Novel proteases: common themes and surprising features

Lucy Vandeputte-Rutten* and Piet Gros†

Proteases perform a wide variety of functions, inside and outside cells, regulating many biological processes. Recent years have witnessed a number of significant advances in the structural biology of proteases, including aspects of intracellular protein and peptide degradation by self-compartmentalizing proteases, activation of proteases in proteolytic cascades of regulatory pathways, and mechanisms of microbial proteases in pathogenicity.

Addresses
Department of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
*e-mail: l.vandeputte@chem.uu.nl
†e-mail: p.gros@chem.uu.nl

Current Opinion in Structural Biology 2002, 12:704–708

0959-440X/02/$ – see front matter

Introduction

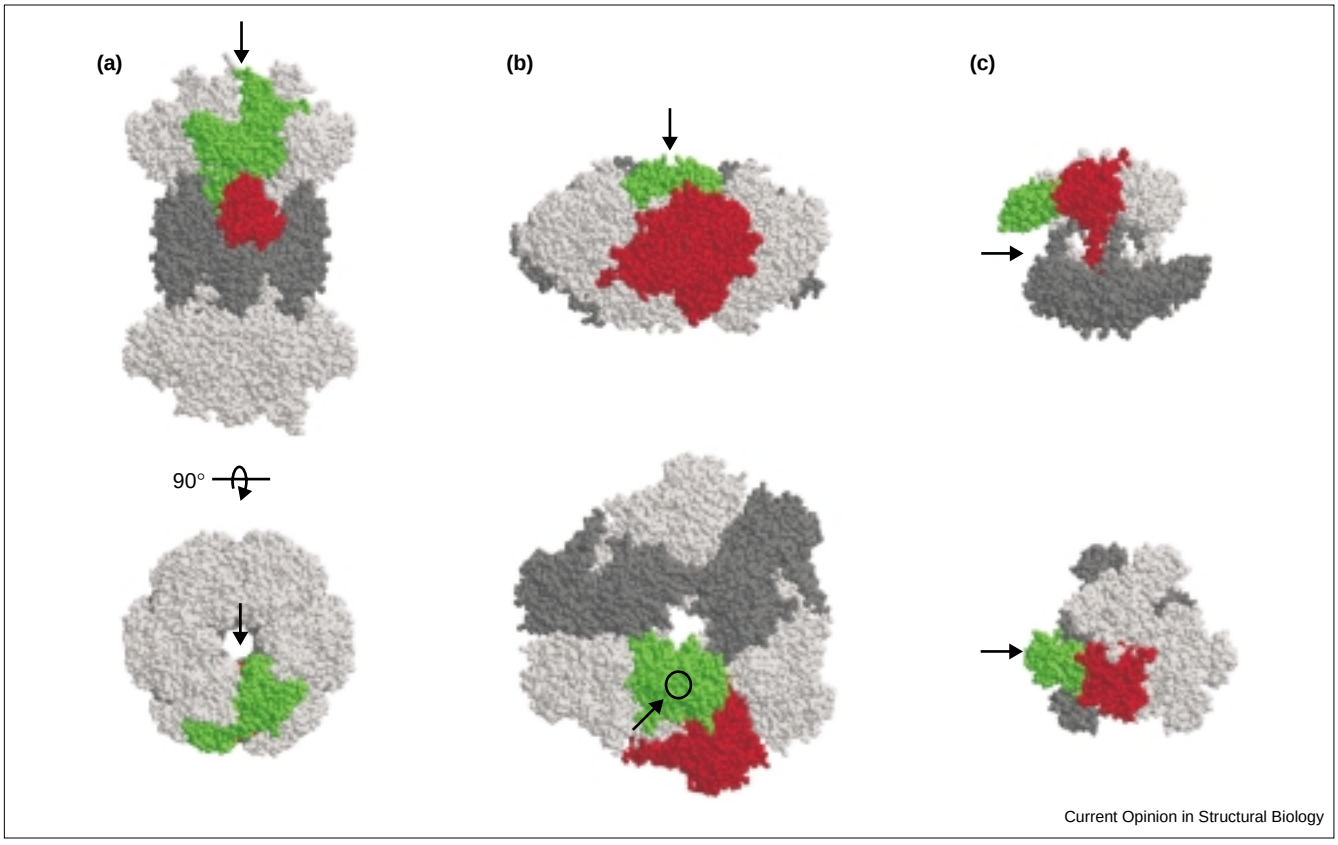
Proteases or peptidases catalyze the cleavage of peptide bonds. They have been categorized into five major classes based on their catalytic mechanisms: serine, threonine, cysteine, aspartate and metallo proteases [1]. Proteolytic reactions are elementary to numerous biological processes and need to be tightly regulated. To accomplish this, a wide variety of protease regulation mechanisms are apparent in nature, for example, high substrate specificity, ATP-driven protein degradation, active site access restriction, and activation cascades. Key examples of processes in which protease activity is important for homeostasis include cell growth, cell death, blood clotting, matrix remodeling and immune defense. In addition, pathogenic viruses and bacteria use proteases for their life cycle and for infection of host cells. Therefore, proteases are important targets for drug design against a diverse set of diseases. In the past two years, several interesting crystal structures of proteases involved in intracellular peptide degradation, regulatory pathways and microbial pathogenicity have been solved. These structures indicate mechanisms for regulating substrate access by multimolecular complexes, the occurrence of surprisingly large structural rearrangements upon protease (in)activation and the existence of hitherto unknown catalytic mechanisms.

Self-compartmentalizing proteases

Proteases involved in cytoplasmic protein degradation must distinguish between unfolded proteins or proteins tagged for degradation, and folded proteins. Structures of such proteases, for example, 20S proteasomes from *Thermoplasma acidophilum* [2] and *Saccharomyces cerevisiae* [3]), as well as the *Escherichia coli* proteasome homolog HslV [4], revealed large multimeric ring structures with a central cavity that can be accessed through axial openings (reviewed in [5]). The active sites, containing catalytic N-terminal threonines, lie on the inside of the cavity. Thus, proteolysis is restricted to proteins or unfolded proteins that can enter the cavity through the narrow pores. In the recent period, structures of several other self-compartmentalizing proteases have been solved, revealing a variety of mechanisms of substrate access.

In 2000, two structures appeared of proteasomal HslV in complex with its ATP-dependent chaperone HslU, those from *E. coli* [6] and *Haemophilus influenza* [7••]. In both structures, HslU forms hexameric rings that bind apical to both sides of the 'double donut' of HslV hexamers. However, the orientation of the HslU hexamers with respect to the HslV double-hexamer rings is opposite in the two models. The most compact organization, with the so-called intermediate domains (or I-domains) pointing outwards, as described by Sousa *et al.* [7••], is favored by electron microscopy studies [8] and small-angle scattering data. Sousa *et al.* [7••] suggest that the I-domains serve to bind substrates and lead them into the interior cavity of the complex, where the active sites are located (Figure 1). The structure, furthermore, provides insight into the allosteric activation of HslV by HslU by transmission of a conformational change to the active site region of the protease; this is supported by the structure of HslUV in complex with a vinyl sulfone inhibitor [9]. A different mechanism of substrate access is demonstrated by the open and closed structures of the serine protease DegP (HtrA) from *E. coli* [10••], a heat shock protein that combines refolding at lower temperatures and protease activities at higher temperatures. DegP assembles as two staggered trimeric rings with a central cavity formed by the six protease domains. Twelve flexible PDZ domains function as mobile 'gatekeepers' for lateral substrate entrance to the central cavity, which harbors the active sites (Figure 1). In contrast, the related mitochondrial serine protease HtrA2/Omi shows a pyramid-shaped trimeric structure [11]. Although this structure corresponds to only one half of DegP, it suggests that substrate access to the active sites is also regulated by the PDZ domains. The structure of the yeast 20S proteasome in complex with the 11S regulator from *Trypanosoma brucei* [12] shows how egress of substrates may be regulated. The 11S regulators bind to the apical parts of the 20S proteasome, inducing a conformational change in the α subunits of the 20S proteasome that creates a more open conformation. This may allow a faster release of substrates, resulting in longer peptides. Peptides produced by proteasomes may be degraded further into dipeptides and tripeptides, for example, by the tricorn protease from *T. acidophilum* [13,14]. The structure of tricorn protease reveals five subdomains, with a serine protease catalytic site, assembled into a hexameric (trimer of dimers) cage-forming structure [15••]. Brandstetter *et al.* [15••]

Figure 1



Orthogonal views of space-filling models of three cage-forming proteases: (a) HslUV [7••], (b) tricorn protease [15••] and (c) DegP [10••]. In the multimeric structures, we highlighted one domain or subunit harboring the catalytic site in red and one domain or subunit involved in substrate entry in green. The arrows indicate the entrance site for substrates (note, for clarity, symmetry-related entrances are not indicated). (a) *E. coli* proteasome homolog HslUV, with the protease HslV 'double donut' in dark gray and the HslU ATP-dependent chaperone in light gray. (b) Tricorn protease from *T. acidophilum*, with the subunits shown alternating in light and dark gray. (c) DegP from *E. coli* (in its open conformation), with trimers in light and dark gray. All molecules are drawn at the same scale.

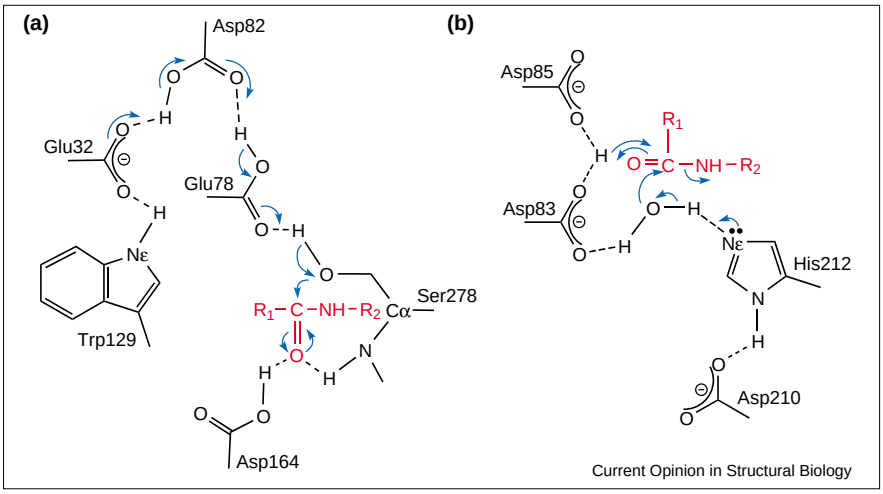
argue that, similar to serine oligopeptidase [16], substrates may be channeled to the active site not through a multimeric pore but through the center of a β-propeller domain (Figure 1).

Proteases in regulatory pathways

Programmed cell death is proteolytically regulated in its initiation and execution phases by cysteine proteases, called caspases (reviewed in [17]). Recent advances include structures of caspase-3 and caspase-7 in complex with XIAP, an endogenous caspase inhibitor critical for suppressing cell death. The structures show binding of the inhibitor in the substrate-binding site, in a reverse orientation with respect to substrate binding [18–20]. In addition, three structures, two of procaspase-7 [21••,*22*••] and one of an activated but uninhibited caspase-7 [21••], provided detailed information about the mechanism of procaspase-7 activation. The zymogen procaspase-7 structure [21••,*22*••] reveals a dimer with catalytically incompetent active sites and nonfunctional substrate-binding sites. The activation loops, which connect the small and large domains, are asymmetrically bound to a central cleft of the procaspase dimer. Cleavage of the loops is needed to flip out the N-terminal end, forming an active site primed for substrate binding [21••]. Additional structural changes induced by inhibitor, or presumably substrate, binding induce flipping out of the C-terminal end, creating a fully competent active site with its characteristic loop bundle. A related activation mechanism induced by dimerization is proposed for the initiator caspase-9 [23]. Surprisingly, the inhibition of the executioner caspase-8 by p35 from baculovirus revealed a covalent complex [24]. Structural rearrangements within the inhibitor block solvent access to the catalytic site, which prevents hydrolysis of the thioester linkage.

Activation of the complement system, part of the mammalian innate immune system that recognizes and eliminates pathogenic microbes, involves serine proteases with multiple regulatory domains. Structures of protease domains from two of the complement activation pathways have been reported [25••,26–28]. The 'classical pathway' of complement starts with the association of the C1 complex with immune complexes formed on pathogenic surfaces. This binding event most probably causes structural

Figure 2



Schematic representations of the proposed novel catalytic mechanisms of **(a)** serine-carboxyl proteases and **(b)** omptins. Catalytic residues are shown in black, substrate in red and the putative electron translocation is depicted by blue arrows. (a) The acylation reaction proposed for kumamolysin [35]. A similar mechanism has been proposed for pepstatin-insensitive carboxyl proteinase, with the exception of residues Trp129 and Glu32, which are not present in this protein [34]. (b) Proposed mechanism for OmpT [36•], which involves a nucleophilic water (that is not visible in the electron density) replacing the role of the serine Oγ in serine proteases.

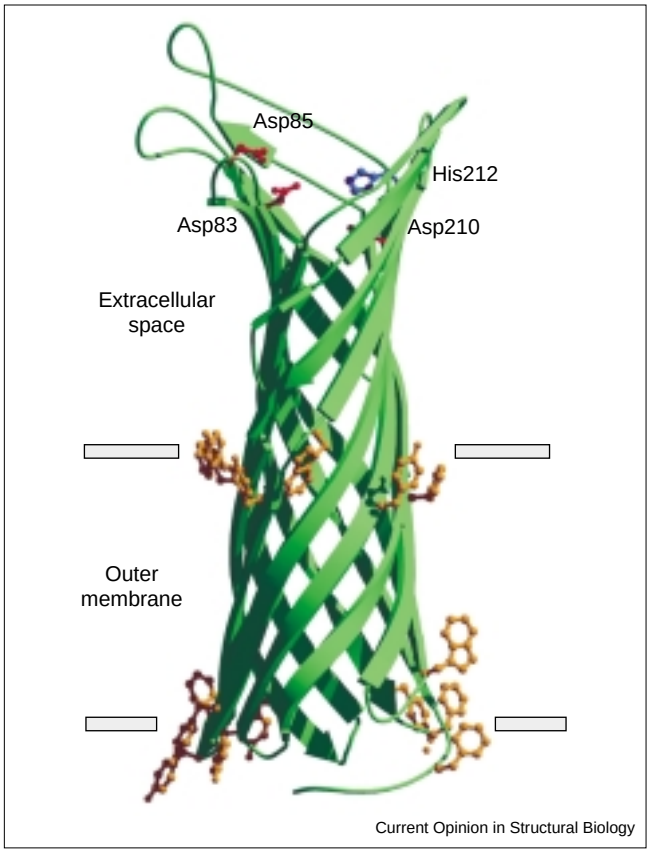
rearrangements in the C1 complex that auto-activate the two C1r proteases in the complex, which subsequently activates the two C1s proteases in the complex. The crystal structure of a C1r fragment, consisting of the protease domain and the two preceding CCP domains, reveals a homodimeric structure arranged head-to-tail with a large central cavity, in agreement with previous electron microscopy data of the C1r2s2 heterotetramer [25••]. The active sites, which have zymogen-like characteristics, are located at either end of the elongated homodimeric structure. Moreover, the active site and the scissile bond of its dimeric partner molecule are approximately 90 Å apart, indicating that large rearrangements are needed to achieve auto-activation. Subsequent rearrangements may be expected that bring the scissile bond of C1s within range of an active site of C1r, thus allowing the second proteolysis step and completing the initiation of complement activation. Inhibition of this type of serine protease by serpins (serine protease inhibitors) also involves major structural rearrangements, as demonstrated by the structure of the covalent complex between trypsin and α_1-antitrypsin [29]. The structure shows a displacement of more than 70 Å of trypsin, and a reversion of the protease towards a zymogen-like state with an incompetent active site and an induced disorder of a large part (~40%) of the protease structure.

Microbial proteases

Infectious microbes use proteases, either secreted or attached to their cell surface, to weaken and invade their hosts. The mechanisms of biological activity of these proteases are diverse in substrate specificities and sites of activity. An example of a highly specific secreted protease is lethal factor from anthrax (*Bacillus anthracis*), which is targeted to host cells by the anthrax protein 'protective antigen'. In a timely publication of the crystal structure of lethal factor, Pannifer *et al.* [30•] showed that this protease, possibly evolved through gene duplication, consists of four domains. The N- and C-terminal domains have a zinc-metalloprotease fold, with the catalytic site residing in the C-terminal domain. Three domains, domains II–IV, form a 40 Å long groove that exhibits high specificity for the N terminus of MAPK kinases; two recent structures of mammalian zinc-metalloproteases also displayed extended peptide-binding grooves formed by helical extensions of the zinc-metalloprotease fold [31,32]. Also within the review period, Kagawa *et al.* [33] reported the structure of SpeB precursor, a virulence factor from *Streptococcus pyogenes*. The structure revealed that this protein is a distant member of the papain superfamily of cysteine proteases. The prosegment, which inhibits activity by displacing the active site histidine, has a novel fold. Unique to cysteine proteases, SpeB has an integrin-binding RGD motif at its surface. Possibly, this site serves to localize SpeB at the host cell surface, where it can cleave fibronectin and vitronectin, two molecules involved in maintaining tissue integrity.

Two new catalytic mechanisms were identified based on structures of microbial proteases [34,35,36•]. The structures of the acidic proteases pepstatin-insensitive carboxyl proteinase from *Pseudomonas* sp 101 [34] and kumamolysin from *Bacillus* novosp MN-32 [35] revealed a fold similar to subtilisins. However, the typical serine, histidine and aspartate triad of serine proteases is substituted by serine, glutamate and aspartate residues, defining a new family of serine-carboxyl proteases. Like in serine proteases, the serine hydroxyl is thought to act as a nucleophile (Figure 2a). The glutamate–aspartate couple, with a short hydrogen-bonding distance between the carboxylate groups, may act as a general base. Comellas-Bigler *et al.* [35] proposed that, in kumamolysin, the proton is shuttled to a third carboxylate group stabilized by a neighboring tryptophan residue. Furthermore, both proteases share a supposedly protonated aspartate residue participating in oxyanion stabilization. A second variation of the serine protease catalytic triad is presented by the structure of the integral membrane protease OmpT from *E. coli* [36•], which exerts its activity

Figure 3



Ribbon drawing of the integral outer membrane protease OmpT from *E. coli* [36•]. The catalytic residues are shown in ball and stick, with the three aspartates (Asp83, Asp85 and Asp210) in red and His212 in blue. The aromatic residues that delineate the lipid boundaries are shown in light brown.

at alkaline pH. This protein is a member of the family of omptins, outer membrane proteases that are virulence factors in several Gram-negative pathogenic bacteria [37]. The OmpT structure consists of a ten-stranded β barrel, a fold not observed in proteases before (Figure 3). The structure and mutagenesis data [38] indicate four residues essential for catalytic activity that combine structural aspects of both serine and aspartate proteases. The authors suggested that a water molecule, activated by a histidine–aspartate and aspartate–aspartate couple, may act as a nucleophile, substituting the role of the catalytic serine in serine proteases. In addition, the aspartate–aspartate couple may be required for proton translocation or stabilization of the oxyanion intermediate (Figure 2b).

Conclusions

The new protease structures of the past two years reflect the diverse biological roles of proteases. They have revealed a variety of mechanisms that determine the proteolytic activity and specificity, including new catalytic mechanisms, different ways of regulating substrate access and proteolytic activation mechanisms involving major structural rearrangements. Given the large number of proteases present in nature and the variations observed so far, it is clear that many surprises may be expected in the future. In particular, the discovery of several intramembrane proteases (e.g. [39] and references therein), which are thought to exert their activity in the lipid bilayer of the membrane, presents an exciting challenge for protease structural biology.

Acknowledgements

We wish to thank EG Huizinga and DAA Vandeputte for valuable comments on the manuscript. LV-R is supported by the Netherlands Organization for Scientific Research (NWO-CW: Pionier 99-402).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Barrett AJ, Rawlings ND, Woessner JF: *Handbook of Proteolytic Enzymes*. London: Academic Press; 1998.

2. Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R: Crystal structure of the 20S proteasome from the archaeon ***T. acidophilum*** at 3.4 Å resolution. *Science* 1995, 268:533-539.

3. Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R: Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 1997, 386:463-471.

4. Bochtler M, Ditzel L, Groll M, Huber R: Crystal structure of heat shock locus V (HslV) from ***Escherichia coli.*** *Proc Natl Acad Sci USA* 1997, 94:6070-6074.

5. Bochtler M, Ditzel L, Groll M, Hartmann C, Huber R: The proteasome. *Annu Rev Biophys Biomol Struct* 1999, 28:295-317.

6. Bochtler M, Hartmann C, Song HK, Bourenkov GP, Bartunik HD, Huber R: The structures of HsIU and the ATP-dependent protease HsIU-HsIV. *Nature* 2000, 403:800-805.

7. •• Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, McKay DB: Crystal and solution structures of an HslUV protease-chaperone complex. *Cell* 2000, 103:633-643.
This paper describes the structure of the HslUV complex, which serves as a prototype of the Clp/Hsp100 family. In the structure, HslU hexameric rings bind to the central double-hexameric ring of the HslV protease, with the I-domains pointing outwards. Putatively, the I-domains bind the substrates for degradation.

8. Rohrwild M, Pfeifer G, Santarius U, Muller SA, Huang HC, Engel A, Baumeister W, Goldberg AL: The ATP-dependent HslVU protease from ***Escherichia coli*** is a four-ring structure resembling the proteasome. *Nat Struct Biol* 1997, 4:133-139.

9. Sousa MC, Kessler BM, Overkleeft HS, McKay DB: Crystal structure of HslUV complexed with a vinyl sulfone inhibitor: corroboration of a proposed mechanism of allosteric activation of HslV by HslU. *J Mol Biol* 2002, 318:779-785.

10. •• Krojer T, Garrido-Franco M, Huber R, Ehrmann M, Clausen T: Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* 2002, 416:455-459.
This paper describes open and closed structures of DegP (HtrA), a multimeric complex that switches between chaperone and protease function depending on temperature. The authors suggest that the PDZ domains swing out and bind substrates for transfer to the catalytic sites, located in the cavity of the hexameric protease structure.

11. Li W, Srinivasula SM, Chai J, Li P, Wu JW, Zhang Z, Alnemri ES, Shi Y: Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. *Nat Struct Biol* 2002, 9:436-441.

12. Whitby FG, Masters EI, Kramer L, Knowlton JR, Yao Y, Wang CC, Hill CP: Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* 2000, 408:115-120.

13. Tamura T, Tamura N, Cejka Z, Hegerl R, Lottspeich F, Baumeister W: Tricorn protease–the core of a modular proteolytic system. *Science* 1996, 274:1385-1389.

14. Tamura N, Lottspeich F, Baumeister W, Tamura T: The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell* 1998, 95:637-648.

15. Brandstetter H, Kim JS, Groll M, Huber R: Crystal structure of the
•• tricorn protease reveals a protein disassembly line. *Nature* 2001, 414:466-470.
This paper describes the structure of the 720 kDa tricorn protease, which degrades peptides produced by the proteasome further to dipeptides and tripeptides. The authors propose that substrates enter and products exit through the seven- and six-bladed β-propeller domains, respectively. Furthermore, they discuss a protein disassembly line, degrading proteins into single amino acid residues, formed by proteasomes, tricorn proteases and interacting factors F1, F2 and F3

16. Fulop V, Szeltner Z, Polgar L: Catalysis of serine oligopeptidases is controlled by a gating filter mechanism. *EMBO Rep* 2000, 1:277-281.

17. Shi Y: Mechanisms of caspase activation and inhibition during apoptosis. *Mol Cell* 2002, 9:459-470.

18. Chai J, Shiozaki E, Srinivasula SM, Wu Q, Datta P, Alnemri ES, Shi Y, Dataa P: Structural basis of caspase-7 inhibition by XIAP. *Cell* 2001, 104:769-780.

19. Huang Y, Park YC, Rich RL, Segal D, Myszka DG, Wu H: Structural basis of caspase inhibition by XIAP: differential roles of the linker versus the BIR domain. *Cell* 2001, 104:781-790.

20. Riedl SJ, Renatus M, Schwarzenbacher R, Zhou Q, Sun C, Fesik SW, Liddington RC, Salvesen GS: Structural basis for the inhibition of caspase-3 by XIAP. *Cell* 2001, 104:791-800.

21. Chai J, Wu Q, Shiozaki E, Srinivasula SM, Alnemri ES, Shi Y: Crystal
•• structure of a procaspase-7 zymogen: mechanisms of activation and substrate binding. *Cell* 2001, 107:399-407.
This paper describes the structures of unprocessed caspase-7 zymogen and processed active caspase-7. Together with known structures of inhibited caspase-7, it shows the structural changes that occur upon proteolytic activation of procaspase-7 needed to form a fully competent active site.

22. Riedl SJ, Fuentes-Prior P, Renatus M, Kairies N, Krapp S, Huber R,
•• Salvesen GS, Bode W: Structural basis for the activation of human procaspase-7. *Proc Natl Acad Sci USA* 2001, 98:14790-14795.
This paper describes the structure of unprocessed caspase-7 zymogen. The structure reveals the structural changes required to generate active caspase-7.

23. Renatus M, Stennicke HR, Scott FL, Liddington RC, Salvesen GS: Dimer formation drives the activation of the cell death protease caspase 9. *Proc Natl Acad Sci USA* 2001, 98:14250-14255.

24. Xu G, Cirilli M, Huang Y, Rich RL, Myszka DG, Wu H: Covalent inhibition revealed by the crystal structure of the caspase-8/p35 complex. *Nature* 2001, 410:494-497.

25. Budayova-Spano M, Lacroix M, Thielens NM, Arlaud GJ,
•• Fontecilla-Camps JC, Gaboriaud C: 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* 2002, 21:231-239.
This paper describes the structure of zymogen C1r together with two of its five preceding regulatory domains. The authors hypothesize that, based on the dimeric structure, mechanical stress due to binding of the C1 complex to its target may induce large structural rearrangements that bring the scissile loops into the active sites for auto-proteolytic activation.

26. Gaboriaud C, Rossi V, Bally I, Arlaud GJ, Fontecilla-Camps JC: Crystal structure of the catalytic domain of human complement c1s: a serine protease with a handle. *EMBO J* 2000, 19:1755-1765.

27. Jing H, Macon KJ, Moore D, DeLucas LJ, Volanakis JE, Narayana SV: Structural basis of profactor D activation: from a highly flexible zymogen to a novel self-inhibited serine protease, complement factor D. *EMBO J* 1999, 18:804-814.

28. Jing H, Xu Y, Carson M, Moore D, Macon KJ, Volanakis JE, Narayana SV: New structural motifs on the chymotrypsin fold and their potential roles in complement factor B. *EMBO J* 2000, 19:164-173.

29. Huntington JA, Read RJ, Carrell RW: Structure of a serpin-protease complex shows inhibition by deformation. *Nature* 2000, 407:923-926.

30. Pannifer AD, Wong TY, Schwarzenbacher R, Renatus M, Petosa C,
• Bienkowska J, Lacy DB, Collier RJ, Park S, Leppla SH *et al.*: Crystal structure of the anthrax lethal factor. *Nature* 2001, 414:229-233.
This paper describes the crystal structure of anthrax lethal factor from *Bacillus anthracis.* The structure reveals a long binding groove specific for the N-terminal tail of members of the MAPK kinase family. It provides structural data for the design of drugs against the acute effects of anthrax infection.

31. Brown CK, Madauss K, Lian W, Beck MR, Tolbert WD, Rodgers DW: Structure of neurolysin reveals a deep channel that limits substrate access. *Proc Natl Acad Sci USA* 2001, 98:3127-3132.

32. Oefner C, D'Arcy A, Hennig M, Winkler FK, Dale GE: Structure of human neutral endopeptidase (Neprilysin) complexed with phosphoramidon. *J Mol Biol* 2000, 296:341-349.

33. Kagawa TF, Cooney JC, Baker HM, McSweeney S, Liu M, Gubba S, Musser JM, Baker EN: Crystal structure of the zymogen form of the group A Streptococcus virulence factor SpeB: an integrin-binding cysteine protease. *Proc Natl Acad Sci USA* 2000, 97:2235-2240.

34. Wlodawer A, Li M, Dauter Z, Gustchina A, Uchida K, Oyama H, Dunn BM, Oda K: Carboxyl proteinase from Pseudomonas defines a novel family of subtilisin-like enzymes. *Nat Struct Biol* 2001, 8:442-446.

35. Comellas-Bigler M, Fuentes-Prior P, Maskos K, Huber R, Oyama H, Uchida K, Dunn BM, Oda K, Bode W: The 1.4 Å crystal structure of kumamolysin. A thermostable serine-carboxyl-type proteinase. *Structure* 2002, 10:865-876.

36. Vandeputte-Rutten L, Kramer RA, Kroon J, Dekker N, Egmond MR,
• Gros P: Crystal structure of the outer membrane protease OmpT from ***Escherichia coli*** suggests a novel catalytic site. *EMBO J* 2001, 20:5033-5039.
This paper describes the first structure of a membrane protease, OmpT, from the outer membrane of *E. coli.* The structure shows a β-barrel fold, typical of outer membrane proteins, with the catalytic site located extracellularly in a groove formed by the extended barrel structure. The catalytic site reveals a novel arrangement of catalytic residues involving one histidine and three aspartate residues.

37. Sodeinde OA, Subrahmanyam YV, Stark K, Quan T, Bao Y, Goguen JD: A surface protease and the invasive character of plague. *Science* 1992, 258:1004-1007.

38. Kramer RA, Vandeputte-Rutten L, de Roon GJ, Gros P, Dekker N, Egmond MR: Identification of essential acidic residues of outer membrane protease OmpT supports a novel active site. *FEBS Lett* 2001, 505:426-430.

39. Wolfe MS, Selkoe DJ: Biochemistry. Intramembrane proteases–mixing oil and water. *Science* 2002, 296:2156-2157.