acad Sci U S A 106(27):11034–11039
22. Ascenzi P, Bocedi A, Bolognesi M, Spallarossa A, Coletta M, De Cristofaro R, Menegatti E (2003) The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein. Curr Protein Pept Sci 4(3):231–251
23. Katz BA, Finer-Moore J, Mortezaei R, Rich DH, Stroud RM (1995) Episelection: novel Ki approximately nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface. Biochemistry 34(26):8264–8280
24. Perona JJ, Craik CS (1995) Structural basis of substrate specificity in the serine proteases. Protein Sci 4(3):337–360. doi:10.1002/pro.5560040301
25. Liu SQ, Meng ZH, Fu YX, Zhang KQ (2010) Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. J Mol Model 16(1):17–28. doi:10.1007/s00894-009-0518-x
26. Rawlings ND, Tolle DP, Barrett AJ (2004) MEROPS: the peptidase database. Nucleic Acids Res 32:D160–D164. doi:10.1093/nar/gkh071
27. Baird TT Jr, Wright WD, Craik CS (2006) Conversion of trypsin to a functional threonine protease. Protein Sci 15(6):1229–1238
28. Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K (2009) Discovery through the computational microscope. Structure 17 (10):1295–1306
29. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652. doi:10.1038/nsb0902-646
30. Case DA, Darden TA, Cheatham TEI, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2010) AMBER 11. University of California, San Francisco

31. McGrath ME, Haymore BL, Summers NL, Craik CS, Fletterick RJ (1993) Structure of an engineered, metal-actuated switch in trypsin. Biochemistry 32(8):1914–1919
32. 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
33. 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
34. Steinbrecher T, Mobley DL, Case DA (2007) Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. J Chem Phys 127(21):214108. doi:10.1063/1.2799191
35. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690
36. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688. doi:10.1002/Jcc.20290
37. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 106 (5):1589–1615. doi:10.1021/Cr040426m
38. Cerutti DS, Freddolino PL, Duke RE, Case DA (2010) Simulations of a protein crystal with a high resolution x-ray structure: evaluation of force fields and water models. J Phys Chem B 114 (40):12811–12824. doi:10.1021/Jp105813j
39. Guliaev AB, Hang B, Singer B (2004) Structural insights by molecular dynamics simulations into specificity of the major human AP endonuclease toward the benzene-derived DNA adduct, pBQ-C. Nucleic Acids Res 32(9):2844–2852. doi:10.1093/Nar/Gkh594
40. Rodriguez B, Yang YN, Guliaev AB, Chenna A, Hang B (2010) Benzene-derived N-2-(4-hydroxyphenyl)-deoxyguanosine adduct: UvrABC incision and its conformation in DNA. Toxicol Lett 193 (1):26–32. doi:10.1016/j.toxlet.2009.12.005
41. Beck DAC, Daggett V (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. Methods 34 (1):112–120. doi:10.1016/j.ymeth.2004.03.008
42. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38
43. Rastelli G, Del Rio A, Degliesposti G, Sgobba M (2010) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J Comput Chem 31(4):797–810. doi:10.1002/Jcc.21372
44. Onufriev A, Case DA, Bashford D (2002) Effective born radii in the generalized born approximation: the importance of being perfect. J Comput Chem 23(14):1297–1304. doi:10.1002/Jcc.10126
45. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23(4):566–579
46. Perona JJ, Tsu CA, McGrath ME, Craik CS, Fletterick RJ (1993) Relocating a negative charge in the binding pocket of trypsin. J Mol Biol 230(3):934–949. doi:10.1006/jmbi.1993.1211
47. Ishida T, Kato S (2004) Role of Asp102 in the catalytic relay system of serine proteases: a theoretical study. J Am Chem Soc 126(22):7111–7118. doi:10.1021/ja030405u
48. Krem MM, Prasad S, Di Cera E (2002) Ser(214) is crucial for substrate binding to serine proteases. J Biol Chem 277(43):40260–40264. doi:10.1074/jbc.M206173200
49. Spraggon G, Hornsby M, Shipway A, Tully DC, Bursulaya B, Danahay H, Harris JL, Lesley SA (2009) Active site conformational changes of prostasin provide a new mechanism of protease regulation by divalent cations. Protein Sci 18(5):1081–1094. doi:10.1002/pro.118
50. Marquardt U, Zettl F, Huber R, Bode W, Sommerhoff C (2002) The crystal structure of human alpha1-tryptase reveals a blocked substrate-binding region. J Mol Biol 321(3):491–502
51. Rohr KB, Selwood T, Marquardt U, Huber R, Schechter NM, Bode W, Than ME (2006) X-ray structures of free and leupeptin-complexed human alphaI-tryptase mutants: indication for an alpha–>beta-tryptase transition. J Mol Biol 357(1):195–209
52. Bah A, Carrell CJ, Chen Z, Gandhi PS, Di Cera E (2009) Stabilization of the E* form turns thrombin into an anticoagulant. J Biol Chem 284(30):20034–20040
53. Gandhi PS, Chen Z, Mathews FS, Di Cera E (2008) Structural identification of the pathway of long-range communication in an allosteric enzyme. Proc Natl Acad Sci USA 105(6):1832–1837
54. Niu W, Chen Z, Gandhi PS, Vogt AD, Pozzi N, Pelc LA, Zapata F, Di Cera E (2011) Crystallographic and kinetic evidence of allostery in a trypsin-like protease. Biochemistry 50(29):6301–6307. doi:10.1021/bi200878c
55. Shia S, Stamos J, Kirchhofer D, Fan B, Wu J, Corpuz RT, Santell L, Lazarus RA, Eigenbrot C (2005) Conformational lability in serine protease active sites: structures of hepatocyte growth factor activator (HGFA) alone and with the inhibitory domain from HGFA inhibitor-1B. J Mol Biol 346(5):1335–1349
56. Forneris F, Ricklin D, Wu J, Tzekou A, Wallace RS, Lambris JD, Gros P (2010) Structures of C3b in complex with factors B and D give insight into complement convertase formation. Science 330 (6012):1816–1820
57. Carvalho AL, Sanz L, Barettino D, Romero A, Calvete JJ, Romao MJ (2002) Crystal structure of a prostate kallikrein isolated from stallion seminal plasma: a homologue of human PSA. J Mol Biol 322(2):325–337