Review

Allostery in trypsin-like proteases suggests new therapeutic strategies

David W. Gohara and Enrico Di Cera

Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, St. Louis, MO 63104, USA

Trypsin-like proteases (TLPs) are a large family of enzymes responsible for digestion, blood coagulation, fibrinolysis, development, fertilization, apoptosis and immunity. A current paradigm posits that the irreversible transition from an inactive zymogen to the active protease form enables productive interaction with substrate and catalysis. Analysis of the entire structural database reveals two distinct conformations of the active site: one fully accessible to substrate (E) and the other occluded by the collapse of a specific segment (E*). The allosteric E*–E equilibrium provides a reversible mechanism for activity and regulation in addition to the irreversible zymogen to protease conversion and points to new therapeutic strategies aimed at inhibiting or activating the enzyme. In this review, we discuss relevant examples, with emphasis on the rational engineering of anticoagulant thrombin mutants.

Allostery in trypsin-like proteases (TLPs)

Approximately 700 proteases are present in the human genome, mediating extracellular matrix remodeling, blood coagulation, immunity, development, protein processing, cell signaling and apoptosis [1]. Four protease families account for over 40% of all proteolytic enzymes in humans. These are the ubiquitin-specific proteases [2], the adamalysins [3], prolyl oligopeptidases [4] and the TLPs [5]. TLPs are phylogenetically grouped into six functional categories: digestion, coagulation and immunity, tryptase, matriptase, kallikrein and granzymes. Trypsins, chymotrypsins and elastases breakdown polypeptides in the digestive system [6]. Thrombin, C1r and cognate proteases define the blood coagulation and complement cascades [7,8]. Tryptases are major components in the secretory granules of mast cells and feature homotetrameric structure [9], whereas matriptases are membrane-bound proteases associated with a variety of cancers [10]. A similar association with cancer is found in the kallikrein family [11], also known for its role in regulation of blood pressure through the kinin system [12]. Granzymes are mediators of directed apoptosis by natural killer cells and cytotoxic T cells that play key roles in the defense against viral infection [13]. Because of their abundance and involvement in health and disease, TLPs represent obvious targets of therapeutic intervention.

TLPs utilize a canonical catalytic triad for activity, composed of the highly conserved residues H57, D102 and S195 (chymotrypsinogen numbering). Catalysis is furthered by a H-bond between D102 and H57 (Figure 1), which facilitates the abstraction of the proton from S195 and generates a pot nucleophile [14]. A reaction pathway involving two tetrahedral intermediates produces hydrolysis of the target peptide bond. Initially, the hydroxyl O atom of S195 attacks the carbonyl of the peptide substrate with H57 acting as a general base. The oxyanion tetrahedral intermediate is stabilized by the backbone N atoms of G193 and S195, which generate a positively charged pocket within the active site known as the oxyanion hole (Figure 1). Collapse of the tetrahedral intermediate generates the acyl-enzyme and stabilization of the newly created N-terminus is mediated by H57. Then, a water molecule displaces the free polypeptide fragment and attacks the acyl-enzyme. The oxyanion hole stabilizes the second tetrahedral intermediate of the pathway and collapse of this intermediate liberates a new C-terminus in the substrate. The catalytic triad and oxyanion hole are assisted in their critical role by other structural determinants (Figure 1). Residue 189 at the bottom of the primary specificity pocket interacts with the substrate residue at the site of cleavage and defines the primary specificity of the enzyme [5,14]. Residues in the 215–217 segment flanking the active site provide additional recognition sites for substrate residues situated immediately upstream of the peptide bond to be cleaved [15].

Activity in TLPs ensues via a common mechanism that involves the irreversible processing of an inactive zymogen precursor [5,14,15]. The zymogen is proteolytically cut at the identical position, between residues 15 and 16, in all members of the family to generate a new N-terminus that ion-pairs with the highly conserved D194 next to the catalytic S195 and organizes both the oxyanion hole and primary specificity pocket for substrate binding and catalysis (Figure 1). The zymogen to protease conversion affords a useful paradigm to explain the onset of catalytic activity as seen in the digestive system, blood coagulation or the complement system and is particularly useful to understand the initiation, progression and amplification of enzyme cascades, where each component acts as a substrate in the inactive zymogen form in one step and as an active enzyme in the subsequent step [16–18]. However, considerable variation in catalytic activity is observed among TLPs following conversion from the inactive zymogen form. Digestive enzymes like trypsin, chymotrypsin and elastase, complement factors C1r and C1 s, and coagulation factors like thrombin are highly active after the zymogen to

Corresponding author: Di Cera, E. (enrico@slu.edu).

0167-7799/$ – see front matter © 2011 Elsevier Ltd. All rights reserved. doi:10.1016/j.tibtech.2011.06.001 Trends in Biotechnology, November 2011, Vol. 29, No. 11

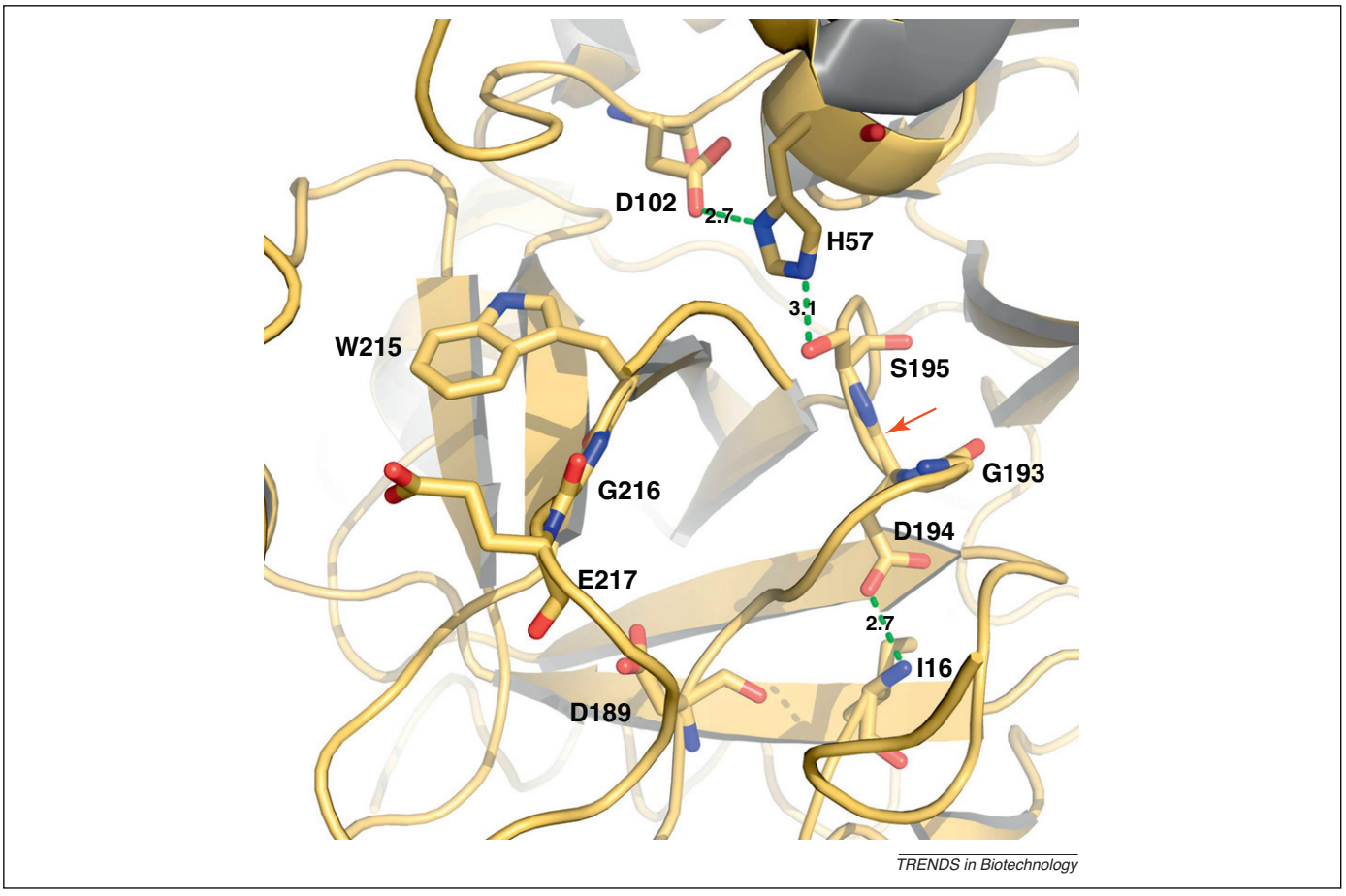

Figure 1. Structure of the active site of a representative trypsin-like protease (TLP). The catalytic residues (H57, D102 and S195), the primary specificity site (D189), the oxyanion hole formed by the backbone N atoms of S195 and G193 (red arrow), residues of the 215–217 segment and the H-bond between the N-terminus of the catalytic chain (I16) and the side chain of D194 are shown. Catalysis requires correct folding of the active site promoted by activation of the zymogen and formation of the I16–D194 H-bond, correct positioning of the catalytic residues for H transfer and correct architecture of the oxyanion hole. Binding of substrate is optimized by interaction with the primary specificity pocket (residue 189) and docking against the 215–217 segment.

protease conversion has taken place. On the other hand, complement factors B and C2 are mostly inactive until binding of complement factors C3b and C4b enable catalytic activity at the site where amplification of C3 activation leads to formation of the membrane attack complex [7,19,20]. Coagulation factor VIIa circulates in the blood as a poorly active protease that acquires full catalytic competence upon interaction with tissue factor that is exposed to the blood stream upon vascular injury [21,22]. Complement factor D assumes an inactive conformation with a distorted catalytic triad [23,24] until binding to C3b and factor B promote substrate binding and catalytic activity [25,26]. The high catalytic activity of trypsin, C1r and thrombin contrasted by the ability of complement factor D or coagulation factor VIIa to remain in a zymogen-like form suggests that the trypsin fold may assume active and inactive conformations even after the zymogen to protease conversion has taken place.

Conformational heterogeneity of trypsin affecting the active site and its immediate vicinity is supported by NMR studies [27] and rapid kinetic measurements [28]. The D216G mutant of αI-tryptase crystallizes in the free form with the 215–217 segment near the active site in equilibrium between open and collapsed conformations [29]. Existence of an allosteric equilibrium between active and inactive forms has been proposed for coagulation factor VIIa [22,30]. A similar equilibrium is supported by rapid kinetics data on thrombin [31,32], factor Xa and activated protein C [33]. Structures of thrombin in the free form reveal a conformation, E, with the active site open [34–36] and an alternative conformation, E*, with the active site blocked by repositioning of the 215–217 segment [37–40]. Conformational heterogeneity of the free form of thrombin is supported by recent NMR studies [41]. Systematic inspection of the entire structural database (Box 1) reveals that the 215–217 segment may assume alternative conformations that control access to the active site. A strong case emerges for TLPs to be considered allosteric enzymes.

Allostery in trypsin-like proteases: the E* and E forms

There are currently 1377 structures of TLPs and zymogens deposited in the Protein Data Bank, but only 51 of them are free of ligands bound to the active site or sites known to influence activity and/or stability (Box 1). These 51 structures depict the protease or zymogen conformation of 21 different members of the trypsin family and were analyzed as a set to establish the 3D architecture of the active site and its substrate accessibility. The high resolution structure 1SHH of thrombin bound to the irreversible inhibitor H-D-Phe-Pro-Arg-CH_2Cl (PPACK) [35] was used as reference for estimating the volume of atoms along the 215–217 segment of the protease/zymogen in the set encroaching

Box 1. Analysis of the current structural database of TLPs and zymogens

All structures belonging to the Trypsin Pfam (PF00089, 1377 total) were downloaded from the RCSB. The structure 1SHH of thrombin bound to PPACK [35] was chosen as reference. All other structures were superimposed on 1SHH by RMSD minimization of the α-carbon atoms using the program LSQKAB from the CCP4 suite [93]. Structures with multiple chains corresponding to the trypsin-like domain were superimposed and analyzed independently (1682 chains total). The bound PPACK in 1SHH (Figure I) was used to determine the accessibility of the active site for each structure in the database along the structurally analogous 215–217 peptide segment. The designation of open (E, Z) versus collapsed (E*, Z*) was assigned based on the degree of structural overlap of the aligned PPACK with atoms in the respective target molecules. Structures where the overlap with PPACK would come from ligands bound to the active site or residues from crystal contacts were excluded from the final statistics. The volume of atomic overlap was determined using standard atomic radii and bonding distances and was calculated as follows. A grid of constant volume was constructed around the region encompassing the 215–217 segment (including side chains) plus PPACK for each of the superimposed structures. Spacing between grid points was set to 0.1 Å. The distance between each grid point and each atom was calculated. Any grid point that fell within the covalent bonding radius of an atom was marked as inside. Based on the positioning of PPACK in the active site, the arginine side chain is located such that in a fully open conformation overlap with the 215–217 segment should not be possible. In cases where overlap of a lower numbered atom (e.g. Cβ) on the side chain occurred, it was assumed that subsequent higher numbered atoms (e.g. Cγ, Cδ...) were also precluded from the area and were therefore counted as part of the overlap volume. Once all of the grid points had been assigned, the number of points marked as inside were counted with each grid point corresponding to a single cube of constant volume ($0.1 \times 0.1 \times 0.1$ Å or 1×10^{-3} Å^3). The sum of all volumes corresponded to the union of space occupied by the 215–217 segment and PPACK, that is the combined volume, was determined. Additionally, for each structure, the volume occupied by the 215–217 segment alone and PPACK alone (a constant) were determined. The sum of the 215–217 segment alone plus PPACK alone minus the combined volume was calculated and assigned as the volume of overlap. In cases where no overlap between the 215–217 segment and PPACK occurs, the sum of the two components minus the combined volume should equal zero. The complete list of structures meeting the criteria for analysis is provided in Table 1.

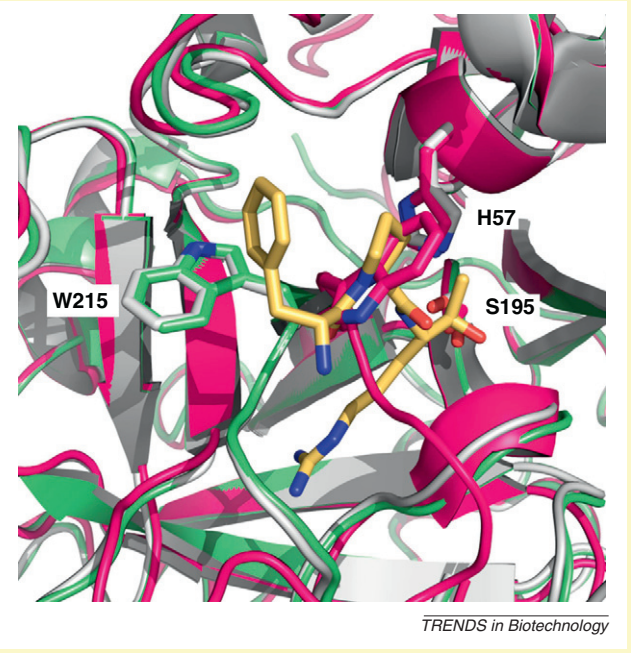

Figure I. The E and E* forms. Overlay of the structures of thrombin in the E (1SGI, green) [35] and E* (3BEI, magenta) [38] form on the structure of thrombin bound to the active site inhibitor PPACK (1SHH, white) [35]. PPACK (yellow) and residues H57, S195 and W215 from the protein are rendered as sticks. In the E form, there is no overlap of residues 215–217 with the space occupied by PPACK in the active site. In the E* form, collapse of W215 and the 215–217 segment produces 159 Å^3 occlusion of the space to be occupied by PPACK, which represents 48% of the total volume of the inhibitor (331 Å^3).

upon PPACK binding to the active site (Box 1). Strikingly, the structures partition into two groups: one with the active site occluded by the side chain of W215 or repositioning of the 215–217 segment as seen in the E* form of thrombin [37–40], and the other with the active site fully open as observed in the E form of thrombin [34–36]. Structures in the E* form feature a collapse of the 215–217 segment that obliterates an average 131 Å^3 of the total 331 Å^3 volume of PPACK bound to the active site. In the E form, the active site is fully open and the 215–217 segment obliterates on average only 1.0 Å^3 of the total volume of PPACK.

The E* form is typified by the structure 3BEI of thrombin in a collapsed conformation [38] and is observed in a number of thrombin mutants [37–40]. Collapse of the 215–217 segment in 3BEI causes an approximate 48% steric blockage in the volume that would be occupied by PPACK in the active site (Figure 2). Importantly, the collapse is removed upon binding of ligands to exosite I located >20 Å away [38], which confirms that the E* and E forms are in equilibrium and their distribution can be shifted by ligand binding to distinct structural domains. Smaller but very significant collapse of the 215–217 segment is seen in thrombin mutants carrying substitutions in this segment [42–44]. The first evidence of collapse of W215 was reported for the S215W mutant of complement factor D [45]. In the wild type, the active site is occluded by a shift of the 215–217 segment that brings S215 within the active site and positions R218 in H-bonding interaction with D189, thereby maintaining the enzyme in a self-inhibited conformation [24,26]. The same self-inhibited form is also seen in the catalytically inactive mutant S195A and R218A [25]. When the atypical S215 is replaced with Trp, the residue found at this position in the majority of TLPs, the collapse becomes even more evident as the indole side chain blocks access to the active site, similar to 3BEI. Most of the collapse is removed when factor D binds to a complex of C3b and factor B [25], thereby enabling factor D to activate its physiological substrate. A more substantial collapse of the 215–217 segment that completely occludes the active site is observed in prostate specific antigen [46], tonin [47] and the prophenoloxidase activating factor II, which is a non-catalytic clip-domain enzyme [48]. αI-tryptase is a tetrameric protease with little enzymatic activity and crystallizes in a collapsed form due to repositioning of W215 in the active site and coupling of D216 immediately downstream with the oxyanion hole [49]. The D216G replacement restores the typical residue found at this position in TLPs and produces a structure where the 215–217 segment can be fit to two conformations in a 3:1 occupancy

Table 1. Structures of TLPs and zymogens in the open (E, Z) and collapsed (E*, Z*) forms[a]

PDB ID		Resolution (Å)	Refs
Collapsed form			
Protease (E*)			
1DST	Complement factor D mutant S215W	2.0	[45]
1DSU, 1HFD	Complement factor D	2.0, 2.3	[24,26]
2XW9	Complement factor D mutant S195A	1.2	[25]
2XWA	Complement factor D mutant R218A	2.8	[25]
1GVZ	Prostate specific antigen	1.42	[46]
1LTO	αI-tryptase	2.2	[49]
2F9O[b]	αI-tryptase mutant D216G	2.1	[29]
1RD3	Thrombin mutant E217K	2.5	[42]
1TQ0, 3EE0	Thrombin mutant W215A/E217A	2.8, 2.75	[43,44]
2GP9, 3BEI	Thrombin mutant D102N	1.87, 1.55	[38,40]
3GIC	Thrombin mutant Δ146–149e	1.55	[37]
3JZ2	Thrombin mutant N143P	2.4	[39]
3EDX, 3HK3, 3HK6	Murine thrombin mutant W215A/E217A	2.4, 1.94, 3.2	[43]
1TON	Tonin	1.8	[47]
1YBW	Hepatocyte growth factor activator	2.7	[50]
2B9L	Prophenoloxidase activating factor II	2.0	[48]
3DFJ, 3E1X	Prostasin	1.45, 1.7	[51,52]
Zymogen (Z*)			
1CHG, 1EX3	Chymotrypsinogen	2.5, 3.0	[53,54]
1DDJ, 1QRZ	Plasminogen	2.0, 2.0	[55,56]
1FDP	Profactor D	2.1	[23]
1GVL	Prokallikrein 6	1.8	[58]
1MD7	Complement profactor C1r	3.2	[60]
1MZA, 1MZD	Progranzyme K	2.23, 2.9	[61]
1SGF	α subunit of nerve growth factor	3.15	[62]
3NXP	Prethrombin-1	2.2	[63]
Open form			
Protease (E)			
1MD8	Complement factor C1r	2.8	[60]
1MH0, 1SGI	Thrombin mutant R77aA	2.8, 2.3	[35,36]
2PGB	Thrombin mutant C191A/C220A	1.54	[34]
2OCV	Murine thrombin	2.2	[64]
1NPM	Neuropsin	2.1	[67]
2F9O[b]	αI-tryptase mutant D216G	2.1	[29]
2G51, 2G52	Trypsin	1.84, 1.84	[68]
2I6Q, 2I6S, 2ODP, 2ODQ	Complement factor C2a	2.1, 2.7, 1.9, 2.3	[65,66]
Zymogen (Z)			
1TGB, 1TGN	Trypsinogen	1.8, 1.65	[69,70]
1ZJK	Zymogen of MASP-2	2.18	[72]
2F83	Coagulation factor XI	2.87	[73]
2OK5	Complement profactor B	2.3	[71]

[a]Structures of TLPs not included in the Table are in the E form but feature Ca^{2+} bound to regions known to affect activity and stability, and/or acetate or sulfate bound at or near the catalytic residues. Examples of such structures are β trypsin, cationic trypsin, trypsins from *Fusarium oxysporum* and *Streptomyces gryseus*, and kallikrein 7. New structures of these enzymes without Ca^{2+} or anionic ligands near the active site will enable unambiguous assignment of the E or E* form.

[b]The structure of αI-tryptase mutant D216G documents both the E and E* conformations in a 3:1 ratio [29].

ratio: one open and the other collapsed as in the wild type [29]. Hepatocyte growth factor activator crystallizes in the free form in a collapsed conformation [50] similar to 3BEI. In prostasin, the 215–217 segment collapses into the active site with the side chain of W215 assuming two possible conformations [51,52], one of which is similar to that of 3BEI [51]. The collapse is removed when Ca^{2+} binds to the primary specificity pocket [52].

Interestingly, collapse of the 215–217 segment into the active site is not unique to the protease form and is documented in the structures of several zymogens (Figure 2). Chymotrypsinogen [53,54] adopts a collapsed form for the active site region. Plasminogen features W215 in a 'foot-in-mouth' conformation that occludes access to the primary specificity pocket [55,56]. The steric hindrance is partially relieved upon binding of streptokinase [57]. In complement profactor D the active site is screened by a collapse of the 215–217 segment [23] and a similar collapse is observed in prokallikrein 6 [58] and complement profactor C1r [59], but is removed in the active C1r [60]. Collapse of residue 215 against the catalytic H57 is observed in progranzyme K [61]. The inactive α-subunit of the 7S nerve growth factor is a zymogen with a collapsed W215 [62]. Finally, in prethrombin-1, W215 assumes the same conformation as in 3BEI and participates in a hydrophobic cluster that occludes the active site [63].

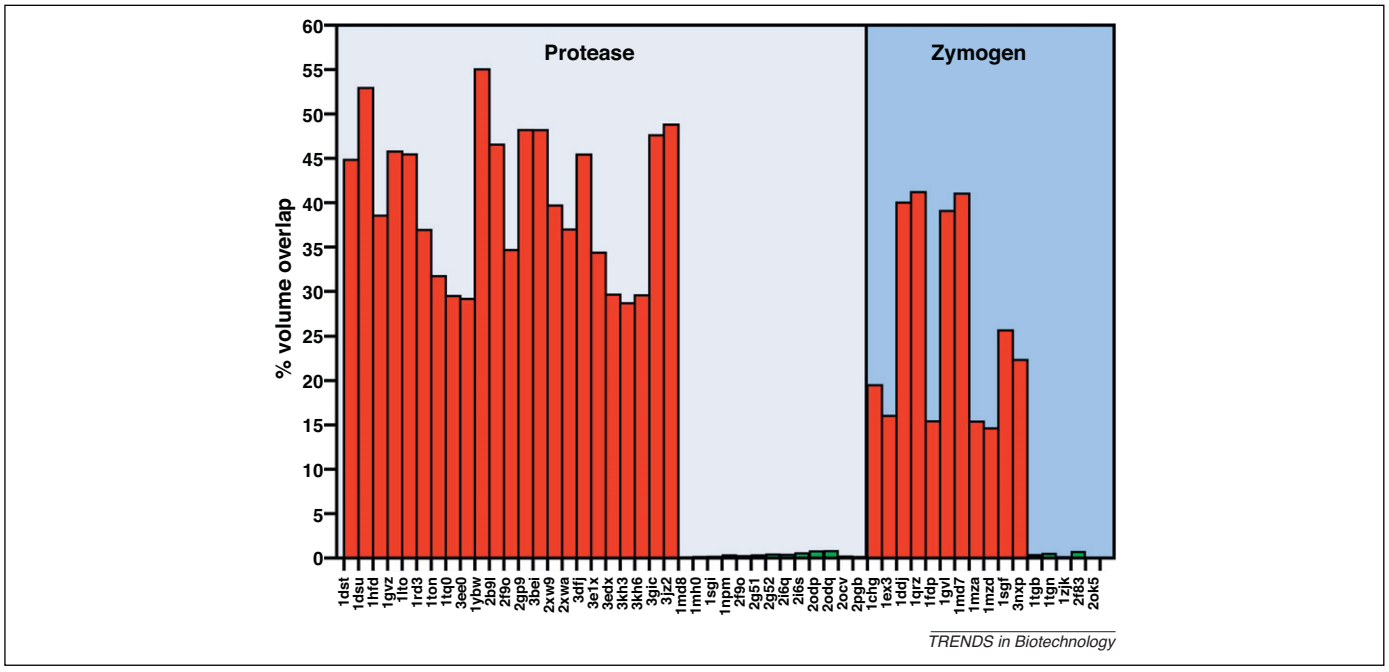

Figure 2. Active site accessibility in the protease and zymogen. Accessibility of the active site for all proteases in the set (Box 1) was calculated in terms of the overlap with PPACK bound to the active site in the structure 1SHH of thrombin and expressed as percentage of the total volume of inhibitor. The structures in the set are all free of ligands bound to the active site or at sites known to affect activity or stability and, therefore, provide relevant sampling of the conformations accessible to the fold in the free form. The analysis identifies two groups in each set: one with considerable blockage of the active site where overlap with the volume to be occupied by PPACK is on average 119 $Å^3$ (36% of total, red bars), and the other with negligible or no overlap averaging 1.0 $Å^3$ (0.3% of total, green bars). The dichotomous distribution supports existence of an equilibrium between mutually exclusive, collapsed and open forms in both the zymogen and protease.

Several structures of TLPs and zymogens solved in the free form assume the E conformation where the active site is wide open to substrate (Figure 2). Thrombin crystallizes in the E form in the presence of a point mutation, R77aA, that prevents auto-proteolysis [35,36], or the double mutation C191A/C220A that perturbs the primary specificity pocket [34]. Wild type murine thrombin crystallizes in the E form [64] and so do C2a [65,66], neuropsin [67] and trypsin [68]. Complement factor C1r crystallizes in the E form [60] but assumes a collapsed conformation in its zymogen form [60]. Several zymogens assume the E conformation with the active site open (Figure 2 and Box 1). They are trypsinogen [69,70], complement profactors B [71] and MASP-2 [72], and coagulation factor XI [73].

The alternative conformations accessible to residue 215 and the 215–217 segment observed in both the protease and zymogen prove that the allosteric E*–E equilibrium originally uncovered in thrombin [31] is a general property of the trypsin fold. Revision of the current zymogen to protease conversion paradigm appears necessary in terms of a more general scheme where both the zymogen and protease are capable of alternative conformations (Figure 3). The E*–E equilibrium in the protease form provides a reversible mechanism of regulation of enzyme activity following the irreversible transition from the zymogen form. When E* is stabilized, the protease possesses low activity and acts as an 'allosteric switch' that can be turned on 'on demand' by binding of specific cofactors or allosteric activators that facilitate conversion to the active E form. Relevant examples are complement factor D that shifts from the E* to the E form upon binding to the complex of C3b and factor B [25] and coagulation factor

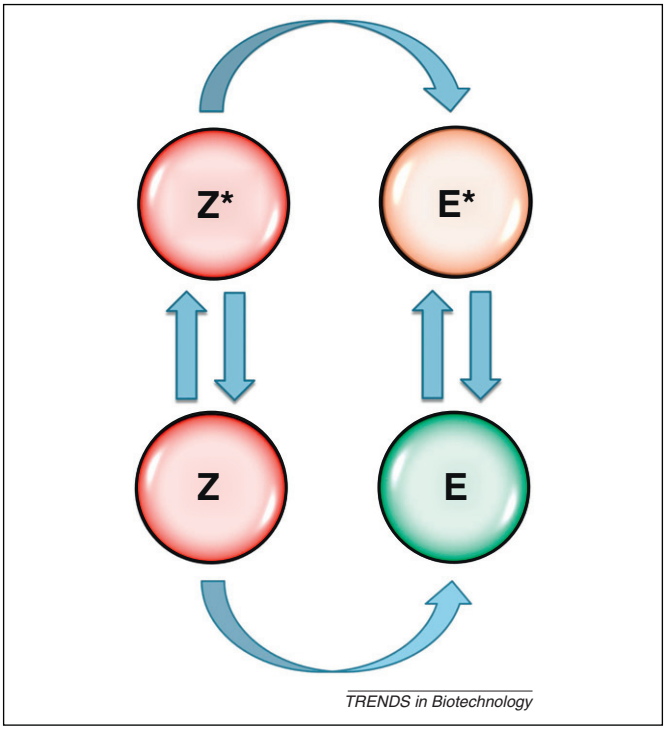

Figure 3. Zymogen to protease conversion scheme. The classical zymogen to protease conversion scheme is extended to account for allostery in the zymogen and protease. Two forms of the zymogen, one with the active site open (Z) and the other with the active site collapsed (Z*) are in equilibrium and each converts irreversibly to the E or E* form of the protease. Both Z and Z* are inactive. Activity of the protease depends on the distribution between the inactive E* and active E form. The Z*–Z equilibrium in the zymogen controls the rate of conversion to the protease. Natural cofactors and synthetic molecules affect activity of the protease or the zymogen to protease conversion by altering the distribution of the various species in the scheme, as discussed in the text.

VIIa that makes a similar transition upon binding to tissue factor [22,30]. Stabilization of E produces a highly active enzyme upon conversion from the zymogen form without requiring macromolecular cofactors, as seen in trypsin, complement factor C1r and thrombin. The allosteric Z*–Z equilibrium in the zymogen influences the rate of conversion to the protease if the Z to E and Z* to E* conversions occur on different time scales. This may enable regulatory control of the kinetics of zymogen to protease conversion *in vivo*.

Exploiting the allosteric E*–E equilibrium in therapeutic applications

The allosteric E*–E equilibrium not only provides a simple mechanism for protease function and regulation, but also affords a new target for therapeutic intervention. Stabilization of E* relative to E has the potential to completely abrogate substrate binding, as observed for competitive inhibition.

Substrate hydrolysis by a TLP undergoing the E*–E equilibrium obeys the following kinetic scheme [39]:

$$E^* \underset{k_r}{\overset{k_{-r}}{\leftrightarrows}} E \underset{k_1[S]}{\overset{k_{-1}}{\leftrightarrows}} ES \xrightarrow{k_2} E + P$$

E* is inactive and the active form E features finite values of the kinetic rate constants for binding (k_1), dissociation (k_{-1}) and hydrolysis (k_2) of substrate S. The ratio $r=k_{-r}/k_r=[E^*]/[E]$ defines the equilibrium constant for the E*–E interconversion. The Michaelis-Menten parameters in the scheme are $k_{cat} = k_2$ and $K_m = \frac{k_{-1}+k_2}{k_1}(1+r)$. The presence of E* has no influence on the value of k_{cat} because this parameter is a property of the enzyme–substrate complex, but it influences K_m, which increases linearly with r. Basically, E* is functionally equivalent to E bound to a competitive inhibitor. As r increases with the population of E*, the value of K_m increases without limits. Hence, any effector/effect that stabilizes E* relative to E has the potential to abrogate activity. By contrast, any effector/effect that stabilizes E relative to E* increases activity by decreasing K_m.

Allosteric inhibition of enzyme function through E* may provide a paradigm shift in the therapeutic control of key proteases. The advantage of an allosteric effector of E* over a competitive inhibitor is that the inhibitory effect is not mediated by binding to the active site of the enzyme, thereby minimizing undesired interference with the activity of other targets. Few but highly relevant examples of allosteric inhibition of TLPs have been reported and fit well into perturbation of the E*–E equilibrium. Allosteric inhibitors of factor VIIa have been designed [74]. They bind away from the active site and stabilize a conformation of factor VIIa where the oxyanion hole is flipped, as observed in the E* form. Allosteric modulators of thrombin inhibiting fibrinogen clotting without interfering with protein C activation have been reported [75]. The molecular mechanism of action of these modulators may be similar to that observed with thrombin mutants specifically engineered for anticoagulant activity.

Owing to its involvement in thrombotic deaths, thrombin remains at the forefront of cardiovascular medicine and a major target of antithrombotic and anticoagulant therapies [76]. Because heparin and direct thrombin inhibitors are plagued with complications related to dosage and bleeding, an entirely new strategy of intervention has been proposed in recent years to take advantage of the protein C anticoagulant pathway [77]. Protein C requires thrombin and the assistance of the cofactor thrombomodulin for activation and activated protein C acts as a potent anticoagulant and cytoprotective agent [78]. A thrombin mutant engineered for exclusive activity toward protein C and devoid of activity toward fibrinogen and the platelet receptor PAR1 could represent an innovative and potentially powerful tool to achieve anticoagulation without disruption of the hemostatic balance. Proof-of-principle that such a strategy could benefit the treatment of thrombotic diseases has come from *in vivo* studies in non-human primates [79–82]. Specifically, the anticoagulant thrombin mutant W215A/E217A (WE) has been shown to elicit antithrombotic and cytoprotective effects more efficacious than the direct administration of activated protein C and safer than the administration of low molecular weight heparins [81,83,84]. WE also improves neurological outcome and reduces cerebral infarct size in a mouse model of ischemic stroke [85], and is effective in the treatment of inflammatory joint disease [86]. Deletion of the autolysis loop also produces a notable anticoagulant effect *in vitro* [87], although the loop is separated from W215 by >20 Å. The mechanism at the basis of the anticoagulant effects observed with these mutants is directly related to the E*–E equilibrium. The mutation stabilizes the E* form, thereby making activity toward fibrinogen and PAR1 vanishingly small, but transition to the active E form ensues upon a concerted action of thrombomodulin and protein C that cannot be reproduced by the procoagulant substrates. The result is a selective anticoagulant effect at the level of the endothelial surface [37,43]. Existing crystal structures of WE [43] and the loopless mutant [37] reveal a conformation for the free enzyme where the 215–217 segment is collapsed (Figure 4) as seen in the E* form (Figure 2 and Box 1).

The allosteric E*–E equilibrium also offers ways to increase the activity of the enzyme by stabilizing the E form, as shown recently for thrombin [88]. Stabilization of Z in the zymogen form can also promote activation. Natural examples of this strategy are provided by cofactor-induced activation of the zymogen, as documented in the interaction of streptokinase with plasminogen [57] or staphylocoagulase with prothrombin [89]. The activation is non-proteolytic and involves an allosteric shift in the conformation of the zymogen toward a state that resembles the E form of the protease, consistent with the scheme shown in Figure 3. This strategy has recently been exploited in other systems with synthetic molecules. Hepatocyte growth factor binds to its target receptor tyrosine kinase, Met, as a single-chain form or as a cleaved two-chain disulfide-linked heterodimer that stimulates Met signaling [90]. Peptides corresponding to the first seven to 10 residues of the cleaved N terminus of the β-chain stimulate Met phosphorylation by the zymogen (single-chain) to levels that are 25% of those stimulated by the two-chain form. Surprisingly, small molecule activators of zymogens have been reported recently for the apoptotic procaspase-3 and procaspase-6 [91] as inducers of autoproteolytic activation by

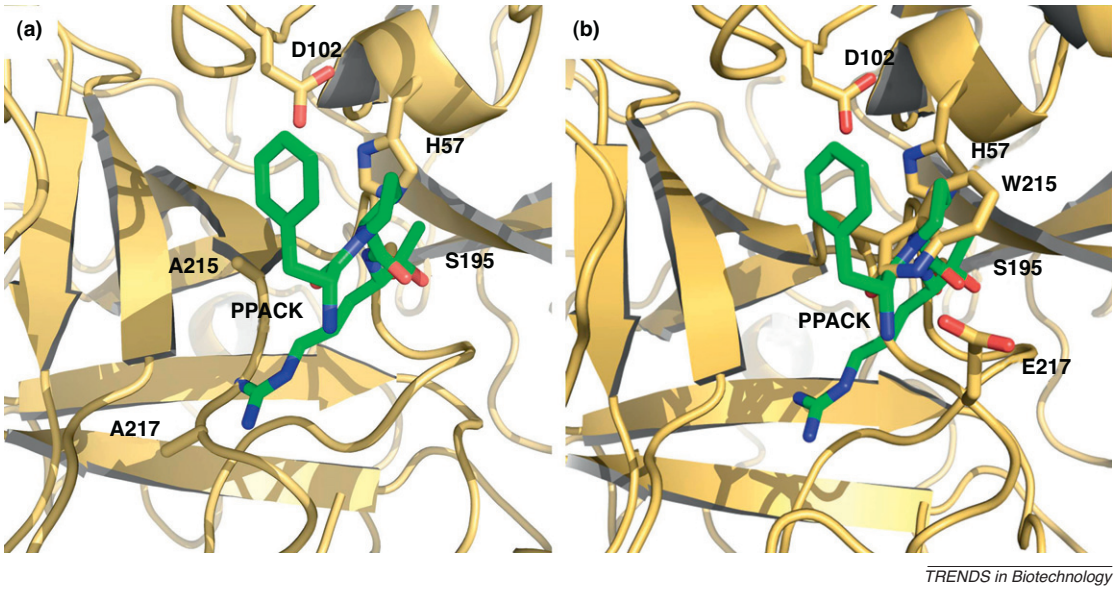

Figure 4. Crystal structures of the anticoagulant thrombin mutants WE [**(a)** 1TQ0] and Δ146–149e [**(b)** 3GIC]. The structures are overlayed with the structure of thrombin bound to PPACK (green, 1SHH) where only PPACK is shown for clarity. The 215–217 segment carrying the W215A/E217A mutation collapses into the active site and clashes with the side chain of Arg at the P1 position of PPACK occluding 29% of the total volume of the inhibitor (311 $Å^3$). In the case of Δ146–149e, the entire 215–217 moves into the active site and occludes 48% of the volume of PPACK. The structures of WE and Δ146–149e offer substantial insight into the mechanism of action of these mutants and show how thrombin can be turned into an anticoagulant with site-directed mutations that stabilize the E* form.

stabilization of a conformation that is both more active and more susceptible to intermolecular proteolysis. Even for this different class of proteolytic enzymes, an allosteric Z*–Z equilibrium can be invoked to explain how stabilization of Z would promote transition to a more competent state and possibly trigger autoproteolysis.

Concluding remarks

The discovery of the allosteric nature of TPLs is a paradigm shift with far reaching consequences on our understanding of their structure, function and regulation. In addition, the allosteric E*–E equilibrium offers a simple conceptual framework for the development of new therapeutic strategies aimed at inhibiting or activating enzyme function. These effects can be achieved by small molecules directed at one of the alternative conformations, or by protein engineering strategies aimed at tipping the equilibrium with site-directed substitutions. Future studies should take advantage of the allosteric nature of TLPs and the potential benefits of interfering with their function through a mechanism that does not involve the active site and thereby improves selectivity. Renewed efforts should be devoted to screenings for small molecules against TLPs aimed at identifying allosteric inhibitors or activators. Similar efforts should be directed to protein therapeutics that take advantage of the E*–E equilibrium. The idea of engineering a TLP acting as an allosteric switch activated on demand by a specific trigger is a challenging goal for biotechnology, but may translate into substantial returns for clinical applications. Anticoagulant thrombin mutants offer a relevant example of this new strategy that expands the current landscape of applications of proteases as therapeutics [92].

Challenges remain also for the structural biologists and enzymologists as they seek unequivocal evidence of a pre-existing equilibrium between alternative conformations in the trypsin fold. Efforts should be directed to crystallization under different solution conditions to promote stabilization and trapping of the open and collapsed conformations of the enzyme and zymogen forms. We still lack evidence that E* and E, or Z* and Z, can be solved crystallographically for the same protein construct and under different solution conditions. Kinetic evidence of the E*–E equilibrium is solid but only for a handful of TLPs amenable to stopped-flow studies. No such evidence currently exists for the Z*–Z equilibrium, nor do we fully understand the possible physiological role of such equilibrium in the zymogen. The field of TLPs is at an important turning point and we predict that some of these open questions will soon be answered with tangible returns for the basic sciences and therapeutic applications.

Acknowledgements

We are grateful to Ms Tracey Baird for her help with illustrations. This work was supported in part by the National Institutes of Health Research Grants HL49413, HL58141, HL73813 and HL95315 (to E.D.C.). E.D.C. holds a patent on the anticoagulant thrombin mutant WE and is a consultant for Verseon.

References

1 Puente, X.S. *et al.* (2005) A genomic view of the complexity of mammalian proteolytic systems. *Biochem. Soc. Trans.* 33, 331–334
2 Ciechanover, A. and Iwai, K. (2004) The ubiquitin system: from basic mechanisms to the patient bed. *IUBMB Life* 56, 193–201
3 Gomis-Ruth, F.X. (2003) Structural aspects of the metzincin clan of metalloendopeptidases. *Mol. Biotechnol.* 24, 157–202
4 Rea, D. and Fulop, V. (2006) Structure–function properties of prolyl oligopeptidase family enzymes. *Cell Biochem. Biophys.* 44, 349–365
5 Page, M.J. and Di Cera, E. (2008) Serine peptidases: classification, structure and function. *Cell Mol. Life Sci.* 65, 1220–1236
6 Rothman, S.S. (1977) The digestive enzymes of the pancreas: a mixture of inconstant proportions. *Annu. Rev. Physiol.* 39, 373–389
7 Gros, P. *et al.* (2008) Complement driven by conformational changes. *Nat. Rev. Immunol.* 8, 48–58

8 Davie, E.W. *et al.* (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* 30, 10363–10370
9 Pereira, P.J. *et al.* (1998) Human beta-tryptase is a ring-like tetramer with active sites facing a central pore. *Nature* 392, 306–311
10 Lin, C.Y. *et al.* (1999) Molecular cloning of cDNA for matriptase, a matrix-degrading serine protease with trypsin-like activity. *J. Biol. Chem.* 274, 18231–18236
11 Diamandis, E.P. *et al.* (2000) The new human kallikrein gene family: implications in carcinogenesis. *Trends Endocrinol. Metab.* 11, 54–60
12 Proud, D. and Kaplan, A.P. (1988) Kinin formation: mechanisms and role in inflammatory disorders. *Annu. Rev. Immunol.* 6, 49–83
13 Barry, M. and Bleackley, R.C. (2002) Cytotoxic T lymphocytes: all roads lead to death. *Nat. Rev. Immunol.* 2, 401–409
14 Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem. Rev.* 102, 4501–4524
15 Perona, J.J. and Craik, C.S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.* 4, 337–360
16 Krem, M.M. and Di Cera, E. (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci.* 27, 67–74
17 Jiang, Y. and Doolittle, R.F. (2003) The evolution of vertebrate blood coagulation as viewed from a comparison of puffer fish and sea squirt genomes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7527–7532
18 Davie, E.W. and Kulman, J.D. (2006) An overview of the structure and function of thrombin. *Semin. Thromb. Hemost.* 32 (Suppl. 1), 3–15
19 Arlaud, G.J. *et al.* (2007) Deciphering complement mechanisms: the contributions of structural biology. *Mol. Immunol.* 44, 3809–3822
20 Ponnuraj, K. *et al.* (2004) Structural analysis of engineered Bb fragment of complement factor B: insights into the activation mechanism of the alternative pathway C3-convertase. *Mol. Cell* 14, 17–28
21 Banner, D.W. *et al.* (1996) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* 380, 41–46
22 Eigenbrot, C. *et al.* (2001) The factor VII zymogen structure reveals reregistration of beta strands during activation. *Structure* 9, 627–636
23 Jing, H. *et al.* (1999) Structural basis of profactor D activation: from a highly flexible zymogen to a novel self-inhibited serine protease, complement factor D. *EMBO J.* 18, 804–814
24 Narayana, S.V. *et al.* (1994) Structure of human factor D. A complement system protein at 2.0 A resolution. *J. Mol. Biol.* 235, 695–708
25 Forneris, F. *et al.* (2010) Structures of C3b in complex with factors B and D give insight into complement convertase formation. *Science* 330, 1816–1820
26 Jing, H. *et al.* (1998) Structures of native and complexed complement factor D: implications of the atypical His57 conformation and self-inhibitory loop in the regulation of specific serine protease activity. *J. Mol. Biol.* 282, 1061–1081
27 Peterson, F.C. *et al.* (2001) High-level bacterial expression and ^{15}N-alanine-labeling of bovine trypsin. Application to the study of trypsin-inhibitor complexes and trypsinogen activation by NMR spectroscopy. *Biochemistry* 40, 6275–6283
28 Gombos, L. *et al.* (2008) Probing conformational plasticity of the activation domain of trypsin: the role of glycine hinges. *Biochemistry* 47, 1675–1684
29 Rohr, K.B. *et al.* (2006) X-ray structures of free and leupeptin-complexed human alphaI-tryptase mutants: indication for an alpha→beta-tryptase transition. *J. Mol. Biol.* 357, 195–209
30 Dickinson, C.D. *et al.* (1996) Identification of surface residues mediating tissue factor binding and catalytic function of the serine protease factor VIIa. *Proc. Natl. Acad. Sci. U.S.A.* 93, 14379–14384
31 Bah, A. *et al.* (2006) Rapid kinetics of Na^+ binding to thrombin. *J. Biol. Chem.* 281, 40049–40056
32 Di Cera, E. (2008) Thrombin. *Mol. Aspects Med.* 29, 203–254
33 Vogt, A.D. *et al.* (2010) Evidence of the E*–E equilibrium from rapid kinetics of Na(+) binding to activated protein C and factor Xa. *J. Phys. Chem. B* 114, 16125–16130
34 Bush-Pelc, L.A. *et al.* (2007) Important role of the Cys-191:Cys-220 disulfide bond in thrombin function and allostery. *J. Biol. Chem.* 282, 27165–27170
35 Pineda, A.O. *et al.* (2004) Molecular dissection of Na^+ binding to thrombin. *J. Biol. Chem.* 279, 31842–31853
36 Pineda, A.O. *et al.* (2002) Crystal structure of the anticoagulant slow form of thrombin. *J. Biol. Chem.* 277, 40177–40180
37 Bah, A. *et al.* (2009) Stabilization of the E* form turns thrombin into an anticoagulant. *J. Biol. Chem.* 284, 20034–20040
38 Gandhi, P.S. *et al.* (2008) Structural identification of the pathway of long-range communication in an allosteric enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1832–1837
39 Niu, W. *et al.* (2009) Mutant N143P reveals how Na^+ activates thrombin. *J. Biol. Chem.* 284, 36175–36185
40 Pineda, A.O. *et al.* (2006) Crystal structure of thrombin in a self-inhibited conformation. *J. Biol. Chem.* 281, 32922–32928
41 Lechtenberg, B.C. *et al.* (2010) NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14087–14092
42 Carter, W.J. *et al.* (2004) Crystal structure of anticoagulant thrombin variant E217K provides insights into thrombin allostery. *J. Biol. Chem.* 279, 26387–26394
43 Gandhi, P.S. *et al.* (2009) Mechanism of the anticoagulant activity of the thrombin mutant W215A/E217A. *J. Biol. Chem.* 284, 24098–24105
44 Pineda, A.O. *et al.* (2004) The anticoagulant thrombin mutant W215A/E217A has a collapsed primary specificity pocket. *J. Biol. Chem.* 279, 39824–39828
45 Kim, S. *et al.* (1995) Crystal structure of a complement factor D mutant expressing enhanced catalytic activity. *J. Biol. Chem.* 270, 24399–24405
46 Carvalho, A.L. *et al.* (2002) Crystal structure of a prostate kallikrein isolated from stallion seminal plasma: a homologue of human PSA. *J. Mol. Biol.* 322, 325–337
47 Fujinaga, M. and James, M.N. (1987) Rat submaxillary gland serine protease, tonin. Structure solution and refinement at 1. 8 A resolution. *J. Mol. Biol.* 195, 373–396
48 Piao, S. *et al.* (2005) Crystal structure of a clip-domain serine protease and functional roles of the clip domains. *EMBO J.* 24, 4404–4414
49 Marquardt, U. *et al.* (2002) The crystal structure of human alpha1-tryptase reveals a blocked substrate-binding region. *J. Mol. Biol.* 321, 491–502
50 Shia, S. *et al.* (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, 1335–1349
51 Rickert, K.W. *et al.* (2008) Structure of human prostasin, a target for the regulation of hypertension. *J. Biol. Chem.* 283, 34864–34872
52 Spraggon, G. *et al.* (2009) Active site conformational changes of prostasin provide a new mechanism of protease regulation by divalent cations. *Protein Sci.* 18, 1081–1094
53 Freer, S.T. *et al.* (1970) Chymotrypsinogen: 2.5-angstrom crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation. *Biochemistry* 9, 1997–2009
54 Pjura, P.E. *et al.* (2000) Protein crystallization by design: chymotrypsinogen without precipitants. *J. Mol. Biol.* 300, 235–239
55 Peisach, E. *et al.* (1999) Crystal structure of the proenzyme domain of plasminogen. *Biochemistry* 38, 11180–11188
56 Wang, X. *et al.* (2000) Human plasminogen catalytic domain undergoes an unusual conformational change upon activation. *J. Mol. Biol.* 295, 903–914
57 Wakeham, N. *et al.* (2002) Effects of deletion of streptokinase residues 48-59 on plasminogen activation. *Protein Eng.* 15, 753–761
58 Gomis-Rüth, F.X. *et al.* (2002) The structure of human prokallikrein 6 reveals a novel activation mechanism for the kallikrein family. *J. Biol. Chem.* 277, 27273–27281
59 Budayova-Spano, M. *et al.* (2002) The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *EMBO J.* 21, 231–239
60 Budayova-Spano, M. *et al.* (2002) Monomeric structures of the zymogen and active catalytic domain of complement protease c1r: further insights into the c1 activation mechanism. *Structure* 10, 1509–1519
61 Hink-Schauer, C. *et al.* (2002) The 2.2-A crystal structure of human pro-granzyme K reveals a rigid zymogen with unusual features. *J. Biol. Chem.* 277, 50923–50933
62 Bax, B. *et al.* (1997) Structure of mouse 7S NGF: a complex of nerve growth factor with four binding proteins. *Structure* 5, 1275–1285
63 Chen, Z. *et al.* (2010) Crystal structure of prethrombin-1. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19278–19283
64 Marino, F. *et al.* (2007) Structural basis of Na^+ activation mimicry in murine thrombin. *J. Biol. Chem.* 282, 16355–16361

65 Krishnan, V. *et al.* (2007) The crystal structure of C2a, the catalytic fragment of classical pathway C3 and C5 convertase of human complement. *J. Mol. Biol.* 367, 224–233
66 Milder, F.J. *et al.* (2006) Structure of complement component C2A: implications for convertase formation and substrate binding. *Structure* 14, 1587–1597
67 Kishi, T. *et al.* (1999) Crystal structure of neuropsin, a hippocampal protease involved in kindling epileptogenesis. *J. Biol. Chem.* 274, 4220–4224
68 Mueller-Dieckmann, C. *et al.* (2007) On the routine use of soft X-rays in macromolecular crystallography. Part IV. Efficient determination of anomalous substructures in biomacromolecules using longer X-ray wavelengths. *Acta Crystallogr. D: Biol. Crystallogr.* 63, 366–380
69 Fehlhammer, H. *et al.* (1977) Crystal structure of bovine trypsinogen at 1-8 A resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J. Mol. Biol.* 111, 415–438
70 Kossiakoff, A.A. *et al.* (1977) Structure of bovine trypsinogen at 1.9 A resolution. *Biochemistry* 16, 654–664
71 Milder, F.J. *et al.* (2007) Factor B structure provides insights into activation of the central protease of the complement system. *Nat. Struct. Mol. Biol.* 14, 224–228
72 Gál, P. *et al.* (2005) A true autoactivating enzyme. Structural insight into mannose-binding lectin-associated serine protease-2 activations. *J. Biol. Chem.* 280, 33435–33444
73 Papagrigoriou, E. *et al.* (2006) Crystal structure of the factor XI zymogen reveals a pathway for transactivation. *Nat. Struct. Mol. Biol.* 13, 557–558
74 Dennis, M.S. *et al.* (2000) Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* 404, 465–470
75 Berg, D.T. *et al.* (1996) Enhanced protein C activation and inhibition of fibrinogen cleavage by a thrombin modulator. *Science* 273, 1389–1391
76 Mackman, N. (2008) Triggers, targets and treatments for thrombosis. *Nature* 451, 914–918
77 Di Cera, E. (2011) Thrombin as an anticoagulant. *Prog. Mol. Biol. Transl. Sci.* 99, 145–184
78 Riewald, M. *et al.* (2002) Activation of endothelial cell protease activated receptor 1 by the protein C pathway. *Science* 296, 1880–1882
79 Gibbs, C.S. *et al.* (1995) Conversion of thrombin into an anticoagulant by protein engineering. *Nature* 378, 413–416
80 Gruber, A. *et al.* (2002) The thrombin mutant W215A/E217A shows safe and potent anticoagulant and antithrombotic effects in vivo. *J. Biol. Chem.* 277, 27581–27584
81 Gruber, A. *et al.* (2007) Relative antithrombotic and antihemostatic effects of protein C activator versus low molecular weight heparin in primates. *Blood* 109, 3733–3740
82 Tsiang, M. *et al.* (1996) Protein engineering thrombin for optimal specificity and potency of anticoagulant activity in vivo. *Biochemistry* 35, 16449–16457
83 Feistritzer, C. *et al.* (2006) Protective signaling by activated protein C is mechanistically linked to protein C activation on endothelial cells. *J. Biol. Chem.* 281, 20077–20084
84 Berny, M.A. *et al.* (2008) Thrombin mutant W215A/E217A acts as a platelet GpIb antagonist. *Arterioscler. Thromb. Vasc. Biol.* 18, 329–334
85 Berny-Lang, M.A. *et al.* (2011) Thrombin mutant W215A/E217A treatment improves neurological outcome and reduces cerebral infarct size in a mouse model of ischemic stroke. *Stroke* 42, 1736–1741
86 Flick, M.J. *et al.* (2011) The development of inflammatory joint disease is attenuated in mice expressing the anticoagulant prothrombin mutant W215A/E217A. *Blood* 117, 6326–6337
87 Dang, Q.D. *et al.* (1997) Selective loss of fibrinogen clotting in a loopless thrombin. *J. Biol. Chem.* 272, 19649–19651
88 Pozzi, N. *et al.* Rigidification of the autolysis loop enhances Na^+ binding to thrombin. *Biophys. Chem.* (in press)
89 Friedrich, R. *et al.* (2003) Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation. *Nature* 425, 535–539
90 Landgraf, K.E. *et al.* (2010) Allosteric peptide activators of pro-hepatocyte growth factor stimulate Met signaling. *J. Biol. Chem.* 285, 40362–40372
91 Wolan, D.W. *et al.* (2009) Small-molecule activators of a proenzyme. *Science* 326, 853–858
92 Craik, C.S. *et al.* (2011) Proteases as therapeutics. *Biochem. J.* 435, 1–16
93 Bailey, S. (1994) The CCP4 suite. Programs for protein crystallography. *Acta Crystallogr. D: Biol. Crystallogr.* 50, 760–763