Journal of Cellular Biochemistry 32:35–49 (1986)
Proteases in Biological Control and Biotechnology 1–15

The Versatility of Proteolytic Enzymes

Hans Neurath

Department of Biochemistry, University of Washington, Seattle, Washington 98195

The growing realization of their physiological importance has generated renewed interest in the study of proteolytic enzymes. Modern methods of protein chemistry and molecular biology have revealed new insights into the protein and gene structure of a variety of protein precursors and their processing by limited proteolysis. Examples are given in this review for transmembrane processes and the role of signal peptidases of both eukaryotic and prokaryotic origin, the processing of prohormones and precursors of growth factors, protein components of blood coagulation, fibrinolysis, and of the complement system, and a group of granulocyte proteases, including the mast cell serine proteases. The relationship of homologous domains found in many of these proteases and their zymogens to protein evolution is a recurrent theme of this discussion.

Key words: model proteases, pancreatic serine proteases, homologous domains, exon/intron distribution, tetrahedral intermediates, limited proteolysis, cotranslational processing, transmembrane processes, signal peptidase, complement system

The growing realization of the physiological importance of proteases, together with the novel methodologies of protein chemistry, molecular biology, and biotechnologies, has generated renewed interest in the study of proteases. Here, some of the major physiological systems in which proteolytic enzymes, their precursors, and inhibitors play a vital role are examined. In this introductory review, I shall examine what we have learned from the study of model proteases and how that knowledge has been subsequently applied to investigate their role in more complex biological processes. Finally, I will outline certain problems and areas of research which deserve greater emphasis in the future. An underlying theme of this discussion will be protein evolution, a topic which, perhaps more than any other, has contributed to our current understanding of the structural and functional relationships among proteolytic enzymes.

MODEL PROTEASES

Much of our current knowledge of the structure, specificity, and mechanism of action of proteases has been derived from investigations of crystalline, well-charac-

Received March 24, 1986; accepted June 3, 1986.

terized enzymes, notably pancreatic trypsin, chymotrypsin, carboxypeptidase A, and elastase [1]. To this small group were subsequently added bacterial subtilisin, penicillopepsin, and the plant protease papain. This repertoire has since been considerably enlarged but it is still restricted to proteases primarily selected for their availability and ease of isolation and crystallization, rather than for their physiological importance or interest. Detailed analysis of amino acid sequences, x-ray structures, kinetics, substrate specificity, and enzyme activation and inhibition led to the identification of the components of the active site and its geometry. From this information, the mechanism of action of many of these model proteases was deduced. As a result, proteases have been grouped into families which by virtue of similar amino acid sequences, conformations, and mechanisms of action are believed to have evolved from common ancestors [2]. These deductions have received considerable support from cloning and sequencing of the parent genes and have given evidence of the occurrence of homologous domains in distantly related or even unrelated proteins [3,4]. The relation of protein domains to the organization of the parent gene, specifically to the exon/intron distribution [5,6], is an important but as-yet-unresolved aspect of protein evolution. Current notions of the role of specific amino acid residues in catalysis, derived from kinetic and mechanistic considerations, are being put to critical tests by the newer method of site-directed mutagenesis [7].

Despite this spectacular progress on many fronts, certain fundamental questions of enzyme action are still being debated. For instance, while the general features of the tetrahedral intermediate of the enzyme-substrate complex of the serine proteases appear well established, there is still no agreement about details of the state of ionization of the amino acid residues of the catalytic triad in the transition state [8,9]. The mechanism of action of carboxypeptidase toward peptide and ester substrates remains in dispute [10]. The chemical changes accompanying the activation of trypsinogen and chymotrypsinogen were elucidated over 30 yr ago but the detailed configurational changes responsible for the generation of catalytic activity are still not resolved with finality. Much less is known in this regard about the activation of procarboxypeptidase, and the mechanism of activation of pepsinogen has only recently been elucidated. It was found to follow a very different course from that of the serine proteases [11]. X-ray comparison of swine pepsinogen and penicillopeptidase, a protease homologous to gastric pepsin, has shown that the primary step is the disruption at low pH of salt bridges that position the amino-terminal region of the proenzyme across the active site of pepsin, followed by the release of the amino-terminal segment and exposure of the catalytic and substrate-binding sites. In this case, the active site of the enzyme is more or less preexistent in the zymogen but sterically blocked by the overlaying activation peptide.

The proteases that have been investigated in such detail are at best prototypes representing a limited number of gene families, in contrast to many physiologically more interesting proteases of undetermined genealogy—orphans, so to speak, in search of their families. Let me just mention the aminopeptidases, the collagenases, the signal peptidases, and many types of tissue proteases, some of which are at least as important physiologically as those that have been so extensively studied.

Processing by Limited Proteolysis

Limited proteolysis is the key to proteolytic processing and regulation [12,13]. It is an important posttranslational modification of proteins because it can affect the

delicate balance among various conformational states of protein substrates and thus can alter their biological functions. In general, proteolysis is limited by the accessibility of peptide bonds to the protease, by the substrate specificity of the enzyme, and by the complementarity of the substrate and the active site of the enzyme in the transition state complex [12]. Limited proteolysis has been a useful tool for the sequence analysis of proteins and for the identification and isolation of functional domains of proteins of known structure [13]. An extreme case of limited proteolysis is the action of several proteases, differing in substrate specificity, on alpha-2 macroglobulin [14], which is exclusively cleaved in a narrow segment of nine amino acid residues near the center of the 1,450-residue peptide chains (the so-called bait region).

Proteolytic processing can occur cotranslationally, posttranslationally, or both [15]. Depending on the processing system, different proteases are involved. The best-studied cotranslational proteolytic processing reaction is the release of signal peptides by signal peptidase during the transfer of proteins across membranes. Posttranslational processing accompanies the cleavage of superchains into their constituent functional proteins (eg, protein hormones and growth factors), macromolecular assembly (eg, collagen fibril formation, biosynthesis of picorna viruses or bacteriophage) and consecutive zymogen activations (cascades) of the more complex blood coagulation and complement systems.

Transmembrane Processes and Signal Peptidases

Secretory proteins of eukaryotic and prokaryotic cells are usually synthesized as precursors containing an amino-terminal extension, the signal peptide, which during translocation across the membrane is cotranslationally cleaved by signal peptidase [16]. Signal peptides characteristically contain some 16–25 amino acid residues, including a hydrophobic core flanked by basic amino acid residues on one side and by polar residues on the other. Some membrane proteins contain another stretch of hydrophobic residues, the "stop transfer sequence," which, as the name implies, directs the termination of the transfer by anchoring the protein chain to the membrane [17]. Protein traffic across membranes has been extensively investigated in the biosynthesis of proteins on polysomes of the RER of eukaryotes, using cell-free translation systems, and across the outer and periplasmic membranes of bacteria. Whereas the isolation and characterization of signal peptidases of eukaryotic systems have met with considerable difficulties [18], the structure of a prokaryotic signal peptidase has been elucidated by DNA cloning and sequencing [19].

An interesting but unresolved problem within the context of the present discussion is the substrate specificity and mechanism of action of the signal peptidases. In particular, one would want to know whether this enzyme recognizes a specific linear amino acid sequence or its conformation or both. Although several hypotheses have been advanced [15,20–23], lack of adequate experimental methods has precluded a critical test. What seems certain is that signal peptidases recognize small residues on the carboxyl side of the susceptible peptide bond, such as glycine, alanine, serine, or threonine. Examination of the signal sequences of some 40 prokaryotic and eukaryotic proteins revealed certain other patterns [21]—for instance, the existence of a recognition sequence of Ala-X-Ala located after the sixth amino acid following the core sequence, or generally, A-X-B, where A and B could be any one of several small residues. Studies of mutants of bacteriophage M 13 procoat protein [24], lipoprotein [25], and beta lactamase [26] have also shown that certain amino acid residues need

to be conserved in order for the signal peptide to be processed. However, the proteolytic removal of the signal peptide is not an absolute requirement for translocation, as in the case of ovalbumin which is not cleaved by the peptidase [27].

Several possible functions have been ascribed to the action of signal peptidases. Among these are rendering translocation irreversible and preventing or minimizing interaction with the internal side of the endoplasmic reticulum; altering the folding of the translocated polypeptide chain in response to the functional requirements of the mature protein; and promoting the affinity of the translocated protein to other proteins within the endoplasmic reticulum and/or the Golgi apparatus. The latter would particularly apply to proproteins such as the precursors of protein hormones and growth factors, which will be considered next.

Processing of Precursors of Protein Hormones and Growth Factors

These two groups of protein precursors and their processing reactions have several common features [28,29]. Following removal of the signal peptide, they are processed while in transit from the RER to their sites of action. The Golgi apparatus and the secretory granules are believed to be the most likely and prevalent sites of proteolytic processing. Among the most thoroughly studied prohormones are the precursors of insulin, the pituitary hormones [eg, the multifunctional proopiomelanocortin (PMOC) and its products, ACTH, the lipotropins, the melanotropin-stimulating hormones (MSH), the endorphins, etc], the precursors of enkephalins, and the neurophysins. Examples of growth factor precursors include the insulinlike growth factors IGF I and II [30–32], nerve growth factors (NGF) [33], the epidermal growth factor (EGF) [34], and the mast cell growth factor [35]. The gene structures of some of these are known and have revealed that in some instances the active principle constitutes but a few percent of the mass of the precursor. While in most cases the norm applies that the propeptide precedes the sequence of the active protein, in the case of the neurophysins, for instance, the pro-sequence is on the carboxyl side relative to the active peptide [36].

Relatively little is known about the substrate specificity of these processing proteases. Proteolytic processing sites usually, but not necessarily, contain two adjacent basic amino acid residues, eg, Arg-Arg, Lys-Arg, or Arg-Lys, the former two being favored over the latter [28]. In some instances (eg, chicken proalbumin, procholecystokinin), proteolytic cleavage occurs adjacent to a single basic residue, Arg or Lys. No other distinguishing sequence requirement for proteolytic processing has been discerned among the precursor proteins that have been studied. In the absence of other structural information one can only assume that the bonds being cleaved are on accessible sites of these protein substrates (eg, on interdomain hinges or external fringes [12]). Three types of enzymes have been isolated, mostly from secretory granules: A trypsinlike enzyme, possibly containing a free sulfhydryl group [15,28], a carboxypeptidase B-like enzyme [37,38], and a thiol protease [39], similar to lysosomal cathepsin B, having a pH optimum of pH 4–6. A yet-different kind of proteolytic processing has been described in the case of promellitin, ie, a dipeptidyl amino transferase which removes sequentially dipeptides until it reaches a resistant amino-terminal Ile-Gly bond in the mellitin sequence [40]. The same type of enzyme is believed to process the precursor of the yeast alpha-mating factor, acting in concert with a trypsinlike enzyme to release four putative tandem copies of the alpha factor

[41]. This small protein is of considerable practical interest for the secretion of heterologous proteins in yeast.

PLASMA PROTEASES

Among the best-characterized regulatory proteases are those of blood plasma, particularly the proteases that participate in blood coagulation and fibrinolysis and those of the complement complex. A wealth of information has been gathered in recent years, largely due to the application of the most advanced methods of protein chemistry, enzymology, and molecular biology. The present discussion will, by necessity, be limited to the most salient aspects of these proteases and their zymogens.

Blood Coagulation and Fibrinolysis

It is now well established that both the intrinsic and the extrinsic pathways of blood coagulation involve a series of consecutive zymogen activation reactions which are fine-tuned by cofactors, protease inhibitors, and proteolytic enzymes [42,43]. The initial phases of the blood coagulation cascades are interrelated to those of the fibrinolytic and the kininogen pathways which are also regulated by the activation of serine protease precursors. Protein and DNA-sequence analyses have demonstrated beyond a doubt that each active protease is a serine protease related to pancreatic trypsin [12,42]. However, the substrate specificities are modulated by protein domains in the zymogen which have no counterpart in trypsin or in any of the pancreatic serine proteases [43]. Thus, while each of these regulatory proteases catalyzes the hydrolysis of small synthetic peptide or ester substrates of trypsin, the physiological protein substrates are cleaved with a high degree of specificity and selectivity. Relatively little is known of the specificity of these proteases toward their physiological substrates. In general, each protease recognizes the zymogen which follows in the cascade, and usually also activates one or more of the preceding zymogens. However, in no known case is the zymogen activated which follows the immediate target zymogen. In large measure, this selectivity must be ascribed to recognition sites in the zymogen precursor which, in contrast to the small hexapeptide of trypsinogen, contain from 150 to 580 amino acid residues and specify the interaction with other macromolecules important for physiological regulation. These precursors contain homologous domains found in other regulatory plasma proteases and others seen in functionally unrelated proteins. A deeper insight into these structures was obtained by a systematic analysis of protein and DNA sequences of the major components of the blood coagulation and fibrinolysis systems by Davie and co-workers [44].

Major sequence homologies, believed to represent structural domains, include the gamma-carboxyl glutamic acid (Gla) domains, kringles (triple-looped, disulfide bonded 70–80-residue peptide segments), domains homologous to epidermal growth factor (EGF domains), and "finger I" and "finger II" domains first observed in fibronectin [45]. A representative list of the type and number of domains occurring in these proteins is given in Table I and their structures are symbolized in Figure 1. Some of these domains have been isolated by limited proteolysis of the parent protein without impairment of function (eg, the kringles of plasminogen activator which are binding sites for fibrin) but in other instances, the functional significance, if any, (eg, EGF domain) remains to be established. Application of current methods of molecular

TABLE I. Occurrence and Distribution of Domains in Coagulation Proteins

Proteins	Domains						
	Protease	Gla	Asp-OH	EGF	Kringle	Finger I	Finger II
Prothrombin	1	1			2		
Factor X	1	1	1	2			
Factor IX	1	1	1	2			
Protein C	1	1	1	2			
Factor XI	2						
Factor XII	1			2	1	1	1
Plasma kallikrein	1						
Plasminogen	1				5		
t-Plasminogen activator	1			1	2	1	
Urokinase	1			1	1		

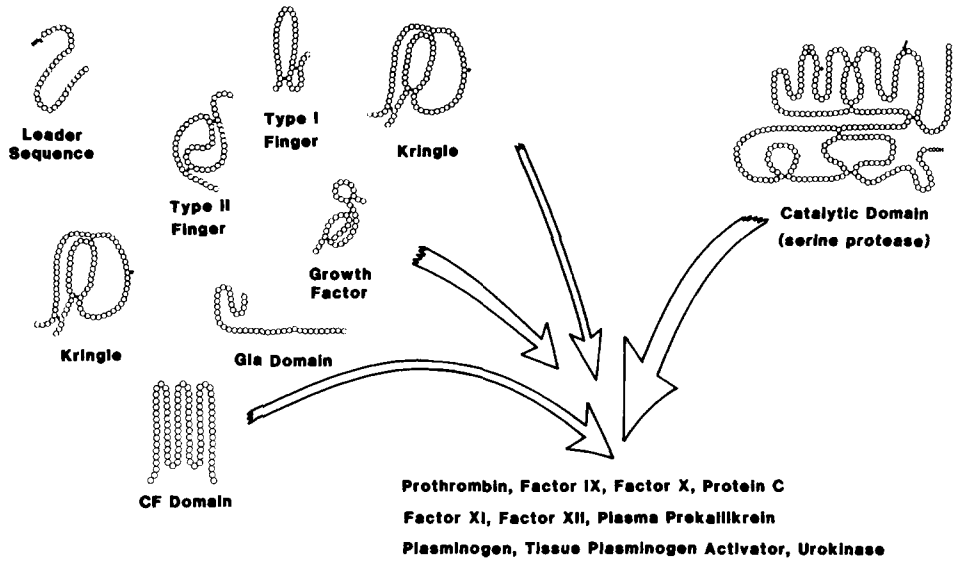
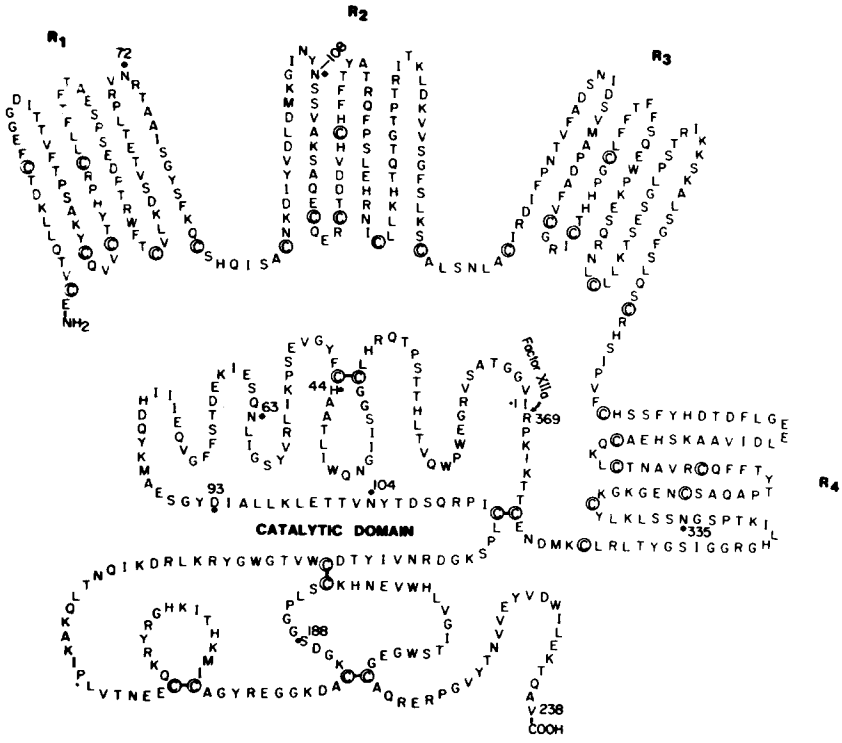


Fig. 1. Schematic representation of domain structures found in serine proteases of blood coagulation and fibrinolysis [courtesy Dr. E.W. Davie].

biology to obtain fusion or deletion proteins promises to provide interesting information in this regard.

Recently, the domain structures of three additional coagulation proteins, human factors XII, XI, and prekallikrein, have been elucidated by sequencing of the corresponding complimentary and genomic DNAs. Factor XII [46] contains domains homologous to analagous structures found in prothrombin, tissue plasminogen activator, urokinase, and plasminogen. Factor XI [47] and prekallikrein [48], by contrast, reveal homologous sequences not previously seen in any of the regulatory plasma proteases (Fig. 2). They contain two pairs of homologous tandem repeats which appear to have diverged from a common ancestor some 280 million years ago. The contribution of the noncatalytic chains of both factor XIa [49] and plasma kallikrein [50] to specificity have been experimentally demonstrated by comparing their activities under physiological conditions to those of the isolated catalytic chains, which were two to three orders lower than those of the complete molecules.



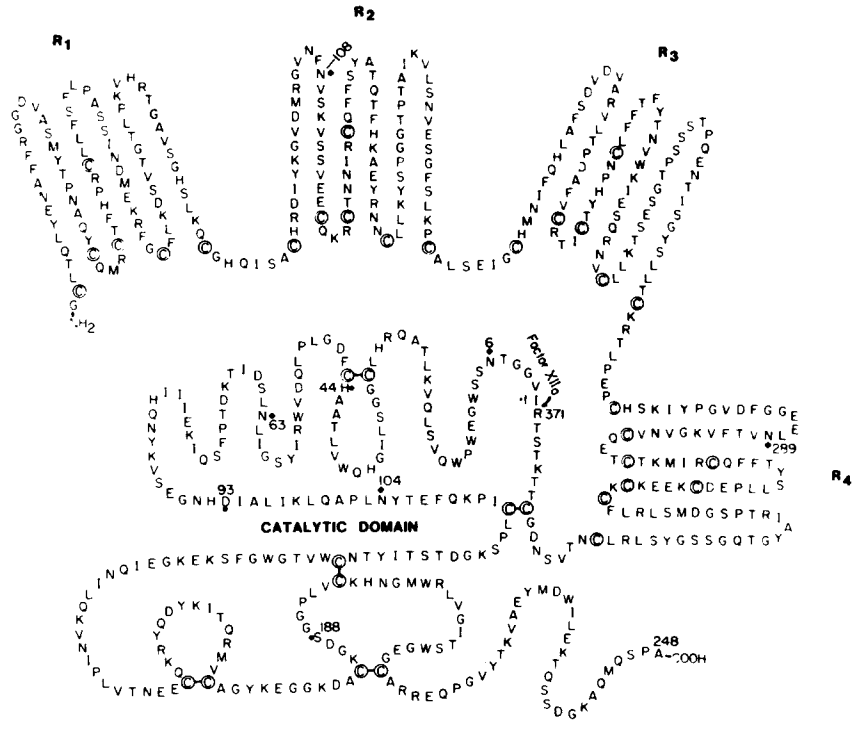


Fig. 2. Domain structures of blood coagulation factor XI and prekallikrein [taken from references 47, 48].

The nonprotease factors V and VIII, which are essential for the activities of factors Xa and IXa, respectively, are substrates for yet-another plasma protease, activated protein C, which does not lie on the direct pathway of blood coagulation [51–53]. It specifically inactivates factors V_a and $VIII_a$ and thus provides yet another redundant regulatory pathway in the blood coagulation cascade. Two other proteins, Z [54] and S [55], appear to be related to the activity of protein C but their functions are not fully understood.

With the exception of some nuclear magnetic resonance (NMR) data on kringle structures [56,57], no direct information bearing on the conformation of these regulatory proteases is available that would independently support the thesis of homologous domains. Comparison by computer simulation of the structures of prothrombin and factors IX and X with that of trypsin [58] indicates that they are mutually compatible and that additions and deletions occur in surface regions that would have no substantial influence on the global conformation of these molecules. This is in agreement with the comparison of the exon/intron splice junctions of the pancreatic serine proteases which also map in these regions [59]. The organization of the genomic DNAs of several plasma proteases and their exon/intron distribution also reveal a common pattern, which surprisingly shows that in all cases the functional amino acid residues of the active site occur in different exons [60]. In general, the case for homologous domains is being argued in terms of linear structures derived either from amino acid or DNA sequencing and clearly requires supporting documentation in terms of both conformation and function.

The Complement System [61,62]

The complement and the blood coagulation systems have certain common features: they are regulated by amplification of a signal which triggers the activation of zymogens by trypsinlike serine proteases, resulting in the programmed interaction of protein complexes along alternate pathways. However, the complement system is more complex and less completely characterized than the coagulation system. The present discussion will be limited to the early phase of the classical pathway, which is best understood, and to the corresponding steps of the alternate pathway. The two pathways converge at the activation of C5, just as the intrinsic and extrinsic pathways of blood coagulation converge at the activation of factor X. As in blood coagulation, the cascade is initiated by the interaction of a multicomponent complex with nonplasma components, in this case, C1 with antigen-antibody aggregates. The trigger is the autocatalytic activation of the zymogen of a trypsinlike enzyme, C1r, a homodimer, which together with C1s and C1q forms a calcium-dependent, pentameric aggregation product (Fig. 3). Activation of C1r occurs by crosswise intermolecular cleavage of an Arg-Ile bond resulting in the formation of two disulfide-bonded chains, a and b, the latter being the catalytic chain [63]. C1r activates only C1s, which in turn activates C2 and C4 but not C3 or C5, the zymogens that follow in the cascade. Both C1r and C1s form stoichiometric, inactive complexes with C1 inhibitor [64]. C1r and C1s are homologous proteins and their sequences are compatible with the three-dimensional structure of pancreatic chymotrypsin. The next major proteolytic activation step is the activation of C4 + C2 to form the heterodimer C4b2a which activates C3. C2 is homologous to protein B [65] of the alternate pathway and C4 is homologous to C3. C2, C4, and B all are components of the human major histocompatibility complex [66]. The activation of protein B constitutes a novel pattern of

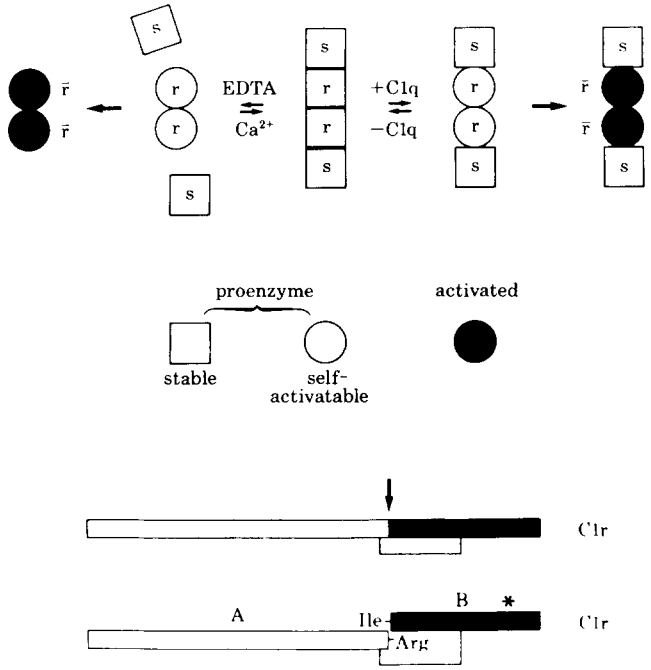


Fig. 3. Activation of complement component C1r as a function of its state of association. The serine protease domain is shown as filled symbols. The site of limited proteolysis is shown at the bottom diagram [taken from 61, 63].

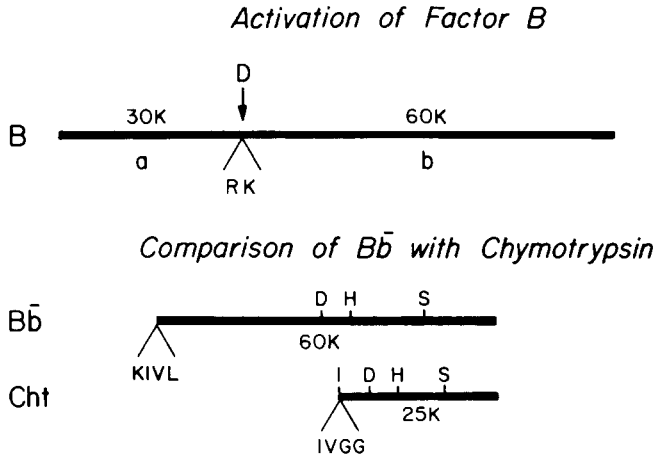


Fig. 4. Activation of complement factor B by factor D (arrow) and comparison of the serine protease domain Bb with that of chymotrypsin. For further details see the text.

zymogen activation. Initial cleavage by protein D produces two separate subunits, the catalytically inactive subunit a with a mass of 30 kD, and the catalytic subunit b, with a mass of 60 kD [65]. Although the carboxyl-terminal fragment of b is homologous to chymotrypsin, the molecule contains about twice as many amino acid residues (Fig. 4). The characteristic amino-terminal segment of chymotrypsin which forms an ion pair between Ile 16 and Asp 194 (chymotrypsinogen numbering system) is replaced by a much longer 222-residue domain which probably interacts with C3

while protein B is still in the zymogen form. The Bb subunit has no convertase activity of its own but only when associated with C3 in the C3B complex. Protein D, previously mentioned, is a serine protease which exhibits exclusive specificity toward an Arg-Lys bond in the C3B zymogen. The complement system, like the coagulation and fibrinolytic systems, is a treasure chest of evolutionary relics. The stems of the six flowers that represent C1q, contain collagenlike sequences near the amino terminus and bind to the C1r-C1s tetramer, to fibronectin and to the C1q receptor. All proteolytic domains are homologous to the pancreatic serine proteases [66]. C3 and C4 contain a thioester bond, as does alpha-2 macroglobulin in an identical Cys-Gly-Glu-Glu-tetrapeptide ring structure [61,67]. During activation of C4, the half-life of this relatively labile bond is reduced from several days to 0.1 ms, probably due to conformational change produced during zymogen activation [62]. The amino-terminal half of the C9 component contains a cystine-rich region with seven repeats of 40 amino acid residues which is homologous to the amino-terminal domain of the low density lipoprotein (LDL) receptor [68]. The exon/intron distribution of the protease domain of factor B is similar to that of the pancreatic serine proteases except for the presence of an additional exon (E) that has no counterpart in the other serine proteases [66] (Fig. 5).

TISSUE PROTEASES

In comparison to the plasma proteases, the proteolytic enzymes of tissues and cellular components constitute a neglected field of research. For the most part, they are considered to be impediments in the isolation of proteins, to be eliminated by inactivation by using mixtures of site-specific protease inhibitors. Notable exceptions are the lysosomal proteases, the neutral proteases of polymorphonuclear leukocytes, and mast cell proteases, all of which are associated with reactions to injury [69–71]. Little is known, however, about the normal physiological role of any of these. Most of these proteolytic enzymes are serine proteases but there are notable exceptions such as the granulocyte collagenase and mast cell carboxypeptidase A. The best-characterized tissue proteases are derived from granulocytes. The granulocyte serine proteases constitute one branch of the ancestral tree of the mammalian serine proteases [72]. Among these, elastase and cathespin G of polymorphonuclear leukocytes have

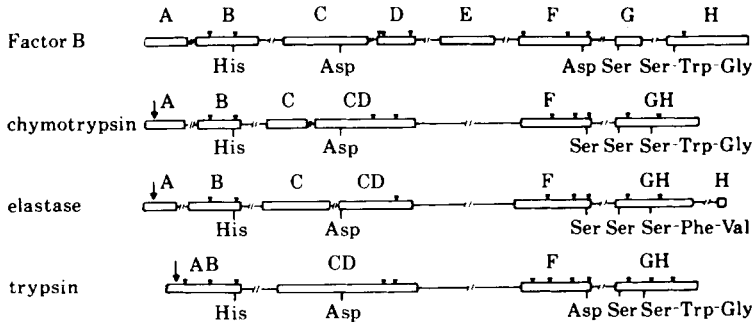


Fig. 5. Comparison of the gene organization of complement factor B with that of pancreatic serine proteases. Exons are shown as rectangles and numbered alphabetically, introns as connecting lines [taken from 66].

received considerable attention [73,74] because of their putative roles in inflammation and emphysema. However, relatively little is known about their normal structure/function relationship. Knowledge of the serine proteases of normal and "atypical" mast cells is more complete [72] although their physiological roles also need to be clarified. Rat tissue contains two types of mast cells: the typical mast cells largely segregated in connective tissue and the atypical mast cells found in mucosal tissue. Both types of mast cells have cell surface receptors specific for IgE, which causes degranulation, and both respond to many types of chemical, physical, and immunologic stimuli resulting from injury, infection, or inflammation. The major proteases isolated from these two types of mast cells are the chymotrypsinlike proteases RMCPI and RMCPII. A comparison of their properties is summarized in Table II. The amino acid sequence of RMCPII [75] and its x-ray structure [76] leave no doubt that it is a serine protease, related to chymotrypsin. Lacking a heparin-binding site, it is released by the mucosal mast cells following the immune-mediated expulsion of parasites from rats and during systemic anaphylaxis [77]. RMCPI, produced by peritoneal and tissue mast cells, is also a chymotrypsinlike enzyme, and partial amino acid sequence analysis (about 70%) shows extensive sequence homology to RMCPII. Unlike the latter, it is tightly associated with high molecular weight heparin, and under physiological conditions it is retained within the matrix even after degranulation. Current studies by R.G. Woodbury and collaborators in our laboratory are directed toward the identification of the physiological substrates of these mast cell proteases and of their precursors as evidenced by gene cloning and DNA sequencing.

CONCLUSIONS AND SUMMARY

We are witnessing a resurrection of research on proteolytic enzymes, engendered largely by the recognition of their vital role in the regulation of many biological

TABLE II. Properties of Rat Mast Cell Proteases RMCPI and RMCPII

Property	RMCPI	RMCPII
Specificity	Chymotrypsinlike	Chymotrypsinlike
Specific activity (BzTyr-OEt, Units/mg)	58.3	8.5
pH Optimum	8–9	8–9
Relative activity toward proteins	High	Low
Inhibition by		
DFP	Yes	Yes
TPCK	Yes	No
PMSF	Yes	No
Alpha-1 antiprotease	Yes	Yes
Extracted by	1 M KCl	0.15 M NaCl
Mr	26,000	24,600
Zymogen form	None	None
Structure	Homologous to RMCPII	35% sequence identity with chymotrypsin X-ray structure known
Possible functions	Protein degradation; regulation of level of vasoactive peptides	Degradation of basement member collagen; regulation of epithelial permeability?

processes. These regulatory proteases are more complex, more versatile, and also more specific than those digestive proteases from which they have evolved. Their greater complexity arises from the presence of noncatalytic domains which have no counterpart among the corresponding digestive proteases. Whereas the activation peptides of digestive proteases have no known functions of their own, and essentially are throw-away pieces, the much larger activation domains of the regulatory proteases confer recognition sites for the interaction with substrates and cofactors. This versatility in structure and function also affects the mechanism and pattern of zymogen activation. Although the primary event of peptide bond cleavage is reasonably well understood, the subtle conformational changes responsible for the generation of catalytic activity by no means follow a common pattern. In the case of trypsinogen and chymotrypsinogen, certain components of the binding sites required for the formation of the tetrahedral transition complex are formed during activation. In the case of pepsinogen, the activation peptide appears to shield the preexistent active site. In the case of the regulatory proteases, essentially no structural information is available to relate the proteolytic peptide bond cleavage of zymogens to the conformational changes required for catalytic activity.

Proteolytic processing plays a vital role in the lifetime of many proteins and is ideally suited to adapt a protein to its physiological environment. The response is essentially irreversible, rapid, and capable of amplifying a signal by consecutive zymogen activations. In this survey, typical examples have been given for processing systems, ranging from the early stages of biosynthesis of nascent polypeptide chains and their translocation, to the mature protein, be it a hormone, a growth factor, a component of the blood coagulation or the complement systems, or as-yet-unidentified physiological substrates of proteases in cells or granules. It is essential that the action of the processing proteases be both specific and of sufficiently short duration to avoid generalized destruction of surrounding tissues. Protein protease inhibitors play an important role in this regard. Although they are generally less specific than the proteases which they inhibit, their combination is more rapid compared to their synthetic low molecular weight analogs, essentially diffusion controlled, and largely irreversible. The association constants range from 10^6 to 10^{11} M^{-1} [78].

The greatest progress in recent years has come from the detailed studies of the primary structures of protein precursors and of the regulatory plasma proteases. Who would have anticipated finding in the nonprotease moieties of these proteases and their precursors domains homologous to those found in growth factors, in fibronectin, or in receptor proteins? What is the role and significance of the trypsinlike gamma subunit in mouse submaxillary gland nerve growth factor [79]? What are the functions of domains I and III in the regulation of the calcium-activated, papainlike tissue protease [80]? While these phenomena are best understood in terms of evolutionary relationships, arising from DNA splicing and shuffling, caution should be observed in ascribing to these substructures functional significance unless and until proven, probably most directly in combination with recombinant DNA techniques. While it would be beyond the scope of this discussion to elaborate current ideas of the origin of homologous protein domains and their role in protein evolution [3,13], prudence needs to be observed when relating one-dimensional patterns of protein and DNA sequences to three-dimensional structure and functions. Although it seems reasonable to assume that homologous catalytic domains of the various regulatory proteases have similar conformations, there are no prototypes of proven structure. The problem, in

a general sense, is that the so-called folding code that dictates the transition from the linear to the globular conformation [81] is a concept that still awaits elaboration to the point where the structure of a polypeptide chain can be predicted from its amino acid sequence. Hence homologous sequences suggest but do not prove homologous domains in the structural, let alone the functional, sense. In this connection it is worthy of note that identical short-range sequences can assume different conformations when part of different proteins [82].

Because so much of current information is derived from DNA sequencing, some thoughts are in order on the relation of the gene organization to protein domains. In proteolytic enzymes and their precursors, homologous domains usually correspond to homologous exons, as was found to be the case for the pancreatic serine proteases [4], the coagulation proteases [60], and certain complement components [66]. Exceptions have been noted—for instance, in the case of kringle 4 of plasminogen—and the general problem of the origin and function of introns remains an unsolved problem beyond the scope of the present discussion.

Following the principle of doing the easy things first, it is not surprising that most of the recent discoveries have emerged from the study of proteases that are more easily isolated and purified than others that are equally important but occur in trace quantities or are not readily extracted from the tissues of origin. How little we know about the proteases involved in differentiation, in tumor invasion and cell migration, in tissue remodelling, or in fertilization! Our knowledge of the physiological role of even those proteases that have been isolated and characterized, eg, neutrophil elastase, cathespin G, the mast cell proteases, or the proteases of macrophages is very fragmentary. Yet, the ultimate goal of biochemical research is to understand the normal physiological function of the molecules being studied, and hence I believe that more efforts should be directed to isolate and identify the regulatory proteases of tissues and to elucidate their action on their physiological substrates.

The ultimate fate of proteins is, of course, their proteolytic degradation. Considering the wealth of information on the mechanism and regulation of protein synthesis, it is indeed amazing how little we know about the regulation of protein degradation, the other side of the coin of protein turnover. Questions of protein traffic across membranes, protease segregation, substrate recognition, and resistance to proteolysis come into play at this stage of homeostasis and need to be examined as controlling factors in the regulation of protein degradation. One can only hope that future research, at long last, will address this problem.

ACKNOWLEDGMENTS

This paper is dedicated to Ephraim Katchalski-Katzir on his 70th birthday. The research carried out in this laboratory has been supported by the National Institutes of Health (GM-15731 and HL-36114).

REFERENCES

1. Neurath H: Fed Proc 44:2907, 1985.
2. Neurath H: Science 224:350, 1984.
3. Bajaj M, Blundell T: Ann Rev Biophys Bioeng 13:453, 1984.
4. Craik CS, Sprang S, Fletterick R, Rutter WJ: Nature 299:180, 1982.

5. Gilbert W: Science 228:823, 1985.
6. Flavell A: Nature 316:574, 1985.
7. Craik CS, Largman C, Fletcher T, Rocznik S, Barr PJ, Fletterick R, Rutter WJ: Science 228:291, 1985.
8. Bachovchin WW: Proc Natl Acad Sci USA 82:7948, 1985.
9. Wang D, Bode W, Huber R: J Mol Biol 185:595, 1985.
10. Vallee BL, Galdes A, Auld DS, Riordan JF: In Spiro TG (ed): "Zinc Enzymes" New York, JW Wiley: 1983.
11. James MNG, Sielecki AR: Nature 319:33, 1986.
12. Neurath H: In Jaenicke R (ed): "Protein Folding" Elsevier/North Holland Biochemical Press, Amsterdam, New York, p. 501 1980.
13. Neurath H: Chemica Scripta 26: (in press).
14. Mortensen SB, Sottrup-Jensen L, Hansen HF, Petersen TE, Magnusson S: FEBS Lett 135:295, 1981.
15. Zimmerman M, Mumford RA, Steiner DF (eds): "Precursor Processing in the Biosynthesis of Proteins." Ann NY Acad Sci 343:1–447, 1980.
16. Wickner WT, Lodish HF: Science 230:400, 1985.
17. Blobel G: Proc Natl Acad Sci USA 77:1496, 1980.
18. Lively MO, Walsh KA: J Biol Chem 258:9488, 1983.
19. Wolfe PB, Wickner WT, Goodman JM: J Biol Chem 258:12,073, 1983.
20. von Heijne G: J Mol Biol 173:243, 1984.
21. Perlman D, Halvorson HO: J Mol Biol 167:391, 1983.
22. Finkelstein AV, Bendzko P, Rapoport TA: FEBS Lett 161:176, 1983.
23. Rapoport TA: FEBS Lett 187:1, 1985.
24. Ghrayeb J, Lunn CA, Inouye S, Inouye M: J Biol Chem 260:10,961, 1985.
25. Kuhn A, Wickner W: J Biol Chem 260:15,914, 1985.
26. Kadonaga JT, Pluckthun A, Knowles JR: J Biol Chem 260:16,192, 1985.
27. Palmiter RD, Gagnon J, Walsh KA: Proc Natl Acad Sci USA 75:94, 1978.
28. Chretien M, Boileau G, Lazure C, Seidah NG: In Cantin M (ed): "Cell Biology of the Secretory Process." Basel, S. Karger: 214–246, 1984.
29. James R, Bradshaw RA: Ann Rev Biochem 53:259, 1984.
30. Rotwein P: Proc Natl Acad Sci USA 83:77, 1986.
31. Zumstein PP, Luethi C, Humbel RE: Proc Natl Acad Sci USA 82:3169, 1985.
32. de Pagter-Holthuizen P, van Schaik FMA, Verduijn GM, van Ommen GJB, Bouma BN, Jansen M, Sussenbach JS: FEBS Lett 195:179, 1986.
33. Isackson PJ, Ullrich A, Bradshaw RA: Biochemistry 23:5997, 1984.
34. Baldwin GS: Proc Natl Acad Sci USA 82:1921, 1985.
35. Yokota T, Lee F, Rennick D, Hall C, Arai N, Mosmann T, Nabel G, Cantor H, Arai K-I: Proc Natl Acad Sci USA 81:1070, 1984.
36. Land H, Schuetz G, Richter D: Nature 295:299, 1982.
37. Hook VYH, Loh YP: Proc Natl Acad Sci 81:2,776, 1984.
38. Hook VYH, Mezey E, Fricker LD, Pruss RM, Siegel RE, Brownstein MJ: Proc Natl Acad Sci USA 82:4,745, 1985.
39. Docherty K, Carroll RJ, Steiner DF: Proc Natl Acad Sci USA 79:4,613, 1982.
40. Kreil G, Haimi L, Suchanek G: Eur J Biochem 111:49, 1980.
41. Kurjan J, Herskowitz I: Cell 30:933, 1982.
42. Davie EW, Fujikawa K, Kurachi K, Kisiel W: Adv Enzymol 48:277, 1979.
43. Davie EW: In Colman RW, Hirsch J, Marder VJ, Salzman EW (eds): "Hemostasis and Thrombosis." 2nd Edition. JB Lippincott Co., in press.
44. Davie EW: Chemica Scripta 26: (in press).
45. Pedersen TE, Skorstengaard K: In McDonagh J (ed): "Fibronectin, Its Role in Coagulation and Fibrinolysis." Marcel Dekker, Inc, 1984.
46. Que BG, Davie EW: Biochem, 25:1525, 1986.
47. Fujikawa K, Chung DW, Hendrickson L, Davie EW: Biochem 25:2417, 1986.
48. Chung DW, Fujikawa K, McMullen BW, Davie EW: Biochem, 25:2410, 1986.
49. Van der Graaf F, Greengard JS, Bouma BN, Kerbiriou DM, Griffin JH: J Biol Chem 258:9669, 1983.

50. Van der Graaf F, Tans G, Bouma BN, Griffin JH: J Biol Chem 257:14,300, 1982.
51. Kisiel W: J Clin Invest 64:761, 1979.
52. Fernlund P, Stenflo J: J Biol Chem 257:12,170, 1982.
53. Foster D, Davie EW: Proc Natl Acad Sci USA 81: 4,766, 1984.
54. Hojrup P, Jensen MS, Petersen TE: FEBS Lett 184:333, 1985.
55. Dahlbaeck B, Lundwall A, Stenflo J: Abstracts XIIth Int Congress on Thrombosis and Haemostasis 331:56, 1985.
56. Banyai L, Varadi A, Patthy L: FEBS Lett 163:37, 1983.
57. Trexler M, Banyai L, Patthy L, Pluck ND, Williams RJP: FEBS Lett 154:311, 1983.
58. Furie B, Bing DH, Feldman RJ, Robison DJ, Burnier JP, Furie BC: J Biol Chem 257:3,875, 1982.
59. Craik CS, Choo QL, Swift GH, Quinto C, MacDonald RJ, Rutter WJ: J Biol Chem 259:14,255, 1984.
60. Roger J: Nature 315:458, 1985.
61. Porter RR, Lachman PJ, Reid KBM (eds): "Biochemistry and Genetics of Complement: Proc R Soc Discussion Meeting." London Royal Soc, 1984, 279–340.
62. Porter RR: CRC Crit Rev Biochem 16:1984.
63. Villiers CL, Arlaud GJ, Colomb MG: Proc Natl Acad Sci USA 82:4,477, 1985.
64. Salvesen GS, Catanese JJ, Kress LF, Travis J: J Biol Chem 260:2,432, 1985.
65. Gagnon J, Christie DL: Biochem J 209:51, 1983.
66. Campbell RD, Bentley DR, Morley BJ: in ref 61, p. 367.
67. Reid KBM, Porter RR: Ann Rev Biochem 50: 433, 1981.
68. Stanley KK, Kocher H-P, Luzio JP, Jackson P, Tschopp J, Dickson J: EMBO J 4:375, 1985.
69. Barrett AJ, McDonald JK (eds): "Mammalian Proteases," Vol 1. Academic Press, London, New York: 1980.
70. Barrett AJ (ed): "Proteinases in Mammalian Cells and Tissues." North Holland Publ Co., Amsterdam, New York: 1977.
71. Havemann K, Janoff A (eds): "Neutral Proteases of Human Polymorphonuclear Leukocytes." Urban & Schwarzenberg, Baltimore, Munich: 1978.
72. Woodbury RG, Neurath H: In Holzer H (ed): "Metabolic Interconversion of Enzymes." Springer Verlag, Berlin, New York: 1981, p. 145.
73. Barrett AJ: In Lorand L (ed): "Methods Enzymol," Vol 80 Part C., New York: 1980, p. 588.
74. Powers JC, Tanaka T, Harper JW, Minematsu Y, Barker L, Lincoln D, Crumley KV, Fraki JE, Schechter NM, Lazarus GG, Nakajima K, Nakashino K, Neurath H, Woodbury RG: Biochem 24:2,048, 1985.
75. Woodbury RG, Katunuma N, Kobayashi K, Titani K, Neurath H: Biochemistry 17:811, 1978.
76. Reynolds RA, Remington SJ, Weaver LH, Fisher RG, Anderson WF, Ammon LH, Matthews BJ: Acta Cryst B41:139, 1985.
77. Woodbury RG, Miller HRP, Huntley JF, Newlands GFJ, Palliser AC, Wakelin D: Nature 312:450, 1984.
78. Katunuma N, Umezawa H, Holzer H (eds): "Proteinase Inhibitors." Springer Verlag, Berlin, New York: 1983.
79. Thomas KA, Baglan NC, Bradshaw RA: J Biol Chem 256:9,156, 1981.
80. Ohno S, Emori Y, Imajoh S, Kawasaki H, Kisaragi M, Suzuki K: Nature 312:566, 1984.
81. Jaenicke R: Angewandte Chemie. Int Ed Engl 23:395, 1984.
82. Wilson IA, Hafi DH, Getzoff ED, Tainer JA, Lerner RA, Brenner S: Proc Natl Acad Sci USA 82:5255, 1985.