Protein Science (1998), 7:815–836. Cambridge University Press. Printed in the USA.

REVIEW

Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes

AMIR R. KHAN AND MICHAEL N.G. JAMES

MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

Abstract

Proteolytic enzymes are synthesized as inactive precursors, or "zymogens," to prevent unwanted protein degradation, and to enable spatial and temporal regulation of proteolytic activity. Upon sorting or appropriate compartmentalization, zymogen conversion to the active enzyme typically involves limited proteolysis and removal of an "activation segment." The sizes of activation segments range from dipeptide units to independently folding domains comprising more than 100 residues. A common form of the activation segment is an N-terminal extension of the mature enzyme, or "prosegment," that sterically blocks the active site, and thereby prevents binding of substrates. In addition to their inhibitory role, prosegments are frequently important for the folding, stability, and/or intracellular sorting of the zymogen. The mechanisms of conversion to active enzymes are diverse in nature, ranging from enzymatic or nonenzymatic cofactors that trigger activation, to a simple change in pH that results in conversion by an autocatalytic mechanism. Recent X-ray crystallographic studies of zymogens and comparisons with their active counterparts have identified the structural changes that accompany conversion. This review will focus upon the structural basis for inhibition by activation segments, as well as the molecular events that lead to the conversion of zymogens to active enzymes.

Keywords: conversion; limited proteolysis; prosegment; proteinase; zymogen

Proteolytic enzymes are essential for a variety of biological processes in organisms ranging from viruses and bacteria to mammals. In addition to their fundamental role in the digestion and catabolism of proteins, these enzymes have evolved to perform many specialized tasks in complex organisms (Neurath, 1984). In mammals, proteolytic action is necessary for such diverse processes as blood coagulation and fibrinolysis (Davie et al., 1991), the controlled degradation of cytosolic and nuclear proteins (Ciechanover, 1994; Palombella et al., 1994; Peters, 1994), regulation of sodium balance and blood pressure (Vallet et al., 1997), and the immune response and apoptosis (Greenberg, 1996; Groettrup et al., 1996). Bacterial enzymes, such as the serine proteinases, share mechanistic similarities with their mammalian counterparts, but have evolved separately and have unrelated amino acid sequences and three-dimensional structures (Matthews, 1977). The viral enzymes, such as the retroviral aspartic proteinases and the picornaviral cysteine proteinases, are structurally related to cellular enzymes. The viral cysteine proteinases are unusual because they have a chymotrypsin-like fold, but the nucleophilic residue is a cysteine (Allaire et al., 1994). The reason why a serine proteinase fold should possess a cysteine nucleophile in viral enzymes is unknown, but it has been postulated that they are the evolutionary precursors of the chymotrypsin family (Brenner, 1988).

All known cellular and bacterial proteolytic enzymes are synthesized as inactive precursors (or zymogens) to prevent unwanted protein degradation. Zymogen conversion to the active enzyme generally occurs by limited proteolysis of an inhibitory "activation segment" within a subcellular compartment or the extracellular milieu, in which the particular enzyme exerts its biological function. Conversion may involve accessory molecules, or the process may be autocatalytic, requiring no additional factors other than a drop in pH. Many zymogens contain N-terminal extensions of the mature enzyme (or prosegments) that prevent access of substrates to the active site. In the context of the present discussion, the "p" suffix following the residue names or numbers will denote those residues comprising the prosegments.

Activation segments have also been observed as insertions within the primary sequence of the mature enzyme, between the catalytic residues (Rudenko et al., 1995). However, the activation segment has never been found at the C-terminus of a zymogen, thus precluding the risk of the active site gaining activity before the synthesis of the polypeptide is complete. Some activation segments have additional roles in protein folding and/or intracellular sorting (e.g., prosubtilisin (Strausberg et al., 1993); procaricain (Groves et al., 1996); α-lytic pro-protease (Baker et al., 1993); yeast procarboxypeptidase Y (Winther & Sorenson, 1991); yeast proproteinase A (Klionsky et al., 1988; Westphal et al., 1996);

Reprint requests to: Michael N.G. James, MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada; e-mail: michael.james@ualberta.ca.

procathepsin L (Cuozzo et al., 1995)). Proteinases from retroviruses and picornaviruses are essential for the cleavage of the large polyprotein precursors of the viruses to their functional units. The proteolytic enzyme is initially synthesized as part of the polyprotein and is flanked by structural and enzymatic modules that are required for replication of the virus. The proteolytic enzyme is excised from the polyprotein by a stepwise and coordinated activation mechanism that has not been characterized in structural detail.

The subject of zymogen activation traces its roots to studies of the pancreatic and gastric enzymes in the 1930s (reviewed by Neurath, 1957). Kunitz and Northrop characterized the biochemical and biophysical properties of the molecular species detected during the conversion of chymotrypsinogen to γ-chymotrypsin, and noticed partial peptide hydrolysis (Kunitz & Northrop, 1935; Kunitz, 1938). In later studies of the conversion of trypsinogen to trypsin by Desnuelle and co-workers, the newly formed N-terminal sequence was found to be Ile-Val-Gly, as characterized by the Edman degradation method (Rovery et al., 1953). These seminal experiments eventually contributed to our understanding of the role of limited proteolysis in the conversion process. During the late 1930s, Roger Herriott analyzed the kinetics of the conversion of porcine pepsinogen to pepsin at low pH and identified an intermediate compound, noting that the loss of pepsinogen was not immediately accompanied by an increase in pepsin (Herriott, 1939). It is remarkable that these careful experiments were performed at a time when the fundamental structure of proteins remained a matter of intense debate (Tanford, 1997). The development of X-ray crystallographic methods for proteins in the 1960s and 1970s expanded our knowledge of the structural basis for the conversion process in the serine proteinases. Crystallographic studies by Kraut and colleagues (chymotrypsinogen) and Bode and Huber (tryspinogen) provided the first glimpse of the structural basis for the inactivity of zymogens. It is remarkable to note that the contributions of Bode and Huber to the field of zymogen activation has continued unabated through two decades to the present (Aviles et al., 1993; Bode & Renatus, 1997).

It is satisfying to observe a rapid increase in the number of zymogen structures determined by X-ray crystallography in recent years. Most families of proteolytic enzymes now have at least one representative zymogen structure in the protein database (Table 1). In some instances, recombinant DNA techniques have been used to mutate active site residues in order to prevent autocatalytic conversion of the zymogen, thus enabling purification and structural analyses. In other instances, the mechanisms of inhibition in zymogens have been studied with a view to the development of novel and specific inhibitors of the corresponding mature enzymes in the pharmaceutical industry (Becker et al., 1995). In this study, the prosegment of pro-stomelysin-1, a cysteine proteinase implicated in tumor invasion, interacts with the active site in the reverse polypeptide direction (N $\rightarrow$ C termini) that would be expected for

Table 1. *X-ray crystallographic structures of zymogens*

Zymogen	Family	Resolution (Å)	PDB	Reference
Bovine chymotrypsinogen	Serine endopeptidase	1.8	2cga	Wang et al., 1985
		2.5	1chg	Freer et al., 1970
Bovine trypsinogen		1.9	1tgn	Kossiakoff et al., 1977
		1.8	1tgb	Fehlhammer et al., 1977
Bovine zymogen E		2.3	1fon	Pignol et al., 1995
Prosubtilisin[a]		2.0	1spb	Gallagher et al., 1995
Human protective protein	Serine carboxypeptidase	2.2	1ivy	Rudenko et al., 1995
Porcine pepsinogen	Aspartic endopeptidase	1.8	2psg	James & Sielecki, 1986
		1.65	3psg	Hartsuck et al., 1992
Human progastricsin		1.62	1htr	Moore et al., 1995
Human procathepsin L	Cysteine endopeptidase	2.2	1cjl	Coulombe et al., 1996
Human procathepsin B		2.5	3pbh	Turk et al., 1996
Rat procathepsin B		2.8	1mir	Cygler et al., 1996
Procaricain from *Carica papaya*		3.2	1pci	Groves et al., 1996
Porcine procarboxypeptidase A1	Zinc carboxypeptidase	2.0	1pca	Guasch et al., 1992
Human procarboxypeptidase A2	Zinc carboxypeptidase	1.8	—	Garcia-Saez et al., 1997
Porcine procarboxypeptidase B		2.3	1nsa	Coll et al., 1991
Human pro-stromelysin-1	Zinc matrix metalloproteinase	1.9	1slm	Becker et al., 1995
Bovine ternary complex	Two serine endopeptidases Zinc carboxypeptidase	2.35	1pyt	Gomes-Ruth et al., 1997
Yeast proteasome[b]	Ntn hydrolase family	2.35	1ryp	Groll et al., 1997

[a]The structure was a noncovalent complex of the prosegment with the active enzyme.

[b]The proteasome contains two catalytic β-subunits that are inactive due to uncleaved N-terminal extensions (prosegments) of their mature segments.

genuine substrates. It is proposed that inhibitors may be designed to exploit this ability of the enzyme to find peptides in the reverse sense. Since zymogen activation encompasses such fields as protein folding, targeting, and catalysis, there is little doubt that studies of zymogens and their conversion will continue to provide new insights into protein structure and function in years to come.

The purpose of this review is to acquaint readers with the molecular events that result in the conversion of zymogens to active enzymes. Consequently, the scope of the review will be limited to published structures of zymogens and the corresponding mature enzymes, as well as any biophysical or biochemical data related to these structural studies. The inhibitory mechanisms utilized by activation segments are diverse, but a common property of zymogens is that the "catalytic machinery" is preformed. Here, the "catalytic machinery" refers to all aspects of the active site (bond-cleavage apparatus, substrate-binding cleft) that are required for productive cleavage of peptide bonds. For example, the positions and conformations of the catalytic triads of the serine and cysteine proteinase zymogens, the two catalytic Asp residues in aspartic proteinase zymogens, and the catalytic Zn^{2+} in metalloproteinase zymogens are virtually identical to their corresponding active forms. However, the conversion process often involves significant conformational changes in regions that are adjacent to the active site, or within the activation segments that are subsequently removed.

This review is organized into the four enzyme superfamilies, as classified by the International Union of Biochemistry (IUBMB) according to their mechanisms of peptide hydrolysis (IUBMB, 1992). The fifth section is devoted to oligomeric complexes of zymogens and the effects upon activation pathways. Within each family, the discussion proceeds roughly in the following manner: (1) a description of the mature enzyme and its mechanism of catalysis, (2) the structure of the zymogen and the mechanism of inhibition, and (3) the molecular events that result in conversion of the zymogen to the active enzyme. The review concludes with a discussion of the future directions in studies of zymogen activation.

1. Serine proteinases and their zymogens

Serine proteinases (EC 3.4.21.-) contain a catalytic serine residue that acts as the nucleophile during catalysis. These enzymes have evolved from their role in digestion to specialized roles such as participation in the cascade of cleavages that sequentially activate zymogens during the formation of a blood clot (Neurath, 1984). The overall folds are categorized as the chymotrypsin-like enzymes, which contain two domains that are each composed of a six-stranded and anti-parallel β-barrel, and the bacterial (subtilisin-like) enzymes. Our discussions will focus upon the chymotrypsin family, although the reader is directed to the recently determined crystal structure of subtilisin in a noncovalent complex with its inhibitory prosegment (Gallagher et al., 1995).

The same catalytic machinery has arisen from an independent evolutionary history in all serine proteinases. During the reaction cycle, the catalytic serine forms a covalently-attached tetrahedral intermediate with the carbonyl-carbon atom of the scissile peptide bond of substrates (Kraut, 1977; Matthews et al., 1977; Steitz & Shulman, 1982; Polgar, 1989). The serine residue is found at the active center as part of a geometrically conserved Ser-His pair (Ser195 and His57 in chymotrypsin; Fig. 1A). The histidine residue acts as a general base to enhance the nucleophilicity of Ser195 during the reaction (Perona & Craik, 1995). In addition, an aspartate residue (Asp102) is often situated in the vicinity of the His and may act to stabilize the required tautomer and rotamer of the imidazolium side chain (Craik et al., 1987; Sprang et al., 1987). Finally, an essential component of the active site is a pair of hydrogen bond donors (the oxyanion hole) that stabilize the developing negative charge on the carbonyl oxygen atom of the tetrahedral intermediate. In chymotrypsin, these are the backbone amide atoms of Gly193 and Ser195 (Robertus et al., 1972). The amides donate their protons to the carbonyl oxygen of the P1 residue (notation of Schechter & Berger, 1967), providing the stabilization of the transition state that is necessary for catalysis. Together, these properties of the active site machinery require a precise conformation for the segment 190–195 in chymotrypsin.

The conversion of chymotrypsin-like zymogens to the active enzymes was the earliest model of the role of limited proteolysis in zymogen activation (Neurath, 1957). The basic principle of conversion, that structural changes in the zymogen are necessary for activation, was established over 40 years ago (Davie & Neurath, 1955). The activation of chymotrypsinogen is initiated by trypsin-mediated cleavage of a peptide bond between Arg15–Ile16. The conversion process can also be triggered by the action of enterokinase and is influenced by pH and the presence of Ca^{2+} ions (reviewed by Neurath, 1957). Limited proteolysis results in a newly-liberated NH_3^+-terminus at Ile 16, which forms an ion-pair with the carboxylate side chain of Asp194, triggering the formation of an active enzyme (Freer et al., 1970; Huber & Bode, 1978). Subsequent cleavages result in the release of the dipeptides Ser14–Arg15 and Thr147–Asn148 by an autocatalytic mechanism. The α- and γ-forms of chymotrypsin are chemically identical but are distinguished by the space group in which the protein forms crystals under the influence of varying pH (α-chymotrypsin, $P2_1$, two molecules in the asymmetric unit; γ-chymotrypsin, $P4_22_12$, one molecule per anymmetric unit (Kunitz, 1938; Desnuelle, 1959; Cohen et al., 1981; Blevins & Tulinsky, 1985; Tsukada & Blow, 1985). The N-terminal peptide (residues 1–13) remains attached to the rest of the enzyme via a disulfide bridge between Cys1 and Cys122. The chemical and structural properties of α- and β-trypsin have also been characterized (Bode & Schwager, 1975; Bode et al., 1976; Bartunik et al., 1989). The two forms of the active enzyme have similar catalytic properties but are distinguished by the lack of cleavage within the autolysis loop, following Lys145, in β-trypsin.

X-ray crystallographic studies of serine proteinases in the 1960s and 1970s established that the overall fold of the zymogen and the active enzyme are identical but that a small region undergoes large structural changes during conversion. Comparisons of the structures of chymotrypsinogen and α-chymotrypsin revealed that the substrate-binding cleft was only partially formed in the zymogen and that the peptide bond between Met192–Gly193 was in the wrong orientation to contribute a proton to the oxyanion hole (Steitz et al., 1969; Freer et al., 1970). Soon afterward, the detailed crystallographic studies of trypsinogen and trypsin in several crystal forms by Bode and Huber, including an isomorphous trigonal form of the zymogen and the active enzyme, provided the fundamental insight into the basis for low reactivity of the zymogen (Fehlhammer et al., 1977; Bode & Huber, 1978; Huber & Bode, 1978). These studies revealed that the zymogen has a pre-formed catalytic traid, but the inactivity arises from a disordered loop (186–194) resulting in a partially obstructed substrate-binding cleft and an immature oxyanion hole. These conclusions were consistent with studies by Neurath and co-workers showing that small active-site inhibitors of serine proteinases (e.g., methanesulfonyl fluoride) acylate the zymogen nearly as efficiently as the mature

enzyme. However, bulky reagents have difficulties in accessing the active site of trypsinogen (vs. trypsin), as evidenced by comparisons of their second-order rate constants (Morgan et al., 1974; Kerr et al., 1975).

Activation mechanism of chymotrypsinogen

In the 1980s, the structure of bovine chymotrypsinogen was solved to 1.8 Å resolution (Wang et al., 1985) and compared to the refined models of γ-chymotrypsin and α-chymotrypsin (Cohen et al., 1981; Blevins & Tulinsky, 1985). These studies revealed that the position and conformation of the active site residues in the zymogen, Ser195 and His57, are indistinguishable from those of the mature enzyme (Fig. 1A,B). Large side-chain movements near the active site (Ile99, Ser214–Gly216) are not observed, as had been suggested previously (Freer et al., 1970). Part of the substrate binding cleft that includes the segments 213–220 and 225–228 is preformed in the zymogen.

The most dramatic conformational changes occur in the segment Ser189–Asp194, which immediately precedes the catalytic Ser195 (Fig. 1B). This segment consists of a loop that is turned "inward" in the zymogen, but following activation, the loop moves "outward" toward the solvent, thus forming the mature substrate binding cleft. The conformational rearrangement of the backbone in this segment is facilitated by cleavage of the peptide bond between

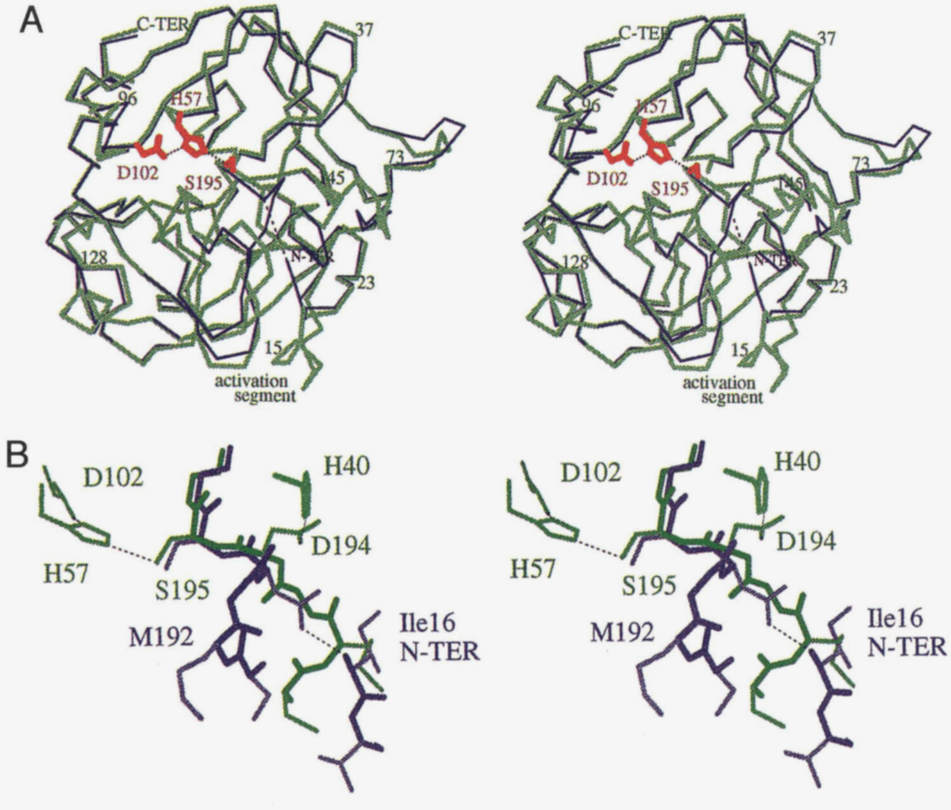

Fig. 1. Structures of serine proteinases and their zymogens. **A:** C_α vector superposition of bovine chymotrypsinogen (thick green; PDB code 2cga, Table 1) with α-chymotrypsin (thin blue, PDB code 5cha (Blevins & Tulinsky, 1985)). The models shown include residues 1–245 for the zymogen; 1–8, 16–146, and 149–245 for α-chymotrypsin. The RMSD for 1,736 backbone atoms was 1.27 Å. The side chains of Ser195, His157, and Asp102 of the zymogen are red. The dashed line between Ile16(C_α) and Asp194(C_α) of α-chymotrypsin helps to visualize the salt bridge between Ile16(N) and the side chain of Asp194 that leads to the conformational changes in the backbone of the segment 190–194. Conformational changes are also observed in the segments 16–19, 144–152, and 217–221. The figure was drawn using BOBSCRIPT (Esnouf, 1997). **B:** Active site of chymotrypsinogen (green) and α-chymotrypsin (purple), shown as stick vectors. The view is a close-up of the active site, in a similar orientation to (A). Hydrogen bonds are shown as dashed lines. The active site residues Ser195, His57, and Asp102 remain unchanged during conversion. However, the backbone conformation of the region from 190–194 is dramatically altered. The side chain of Asp194 hydrogen bonds to the side chain of His40 in the zymogen. However, the side chain of Asp194 rotates to a new position, forming a salt bridge with the newly cleaved α-amino group of Ile16 in the active enzyme. **C:** Ribbon model of human protective pro-protein (PDB code 1ivy) in stereo. The "cap" region (residues 182–302) is shown in green ribbons, while the "core" region is represented in grey. The activation segment that is removed during conversion, Met285–Arg298, is red. The side chains of the catalytic residues (Ser150, His429, Asp372) are also red. The side chain of Asn275, which occupies the S1 binding pocket, as well as two pairs of salt-bridged residues (Arg262–Asp300, Glu264–Arg298) adjacent to the activation segment are also shown. The figure was drawn using BOBSCRIPT (Esnouf, 1997) and rendered with Raster3D (Merritt & Murphy, 1994). **D:** Close-up of the active site and the activation segment of pro-HPP. The orientation and color scheme are identical to (C). Hydrogen bonds and salt bridges are shown as dashed lines. The sites of limited proteolysis are Arg284–Met285 and Arg298–Met299, resulting in the removal of the region highlighted in red. However, this region itself does not block the active site, suggesting that conformational changes that uncover the active must take place during conversion. Following limited proteolysis and removal of Met285–Arg298, the complex of the N- and C-terminal polypeptides is stabilized by a pair of disulfide bonds (not shown). (*Figure continues on facing page.*)

Arg15–Ile16. His40, at one time implicated in a push-pull mechanism of proteolytic hydrolysis with His57 (Polgar & Halasz, 1982), forms a hydrogen-bonded ion-pair with the Asp194 carboxylate in the zymogen. In the active enzyme, Asp194 changes its allegiance and forms an ion-pair with the newly liberated N-terminus at Ile16. The overall change at Asp194 involves a rotation at the C^{α}-C bond of about 180°. The resulting conformational changes in the backbone orient the Gly193(N) in the direction of the substrate binding site, where it serves as a hydrogen-bond donor to the substrate carbonyl oxygen in the oxyanion hole. Another consequence of the conformational change is the movement of the side chain of Met192 from a buried position in the zymogen to a completely exposed location in the mature enzyme (Kerr et al., 1976; Wang et al., 1985). The C_{α} atom of Met192 differs by 7.3 Å from its position in the zymogen as compared to mature chymotrypsin. In summary, conversion of the zymogen to the mature enzyme involves the completion of a partially formed substrate-binding cleft and oxyanion hole.

Comparisons of the structure of bovine chymotrypsinogen to that of bovine trypsinogen (Fehlhammer et al., 1977; Wang et al., 1985) indicate minor differences in the regions that are relevant to the conversion process. The differences include the loop between residues Ile16–Gly19, which appears to be strained in chymotrypsinogen due to the close approach of Val17 (C_{β}) to Asn18(N); the equivalent loop in trypsinogen is disordered. In both zymogens, limited proteolysis and formation of a salt bridge involving the newly liberated N-terminus and an internal Asp residue is essential for conversion. However, it is possible to generate a functional serine proteinase without limited proteolysis of an activation segment (Bolognesi et al., 1982). The structures of the bacterial serine proteinases from *Streptomyces griseus* and α-lytic protease have revealed that these enzymes do not have zymogen precursors that must be cleaved to the active form. The active conformation is triggered by substituting the side chain of another residue (Arg138) to provide the counter-ion for Asp194 (Brayer et al., 1979). This salt bridge is buried internally in these enzymes, suggesting that its formation takes place during the folding of the polypeptide (Brayer et al., 1979). Also, trypsinogen can be converted to a partially active enzyme upon addition of Ile-Val dipeptides without any proteolytic cleavages. The N-terminus of the dipeptide is able to mimic the cleaved N-terminus at Ile16, thus forming a salt bridge with Asp194 and triggering the conversion process (Bode et al., 1978).

Detailed structural and thermodynamic studies of trypsinogen complexes with various peptides have indicated that the conformational transition is achieved by the movement of Ile16 toward a

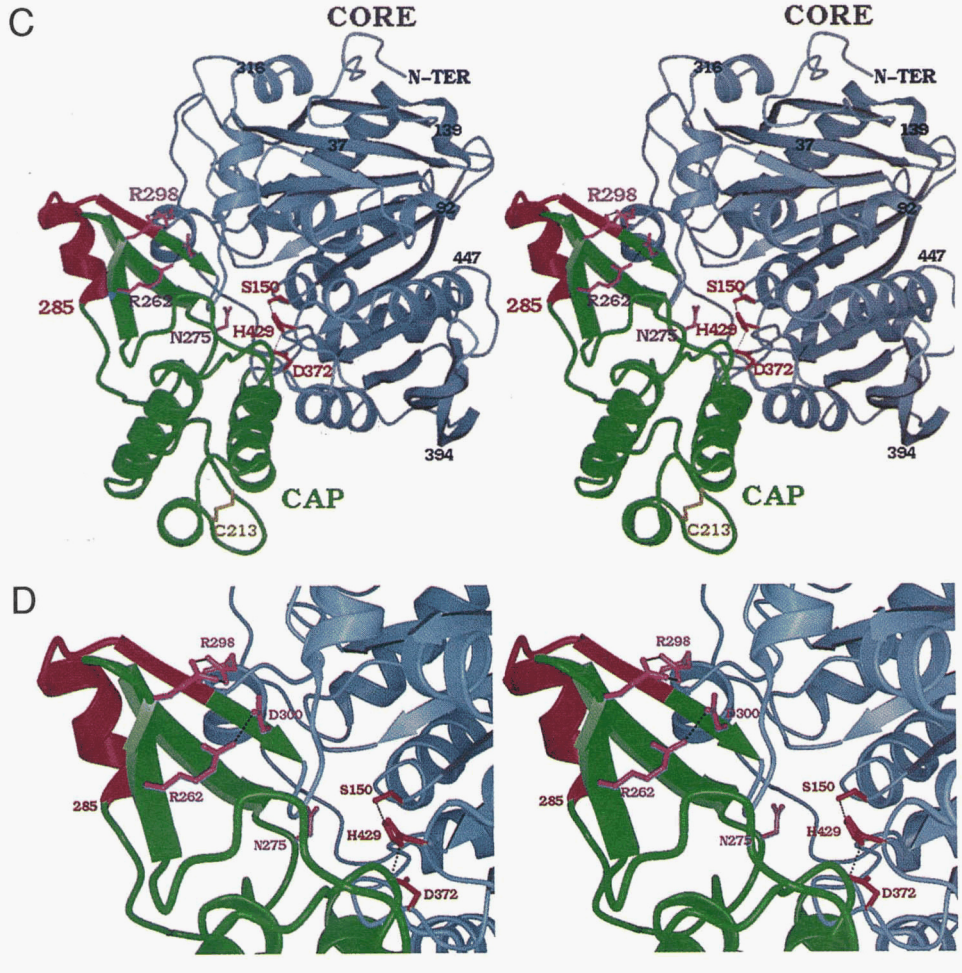

Fig. 1. *Continued.*

pocket (the "Ile16 cleft"), and formation of the salt bridge between Ile16(N) and the side-chain carboxylate of Asp194 (Bode, 1979). The Ile16 cleft is solvent-filled in trypsinogen, and during the conversion process, it has the correct size and shape to accommodate the side chain of Ile16. Alternatively, the binding of pancreatic trypsin inhibitor to trypsinogen is found to be sufficient to drive the conversion of trypsinogen to a "trypsin-like" conformation. Based upon these observations, a model has been proposed in which the zymogen is able to alternate between inactive and active conformations such that limited proteolysis of the activation segment forces the zymogen irreversibly into the active state (Bode, 1979). A residual activity in chymotrypsinogen toward small ester-based substrates is suggestive that the active site requires very subtle changes and is poised for catalysis (Kerr et al., 1976). In addition, trypsinogen is able to self-activate in the presence of Ca^{2+} ions, and the zymogen is further able to activate chymotrypsinogen in *trans*, suggesting partial catalytic activity inherent within these zymogens (Kay & Cassell, 1971).

Regulation of zymogen activation

The regulation of serine proteinase activity represents perhaps the most prominent example of zymogen conversion that has been exploited for a wide range of biological processes. In several examples, the zymogen forms of the serine proteinases have significant catalytic activity that can be further enhanced by interactions with nonproteolytic cofactors under physiological conditions. One example is the tissue-type plasminogen activator (t-PA), a multidomain protein that contains a chymotrypsin-like proteinase module. The zymogen is converted to a two-chain (active) form by limited proteolysis at an Arg-Leu peptide bond that corresponds to the Arg15–Ile16 bond of chymotrypsinogen. However, unlike most serine proteinase zymogens, single-chain t-PA has significant (10–20%) catalytic activity relative to the two-chain form (Madison et al., 1993). The catalytic activity can be further stimulated upon interactions with fibrin via its finger domains at the sites of tissue injury (Lamba et al., 1996). The X-ray structure of the single chain (zymogen) form of t-PA has revealed that a salt-bridge interaction between Asp194 and Lys156 (chymotrypsin numbering) contributes to the "active-like" conformation (Renatus et al., 1997). This structural property of t-PA is reminiscent of the molecular basis for the activity observed in α-lytic protease. Although many serine proteinases have the equivalent Lys156, additional structural determinants have been identified in t-PA that shield solvent from the Lys residue and confer the propensity for salt-bridge formation with Asp194 (Bode & Renatus, 1997; Renatus et al., 1997).

Another notable exception to the general principles of zymogen conversion is the activation pathway of the serine proteinase, complement factor D. The active form of this chymotrypsin-like enzyme cleaves another serine proteinase, factor B, thereby triggering the alternative pathway of the complement system (Muller-Eberhard, 1988). Unlike the chymotrypsin-like enzymes, the N-terminal activation peptide of pro-factor D is cleaved and removed within the secretory pathway of the adipocytes that synthesize and secrete the enzyme into the blood (Steiner et al., 1992; White et al., 1992; Yamauchi et al., 1994). However, this circulating form of factor D is inactive (Yamauchi et al., 1994), and biochemical studies *in vitro* have shown that mature factor D alone is a poor catalyst of the hydrolysis of peptide thioester substrates (Kam et al., 1987; Kim et al., 1994). X-ray crystallographic studies of factor D have attributed the observed kinetic properties to a distorted catalytic triad and an unusual conformation for the loop 214–220 that results in a narrow substrate binding cleft (Narayana et al., 1994). Factor D is activated upon Mg^{2+}-dependent association with factor B, its only known substrate, together with the major fragment of C3 cleavage (C3b). The formation of the resulting complex, Mg^{2+}-factor D-C3b, is predicted to induce conformational changes at the active site of factor D, leading to a realignment of the catalytic triad and acquisition of proteolytic activity (Narayana et al., 1994; Volanakis & Narayana, 1996). In summary, the low intrinsic activity of factor D, its limited substrate specificity, and its requirement for cofactors prevent inappropriate proteolysis in the circulatory system.

In general, limited proteolysis results in the irreversible conversion of zymogens to active enzymes. In response to unchecked proteolytic activity, nature has developed an elaborate system of "protein-proteinase" inhibitors to turn the enzymes "off." For the serine proteinases, the two families of these inhibitors are the substrate-like canonical inhibitors (Laskowski & Kato, 1980; Bode & Huber, 1992) and the "serine proteinase inhibitors" (serpins; Potempa et al., 1994; Wright, 1996). Examples of serpins include α_1-antitrypsin (the archetypal member), C1-inhibitor (which attenuates the complement cascade), antithrombin, and the "plasminogen activator inhibitor 1" (PAI-1) that inhibits both the single-chain and two-chain forms of t-PA. The critical balance of proteolytic activity and inhibitor is demonstrated by the link between natural variants of serpins and diseases such as emphysema (α_1-antitrypsin), thromboembolic disease (antithrombin), and hereditary angioedema (C1-inhibitor; Stein & Carrell, 1995).

Human protective protein

The human protective protein (HPP, cathepsin A; EC 3.4.16.5) is a serine carboxypeptidase that is related in sequence and structure to the wheat and yeast serine carboxypeptidases (Liao et al., 1992; Endrizzi et al., 1994). The zymogen form of the protein associates with β-galactosidase and neuraminidase as a multienzyme complex in lysosomes, "protecting" these glycosidases from degradation (d'Azzo et al., 1982). The physiological function of the mature enzyme is unclear; however, a role in the deactivation of bioactive peptides such as endothelin I has been suggested (Jackman et al., 1992). The enzyme is synthesized as a 452-residue zymogen (pro-HPP) with an internal activation segment (residues Met285–Arg298; Fig. 1C). The catalytic triad (Ser150, His429, and Asp372) is preformed in the zymogen and does not undergo substantial changes during conversion, as observed in the chymotrypsinogen counterpart. However, unlike chymotrypsinogen, the structure of pro-HPP has revealed that both the oxyanion hole and the substrate binding cleft are also fully formed (Rudenko et al., 1995). The zymogen is inactive because substrates are unable to gain access to the catalytic cleft (Fig. 1D). Conversion to the active enzyme takes place in lysosomes and involves a trypsin-like cleavage and removal of the activation segment (Met285 to Arg298) to form the mature enzyme. The N- and C-terminal fragments of the mature enzyme remain attached to each other by disufide bonds between Cys253–Cys303 and Cys60–Cys334 (Bonten et al., 1995). Remarkably, the activation peptide itself does not block the active site, but probably influences the position and conformation of an adjoining region that resides in the active site cleft (Fig. 1D). These observations suggest that conformational changes within the mature segment of the zymogen must accompany the conversion process (Rudenko et al., 1995).

The overall structure of pro-HPP is organized into two domains, a "cap" domain and a "core" domain (Fig. 1C), that are typical of the family of α/β hydrolases (Ollis et al., 1992). The "cap" domain (residues 182–302) contains a helical region (Hα1–Hα3, residues 182–253) and a "maturation" region consisting of a three-stranded mixed β-sheet (Mβ1–Mβ3, residues 254–302). The remainder of the zymogen is the core domain that forms the scaffold for the active site and contributes the catalytic residues Ser150, His429, and Asp372. The core domain contains a central 10-stranded β-sheet that is flanked on both sides by a total of 10 α-helices. The zymogen is inactive because the preformed active site in the core domain is blocked by residues from a loop (Asn275–Phe277) that follows Mβ2 in the maturation region. These residues form hydrophobic interactions with the core domain, and in particular, Asn275 establishes contacts with a part of the S1 binding pocket (Fig. 1D).

The *in vivo* activation process in the lysosomes can be mimicked *in vitro* by trypsin cleavages after residues Arg284 and Arg298 (Bonten et al., 1995). The activation peptide is located on an exposed loop that is adjacent to the inhibitory interactions between the cap and the core domains. The structure of the zymogen indicates that cleavage and removal of the activation peptide would not directly result in an accessible active site (Rudenko et al., 1995). A model has been proposed for the pH-dependent conversion process that involves the disruption of several salt bridges within the maturation region (Arg262 to Asp300, Glu264 to Arg298). These disruptions would likely cause local conformational changes in the region of the mixed β-sheet, thus enabling limited proteolysis and removal of the peptide Met285–Arg298. Subsequently, a conformational change in the flanking maturation region (residues 254–284, 299–302) is proposed to allow accessibility of substrates to the active site (Rudenko et al., 1995). The detailed changes in conformation leading to mature HPP are unknown and await structural studies of the active enzyme. The residues that form the salt bridge interactions are also conserved in mouse and chicken protective proteins, suggesting a common conversion pathway (Galjart et al., 1991). However, this pH-dependent trigger model requires further biochemical verification, such as a detailed pH-dependent analysis of the conversion process (Rudenko et al., 1995).

2. *Aspartic proteinases and their zymogens*

Aspartic proteinases (EC 3.4.23.-) are found in viruses, fungi, yeast, plants, and mammals. The common feature of these enzymes is an active site cleft that contains two catalytic aspartate residues (reviewed by Davies, 1990). Some of the biological functions of aspartic proteinases include: (1) site-specific proteolysis of the retroviral precursor polyprotein to the structural, regulatory, and enzymatic proteins that are necessary for replication of the virion (e.g., HIV-1 protease (Wlodawer & Erickson, 1993)); (2) a role in tissue invasion and virulence for fungal pathogens of humans (Cutfield et al., 1995); (3) cleavage of hemoglobin in the digestive vacuoles of the malarial parasite (plasmepsins (Francis et al., 1997)); (4) regulation of blood pressure (renin (Sielecki et al., 1989)); and (5) digestion of dietary proteins in the low pH environment of the stomach in mammals. The aspartic proteinases have a long history in the field of human cuisine. Chymosin has been the active ingredient in cheese-making for thousands of years, and aspartic proteinases are also used for the production of soya sauce, a practice that apparently originated during the Zhou dynasty (Hofmann, 1989; Davies, 1990).

The overall fold of the aspartic proteinases has been conserved in all organisms despite significant divergence of their primary sequences (Blundell & Johnson, 1993). The structure consists of two β-barrel domains that are connected by a central β-sheet (Fig. 2A–C). The viral aspartic proteinases have evolved as a compact single β-barrel domain that dimerizes to form the active enzyme (Miller et al., 1989; Navia et al., 1989), consistent with the prediction that cellular aspartic proteinases have emerged by gene duplication of a single domain (Tang et al., 1978). The active site is composed of two aspartate residues (Asp32 and Asp215 in pepsin; Asp32 and Asp217 in gastricsin) that are located on a pair of similarly-folded loops which reside in a prominent cleft between the domains. A water molecule is hydrogen-bonded to the two active site Asp residues at the base of an extended hydrophobic groove that is formed by the juxtaposition of the two domains. During catalysis, the water is de-protonated and subsequently attacks the carbonyl carbon of the scissile bond of substrates to form a tetrahedral intermediate (Davies, 1990; James et al., 1992). Unlike the serine proteinases, the catalytic Asp residues do not form a covalent complex with their substrates.

Whereas the structures of the active enzymes are known from many organisms, the only zymogen structures that have been determined are from mammalian sources. These zymogens are porcine pepsinogen (pPGN), human pepsinogen A (hPGA), and human progastricsin (hPGC; Fig. 2A), which are components of the gastric juices of the respective species (James & Sielecki, 1986; Moore et al., 1995; Bateman et al., 1998). The sequence identities are about 50% between the mature forms of these enzymes, pepsin and gastricsin. Finally, the structure of an intermediate form of human gastricsin has been solved by X-ray crystallography (Fig. 2B), thus providing molecular details of various stages in the activation pathway (Khan et al., 1997).

The zymogen forms of gastric proteinases contain a positively-charged N-terminal prosegment that wraps around the central portion of the enzyme, forming salt-bridge interactions with the negatively charged mature segment (Fig. 2A). The prosegments of the gastric zymogens vary between 43 to 47 residues in length and are related in both sequence and structure (Fig. 2D). The prosegment is not an independently folding unit, and once removed, the peptide is autocatalytically degraded by the active enzyme. In the structure of human progastricsin, a Lys residue (Lys37p in hPGC; "p" suffix denotes the prosegment) forms hydrogen-bonded salt bridges to the active site Asp residues (Asp32, Asp217). These interactions position the prosegment region Pro34p to Arg39p (a 3_{10}-helix) in front of the preformed active site, thereby preventing the approach of substrates. The conversion process is initiated by disruption of these salt bridges at low pH and subsequent limited proteolysis of the prosegment. No other accessory molecules are required for conversion.

Comparisons of the structures of the zymogens (pepsinogen and progastricsin) and their active counterparts have revealed that upon proteolytic removal of the prosegment, the N-terminus (Ser1 to Met10 in gastricsin) of the mature segment undergoes a major conformational rearrangement. In the zymogen, this region consists of a loop and 3_{10} helix that protrudes into the active site cleft. However, in the mature enzyme, residues Ser1–Met10 swing around to the back of the molecule, forming the first strand of the six-stranded β-sheet (Fig. 2C). In the zymogen, prosegment residues Ala1p to Lys10p comprise the equivalent β-strand (Fig. 2A), but the prosegment is removed during the conversion process.

Biochemical and structural studies of activation intermediates

The activation process of the aspartic proteinases has been extensively studied using biochemical and biophysical techniques. Kinetic studies by Roger Herriott in the late 1930s showed that the conversion of pepsinogen into pepsin at low pH is an autocatalytic reaction that involves the formation of intermediates (Herriott, 1939). Further studies using spectroscopic techniques in the 1970s established that conformational changes take place during conversion. Upon initial exposure of the zymogen to acid, a conformational changes occurs without proteolytic cleavage in the 5 ms to 2 s timescale (Auer & Glick, 1984). This intermediate of hPGC has been termed intermediate 1 (Foltmann & Jensen, 1982), and the transition can be reversed by rapidly returning the pH of the solution to neutrality. It is predicted that the helical regions of the prosegment become unraveled in intermediate 1, thus resulting in altered migration of the protein upon gel electrophoresis under nondenaturing conditions.

The first hydrolytic event during activation of hPGC is the unimolecular cleavage of the peptide bond between Phe26p and Leu27p (Foltmann & Jensen, 1982). Subsequently, intermolecular cleavage at the pro-mature junction (Leu43 to Ser1) results in a transient intermediate (intermediate 2, or hGSi) in which residues Ala1p to Phe26p are noncovalently associated with the mature enzyme (Ser1 to Ala329), while the remainder of the prosegment (Leu27p to Leu43p) is removed (Foltmann & Jensen, 1982). The intermediate complex can be stabilized by transferring the contents of the reaction vessel to neutral pH, thereby preventing further activation events and inhibiting proteolytic activity.

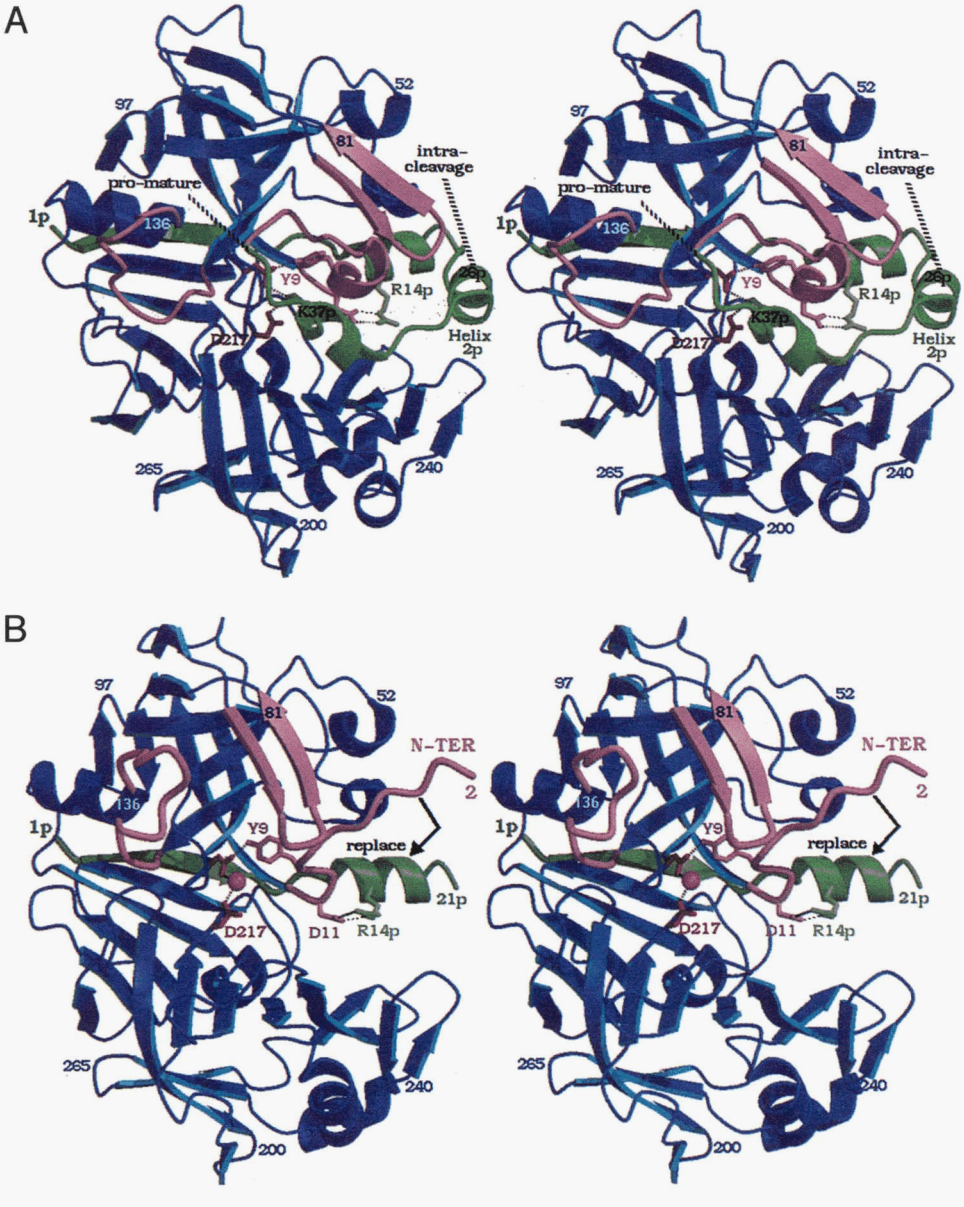

Fig. 2. *Figure continues on facing page.*

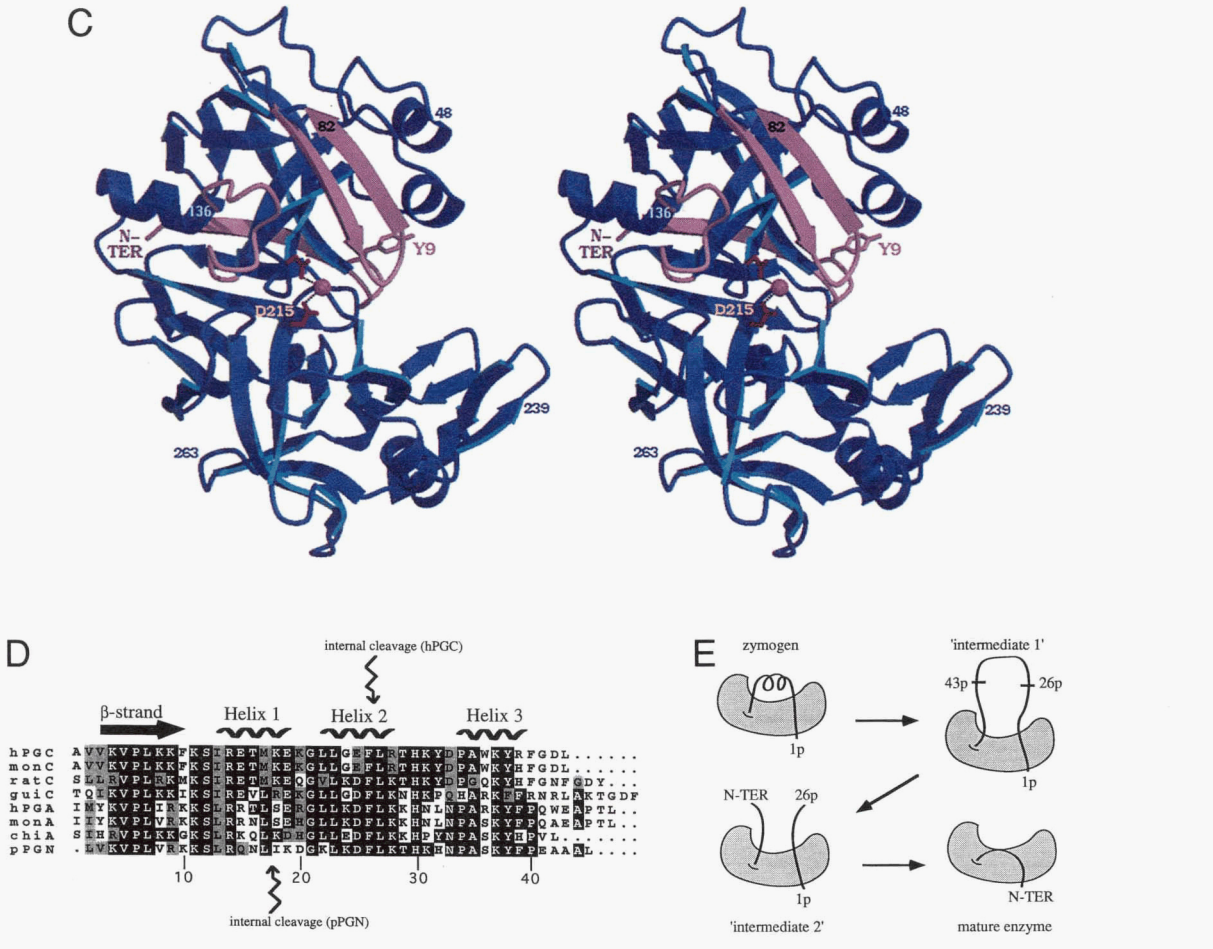

Fig. 2. Activation pathway for the gastric aspartic proteinases. The conversion process can be followed by noting the gradual loss of the green prosegment from Figures 2A–C. **A:** Ribbon model of human progastricsin (PDB code 1htr). The prosegment, 1p–43p, is green; the mature segment (gastricsin) is blue. The regions within the mature segment that undergo conformational changes (residues 1–10, 71–81, 124–135) are emphasized in pink. The pro-mature junction is marked with the dashed line. A site of intramolecular cleavage within the helical region of the prosegment (helix 2p) is indicated. Active site residues Asp32 and Asp217 are red. Lys37p (green), located within helix 3p, forms salt bridges with the active site aspartates (dashed lines). Tyr9 hydrogen bonds with Asp32 and resides in the approximate S1 binding pocket. The side chains of Arg14p and Asp11 form a salt bridge, further stabilizing the prosegment and its association with the mature segment. **B:** Structure of activation "intermediate 2" (PDB code 1avf), the noncovalent complex of a prosegment peptide (Ala1p–Phe26p) with the mature enzyme. The color scheme is identical to Figure 2A. The remaining prosegment interacts in a similar fashion to the equivalent region in the zymogen. However, the N-terminus of mature gastricsin has moved by about 30 Å from its position in the zymogen, as indicated by residue 2 (valine). A water molecule (sphere) now resides between the catalytic aspartate residues. However, Tyr9 remains hydrogen bonded to Asp32, partially obstructing the S1 binding site. The next stage in conversion is the dissociation of the prosegment peptide and its replacement by the N-terminus of mature gastricsin (arrow). **C:** Structure of active human pepsin (PDB code 1pso). The prosegment peptide has dissociated, permitting the mature N-terminus to occupy the first β-strand in the six-stranded β-sheet at the back of the enzyme. The active site is fully uncovered, as evidenced by the position and orientation of Tyr9. The equivalent Asp217 of gastricsin is numbered Asp215 in pepsin. The structure of mature gastricsin has not been determined, but is expected to closely resemble that of pepsin. **D:** Sequence alignment of the prosegments of mammalian gastric aspartic proteinases. Conserved residues are highlighted in bold. Human progastricsin numbering is given at the bottom of the figure. The prosegments are predicted to have a similar organization of secondary structure elements, based upon the crystal structures of human progastricsin (Moore et al., 1995) and porcine pepsinogen (James & Sielecki, 1986). The helical regions are predicted to uncoil during the initial events of the conversion process. Aside from the pro-mature junction, the internal sites of cleavage for human progastricsin (top; Phe26p–Leu27p) and porcine pepsinogen (bottom; Leu16p–Ile17p) are indicated within the helical region of the prosegment. Sequences are: hPGC, human progastricsin (Hayano et al., 1988; Taggart et al., 1989); monC, monkey progastricsin (Kageyama & Takahashi, 1986a); ratC, rat progastricsin (Ichihara et al., 1986); guiC, guinea pig progastricsin (Kageyama et al., 1992); hPGA, human pepsinogen (Sogawa et al., 1983); monA, monkey pepsinogen (Kageyama & Takahashi, 1986b); chiA, chicken pepsinogen (Baudys & Kostka, 1983); pPGN, porcine pepsinogen (Foltmann, 1988). **E:** Schematic representation of the activation pathway. The view has been rotated from Figure 2A–C so that the N- and C-terminal lobes are on the left and right, respectively. The known structures are the zymogen, "intermediate 2," and the mature enzyme. In step 1, the drop in pH causes the three helices of the prosegment to uncoil, exposing the active site and initiating the autocatalytic conversion process. The first two autolytic cleavages take place at Phe26p–Leu27p and Leu43p–Ser1 to form "intermediate 2," the noncovalent complex of the prosegment peptide (1p–26p) and mature gastricsin. In the final stage, resides Ser1 to Met10 (N-TERM) pivot around to the "backside" of the enzyme and replace the prosegment β-strand that dissociates from mature gastricsin.

The structure of intermediate 2 reveals that the prosegment no longer obstructs the active site, as it does in the zymogen (Khan et al., 1997; Fig. 2B). Strikingly, a water molecule now resides between the catalytic Asp residues, characteristic of a mature active site with the nucleophilic water poised for catalysis. The N-terminal part of the prosegment (Ala1p to Gly21p) is in the same position and conformation as was observed in the zymogen. This region contains the β-strand (Val3p to Lys7p) and a helical region (Ile13p to Lys20p) that interacts via hydrogen bonds, hydrophobic interactions and salt bridges with the mature portion of the enzyme. Residues Leu22p to Phe26p are not visible in the electron density and are presumably disordered. The newly formed N-terminus of mature gastricsin (Ser1 to Ala8) has rotated away from the active site but has not taken its final position as part of the β-sheet at the back of the enzyme. This is because the prosegment β-strand presumably remains fixed during the first two cleavages, preventing the rearrangement until the latter stages of conversion. Consequently, residues Met7 to Asp11 are positioned adjacent to the active site in the S1–S3 binding pocket and interfere with substrate binding, thus rendering "intermediate 2" inactive.

Activation mechanism

Based upon the biochemical and structural data, a mechanism for acid activation of gastric aspartic proteinases has been proposed (Fig. 2E). Exposure of the zymogen to low pH results in protonation of the carboxylate side chains of Asp and Glu residues (Glick et al., 1989), thus destabilizing several salt-bridge interactions between the prosegment and the mature segment. The inhibitory salt bridges between Lys37p and the catalytic aspartate residues are disrupted, leading to conformational changes that uncover the active site. Lys37p is located on a well-ordered 3_{10}-helix in the zymogen (Pro34p to Arg39p), but the flanking regions (residues Thr29p to Asp33p and Phe40p to Leu43p) are highly mobile and have few interactions with the mature enzyme (Moore et al., 1995). These observations suggest that the disruption of salt bridges between Lys37p and the catalytic dyad, Asp32/Asp217, will destabilize the entire C-terminal region of the prosegment. In contrast to this 3_{10}-helix in hPGC, the prosegment β-strand (Val3p to Lys8p) is tethered to the mature enzyme by hydrogen bonds and hydrophobic interactions that are pH independent and would be relatively unaffected upon initial exposure of the zymogen to an acidic medium. While the prosegment β-strand remains fixed in the zymogen, the helical regions in the prosegment uncoil. All of these helical regions are structurally conserved in the known structures of zymogens (Fig. 2D). Disruption of the helical structure results in the uncovering of the preformed active site. Autocatalytic cleavages (both intra- and intermolecular) take place within the helical region and the pro-mature junction, resulting in removal of the C-terminal part of the prosegment (Leu27p to Leu43p). The final stages of the conversion process involve a dissociation of the prosegment β-strand and its replacement by the N-terminus of mature gastricsin (Fig. 2C).

The sites of internal cleavages in the prosegments and the kinetics of activation are variable among the gastric aspartic proteinases. The differences are partly explained by the positions of the cut sites that are recognized by the S1-S1′ specificity pocket, which is ideally a pair of bulky hydrophobic amino acids (Dunn, 1997; Fig. 2D). However, given their sequence and structural similarities, the prosegments appear to share the common features of an acid-labile C-terminal helical region that rapidly clears the active site to confer catalytic activity to precursors of the mature enzyme. The prosegment also contains a more stable N-terminal region composed of a β-strand that likely dissociates as the final step in the conversion process, permitting refolding of the N-terminus of the mature enzyme.

The biological implications of the conversion mechanism are not evident, but deserve some speculation. The structural characteristics of the zymogen ensure stability and inactivity following synthesis at neutral pH within the cell and subsequent transport to the stomach via digestive vacuoles. However, the activation pathway is able to be triggered in the stomach by the disruption of acid-labile salt bridges within the prosegment. Subsequent hydrolysis of the prosegment and its replacement by the N-terminus of the mature segment are likely to ensure that activation is irreversible, thus preventing the released prosegment peptides from acting as competitive inhibitors of the active enzyme.

3. *Cysteine proteinases and their zymogens*

Cysteine proteinases are present in both prokaryotic and eukaryotic organisms. These enzymes contains a Cys-His pair at the active site that is analogous to the Ser-His dyad in the serine proteinases. The catalytic mechanism involves nucleophilic attack on the scissile bond by the thiolate form of the Cys side chain (reviewed by Storer & Menard, 1994), thereby forming a covalently attached tetrahedral intermediate, as observed for the serine proteinases. The largest group of cysteine proteinases, as identified by sequence similarities, is the papain superfamily (EC 3.4.22; Berti & Storer, 1995) for which many crystal structures have been determined. The active enzymes of this superfamily share a common fold consisting of two domains: an N-terminal, mainly helical domain, and a C-terminal, predominantly β-sheet domain (Baker & Drenth, 1987). The catalytic Cys25 and His163 are contributed by the N- and C-terminal domains, respectively, and are located at the junction of the two domains, where a substrate binding cleft is formed (Fig. 3A,B). In mammals, the papain-like enzymes are components of the protein degradation machinery in lysosomes (Kirschke & Barrett, 1987).

The cathepsins B and L (catB, catL) are papain-like enzymes that are synthesized as zymogens (procatB, procatL) containing N-terminal inhibitory prosegments. Whereas the mature enzymes have about 25% sequence identity, the prosegments have limited sequence similarities between the B and L families. The prosegments range in length from 60 residues in procatB to about 100 in procatL. Aside from inhibiting the enzymatic activity, the prosegments play a role in the folding and stability of the enzyme during synthesis and transport at neutral pH (Tao et al., 1994). A region of the prosegment forms the recognition site for modification with mannose-6-phosphate that is necessary for subsequent targeting of the zymogen to lysosomes (McIntyre et al., 1994). In contrast to the aspartic proteinases, the intact prosegments of the cysteine proteinase zymogens are potent and specific inhibitors of the mature enzymes (Carmona et al., 1996).

Structures of inhibitory prosegments

The structures of human procatL and rat procatB/L have been determined by X-ray crystallography (Coulombe et al., 1996; Cygler et al., 1996; Turk et al., 1996; Fig. 3A,B). The structure of procaricain from *Carica papaya*, a papain-like cysteine proteinase, has also been determined and is structurally similar to the pro-

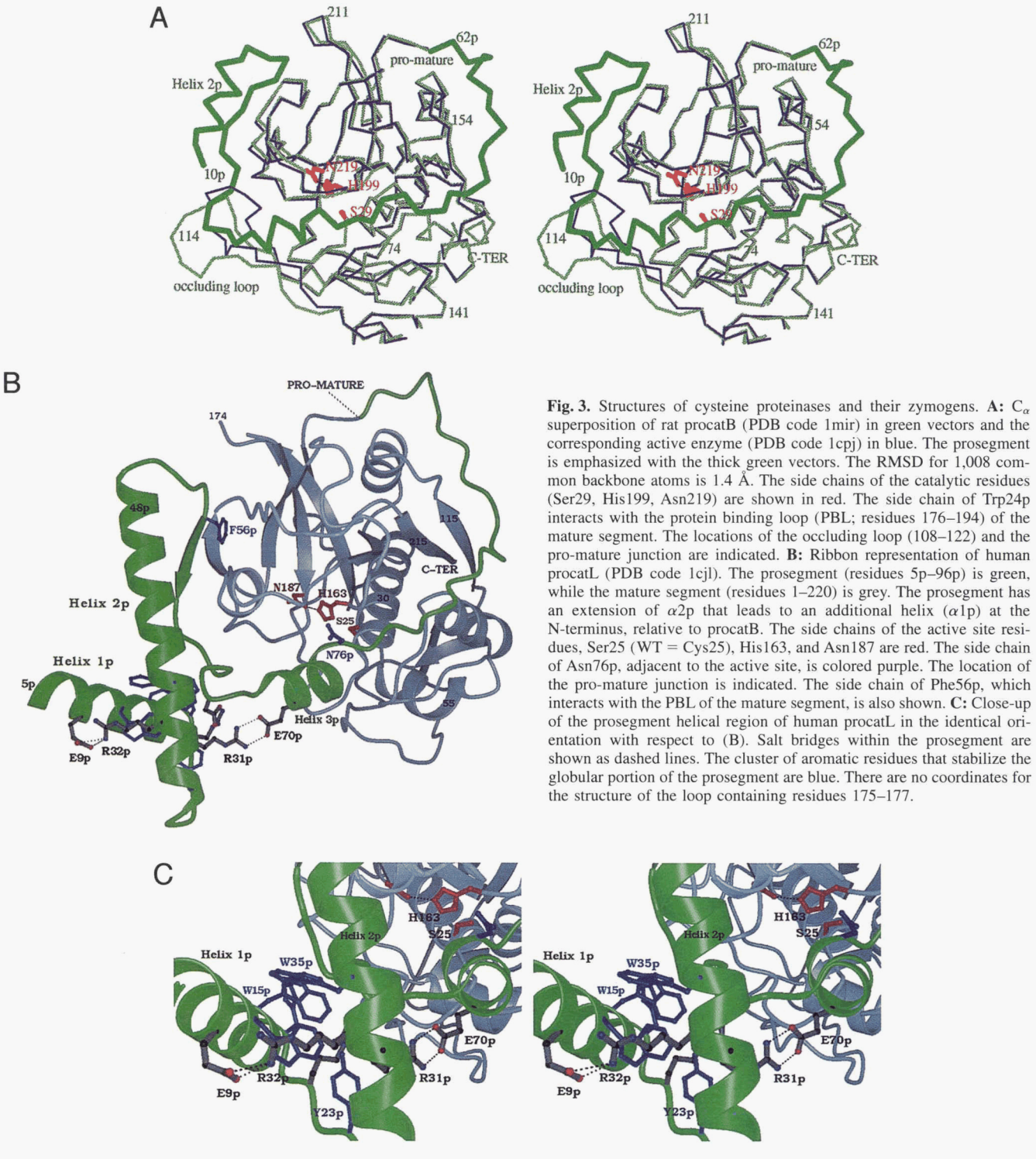

Fig. 3. Structures of cysteine proteinases and their zymogens. **A:** C_α superposition of rat procatB (PDB code 1mir) in green vectors and the corresponding active enzyme (PDB code 1cpj) in blue. The prosegment is emphasized with the thick green vectors. The RMSD for 1,008 common backbone atoms is 1.4 Å. The side chains of the catalytic residues (Ser29, His199, Asn219) are shown in red. The side chain of Trp24p interacts with the protein binding loop (PBL; residues 176–194) of the mature segment. The locations of the occluding loop (108–122) and the pro-mature junction are indicated. **B:** Ribbon representation of human procatL (PDB code 1cjl). The prosegment (residues 5p–96p) is green, while the mature segment (residues 1–220) is grey. The prosegment has an extension of α2p that leads to an additional helix (α1p) at the N-terminus, relative to procatB. The side chains of the active site residues, Ser25 (WT = Cys25), His163, and Asn187 are red. The side chain of Asn76p, adjacent to the active site, is colored purple. The location of the pro-mature junction is indicated. The side chain of Phe56p, which interacts with the PBL of the mature segment, is also shown. **C:** Close-up of the prosegment helical region of human procatL in the identical orientation with respect to (B). Salt bridges within the prosegment are shown as dashed lines. The cluster of aromatic residues that stabilize the globular portion of the prosegment are blue. There are no coordinates for the structure of the loop containing residues 175–177.

cathepsins (Groves et al., 1996). Recombinant human procatL was mutated at the active site cysteine (Cys25 → Ser) in order to prevent autocatalytic conversion of the zymogen to the mature enzyme, thereby facilitating crystallization. Several other mutations relative to the wild-type enzyme were also introduced into the molecule, including a Thr110 → Ala substitution that abolished a site of O-linked glycosylation (Coulombe et al., 1996).

The 96-residue prosegment of human procatL has an N-terminal globular domain (1p to 75p) that contains three α-helices (α1p–α3p) and their connecting loops (Fig. 3B). This domain is followed by a segment with an extended structure (76p to 96p) that passes across the substrate binding cleft. The direction of the polypeptide across the binding cleft is in the reverse orientation (N → C-terminal) with respect to substrates that are cleaved. The globular domain is

stabilized by extensive interactions among the α-helices that include several salt bridges and a remarkable hydrophobic core composed of interdigitating aromatic residues (Trp12p, Trp15p, Trp35p, His19p, Tyr23p; Fig. 3C) and a pair of interacting methionine residues (Met39p and Met60p).

Interactions between the prosegment and the mature enzyme are grouped together in two distinct regions. The first set of interactions are between the α2p–α3p loop and the "prosegment binding loop" (PBL) of the mature enzyme (residues His140 to Asp155). Within this interface, a short two-stranded anti-parallel β-sheet is formed (Phe56p to Met60p and Phe147 to Ile150). A cluster of aromatic residues are contributed from the β-sheet (Phe56p, Tyr146, Tyr151) and extend toward the active site via the side chains of Phe63p and Phe71p, as well as several aromatic residues in the mature segment. The second group of interactions take place between the C-terminal region of the prosegment (Val74p to Gln79p) and the substrate binding cleft. By analogy to papain, the S1 to S3 subsites are occupied by the C-terminal residues Gly77p to Gln79p, in the opposite orientation that would be expected for a genuine substrate. Gly77p brings the prosegment deep inside the catalytic cleft, such that part of the oxyanion hole (the amide proton of Ser25) forms a hydrogen bond to Asn76p(O). The S′ side of the substrate-binding cleft is occupied by residues from the end of α3p, including Phe71p, Met75p, and Asn76p.

The folds of the prosegments of procathepsins B and L are similar, despite a 30-residue deletion from the N-terminus of procatB (Fig. 3A,B). The prosegment of procatB is missing α1p and the N-terminal portion of α2p. In addition, the orientations of helices α2p and α3p are different as a consequence of the "occluding loop" (residues 108–122; Fig. 3A) that is found only in the cathepsin B family. This loop imparts the ability to cleave dipeptide units from the C-terminus of substrates (Kirschke & Barrett, 1987), possibly as a result of stabilizing interactions between the occluding loop and the negative charge at the P2′ carboxylate of the substrate (Cygler et al., 1996). Aside from these differences, the overall mode of inhibition by the prosegments of procatB and procatL is preserved, suggesting a common structural motif for all papain-like zymogens. The common features are the interactions between the PBL and the prosegment, including a conserved aromatic residue in the prosegment (Phe56p in procatL, Trp24p in procatB) that mediates prosegment-PBL contacts within this interface. In addition, both zymogens have a Gly residue (Gly77p, Gly43p) that is nearest to the catalytic Cys residue. The small size of this residue allows the C-terminal regions of the prosegment to push deep into the catalytic cleft and interact closely with the substrate binding sites *via* hydrogen bonds and hydrophobic interactions.

Activation mechanisms

The zymogens are targeted to the low-pH environment of the lysosomes where they are converted to the active enzymes by limited proteolysis and removal of the prosegment by an autocatalytic mechanism. This property of the conversion mechanism is shared by the aspartic proteinase zymogens. Biochemical and kinetic studies of the activation process of the related zymogen, propapain, have shown that the optimal pH for activation is 3.3, and that the initial cleavage reaction is intramolecular (Vernet et al., 1991). Based upon alignments of procathepsin L sequences, the pH dependence of the conversion process may be regulated by the conserved salt bridges between Asp65p and Arg21p, as well as between Glu70p and Arg31p within the prosegment. Mutagenesis studies have also confirmed the essential role of the equivalent Asp65p residue of papain in folding of the prosegment (Vernet et al., 1995).

Disruption of the salt bridges by protonation of the carboxylate groups at the lower pH could conceivably trigger the disruption of the hydrophobic core of the prosegment, leading to dissociation of the prosegment from the active site, and thereby initiating the process of autocatalytic conversion. The pro-mature junction resides on an exposed loop that would be accessible to proteolytic cleavage. The segment preceding Asp65p of the prosegment is a four-residue motif (**G/A-X-N-X-F-X-D**^{65p}) that is conserved in the prosegments of many papain-like zymogens. Mutations within this region of propapain are observed to alter the pH dependence of the intramolecular cleavage reaction (Vernet et al., 1995). For example, the mutant Phe63p → His, which is predicted to confer a more positively charged character to this region, increases the pH optimum for the activation process to 4.3. However, these several salt bridges are not conserved in procatB, suggesting that alternative salt-bridge interactions are responsible for initiating the pH-dependent conversion in that zymogen (Coulombe et al., 1996).

The structure of rat procatB has been compared directly to the structure of mature rat cathepsin B, which was also determined by X-ray crystallography (Jia et al., 1995). The mature segments in both structures are virually identical (Fig. 3A). The active site machinery, including the oxyanion hole, is preformed and does not undergo conformational changes upon conversion of the zymogen. The only differences are found at the occluding loop, between residues 105 to 125. In the zymogen, a part of the prosegment (between β1p and α2p) is wedged into a crevice that is formed between the occluding loop and the PBL. However, with the departure of the prosegment in the mature enzyme, the occluding loop undergoes conformational changes that result in displacements of atoms by up to 14 Å. The loop shifts toward the active site with a disulfide bridge (Cys108–Cys119) acting as a pivot and is now positioned to interact with substrates. In contrast, human procatL and other papain-like cysteine proteinases do not generally have occluding loops. Therefore, the mature segment of procatL is not expected to undergo any significant conformational changes during conversion (Cygler et al., 1996).

The prosegments of cysteine proteinase zymogens may have functions beyond the inhibition of the enzyme. A nine-residue, positively charged region of the prosegment of mouse procatL mediates interactions of the zymogen with membranes at pH 5 (McIntyre et al., 1994). This region of the prosegment also has similarities with yeast vacuolar targeting sequences (Klionsky et al., 1988; Valls et al., 1990). The equivalent region of human procatL is localized to the loop connecting α1p and α2p (Lys16p to Gly24p). In addition, polyanions such as dextran sulfate accelerate the rate of conversion of procatL to the active enzyme *in vitro* (Mason & Massey, 1992). These observations suggest that membrane interactions, mediated by the prosegment, may be important during lysosomal targeting and activation of procathepsin L.

4. Zinc metalloproteinases and their zymogens

The Zn^{2+}-carboxypeptidase family (EC 3.4.17.-) contains a Zn^{2+} ion at the active center of the enzyme that is directly involved in the catalytic mechanism. These enzymes cleave peptide bonds at the C-termini of polypeptide substrates. The pancreatic carboxypeptidases are digestive enzymes that degrade proteins in the

alimentary tract of mammals (Puigserver et al., 1986). The three-dimensional structures of the pancreatic enzymes have been determined by Lipscomb and co-workers (Quiocho & Lipscomb, 1971; Rees et al., 1983; Christianson & Lipscomb, 1989). The structure of carboxypeptidase A consists of a central twisted parallel/anti-parallel (mixed) β-sheet, composed of eight strands, that is flanked on both sides by α-helices (Fig. 4A,B). The active site cleft is formed in a shallow groove on one side and is bounded by a strand of the β-sheet, two α-helices and a loop that partially covers the cleft. The Zn^{2+} ion is coordinated by three protein residues (His69, His196, Glu72) that are contributed by turns and loops which connect the secondary structures. The mechanism of peptide hydrolysis involves the activation of a water molecule by Glu270 (procarboxypeptidase A numbering) and subsequent nucleophilic attack of the scissile bond by the hydroxide (Matthews, 1988; Hanson et al., 1989). Positively charged residues, including the Zn^{2+} ion itself, assist the hydrolytic reaction by neutralizing the developing negative charge of the tetrahedral intermediate, analogous to the oxyanion hole of the serine and cysteine proteinases.

Structures of pancreatic zymogens

The first structures of the Zn^{2+}-carboxypeptidase zymogens, porcine pancreatic procarboxypeptidases A and B (pro-CPA and pro-CPB), were determined by Huber and colleagues using X-ray crystallography (Coll et al., 1991; Guasch et al., 1992; Fig. 4A,B). The mature enzymes CPA and CPB are distinguished by their substrate preferences, with CPA cleaving aliphatic residues and CPB cleaving basic residues at the P1′ position (Folk, 1956; Neurath, 1959). The zymogens are synthesized with N-terminal prosegments of about 95 residues sharing 26% sequence identity. The prosegments have identical folds, superimposing with a root-mean-square (RMS) deviation of 0.29 Å for the conserved secondary structure elements (68 C_α atoms). The high degree of structural similarity is remarkable considering the low sequence identities in the prosegment, suggesting a strong selective pressure during evolution for preserving the mode of inhibition.

The prosegments consist of a globular domain that contains two α-helices stacked against a four-stranded anti-parallel β-sheet, with the helical axes approximately parallel to the strands of the sheet (Fig. 4A,B). This domain is followed by a connecting segment that leads to the mature enzyme. The overall domain forms a two-layer structure with an open-faced β-sheet sandwich antiparallel-α/antiparallel-β topology. One side of the β-sheet faces the mature enzyme, while the opposite face packs against the two α-helices. The residues within the substrate binding cleft and the catalytic Zn^{2+} ion of the zymogen are in the same positions and have the same conformations as in the mature enzyme. However, the globular prosegment blocks the preformed catalytic cleft, thereby preventing access of substrates to the active site. The most extensive shielding occurs at the S2 to S4 substrate-binding sites in the enzyme from the strands β2p/β3p and their connecting loop. Interactions between the prosegment and the mature enzyme, many of them water-mediated, involved residues Asp36p and Trp38p of the prosegment (pro-CPA numbering) that are found at the end of β2p (Fig. 4C). The residues Asp36p and Trp38p cover subsites S2 and S3 and are conserved in zymogens from several species (Aviles et al., 1993).

The third helix of the prosegment (α3p) marks the start of the connecting region that links the globular domain to the mature enzyme. A second set of interactions involving α3p (85p–91p) and several residues in the mature enzyme portion are observed here. These interactions involve hydrogen bonding and hydrophobic contacts. The prosegment helix α3p is followed by a loop that contains the pro-mature junction (Arg99p-Thr1). This peptide bond is the site of trypsin cleavage during conversion of the zymogen.

Activation mechanism

Comparisons of the porcine zymogens with the structure of the mature enzyme from cattle have been discussed previously in a review (Aviles et al., 1993). Activation is initiated by trypsin cleavage at the pro-mature junction. However, the kinetics of the activation process vary between pro-CPA and pro-CPB, and are also dependent upon environmental conditions such as pH, ionic strength and quaternary structures (Vendrell et al., 1990; Burgos et al., 1991; see Multi-molecular assemblies below). The pro-mature junction in pro-CPA (Arg99p-Ala1) is localized to a flexible loop (high *B*-factors) that may facilitate recognition and cleavage by trypsin. Cleavage at this loop may destabilize the preceding helix, α3p, thus promoting release of the prosegment from its C-terminus (Guasch et al., 1992; Aviles et al., 1993). Comparisons between the zymogens and mature enzymes show that the mature segment does not undergo significant conformational changes during conversion (Fig. 4A).

The prosegment is further degraded by trypsin at internal sites, and the newly liberated mature enzyme is also involved in the process by degrading the prosegment from the C-termini (Wintersberger et al., 1962; Vendrell et al., 1990; Burgos et al., 1991). The internal cleavages occur at sites that are not accessible in the zymogen (pro-CPA, Arg74p-Tyr75p; pro-CPB, Arg83p-Ser84p), suggesting that structural changes must take place within the prosegment during conversion. In contrast to pro-CPB, the prosegment of pro-CPA is released and degraded more slowly, and its ability to inhibit mature CPA may further slow the conversion process. The molecular basis for these differences has been attributed to stronger interactions between α3p of the prosegment and the mature enzyme in pro-CPA (Aviles et al., 1993).

Pro-stromelysin-1

Pro-stromelysin-1 (EC 3.4.24.17) is a Zn^{2+}-endopeptidase and a member of the family of matrix metalloproteinases (MMPs). The MMPs function at neutral pH and are components of the tissue remodeling machinery due to their ability to degrade connective tissue. As a consequence of their link to pathological conditions such as arthritis and tumor invasion, these enzymes are attractive candidates for structure-based drug design (Murphy et al., 1991; Docherty et al., 1992). Pro-stromelysin-1 is synthesized with an inhibitory N-terminal prosegment of 82 residues. In addition to the catalytic domain, pro-stromelysin-1 also contains a C-terminal domain that is believed to mediate interactions with protein-proteinase inhibitors and macromolecular substrates (Becker et al., 1995).

The structure of the catalytic domain of human pro-stromelysin has been determined by X-ray crystallography to 1.9 Å resolution (Becker et al., 1995; Fig. 4D). The structure of the mature segment of pro-stromelysin-1 bears no sequence or structural resemblance to the pancreatic metalloenzymes apart from the presence of a catalytic Zn^{2+} at the active site. The overall fold consists of a five-stranded mixed β-sheet composed of one anti-parallel and four parallel β-strands, as well as three α-helices stacked against one face of the β-sheet (Becker et al., 1995). The active site cleft

resides on the opposite side in a shallow groove. In addition to the catalytic Zn^{2+} that is coordinated by three His side chains in the active enzyme, there is also a structural Zn^{2+} in the molecule (not shown).

The structure of the prosegment (1p–82p) consists of an N-terminal globular domain that is composed of three α-helices (α1p to α3p; residues 16p–70p) followed by the mature segment (residues 1–168; Fig. 4D). The first 15 residues in the prosegment, as well as the segment 31p–39p, are not visible in the electron density. The C-terminal part of the prosegment forms a short β-strand (73p–76p) that interacts with a partner strand (residues 139–141) in the substrate-binding cleft. In contrast to the Zn^{2+}-pro-

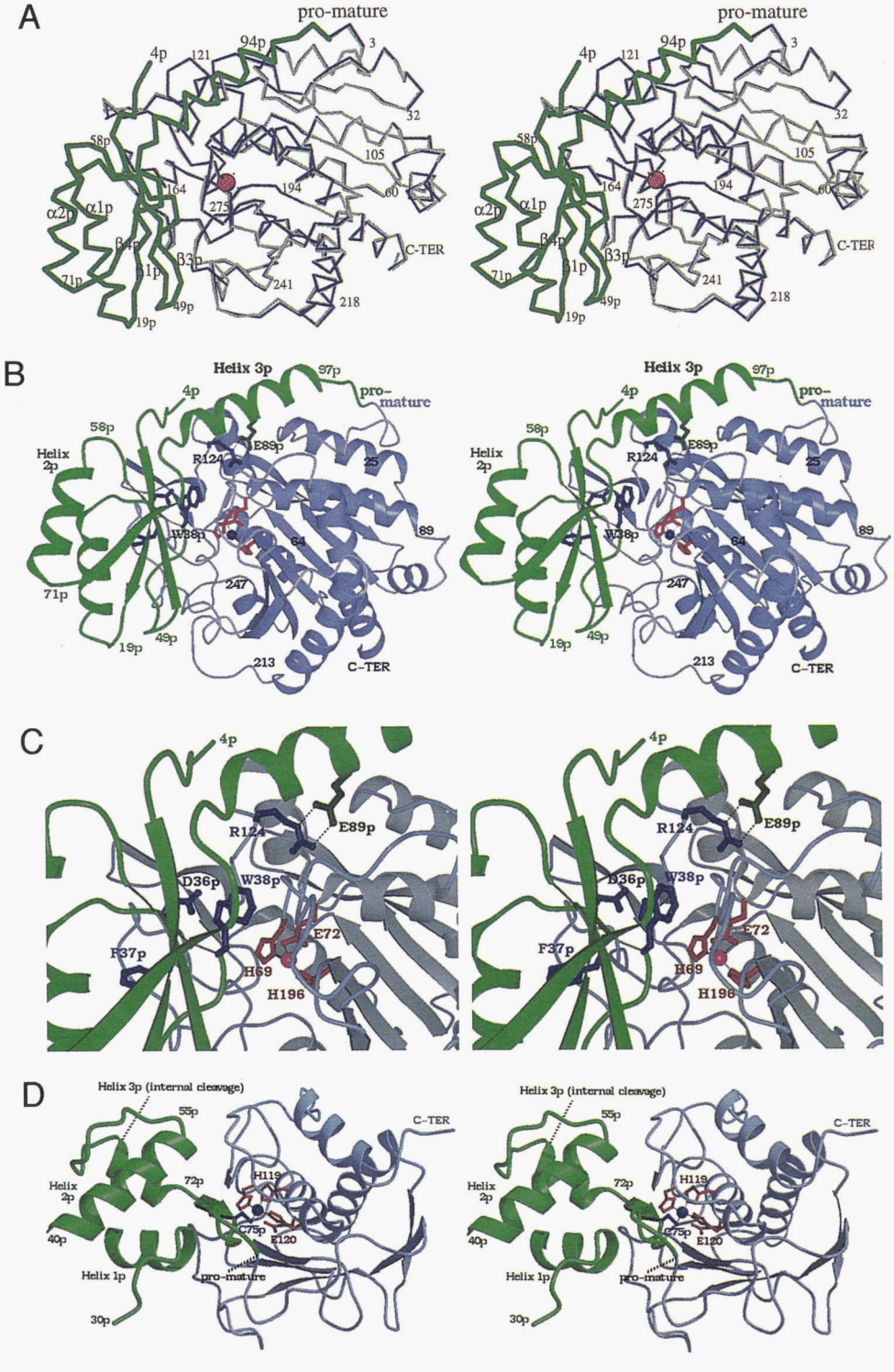

Fig. 4. *See caption on facing page.*

carboxypeptidases, the catalytic Zn^{2+} ion is directly coordinated by a residue from the prosegment in pro-stromelysin-1. The sulfur atom from the side chain of Cys75p in the β-strand coordinates the Zn^{2+} ion, stabilizing the interactions in this region as well as preventing catalytic activity (Fig. 4D). Strikingly, the direction of the prosegment at the active site is in the reverse direction (N → C termini) relative to genuine substrates. This latter feature is reminiscent of the cysteine proteinase zymogens.

Conversion of the zymogen to the active enzyme involves limited proteolysis at an exposed loop that contains the pro-mature junction (His82p-Phe1), and cleavage sites within the prosegment (Glu68–Val69, helix 3p) have also been identified (Nagase et al., 1991; Fig. 4D). Multiple mechanisms are capable of initiating the cleavage events, including other proteolytic enzymes. Heat and mercurial agents (e.g., 4-amino-phenylmercuric acetate) are also able to trigger the conversion process via an auto-proteolytic pathway (Okada et al., 1988; Okada & Nakanashi, 1989; Nagase et al., 1990; Koklitis et al., 1991). However, the positions and conformations of the active site residues are unchanged during conversion. The only significant conformational changes take place in the loop that contains the pro-mature junction. Following cleavage, the loop undergoes a conformational rearrangement that results in salt bridge formation between the newly formed N-terminus at Phe1 and Asp237 in the mature segment. This rearrangement resembles the molecular events involving the N-terminus of serine proteinases during their conversion process. However, the salt bridge is 12 Å away from the catalytic Zn^{2+} ion and is not expected to affect the active site significantly. In contrast, ion-pair formation is critical for the formation of the oxyanion hole in the serine proteinases. The salt bridge in pro-stromelysin-1 may function to swing the loop away from the substrate binding cleft, thus indirectly enhancing catalytic activity (Reinemer et al., 1994).

The structure of pro-stromelysin-1 is an excellent example of how the same molecular mechanisms are utilized in different structural contexts. In pro-stromelysin-1, aspects of cysteine proteinase zymogens (a reverse-orientation of the prosegment in the substrate binding cleft) and serine proteinase zymogens (a salt bridge at the newly formed N-terminus) are utilized for inhibition and during the conversion process. The detailed structural events taking place during conversion of pro-stromelysin to the mature enzyme have not been well characterized. However, the sites of internal cleavages in the prosegment indicate that conformational changes must accompany conversion, since several of these sites are inaccessible in the zymogen.

5. Multi-molecular assemblies

Zymogen conversion frequently occurs in the context of homo- or hetero-oligomeric states of the enzyme under physiological conditions. One role ascribed to zymogens is the stabilization of other proteins during cellular transport (Rudenko et al., 1995; Stevens et al., 1996). Zymogen conversion in oligomeric states may also be exploited as a regulatory mechanism for physiological processes. For example, the C1 component of complement is a multi-molecular complex that contains a tetrameric arrangement of serine proteinase zymogens, C1s-C1r-C1r-C1s, that are associated with the stalk region of the lollipop-shaped C1q component. The activation process is initiated when the globular head of C1q attaches to antibody. A conformational signal is transmitted through its stalk region, thereby triggering the autocatalytic conversion of C1r (inactive) → C1*r* (active), followed by cleavage of C1s (inactive) → C1*s* (active) by C1*r* (Liszewski et al., 1996). The newly-activated serine proteinase, C1*s*, proceeds to cleave other serine proteinase zymogens during the complement cascade.

The conversion process in oligomeric assemblies will be affected if the sites of limited proteolysis during activation are inaccessible (Neurath & Walsh, 1976). One example is the structure of a ternary complex of two serine proteinase zymogens, bovine chymotrypsinogen C (CTGN-C) and pro-proteinase E (pro-PE), with the metalloenzyme zymogen procarboxypeptidase A (pro-CPA; Gomes-Ruth et al., 1997). The overall complex is arranged with pro-CPA at the base of a Y-shaped structure, flanked by the two serine proteinase zymogens (Fig. 5). The serine proteinase zymogens do not interact with each other. Most of the interactions between pro-CPA/bCGN-C and pro-CPA/pro-PE are mediated by the 99-residue prosegment of pro-CPA, revealing another physiological role of its prosegment.

The structure of the prosegment of bovine pro-CPA, as part of the ternary complex, is similar to that described previously for porcine pancreatic pro-CPA. However, the trypsin-sensitive cleavage site (Arg99p-Ala1) in the tenary complex, which would be accessible in the monomer, is partially covered in the complex because of interactions with chymotrypsinogen C (Fig. 5). The internal cleavage site, Arg74p-Tyr75p (β4p), is even more inac-

Fig. 4. (*See figure on facing page.*) Structures of metalloproteinases and their zymogens. **A:** Superposition of porcine procarboxypeptidase A (green, PDB code 1pca) and bovine carboxypeptidase A (blue, PDB code 2ctb), drawn as C_α vectors. The RMSD for 1,211 common backbone atoms was 0.39 Å. The prosegment (4p–99p) is emphasized in thick lines. The location of the pro-mature junction is indicated by "N-TER" and its position remains unchanged during conversion. The position of the catalytic Zn^{2+} is shown as a red sphere; it is coordinated by the side chains of His69, Glu72, and His196. **B:** Stereo ribbon model of porcine pro-CPA in an identical orientation relative to (A). The prosegment is green and the mature segment is grey; the side chains of Glu89p and Arg124 form a salt bridge. The side chains of D36p, Phe37p, and Trp38p are purple. The catalytic Zn^{2+} (sphere) and its coordinating ligands (red) are shown. **C:** Close-up of the ribbon figure drawn in (B). The catalytic Zn^{2+} ion (sphere) is coordinated by the side chains of His69, Glu72, and His196, shown in red. The general acid-base residue Glu270 (not shown), which would reside below His196 in the figure, does not interact with the prosegment. **D:** Stereo ribbon model of pro-stromelysin-1 (PDB code 1slm). The prosegment is green (16p–82p) and the mature segment is grey (residues 1–168). The first 15 residues of the prosegment and the linker between α1p and α2p (residues 31p–39p) are disordered in the structure. The catalytic Zn^{2+} (sphere) is coordinated by the side chains of His119, His123, and His129. Cys75p from the β-strand of the prosegment provides the fourth ligand in the zymogen. The side chain of the general acid base catalyst, Glu120, is situated 3.7 Å away from the S_γ atom of Cys75p. A site of autocatalytic cleavage takes place within the second turn of α3p (Glu68p–Val69p), which is inaccessible in the zymogen. The cleavage event is intramolecular and occurs following treatment with 4-aminophenylmercuric acetate (Nagase et al., 1991). [Note that the mature segment numbering has been modified in order to remain consistent with the remainder of this review. In the literature, the residue numbers increase continuously through the pro-mature junction, such that the three His ligands are numbered His201, His205, and His211.]

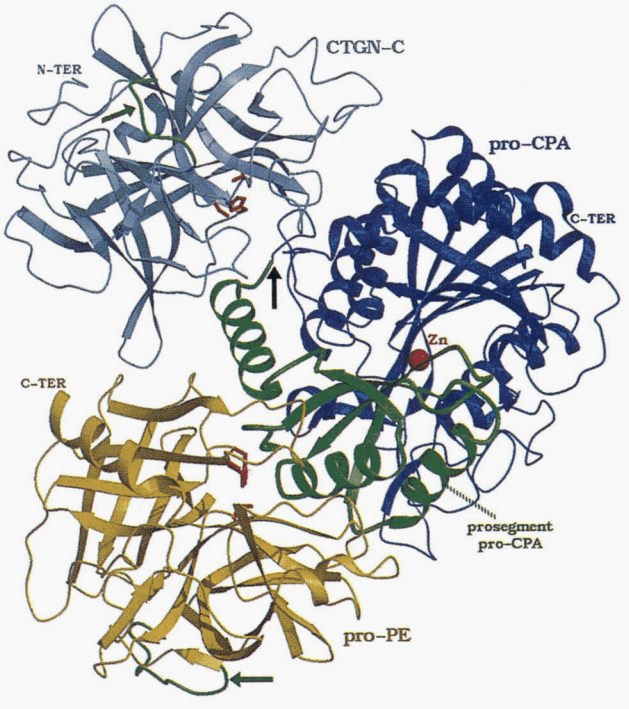

Fig. 5. Ribbon model of the bovine ternary complex (PDB code 1pyt) of procarboxypeptidase A (pro-CPA, blue), chymotrypsinogen C (CTGN-C, grey), and pro-proteinase E (pro-PE, yellow). The prosegment of pro-CPA is colored green and is responsible for the stabilization of the ternary complex. The catalytic Zn^{2+} ion of pro-CPA is represented as a red sphere. The pro-mature junction in pro-CPA is indicated by the black arrow and is partially obscured by interactions with CTGN-C. The activation loops of the two serine proteinases are green (green arrows) and are exposed in the complex. The side chains of their catalytic residues are colored red.

cessible, suggesting that one reason for the formation of an oligomeric complex is to stabilize and protect the prosegment from the destructive action of proteolytic enzymes during passage of the complex through the duodenum to the pancreas (Uren & Neurath, 1972; Gomes-Ruth et al., 1997). Another suggestion is that the oligomeric complexes allow coordination and timing for the onset of catalytic activities. In contrast, the activation sites for both serine proteinase zymogens are accessible to tryptic cleavage. A two-step activation scheme has been proposed in which the chymotrypsin activity is expressed first, followed by dissociation of the complex and subsequent conversion of pro-CPA (Kerfelec et al., 1985). However, the mechanism of dissociation of the complex, as well as the physiological importance of sequential activation, remains unknown. It is nevertheless interesting to note that procarboxypeptidases are often stored and secreted as a ternary complex in ruminant species (e.g., cattle, goats, camels) but are found as monomers or binary complexes in nonruminant species (Kerfelec et al., 1985; Burgos et al., 1989; Pascual et al., 1990).

Another example of the importance of multi-subunit interactions is the activation of the catalytic β-subunits of the yeast proteasome. The proteasome is a member of the N-terminal nucleophile (Ntn) hydrolase family in which the N-terminal residue (Thr1 in the yeast proteasome) initiates nucleophilic attack, *via* its side-chain O_γ atom, and subsequently cleaves peptide substrates (Brannigan et al., 1995; Groll et al., 1997). The β-subunits are synthesized with an inhibitory N-terminal prosegment that is also essential for the formation of the complex *via* inter-subunit contacts during assembly (Seemuller et al., 1996; Groll et al., 1997). Seven β-subunits together with seven α-subunits form one-half of the 28-subunit cylinder that is responsible for the regulated degradation of proteins in the cytosol and nucleus. However, in this case the acquisition of catalytic activity is coupled to the formation of the complex (Chen & Hochstrasser, 1996; Seemuller et al., 1996). This conversion mechanism prevents the hazardous effects of proteolysis in the cytosol until the active site is safely sequestered within the proteolytic chamber of the proteasome.

The conversion process for the yeast proteasome is autocatalytic and involves limited proteolysis and removal of the N-terminal prosegments of the catalytic β-subunits. Surprisingly, the structure of the active proteasome has revealed that two of the five β-subunits that are synthesized with a prosegment are not fully processed and are therefore inactive. These observations suggest that conversion involves stepwise proteolytic cleavages, and that the two inactive β-subunits may represent intermediates during the activation pathway that fail to complete the conversion process (Groll et al., 1997).

Discussion

General properties of activation mechanisms

Most of the zymogen structures discussed in this review possess an active site structure that is virtually indistinguishable from their mature counterparts. The two notable exceptions are the chymotrypsin-like serine proteinases, which require a backbone conformational change at the substrate-binding cleft, and the "Ntn hydrolase" family. A recurring theme for many zymogens is the presence of a preformed active site that is rendered sterically inaccessible to substrates by a competitive inhibitory prosegment (Bode & Huber, 1992; Sohl et al., 1997). However, the prosegment-active site interactions are generally loose and poised for disruption, unlike small molecule or peptide inhibitors of the active enzymes that bind tightly and often mimic the cleavage transition state. The prosegment of procathepsin L has high backbone *B*-factors on the C-terminal side of the stretch of residues (Met75p-Gln79p) that form contacts with the substrate-binding cleft (Fig. 6A). The prosegments of gastric aspartic proteinases and pro-stromelysin are well ordered within segments that contact the active site, but the flanking regions have high *B*-factors that likely contribute to the disruption of the prosegment-active site interactions (Fig. 6B,C). In the structure of pro-carboxypeptidase A, the prosegment interacts with the S2 to S4 portion of the substrate-binding cleft. The substrate-binding residue Arg127 (Kim & Lipscomb, 1991) and the general acid-base catalyst, Glu270, do not directly contact the prosegment. Finally, a general feature of the inhibitory interactions is the strategic positioning of prosegments to prevent self-cleavage prior to the appropriate signal. For example, the prosegments of papain-like cysteine proteinases extend across the active site in the reverse N $\rightarrow$ C-terminal orientation that is necessary for productive cleavage. The prosegments of aspartic proteinases interact with the active site via a 3_{10}-helix, distinct from the extended conformations of substrate-like peptides that allow intimate contacts with the narrow catalytic cleft.

Table 2 summarizes the properties of zymogen structures and the mechanisms of activation. The mature segments of zymogens may be subjected to large conformational changes during conversion, or they may remain unchanged. The gastric aspartic protein-

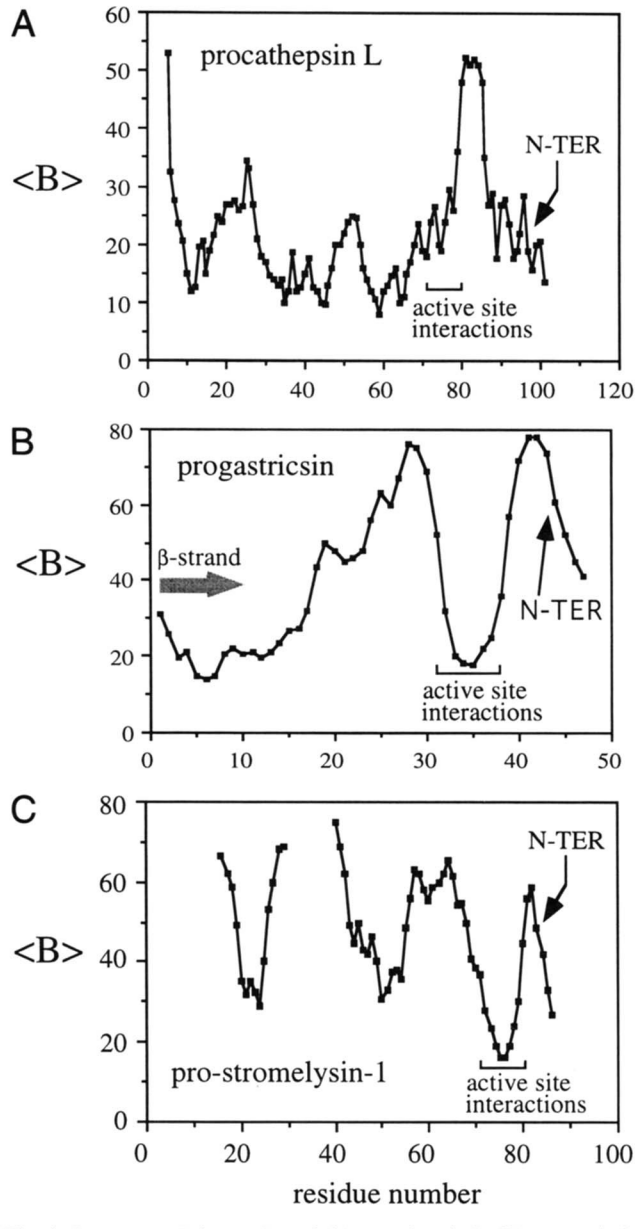

Fig. 6. Prosegment *B*-factor plots of (**A**) procathepsin L, (**B**) progastricsin, and (**C**) pro-stromelysin-1. The average backbone *B*-factor, $\langle B \rangle$, is plotted against the prosegment residue number. The regions of the prosegment that interact with the active site cleft are indicated. The beginning of the mature segment is labeled "N-TER."

ases and subtilisin undergo a major rearrangement of the first ten residues at the mature N-terminus following limited proteolysis at their pro-mature junction (James & Sielecki, 1986; Gallagher et al., 1995; Khan et al., 1997). A common feature of zymogens that are activated in low pH compartments is the critical role of salt bridges in maintaining the stability of the prosegment and its interactions with the mature segment. These interactions may be essential for the stability of zymogens in a neutral pH environment. For example, the active forms of gastric aspartic proteinases are enriched in negatively charged amino acids and become denatured upon incubation at neutral pH, presumably as a consequence of electrostatic repulsions between their ionized side chains. In the zymogens, the prosegments are positively charged and form several salt bridges with the mature segment, thus protecting the enzyme during its passage through a cellular environment.

The prosegments of bacterial and yeast zymogens play additional roles in folding and/or targeting. When added as separate components, the prosegments of α-lytic protease and subtilisin are able to refold the denatured enzymes (Baker et al., 1992; Strausberg et al., 1993). The noncovalent complex of subtilisin with its prosegment suggests that the prosegment may catalyze the folding reaction by stabilizing a central "$\alpha\beta\alpha$" structure in subtilisin (Bryan et al., 1995). Studies relating to the folding and conversion of subtilisin, as well as future structural characterization of α-lytic pro-protease, are likely to have wider implications in the field of protein folding.

Upon transport of zymogens to their appropriate compartments, prosegments are removed by limited proteolysis, often by an autocatalytic mechanism (e.g., zymogen forms of subtilisin, α-lytic protease, gastric aspartic proteinases, cysteine proteinases, yeast proteasome, pro-stromelysin-1). Autocatalytic conversion is perhaps the most economical mechanism since no other enzymatic cofactors are necessary for activation. In this case, conversion typically requires a conformational change in the prosegment in order to uncover the active site and thereby generate active forms of the zymogen. Limited proteolysis and removal of the prosegment may involve unimolecular and/or bimolecular cleavage events. Interestingly, the mechanism of release of the cysteine proteinase from the polyprotein precursor of picornaviruses has been reported to involve intramolecular cleavage (Palmenberg & Rueckert, 1982). The detailed kinetics of the viral polyprotein conversion pathways are unknown, but the autoproteolytic sequence of cleavages may be regulated to release the desired proteins at the appropriate time in the infectious cycle (Palmenberg, 1990).

Whether the mechanism is autocatalytic or dependent upon other enzymes, the conversion process frequently proceeds through stepwise proteolytic cleavages to remove the entire prosegment (e.g., activation pathways of gastric aspartic proteinases, papain-like cysteine proteinases, yeast proteasome, yeast carboxypeptidase Y, human protective protein, matrix metalloproteinases). The structures of activation intermediate 2 of human gastricsin and the partially processed β-subunits of the yeast proteasome have provided insights into their respective activation pathways. Future biochemical and structural studies of activation intermediates will help elucidate the activation pathways of other zymogens. Following conversion, the discarded prosegments are typically degraded, often by the newly activated enzyme. Digestion of the prosegment of subtilisin is rationalized as a means of recycling the amino acids for subsequent use by the bacterium (Gallagher et al., 1995). It can further be reasoned that the hydrolysis of prosegments ensures that the conversion process is irreversible and that the prosegments do not act as competitive inhibitors of the active enzymes.

Regulation of proteolytic activity and future prospects

Following the conversion process, there exists an elaborate collection of specialized proteins that specifically inhibit the active enzymes. The importance of these "protein-proteinase" inhibitors range from attenuation of the serine proteinase cascades (blood coagulation, complement activation) to a role in the defense mechanism against foreign pathogens. The protein-proteinase inhibitors of serine proteinases have been discussed in the section "Serine proteinases and their zymogens." The papain-like cysteine protein-

Table 2. *Inhibitory interactions and activation mechanisms*

Enzyme class	Mature active site machinery?	Inactivation mechanism	Additional roles of the prosegments	Conversion process
Serine proteinase	No (chymotrypsin-like enzymes) Yes (human protective pro-protein) Yes (subtilisin)	1. Chymotrypsin-like-immature substrate-binding cleft 2. Subtilisin-77-residue prosegment that sterically blocks the active site –Prosegment is structurally similar to prosegments of pancreatic metalloproteinases, despite lack of sequence similarities	*Bacterial/yeast* –α-lytic protease and yeast procarboxypeptidase Y-folding and targeting –Subtilisin-essential for the folding reaction	1. Chymotrypsin-like –Limited proteolysis of a dipeptide and formation of salt bridge with new N-terminus –Conformational change at the substrate-binding cleft 2. Subtilisin-limited proteolysis of prosegment, rearrangement at the N-terminus
Aspartic proteinase	Yes	–Steric block of the active site by the ~45 residue prosegment –Salt bridges with active-site Asp residues	–Folding?/stabilization during transport to stomach (gastric enzymes) –Yeast-folding and targeting to vacuoles	–Disruption of salt bridges and proteolysis/removal of ~45 residue prosegment, refolding of the mature N-terminus
Cysteine proteinase (papain-like)	Yes	–Steric block of the active site by the 60/100 residue prosegment –Reverse oriented prosegment peptide across active site, relative to substrates	–Folding/stability during transport to lysosomes –Mediate modification by mannose-6-phosphate –Nine-residue segment mediates membrane interactions	–Disruption of salt bridges and limited proteolysis of the prosegment –Mature segment does not undergo significant conformational changes
Metalloproteinase	Yes	–Steric block of the active site by ~95 residue prosegment –Pro-stromelysin-1, steric block by 82-residue prosegment, which passes through active site in reverse sense relative to genuine substrates	–Bovine ternary complex-prosegment mediates assembly of the oligomeric complex	–Limited proteolysis of prosegment by trypsin –Mature segment does not undergo significant conformational changes

ases are strongly but reversibly inhibited by the cystatins, a superfamily of homologous proteins that includes chicken cystatin as the archetypal member (Bode et al., 1988). Small protein-proteinase inhibitors of metalloproteinases are also known (Rees & Lipscomb, 1982; Birkedal-Hansen et al., 1990), such as the 39-residue potato carboxypeptidase inhibitor that may be important for plant defense against fungi (Ryan, 1989). Recently, the first protein-proteinase inhibitor of aspartic proteinases from the intestinal parasitic nematode *Ascaris suum* has been characterized (Martzen et al., 1990). Some of these protein-proteinase inhibitors have been found to mimic substrates, but they subsequently trap the enzyme into a slow and inefficient cleavage reaction. However, the common feature of protein-proteinase inhibitors is that the inactivation mechanisms are distinct from the strategies utilized by prosegments in zymogen structures.

The recent determinations of zymogen structures highlight the rich variety of inhibitory mechanisms and activation pathways. However, for the most part, the pathways linking the zymogen → mature enzyme remain a "black box," obscuring the detailed conformational changes and energetic considerations that remain speculative. The structures described here, as well as those listed in Table 1, provide a rational basis for addressing these many remaining questions using a variety of biochemical, biophysical, and structural methods in the future. Finally, proteolytic processing (e.g., growth factor maturation, protein degradation) has recently been implicated in the signaling pathways that lead to pattern formation for multicellular organisms across the evolutionary spectrum. Proteolytic enzymes such as the "astacins" and "reprolysins" (Bond & Beyon, 1995; Pan & Rubin, 1997; Weinmaster, 1998) and the serine proteinase "hepsin" (Vu et al., 1997) are synthesized as zymogens that are targeted to the extracellular milieu, and in some cases, remain tethered to the cell *via* a transmembrane anchor. The activation pathways of these zymogens may represent an additional level of regulation in cell-cell signaling. Clearly, the future holds many more surprises in the field of zymogen activation.

Acknowledgments

This review was supported by the MRC of Canada through a grant to the Group in Protein Structure and Function. A.R. Khan was supported by a Medical Research Council of Canada Studentship. We wish to express thanks to Dr M. Fujinaga for a careful review of the manuscript.

References

Allaire M, Chernaia M, Malcolm BA, James MNG. 1994. Picornaviral 3C cysteine proteinases have a fold similar to the chymotrypsin-like serine proteinases. *Nature (London) 369*:72–77.

Auer HE, Glick DM. 1984. Early events of pepsinogen activation. *Biochemistry 23*:2735–2739.

Aviles FX, Vendrell J, Guasch A, Coll M, Huber R. 1993. Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. *Eur J Biochem 211*:381–389.

Baker D, Shiau AK, Agard DA. 1993. The role of pro regions in protein folding. *Curr Op Cell Biol 5*:966–970.

Baker D, Sohl JL, Agard DA. 1992. A protein-folding reaction under kinetic control. *Nature 356*:263–265.

Baker EN, Drenth J. 1987. The thiol proteases: Structure and mechanism. In: Jurnak FA, McPherson A, eds. *Biological macromolecules and assemblies*, Vol. 3. New York: John Wiley and Sons. pp 313–368.

Bartunik HD, Summers LJ, Bartsch HH. 1989. Crystal structure of bovine β-trypsin in a crystal form with low molecular packing density. Active site geometry, ion pairs and solvent structure. *J Mol Biol 210*:813–828.

Bateman KS, Cherney MM, Tarasova NI, James MNG. 1998. Crystal structure of human pepsinogen A. In: James MNG, ed. *The aspartic proteinases: Retroviral, fungal, plant and mammalian*. New York: Plenum Press. In press.

Baudys M, Kostka V. 1983. Covalent structure of chicken pepsinogen. *Eur J Biochem 136*:89–99.

Becker JW, Marcy AI, Rokosz LL, Axel MG, Burbaum JJ, Fitzgerald PMD, Cameron PM, Esser CK, Hagmann WK, Hermies JD, Springer JP. 1995. Stromelysin-1: Three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Protein Sci 4*:1966–1976.

Berti PJ, Storer AC. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J Mol Biol 246*:273–283.

Birkedal-Hansen H, Werb Z, Welgus H, Van Wart H. 1990. In: James MNG, ed. *Matrix metalloproteinases and inhibitors*. Stuttgart, Germany: Gustav Fischer Verlag.

Blevins RA, Tulinsky A. 1985. The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. *J Biol Chem 260*:4624-4275.

Blundell TL, Johnson MS. 1993. Catching a common fold. *Protein Sci 2*:877–883.

Bode W. 1979. The transition of trypsinogen to a trypsin-like state upon strong ligand binding. II. The binding of pancreatic trypsin inhibitor and of isoleucine-valine and of sequentially related peptides to trypsinogen and to *p*-guanidinobenzoate-trypsinogen. *J Mol Biol 127*:357–374.

Bode W, Engh R, Musi D, Thiele U, Huber R, Karshikov A, Brzin J, Kos J, Turk V. 1988. The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J 7*:2593–2599.

Bode W, Huber R. 1978. Crystal structure analysis and refinement of two variants of trigonal trypsinogen. *FEBS Lett 90*:265–269.

Bode W, Huber R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem 204*:433–451.

Bode W, Renatus M. 1997. Tissue-type plasminogen activator: Variants and crystal/solution structures demarcate structural determinants of function. *Curr Op Struct Biol 7*:865–872.

Bode W, Schwager P. 1975. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site and active site at pH 7.0. *J Mol Biol 98*:693–717.

Bode W, Schwager P, Huber R. 1976. In: Ribbons DW, Brew W, eds. *Proteolysis and physiological regulation*. Miami Winter Symposia, Vol. 11. New York: Academic Press. pp 43–76.

Bode W, Schwager P, Huber R. 1978. The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with ILE-Val at 1.9 Å resolution. *J Mol Biol 118*:99–112.

Bolognesi M, Gatti G, Menegatti E, Guarneri M, Marquart M, Papamokos E, Huber R. 1982. Three-dimensional structure of the complex between pancreatic secretory trypsin inhibitor (Kazal type) and trypsinogen at 1.8 Å resolution. *J Mol Biol 162*:839–868.

Bond JS, Beyon RJ. 1995. The astacin family of metalloendopeptidases. *Protein Sci 4*:1247–1261.

Bonten EJ, Galjart NJ, Willemsen R, Usmany M, Vlak JM, d'Azzo A. 1995. Lysosomal protective protein/cathepsinA: Role of the 'linker' domain in catalytic activation. *J Biol Chem 270*:26441–26445.

Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, Murzin AG. 1995. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature 378*:416–419.

Brayer GD, Delbaere LT, James MNG. 1979. Molecular structure of the α-lytic protease from *Myxobacter 495* at 2.8 Å resolution. *J Mol Biol 131*:743–775.

Brenner S. 1988. The molecular evolution of genes and proteins: A tale of two serines. *Nature 334*:528–530.

Bryan P, Wang L, Hoskins J, Ruvinov S, Strausberg S, Alexander P, Almog O, Gilliland G, Gallagher T. 1995. Catalysis of a protein folding reaction: Mechanistic implications of the 2.0 Å structure of the subtilisin-prodomain complex. *Biochemistry 34*:10310–10318.

Burgos FJ, Pascual R, Salva M, Cuchillo CM, Aviles FX. 1989. The separation of pancreatic procarboxypeptidases by high-performance liquid chromatography and chromatofocusing. *J Chromatogr 481*:233–243.

Burgos FJ, Salva M, Villegas V, Soriano F, Mendez E, Aviles FX. 1991. Analysis of the activation process of porcine procarboxypeptidase B and determination of the sequence of its activation segment. *Biochemistry 30*:4082–4089.

Carmona E, Dufour E, Plouffe C, Takebe S, Mason P, Mort JS, Menard R. 1996. Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry 35*:8149–8157.

Chen P, Hochstrasser M. 1996. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell 86*:961–972.

Christianson DW, Lipscomb WN. 1989. Carboxypeptidase A. *Acc Chem Res 22*:62–69.

Ciechanover A. 1994. The ubiquitin-proteasome proteolytic pathway. *Cell 79*: 13–21.

Cohen GH, Silverton EW, Davies DR. 1981. Refined crystal structure of γ-chymotrypsin at 1.9 Å resolution. Comparison with other pancreatic serine proteases. *J Mol Biol 148*:449–479.

Coll M, Guasch A, Aviles FX, Huber R. 1991. Three-dimensional structure of porcine procarboxypeptidase B: Structural basis of its inactivity. *EMBO J 10*:1–9.

Coulombe R, Grochulski P, Sivaraman J, Menard R, Mort JS, Cygler M. 1996. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J 15*:5492–5503.

Craik CS, Roczniak S, Largman C, Rutter WJ. 1987. The catalytic role of the active site aspartic acid in serine proteases. *Science 239*:909–913.

Cuozzo JW, Tao K, Wu Q, Yong W, Sahagian GG. 1995. Lysine-based structure in the proregion of procathepsin L is the recognition site for mannose phosphorylation. *J Biol Chem 270*:15611–15619.

Cutfield SM, Dodson EJ, Anderson BF, Moody PC, Marshall CJ, Sullivan PA, Cutfield JF. 1995. The crystal structure of a major secreted aspartic proteinase from Candida albicans in complexes with two inhibitors. *Structure 3*:1261–1271.

Cygler M, Sivaraman J, Grochulski P, Coulombe R, Storer AC, Mort JS. 1996. Structure of rat procathepsin B. Model for inhibition of cysteine protease activity by the proregion. *Structure 4*:405–416.

Davie EW, Fujikawa K, Kisiel W. 1991. The coagulation cascade: Initiation, maintenance and regulation. *Biochemistry 30*:10363–10370.

Davie EW, Neurath H. 1955. Identification of a peptide released during autocatalytic activation of trypsinogen. *J Biol Chem 212*:515–529.

Davies DR. 1990. The structure and function of the aspartic proteinases. *Ann Rev Bioph Chem 19*:189–215.

d'Azzo A, Hoogeveen A, Reuser AJJ, Robinson D, Galjaard H. 1982. Molecular defect in combined β-galactosidase and neuraminidase deficiency in man. *Proc Nat Acad Sci (USA) 79*:4535–4539.

Desnuelle P. 1959. In: Boyer RD, Lardy H, Myrback K, eds. *The enzymes*, Vol. 4. New York: Academic Press. pp 93–118.

Docherty AJP, O'Connell J, Crabbe T, Angal S, Murphy G. 1992. The matrix metalloproteinases and their natural inhibitors—Prospects for treating degenerative tissue diseases. *Trends Biotechnol 10*:200–207.

Dunn B. 1997. Splitting image. *Nature Struct Biol 4*:969–972.

Endrizzi JA, Breddam K, Remington SJ. 1994. 2.8 Å-structure of yeast carboxypeptidase. *Biochemistry 33*:11106–11120.

Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics 15*:133–138.

Fehlhammer H, Bode W, Huber R. 1977. Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J Mol Biol 111*:415–438.

Folk JE. 1956. A new pancreatic carboxypeptidase. *J Am Chem Soc 78*:3541–3542.

Foltmann B. 1988. Aspartic proteinases: Alignment of amino acid sequences. In: Foltmann B, ed. *Proceedings of the 18th Linderstrom-Lang Conference, 4–8 July 1988*. Elsinore, Denmark. pp 7–20.

Foltmann B, Jensen AL. 1982. Human progastricsin. Analysis of intermediates during activation into gastricsin and determination of the amino acid sequence of the propart. *Eur J Biochem 128*:63–70.

Francis SE, Banerjee R, Goldberg DE. 1997. Biosynthesis and maturation of the malarial aspartic hemoglobinases plasmepsins I and II. *J Biol Chem 272*:14961–14968.

Freer ST, Kraut J, Robertus JD, Wright HT, Xuong NH. 1970. Chymotrypsinogen: 2.5 Å crystal structure, comparison with α-chymotrypsin, and implications for zymogen activation. *Biochemistry 9*:1997–2009.

Galjart NJ, Moreau H, Willemsen R, Gillemans N, Bonten EJ, d'Azzo A. 1991. Human lysosomal protective protein has a cathepsin A-like activity distinct from its protective function. *J Biol Chem 266*:14754–14762.

Gallagher T, Gilliland G, Wang L, Bryan P. 1995. The prosegment BPN′ complex: Crystal structure of a specific foldase. *Structure 3*:907–914.

Garcia-Saez I, Reverter D, Vendrell J, Aviles FX, Coll M. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J 16*:6906–6913.

Glick DM, Shalitin Y, Hitt CR. 1989. Studies of the irreversible step of pepsinogen activation. *Biochemistry 28*:2626–2630.

Gomes-Ruth FX, Gomez-Ortiz M, Vendrell J, Ventura S, Bode W, Huber R, Aviles FX. 1997. Crystal structure of an oligomer of proteolytic zymogens: Detailed conformational analysis of the bovine ternary complex and implications for their activation. *J Mol Biol 269*:861–880.

Greenberg AH. 1996. Granzyme B-induced apoptosis. *Adv Exp Med & Biol 406*:219–228.

Groettrup M, Soza A, Jucjelkorn U, Kloetzel PM. 1996. Peptide antigen production by the proteasome: Complexity provides efficiency. *Immunology Today 17*:429–435.

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

Groves MR, Taylor MAJ, Scott M, Cummings NJ, Pickersgill RW, Jenkins JA. 1996. The prosequence of procaricain forms an α-helical domain that prevents access to the substrate-binding cleft. *Structure 4*:1193–1203.

Guasch A, Coll M, Aviles FX, Huber R. 1992. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J Mol Biol 224*:141–157.

Hanson JE, Kaplan AP, Bartlett PA. 1989. Phosphonate analogues of carboxypeptidase A substrates are potent transition-state analogue inhibitors. *Biochemistry 28*:6294–6305.

Hartsuck JA, Koelsch G, Remington SJ. 1992. The high-resolution structure of porcine pepsinogen. *Proteins Struct Funct Genet 13*:1–25.

Hayano T, Sogawa K, Ichihara Y, Fuji-Kuriyama Y, Takahashi K. 1988. Primary structure of human pepsinogen C gene. *J Biol Chem 263*:1382–1385.

Herriott RM. 1939. Kinetics of the formation of pepsin from swine pepsinogen and identification of an intermediate compound. *J Gen Physiol 22*:65–78.

Hofmann T. 1989. *Zoological research*. Beijing (in Chinese).

Huber R, Bode W. 1978. Structural basis of the activation and action of trypsin. *Acc Chem Res 11*:114–122.

Ichihara Y, Sogawa K, Morohashi K, Fuji-Kuriyama Y, Takahashi K. 1986. Nucleotide sequence of a nearly full-length cDNA coding for pepsinogen of rat gastric mucosa. *Eur J Biochem 161*:7–12.

International Union of Biochemistry and Molecular Biology. 1992. *Enzyme nomenclature: Recommendations of the nomenclature committee of the International Union*. San Diego, CA: Academic Press.

Jackman HL, Morris PW, Deddish PA, Skidgel RA, Erdos EG. 1992. Inactivation of endothelin I by deamidase (lysosomal protective protein). *J Biol Chem 267*:2872–2875.

James MNG, Sielecki AR. 1986. Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature (London) 319*: 33–38.

James MNG, Sielecki AR, Hayakawa K, Gelb MH. 1992. Crystallographic analysis of transition state mimics bound to pencillopepsin: Difluorstatine- and difluorostatone-containing peptides. *Biochemistry 31*:3872–3886.

Jia Z, Hasnain S, Hirama T, Lee X, Mort JS, To R, Huber CP. 1995. Crystal structure of rat cathepsin B and a cathepsin B-inhibitor complex: Implications for structure-based inhibitor design. *J Biol Chem 270*:5527–5533.

Kageyama T, Ichinose M, Tsukada S, Miki K, Kurokawa K, Koiwai O, Tanji M, Yakabe E, Athauda SBP, Takahashi K. 1992. Gastric procathepsin E and progastricsin from guinea pig. *J Biol Chem 267*:16450–16459.

Kageyama T, Takahashi K. 1986a. The complete amino acid sequence of monkey pepsinogen A. *J Biol Chem 261*:4395–4405.

Kageyama T, Takahashi K. 1986b. The complete amino acid sequence of monkey pepsinogen. *J Biol Chem 261*:4406–4419.

Kam CM, McCrae BJ, Harper JW, Niemann MA, Volanakis JE, Powers JC. 1987. Human complement proteins D, C2 and B. Active site mapping and peptide thioester substrates. *J Biol Chem 262*:3444–3451.

Kay J, Cassell B. 1971. The autoactivation of trypsinogen. *J Biol Chem 246*:6661–6665.

Kerfelec B, Chapus C, Puigserver A. 1985. Existence of ternary complexes of procarboxypeptidase A in the pancrease of some ruminant species. *Eur J Biochem 151*:515–519.

Kerr MA, Walsh KA, Neurath H. 1975. Catalysis by serine proteases and their zymogens. A study of acyl intermediates by circular dichroism. *Biochemistry 14*:5088–5094.

Kerr MA, Walsh KA, Neurath H. 1976. A proposal for the mechanism of chymotrypsinogen activation. *Biochemistry 15*:5566–5570.

Khan AR, Cherney MM, Tarasova NI, James MNG. 1997. Structural characterization of activation intermediate 2 on the pathway to human gastricsin. *Nature Struct Biol 4*:1010–1015.

Kim H, Lipscomb WN. 1991. Comparison of the structures of three carboxypeptidase A-phosphonate complexes determined by X-ray crystallography. *Biochemistry 30*:8171–8180.

Kim S, Narayana SVL, Volanakis JE. 1994. Mutational analysis of the substrate binding site of human complement factor D. *Biochemistry 33*:14393–14399.

Kirschke H, Barrett AJ. 1987. Chemistry of lysosomal proteases. In: Glaumann H, Ballard AJ, eds. *Lysosomes: Their role in protein breakdown*. London, UK: Academic Press. pp 193–238.

Klionsky DJ, Banta LM, Emr SD. 1988. Intracellular sorting and processing of a yeast vacuolar hydrolase: Proteinase A propeptide contains vacuolar targetting information. *Mol Cell Biol 8*:2105–2116.

Koklitis PA, Murphy G, Sutton C, Angal S. 1991. Purification of recombinant human prostromelysin. Studies on heat activation to give high-Mr and low-Mr active forms, and a comparison of recombinant with natural stromelysin activities. *Biochem J 276*:217–221.

Kossiakoff AA, Chambers JL, Kay LM, Stroud RM. 1977. Structure of bovine trypsinogen at 1.9 Å resolution. *Biochemistry 16*:654–664.

Kraut J. 1977. Serine proteases: Structure and mechanism of catalysis. *Ann Rev Biochem 46*:331–358.

Kunitz M. 1938. Formation of new crystallographic enzymes from chymotrypsin. *J Gen Physiol 22*:207–237.

Kunitz M, Northrop JH. 1935. Crystalline chymo-trypsin and chymo-trypsinogen. I. Isolation, crystallization, and general properties of a new proteolytic enzyme and its precursor. *J Gen Physiol 18*:433–456.

Lamba D, Bauer M, Huber R, Fischer S, Rudolph R, Kohnert U, Bode W. 1996. The 2.3 Å crystal structure of the catalytic domain of recombinant two-chain human tissue-type plasminogen activator. *J Mol Biol 258*:117–135.

Laskowski M Jr, Kato I. 1980. Protein inhibitors of proteinases. *Ann Rev Biochem 49*:593–626.

Liao DI, Breddam K, Sweet B, Bullock T, Remington SJ. 1992. Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry 31*:9796–9812.

Liszewski MK, Farries TC, Lublin DM, Rooney IA, Atkinson JD. 1996. Control of the complement system. *Adv Immunology 61*:201–283.

Madison EL, Kobe A, Gething M-J, Sambrook JF, Goldsmith EJ. 1993. Converting tissue plasminogen activator to a zymogen: A regulatory triad of Asp-Ser-His. *Science 262*:419–421.

Martzen MR, McMullen BA, Smith NE, Fujikawa K, Peanasky RJ. 1990. Primary structure of the major pepsin inhibitor from the intestinal parasitic nematode *Ascaris suum*. *Biochemistry 29*:7366–7372.

Mason RW, Massey SD. 1992. Surface activation of pro-cathepsin L. *Biochem Bioph Res Commun 189*:1659–1666.

Matthews BH. 1988. Structural basis of the action of thermolysin and related zinc peptidases. *Acc Chem Res 21*:333–340.

Matthews BW. 1977. In: Neurath H, Hill RL, eds. *The proteins*, 3rd ed. Vol. 3. New York: Academic Press. pp 404–590.

Matthews DA, Alden RA, Birktoft JJ, Freer ST, Kraut J. 1977. Re-examination of the charge-relay system in subtilisin and comparison with other serine proteases. *J Biol Chem 252*:8875–8883.

McIntyre GF, Godbold GD, Erickson AH. 1994. The pH-dependent membrane association of procathepsin L is mediated by a 9-residue sequence within the propeptide. *J Biol Chem 269*:567–572.

Merritt EA, Murphy MEP. 1994. Raster3D version 2.0: A program for photorealistic molecular graphics. *Acta Crysallogr D50*:869–873.

Miller M, Jaskolski M, Rao JKM, Leis J, Wlodawer A. 1989. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature 337*:576–579.

Moore SA, Sielecki AR, Chernaia MM, Tarasova NI, James MNG. 1995. Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J Mol Biol 247*:466–485.

Morgan PH, Walsh KA, Neurath H. 1974. Inactivation of trypsinogen by methane sulfonyl fluoride. *FEBS Lett 41*:108–110.

Muller-Eberhard HJ. 1988. Molecular organization and function of the complement system. *Annu Rev Biochem 57*:321–347.

Murphy GJP, Murphy G, Reynolds JJ. 1991. The origin of matrix metalloproteinases and their familial relationships. *FEBS Lett 289*:4–7.

Nagase H, Enghild JJ, Suzuki K, Salvesen G. 1990. Stepwise activation mechanisms of the precursor of matrix metalloproteinase 3 (stromelysin) by proteinases and (4-aminophenyl)mercuric acetate. *Biochemistry 29*:5783–5789.

Nagase H, Suzuki K, Enghild JJ, Salvesen G. 1991. Stepwise activation mechanisms of the precursors of matrix metalloproteinases I (tissue collagenase) and 3 (stromelysin). *Biomed Biochim Acta 50*:749–754.

Narayana SVL, Carson M, El-Kabbani O, Kilpatrick JM, Moore D, Chen X, Bugg CE, Volanakis JE, DeLucas LJ. 1994. Structure of human factor D. A complement system protein at 2.0 Å resolution. *J Mol Biol 235*:695–708.

Navia MA, Fitzgerald PMD, Mckeever BM, Leu C-T, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP. 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature 337*:615–620.

Neurath H. 1957. The activation of zymogens. *Adv Protein Chem 12*:319–386.

Neurath H. 1959. Carboxypeptidases A and B. In: Boyer PD, Lardy H, Myrback K, eds. *The enzymes*, Vol. 4. New York: Academic Press. pp 11–37.

Neurath H. 1984. Evolution of proteolytic enzymes. *Science 224*:350–356.

Neurath H, Walsh KA. 1976. Role of proteolytic enzymes in biological regulation (a review). *Proc Natl Acad Sci (USA) 73*:3825–3832.

Okada Y, Harris ED Jr, Nagase H. 1988. The precursor of a metalloendopeptidase from human rheumatoid sinovial fibroblasts. Purification and mechanisms of activation by endopeptidases and 4-aminophenylmercuric acetate. *Biochem J 254*:731–741.

Okada Y, Nakanashi I. 1989. Activation of matrix metalloproteinase 3 (stromelysin) and matrix metalloproteinase 2 ("gelatinase") by human neutrophil elastase and cathepsin G. *FEBS Lett 249*:353–356.

Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschueren KHG, Goldman A. 1992. The α/β hydrolase fold. *Protein Eng 5*:197–211.

Palmenberg AC. 1990. Proteolytic processing of picornaviral polyproteins. *Ann Rev Microbiol 44*:603–623.

Palmenberg AC, Rueckert RR. 1982. Evidence for intramolecular cleavage of picornaviral replicase processing enzyme. *J Virol 41*:244–249.

Palombella VJ, Rando OJ, Goldberg AL, Maniatis T. 1994. The ubiquitin-proteasome pathway is required for processing the NF-κB1 precursor protein and the activation of NF-κB. *Cell 78*:773–785.

Pan D, Rubin GM. 1997. Kuzbanian controls proteolytic processing of Notch and mediates lateral inhibition during Drosophila and vertebrate neurogenesis. *Cell 90*:271–280.

Pascual R, Vendrell J, Aviles FX, Bonicel J, Wicker C, Puigserver A. 1990. Autolysis of proproteinase E in bovine procarboxypeptidase A ternary complex gives rise to subunit III. *FEBS Lett 277*:37–41.

Perona JS, Craik CS. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci 4*:337–360.

Peters JM. 1994. Proteasomes: Protein degradation machines of the cell. *Trends Biochem Sci 19*:377–382.

Pignol D, Granon S, Chapus S, Chapus C, Fontecilla-Camps JC. 1995. Crystallographic study of a cleaved, non-activatable form of porcine zymogen E. *J Mol Biol 252*:20–24.

Polgar L. 1989. Structure and function of serine proteases. In: *Mechanisms of protease action*. Boca Raton, Florida: CRC Press. Chapter 3.

Polgar L, Halasz D. 1982. Current problems in mechanistic studies of serine and cysteine proteinases. *Biochem J 207*:1–10.

Potempa J, Korzus E, Travis J. 1994. The serpin superfamily of proteinase inhibitors: Structure, function, and regulation. *J Biol Chem 269*:15957–15960.

Puigserver A, Chapus C, Kerfelec B. 1986. In: Desnuelle P, Sjostrom H, Noren O, eds. *Molecular and cellular basis of digestion*. Amsterdam: Elsevier. pp 235–247.

Quiocho FA, Lipscomb WN. 1971. Carboxypeptidase A: A protein and an enzyme. *Adv Protein Chem 25*:1–78.

Rees DC, Lewis M, Lipscomb WN. 1983. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J Mol Biol 168*:367–387.

Rees DC, Lipscomb WN. 1982. Refined crystal structure of potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol 160*:475–498.

Reinemer P, Grams F, Huber R, Kleine T, Schnierer S, Piper M, Tschesche H, Bode W. 1994. Structural implications for the role of the N-terminus in the superactivation of collagenases. *FEBS Lett 338*:227–233.

Renatus M, Engh RA, Stubbs MT, Huber R, Fischer S, Kohnert U, Bode W. 1997. Lys156 promotes the anomalous proenzyme activity of t-PA: X-ray crystal structure of single-chain human t-PA. *EMBO J 16*:4797–4805.

Robertus JD, Kraut J, Alden RA, Birktoft JJ. 1972. Subtilisin: A stereochemical mechanism involving transition-state stabilization. *Biochemistry 11*:4293–4303.

Rovery M, Fabre C, Desnuelle P. 1953. Studies of the activation of bovine chymotrypsinogen and trypsinogen by determination of the N-terminal residues of the zymogens and their corresponding active enzymes (in French). *Biochim Biophys Acta 12*:547–559.

Rudenko G, Bonten E, d'Azzo A, Hol WGJ. 1995. Three-dimensional structure of the human "protective protein": Structure of the precursor form suggests a complex activation mechanism. *Structure 15*:1249–1259.

Ryan CA. 1989. Proteinase inhibitor gene families: Strategies for transformation to improve plant defenses against herbivores. *Bioessays 10*:20–24.

Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157–162.

Seemuller E, Lupas A, Baumeister W. 1996. Autocatalytic activation of the 20S proteasome. *Nature 382*:468–470.

Sielecki AR, Hayakawa K, Fujinaga M, Murphy ME, Fraser M, Muir AK, Carilli CT, Lewicki JA, Baxter JD, James MNG. 1989. Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5 Å resolution. *Science 243*:1346–1351.

Sogawa K, Fuji-Kuriyama Y, Mizukami Y, Ichihara Y, Takahashi K. 1983. Primary structure of human pepsinogen gene. *J Biol Chem 258*:5306–5311.

Sohl JL, Shiau AK, Rader SD, Wilk BJ, Agard DA. 1997. Inhibition of α-lytic protease by pro region C-terminal steric occlusion of the active site. *Biochemistry 36*:3894–3902.

Sprang S, Standing T, Fletterick RJ, Stroud RM, Finer-Moore J, Xuong NH, Hamlin R, Rutter WJ, Craik CS. 1987. The three-dimensional structure of Asn102 mutant of trypsin: Role of Asp102 in serine protease catalysis. *Science 237*:905–909.

Stein PE, Carrell RW. 1995. What do dysfunctional serpins tell us about molecular mobility and disease? *Nature Struct Biol 2*:96–113.

Steiner DF, Smeekens SP, Ohagi S, Chan SJ. 1992. The new enzymology of precursor processing endoproteases. *J Biol Chem 267*:23435–23438.

Steitz TA, Henderson R, Blow DM. 1969. Structure of crystalline α-chymotrypsin. 3. Crystallographic studies of substrates and inhibitors bound to the active site of α-chymotrypsin. *J Mol Biol 46*:337–348.

Steitz TA, Shulman RG. 1982. Crystallographic and NMR studies of the serine proteases. *Ann Rev Bioph Bioeng 11*:419–444.

Stevens RL, Qui D, McNeil HP, Friend DS, Hunt JE, Austen KF, Zhang J. 1996. Transgenic mice that possess a disrupted mast cell protease 5 (mMCP-5) gene cannot store carboxypeptidase A (mMC-CPA) protein in their granules. *FASEB J 10*:A1307.

Storer AC, Menard R. 1994. Catalytic mechanism in papain family of cysteine peptidases. *Meth Enzymol 244*:486–500.

Strausberg S, Alexander P, Wang L, Schwarz F, Bryan P. 1993. Catalysis of a protein folding reaction: Thermodynamic and kinetic analyses of subtilisin BPN′ interactions with its propeptide fragment. *Biochemistry 32*:8112–8119.

Taggart RT, Cass LG, Mohandas TK, Derby P, Barr PJ, Pals G, Bell GI. 1989. Human pepsinogen C (progastricsin). Isolation of cDNA clones, localization to chromosome 6, and sequence homology with pepsinogen A. *J Biol Chem 264*:375–379.

Tanford C. 1997. How protein chemists learned about the hydrophobic factor. *Protein Sci 6*:1358–1366.

Tang J, James MNG, Hsu IN, Jenkins JA, Blundell TL. 1978. Structural evidence for gene duplication in the evolution of the acid proteases. *Nature 271*:618–621.

Tao K, Stearns NA, Dong J, Wu Q, Sahagian GG. 1994. The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch Biochem Biophys 311*:19–27.

Tsukada H, Blow DM. 1985. Structure of α-chymotrypsin at 1.68 Å resolution. *J Mol Biol 184*:703–711.

Turk D, Podobnik M, Kuhelj R, Dolinar M, Turk V. 1996. Crystal structures of human procathepsin B at 3.2 Å and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett 384*:211–214.

Uren JR, Neurath H. 1972. Mechanism of activation of bovine procarboxypeptidase AS5. Alterations in primary and quaternary structure. *Biochemistry 11*:4483–4492.

Vallet V, Ahmed C, Gaeggler H-P, Horisberger J-D, Rossier BC. 1997. An epithelial serine protease activates the amiloride-sensitive sodium channel. *Nature 389*:607–610.

Valls LA, Winther JR, Stevens TH. 1990. Yeast carboxypeptidase Y vacuolar targetting signal is defined by four propeptide amino acids. *J Cell Biol 111*:361–368.

Vendrell J, Cuchillo CM, Aviles FX. 1990. The tryptic activation pathway of monomeric procarboxypeptidase A. *J Biol Chem 256*:6949–6953.

Vernet T, Berti PJ, de Montigny C, Musil R, Tessier DC, Menard R, Magny MC, Storer AC, Thomas DY. 1995. Processing of the papain precursor. The ionization of a conserved amino acid motif within the pro region participates in the regulation of intramolecular processing. *J Biol Chem 270*:10838–10846.

Vernet T, Khouri HE, Laflamme P, Tessier DC, Gour-Salin B, Storer AC, Thomas DY. 1991. Processing of the papain precursor. Purification of the zymogen and characterization of its mechanism of processing. *J Biol Chem 266*:21451–21457.

Volanakis JE, Narayana SVL. 1996. Complement factor D, a novel serine protease. *Protein Sci 5*:553–564.

Vu TKH, Liu RW, Haaksma CJ, Tomasek JJ, Howard EW. 1997. Identification and cloning of the membrane-associated serine protease, hepsin, from mouse pre-implantation embryos. *J Biol Chem 272*:31315–31320.

Wang D, Bode W, Huber R. 1985. Bovine chymotrypsinogen A. X-ray crystal structure analysis and refinement of a new crystal form at 1.8 Å resolution. *J Mol Biol 185*:595–624.

Weinmaster G. 1998. Reprolysins and astacins ... Alive, Alive-O. *Science 279*:336–337.

Westphal V, Marcusson EG, Winther JR, Emr SD, van den Hazel HB. 1996. Multiple pathways for vacuolar sorting of yeast proteinase A. *J Biol Chem 271*:11865–11870.

White RT, Damm D, Hancock N, Rosen BS, Lowell BB, Usher P, Flier JS, Spiegelman BM. 1992. Human adipsin is identical to complement factor D and is expressed at high levels in adipose tissue. *J Biol Chem 267*:9210–9213.

Wintersberger E, Cox DJ, Neurath H. 1962. Bovine pancreatic procarboxypeptidase B. I. Isolation, properties and activation. *Biochemistry 1*:1069–1078.

Winther JR, Sorenson P. 1991. Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of enzymatic activity. *Proc Natl Acad Sci (USA) 88*:9330–9334.

Wlodawer A, Erickson JW. 1993. Structure-based inhibitors of HIV-1 protease. *Ann Rev Biochem 62*:543–585.

Wright HT. 1996. The structural puzzle of how serpin serine proteinase inhibitors work. *Bioessays 18*:453–464.

Yamauchi Y, Stevens JW, Macon KJ, Volanakis JE. 1994. Recombinant and native zymogen forms of human complement factor D. *J Immunol 152*:3645–3653.