16. A. Renzini and M. Voli, *Astron. Astrophys.* **94**, 175 (1981); I. Iben, Jr., and A. Renzini, *Annu. Rev. Astron. Astrophys.* **21**, 271 (1983).
17. L. H. Aller, *Int. Astron. Union Symp.* **103**, 1 (1983); J. B. Kaler, *ibid.*, p. 245; M. Peimbert, in *Physical Processes in Red Giants* (Reidel, Dordrecht, Netherlands, 1981), p. 409.
18. S. Torres-Peimbert, in *Stellar Nucleosynthesis* (Reidel, Dordrecht, Netherlands, in press).
19. M. Peimbert, *Int. Astron. Union Symp.* **76**, 215 (1978); _____ and S. Torres-Peimbert, *ibid.* **103**, 233 (1983).
20. M. J. Barlow, S. Adams, M. J. Seaton, A. J. Willis, A. R. Walker, *ibid.*, p. 538.
21. M. Peimbert, *Int. Astron. Union Symp.*, in press.
22. A. Renzini, in *Stellar Nucleosynthesis* (Reidel, Dordrecht, Netherlands, in press).
23. D. Reimers and D. Koester, *Astron. Astrophys.* **116**, 341 (1982).
24. M. Peimbert, *Int. Astron. Union Coll.* **45**, 149 (1977).
25. L. Searle, *Astrophys. J.* **168**, 327 (1971); M. Peimbert, *Int. Astron. Union Symp.* **84**, 307 (1979); P. A. Shaver, R. X. McGee, L. M. Newton, A. C. Danks, S. R. Pottasch, *Mon. Not. R. Astron. Soc.* **204**, 53 (1983); H. E. Bond, in preparation.
26. R. J. Talbot, Jr., and W. D. Arnett, *Astrophys. J.* **186**, 51 (1973).
27. B. M. Tinsley, *Fundam. Cosmic Phys.* **5**, 287 (1980); B. E. J. Pagel and M. G. Edmunds, *Annu. Rev. Astron. Astrophys.* **19**, 77 (1981); B. A. Twarog, *Int. Astron. Union Symp.*, in press.
28. L. Searle and W. L. W. Sargent, *Astrophys. J.* **173**, 25 (1972).
29. R. B. Larson, *Nature (London) Phys. Sci.* **236**, 21 (1972).
30. R. J. Talbot, Jr., and W. D. Arnett, *Astrophys. J.* **190**, 605 (1974).
31. M. G. Edmunds and B. E. J. Pagel, *Mon. Not. R. Astron. Soc.* **185**, 77 (1978).
32. A. Serrano and M. Peimbert, *Rev. Mex. Astron. Astrofis.* **8**, 117 (1983).
33. R. E. Williams, *Astrophys. J.* **261**, L77 (1982).
34. B. A. Twarog, *ibid.* **242**, 242 (1980).
35. M. Tosi, *ibid.* **254**, 699 (1982).
36. I. F. Mirabel, *ibid.* **256**, 112 (1982); J. M. Van der Hulst *et al.*, *ibid.* **264**, L37 (1982); I. F. Mirabel and R. Morras, *ibid.*, in press.
37. D. A. Hunter, J. S. Gallagher, D. Rautenkranz, *Astrophys. J. Suppl. Ser.* **49**, 53 (1982).
38. J. S. Gallagher, D. A. Hunter, A. V. Tutukov, *Astrophys. J.*, in press.
39. J. I. Silk, *ibid.* **214**, 718 (1977); V. C. Reddish, *Stellar Formation* (Pergamon, Oxford, 1978), p. 208.
40. S. Adams, M. J. Seaton, I. D. Howarth, M. Aurriere, J. R. Walsch, *Mon. Not. R. Astron. Soc.*, in press.

Evolution of Proteolytic Enzymes

Hans Neurath

Complex molecules, such as the proteins, are the products of evolutionary processes in which, according to Jacob (*1*), "constraints, at every level, specify the rules of the game and define what is possible with these systems; and historical circumstances determine the actual course of events." Proteolytic enzymes, or proteases, are enzymes that catalyze the cleavage of peptide bonds in other proteins. They are presumed to have arisen in the earliest phases of biological evolution since even the most primitive organisms must have required them for digestion and for the metabolism of their own proteins. In fact, present digestive proteases can be shown to have a common ancestry with those of microbial origin, and they are considered to have originated some billion years ago (*2*). If one accepts the generally held view that biological evolution proceeded from the relatively simple to the more complex levels of organization, one would expect that in the course of evolutionary development, the very proteases that originally served purely digestive functions have adapted to the more specific and complex tasks of physiological regulation (*3*). In this process, they acquired a higher degree of specialization by restricting their action to a select number of peptide bonds located at specific sites in specific protein substrates. Such limited proteolysis does not destroy the protein substrate altogether but modifies its properties and physiological role. Many biological processes are regulated by this type of protease action, for example, blood coagulation and fibrinolysis, the release of protein hormones from precursor molecules, the transport of secretory proteins across membranes, the assembly of macromolecular structures such as collagen fibers or certain viruses, fertilization, and the control of proteolytic digestion itself (*3*). In this article, I examine the changes in molecular structure that have accompanied the evolution of proteolytic enzymes from primitive organisms, serving relatively primitive functions, to the diverse and more complex functions that they fulfill in higher organisms including man.

Summary. Proteolytic enzymes have many physiological functions, ranging from generalized protein digestion to more specific regulated processes such as the activation of zymogens, blood coagulation and the lysis of fibrin clots, the release of hormones and pharmacologically active peptides from precursor proteins, and the transport of secretory proteins across membranes. They are present in all forms of living organisms. Comparisons of amino acid sequences, three-dimensional structures, and enzymatic reaction mechanisms of proteases indicate that there are distinct families of these proteins. Changes in molecular structure and function have accompanied the evolution of proteolytic enzymes and their inhibitors, each having relatively simple roles in primitive organisms and more diverse and more complex functions in higher organisms.

Proteolytic enzymes are not only a physiological necessity but also a potential hazard, since, if uncontrolled, they can destroy the protein components of cells and tissues. Hence a discussion of protease evolution necessitates a consideration of the mechanisms whereby the action of proteolytic enzymes themselves is regulated. Two principal regulatory mechanisms, serving the same end, have been devised by nature: (i) activation of inactive protease precursors (zymogens) by limited proteolysis (*4*) and (ii) inactivation of proteases by forming complexes with protein inhibitors (*5*). The physiological importance of both mechanisms has been demonstrated in certain disease states that are related to deficiencies of functional zymogens or protease inhibitors, respectively. For instance, several types of familial hemophilias are due to deficiencies of one or the other of plasma proteases that normally activate zymogens in the blood coagulation cascade (*6*). In another case, familial emphysema, a genetic mutation in α_1-inhibitor impairs its secretion from the liver, resulting in abnormally low plasma concentrations and consequently in destruction of connective tissue of the pulmonary alveoli by leukocyte elastase (*7*).

Hans Neurath is Professor Emeritus in the Department of Biochemistry, University of Washington, Seattle 98195. This article is based on the text of the Margaret Hastings Distinguished Lecture, which was presented at the Scripps Clinic and Research Foundation, La Jolla, California, 22 April 1983.

Peptide Bond Cleavage

The basic function of proteases is the cleavage of peptide bonds. There are only a few ways in which this can be accomplished under physiological conditions. The most common mechanism is the polarization of the peptide bond by nucleophilic attack on the carbon-oxygen bond (either directly or mediated by a water molecule) assisted by the donation of a proton to the peptide nitrogen (Fig. 1). In proteolytic enzymes, certain amino acid residues fulfill the functions of nucleophiles and others act as proton donors. Each group or family of proteases has its characteristic set of such functional amino acid residues arranged in a particular configuration to form the active site (Table 1). Members of a family are generally believed to have evolved from a common ancestor (*8*). However, the serine proteases include two distinct families and these families presumably have evolved independently from each other; for example, the mammalian serine proteases (such as chymotrypsin) differ from certain bacterial serine proteases (such as subtilisin) in amino acid sequence and three-dimensional conformation to the extent that common ancestry is not likely (*9*). Analogously, the mammalian pancreatic carboxypeptidases and the bacterial enzyme thermolysin, each containing a catalytic indispensable zinc atom, have similar active-site configurations but bear no other structural or evolutionary relation to each other (*10*). Many other protease families exist, such as the aminopeptidases (*11*), the collagenases (*12*), and the calcium-activated proteases (*13*); however, classification in terms of structure and mechanism of action remain to be elucidated.

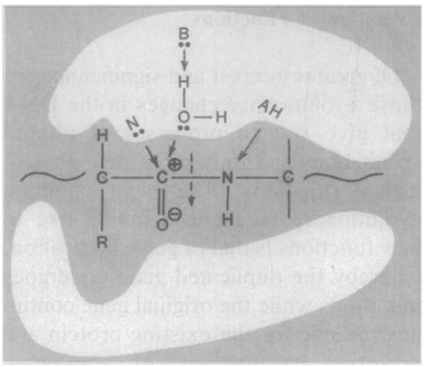


Fig. 1. Mechanism of peptide bond cleavage by attack on the carbonyl carbon atom C by a nucleophile, N: or by a base B: through a water molecule, assisted by the donation of a proton through a donor AH to the peptide nitrogen. In proteases, various amino acid side chains fulfill the functions of nucleophiles and proton donors (see Table 1).

Evolution

In general, proteins are said to have arisen from a common ancestor by divergent evolution if they have similar functions, similar active sites, similar amino acid sequences, and similar three-dimensional structures (*14*). Proteins that are the products of convergent or independent evolution also have similar functions and usually similar active sites, but they have dissimilar amino acid sequences and dissimilar three-dimensional structures. Examples of well-characterized proteins that are products of divergent evolution are the cytochrome c family (*15*), the globins (*16*), the mammalian serine proteases, and the more recently studied cyclic nucleotide-dependent protein kinases (*17*). Other proteins appear to be so distantly related to each other that similarities in amino acid sequences and three-dimensional structure are barely discernible. For instance, Matthews (*18*), Steitz (*19*), and their associates have determined the structure of two DNA binding proteins: the *cro* protein which acts as a repressor of bacteriophage lambda by preventing the transcription by RNA polymerase, and the catabolite activator protein (CAP) of *Escherichia coli* which activates transcription by RNA polymerase (but can also act as a repressor). These two proteins have dissimilar amino acid sequences and dissimilar three-dimensional structures except for a segment in each, composed of a domain of two helical segments, which shows striking structural homology. This is believed to represent the "active site" that binds to the DNA. Partial three-dimensional similarity also exists among three lysozymes that hydrolyze glycosidic linkages in polysaccharides. Thus hen egg-white lysozyme, bacteriophage T4 lysozyme, and, as recently reported (*20*), goose egg-white lysozyme show no similarities in amino acid sequence. They have a common partial structure, which includes the catalytically active site; but each has also a structural component that has no counterpart in any of the other species. Common ancestry or convergence could account for the observed relationships but the exact pathway of evolution is unknown.

Sometimes, discoveries of evolutionary relationships are the result of serendipity, as in the case of hen egg-white lysozyme and α-lactalbumin, whose sequences are homologous but their physiological functions differ (*21*). Lactalbumin has no catalytic activity of its own but alters the specificity of galactosyltransferase so as to produce lactose instead of *N*-acetyllactosamine. At times, proteins of similar functions are found in such widely different organisms that it is indeed surprising to find a homologous relation between them. An example is the recent discovery (*22*) that the plant enzyme papain and the mammalian lysosomal enzymes cathepsin B and H show a high degree of sequence identity. If we allow for deletions and insertions, the peptide chains of the two cathepsins seem to fit the known three-dimensional structure of papain. The evolutionary pathway leading from a common ancestor to the plant and mammalian enzymes

Table 1. Families of proteolytic enzymes.

Family*	Representative proteases	Characteristic active site residues†
Serine proteases I	Chymotrypsin (*27*) Trypsin (*27*) Elastase (*27*) Pancreatic kallikrein (*28*)	Asp^{102}, Ser^{195}, His^{57}
Serine proteases II	Subtilisin (*70*)	Asp^{32}, Ser^{221}, His^{64}
Sulfhydryl proteases	Papain (*71*) Actinidin (*72*) Rat liver cathepsins B and H (*22*)	Cys^{25}, His^{159}, Asp^{158}
Acid proteases	Penicillopepsin (*73*) Pepsin (*74*) *Rhizopus chinesis* (*75*) and *Endothia parasitica* (*76*), acid proteases Renin (*77*)	Asp^{33}, Asp^{213}
Metalloproteases I	Bovine carboxypeptidases A (*78*) and B (*79*)	Zn, Glu^{270}, Tyr^{248}
Metalloproteases II	Thermolysin (*80*)	Zn, Glu^{143}, His^{231}

*This table includes only enzymes of known amino acid sequence and three-dimensional structure, except for rat liver cathepsins B and H, for which the three-dimensional structure has been surmised by analogy to that of papain (*22*). †The numbering of residues corresponds to the amino acid sequence of the underlined enzymes listed in column 2.

is unknown. In other instances, proteins fulfill analogous function in various phyla and species, and the evolutionary relationships derived from structural analysis at the molecular level agree with the phylogenetic relationship derived from taxonomical evidence. This was originally shown to be true of cytochrome c (*15*), the myoglobin-hemoglobin group (*16*), and the fibrinopeptides (*23*); it also applies to proteolytic enzymes (*24*). More recently, Robert Zwilling of the University of Heidelberg and our laboratory have attempted to fill the gap in the evolutionary history of trypsin by determining the amino acid sequence of an analogous protease originating in the hepato-pancreas of the crayfish (*Astacus fluviatilis*), an invertebrate species considered to be on the evolutionary pathway from which decapode crustacea and mammals diverged some 700 million years ago (*25*). The amino acid sequence of the invertebrate enzyme shows approximately 40 percent identity with that of bovine trypsin on one hand and a trypsin of microbial origin (*Streptomyces griseus*) on the other. All functional amino acid residues of bovine trypsin are found in analogous positions in the invertebrate and bacterial enzymes. Hence, crayfish trypsin may be considered a missing link between prokaryotic and vertebrate serine proteases. Comparison of bovine and crayfish carboxypeptidase B reveals a similar relationship in amino acid sequence, suggesting that trypsin and carboxypeptidase B, acting in tandem, have evolved together at similar rates (*26*).

Evolution of Functions

Of greater interest and significance are those evolutionary changes in the DNA that give rise to proteases of altered specificities and thus serve new physiological functions. The most common evolutionary mechanism giving rise to new functions is that of gene duplication, whereby the duplicated gene undergoes mutations while the original gene continues to code for the existing protein and preserves its function. An alternative route of evolution is the rearrangement and fusion of gene segments, producing hybrids containing portions of the coding elements of two, often unrelated, proteins. A typical example of gene duplication is seen in the family of pancreatic serine proteases—trypsin, chymotrypsin, elastase, and thrombin. These four enzymes have evolved from the same archetype, and the differences in their specificity can be ascribed primarily to changes in the amino acid residues that line the so-called tosyl pocket, namely, the part of the active site that binds the amino acid side chain which contributes the carboxyl group to the susceptible peptide bond in protein substrates (*27*). Such substitutions are the result of point mutations in the genomic DNA. X-ray structure analyses of pancreatic trypsin, chymotrypsin, elastase (*27*), and that recently reported for pancreatic kallikrein (*28*)—a trypsin-like enzyme that releases vasodepressor peptides (kinins) from larger protein precursors—demonstrate three-dimensional homology among all of them.

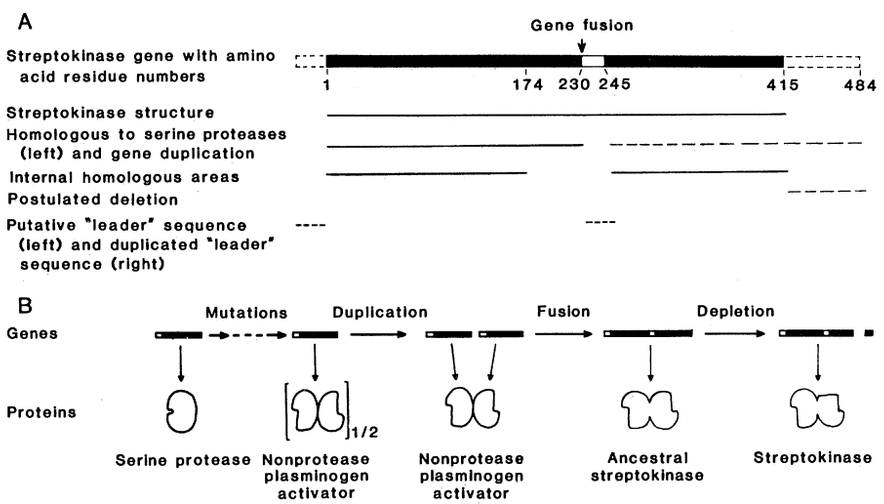


Fig. 2. Proposed scheme of the evolution of streptokinase gene and protein. (A) Homologous sequences (solid lines) and (B) the suggested evolution of the streptokinase gene and protein. The original structural gene, coding for the ancestral protease probably corresponded to the amino-terminal 230 residues of streptokinase. Duplication and fusion of the gene gave rise to an ancestral streptokinase and deletions from the carboxyl-terminal half of about 55 residues and of 15 residues from the interior produced a molecule similar to the current streptokinase molecule. The replacement of histidine at the active site by a nonfunctional glycine residue rendered streptokinase inactive as a serine protease. [Courtesy of *Biochemistry* (*45*)]

More profound changes in structure and specificity accrue from the addition or deletion of DNA segments to or from the original gene. This is seen generally when proteases that fulfill regulatory functions are compared to those serving primarily digestive purposes. Regulatory serine proteases usually contain two polypeptide chains instead of one. One of them displays sequence homology to digestive serine proteases (such as trypsin) and contains the catalytically functional amino acid residues. The other chain has no counterpart in the more primordial proteases. This is true of enterokinase, a large, regulatory protease of the brush border of the small intestine, whose sole function is the conversion of trypsinogen to trypsin. This enzyme recognizes the tetra-aspartyl sequence Val-Asp-Asp-Asp-Asp-Lys- - - in the amino terminus of trypsinogen and cleaves the lysyl bond during activation. The same bond can also be split by trypsin, but 1000 times more slowly because the same four aspartic acid residues that are required for the recognition by enterokinase have a retarding effect on trypsin (*29*). Two-chain structures are also characteristic of the regulatory proteases that participate in the blood coagulation process (as discussed below).

Regulatory serine proteases that have evolved from a common archetype vary in size and complexity. Analysis for homology indicates a common evolutionary origin, including the digestive pancreatic proteases, the hepatic blood coagulation proteases (*30*), and the more recently characterized granulocyte proteases, such as those found in leukocytes and mast cells (*31*). The acrosomal enzyme from the sperm (*32*) and certain proteases of the complement system (*33*) belong to the same family.

Zymogen Activation

One naturally occurring mechanism for inhibiting the activity of a nascent protease is the addition of an amino-terminal extension to the genome of the protease. Proteolytic cleavage of the peptide bond between the activation peptide and the enzyme essentially reverses this evolutionary addition and restores enzymatic activity (*4*). Although the basic principles of the chemical changes in covalent structure responsible for zymogen activation were established some 30 years ago (*34*), the fine details of the attendant conformational changes are still not fully understood. The structural changes that accompany the zymogen-enzyme conversion are

barely within the resolving power of instrumentation, such as x-ray analysis, nuclear magnetic resonance spectroscopy, or neutron scattering. The most that can be said in a generalized way is that the zymogen lacks the structural attributes required for formation of the enzyme-substrate complex or for the complex to reach the transition state (*35*).

The size, amino acid composition, and sequence of the activation peptides of homologous enzymes such as the serine proteases can vary over wide limits. The activation peptides of trypsins of various species are homologous, but those of the homologous serine proteases, chymotrypsin, elastase, and thrombin are different (*36*). Even in the case of the two rat elastases, which have 58 percent sequence identity, the corresponding activation peptides show no similarity to each other (*37*). In the extreme, the activation peptides of homologous serine proteases may be as short as six amino acid residues (for example, bovine trypsinogen), whereas in others they may exceed the size of the enzyme itself, as in prothrombin (320 residues) (*38*) or in plasminogen (560 residues) (*39*).

The physiological role of the activation peptide, if any, is not well understood. In many cases, as in trypsinogen, chymotrypsinogen, or proelastase, it is probably without function and once removed can be considered a throwaway piece. In the case of porcine procarboxypeptidase A, it has been claimed that the ~100-residue activation peptide constitutes a domain of its own and after isolation is a potent inhibitor of carboxypeptidase A (*40*). In the case of the more complex enzymes involved in blood coagulation or fibrinolysis, that is, thrombin, plasmin, and urokinase, the nonprotease regions of the respective zymogens are involved in interactions with cofactors required for the activation of the corresponding zymogens (*30*). Plasminogen, tissue plasminogen activator, and prothrombin contain triple-loop, three-disulfide-bridge structures in their nonprotease segments.

These so-called "kringles" (*38*) (five in plasminogen, two in prothrombin) show an even greater degree of homology to each other than do the corresponding proteases, and each kringle is believed to constitute an independent substructure, or domain (*41*). It has been suggested that in the genomic DNA the intervening sequences or introns occur at splicing junctions coding for domains (*42*). However, there is no relation of the 11 introns in the genomic DNA of prothrombin to the presumed domain structure of the zymogen (*43*). In another case, that of rat preproelastase (*44*), no correlation could be discerned between the intron-exon distribution of the genomic DNA and the domains circumscribed by the signal peptide, the activation peptide, and the enzyme, respectively.

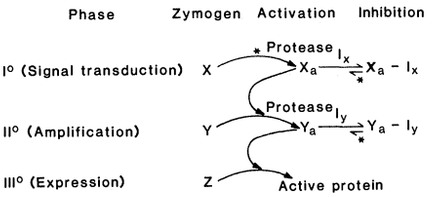


Fig. 3. Generalized scheme of consecutive zymogen activation reactions (cascade) and their control by proteases and protease inhibitors. The zymogens (X, Y, and Z) are activated sequentially in processes such as blood coagulation. Protease intermediates (X_a, Y_a) may be inactivated by specific protease inhibitors (I_x, I_y) to limit their action. A physiological signal (*) initiates the cascade by converting X to X_a. [Courtesy of the *Proceedings of the National Academy of Sciences U.S.A.* (*3*)]

While the activation of zymogens by the cleavage of a specific peptide bond is the rule, some zymogens become activated by forming strong stoichiometric complexes with nonprotease proteins, which, in a second, bimolecular reaction activate another zymogen molecule. Examples of that type are the interaction of plasminogen with streptokinase or staphylokinase, or the activation of prothrombin by reaction with staphylocoagulase (*39*). Although neither streptokinase nor staphylokinase is a protease, the amino-terminal 245 residues of the 415-residue polypeptide chain shows sequence homology to several serine proteases (*45*), including trypsin and *S. griseus* proteases A and B. Streptokinase is enzymatically inactive because His^{57} of the catalytic triad is replaced by glycine, whereas Asp^{102} and Ser^{195} are at the required positions. Streptokinase also shows sequence homology between the 173 amino-terminal residues and a carboxyl-terminal 162-residue region, suggesting that the molecule arose by gene duplication and fusion (in Fig. 2), which, in turn, suggests a mechanism of its evolution.

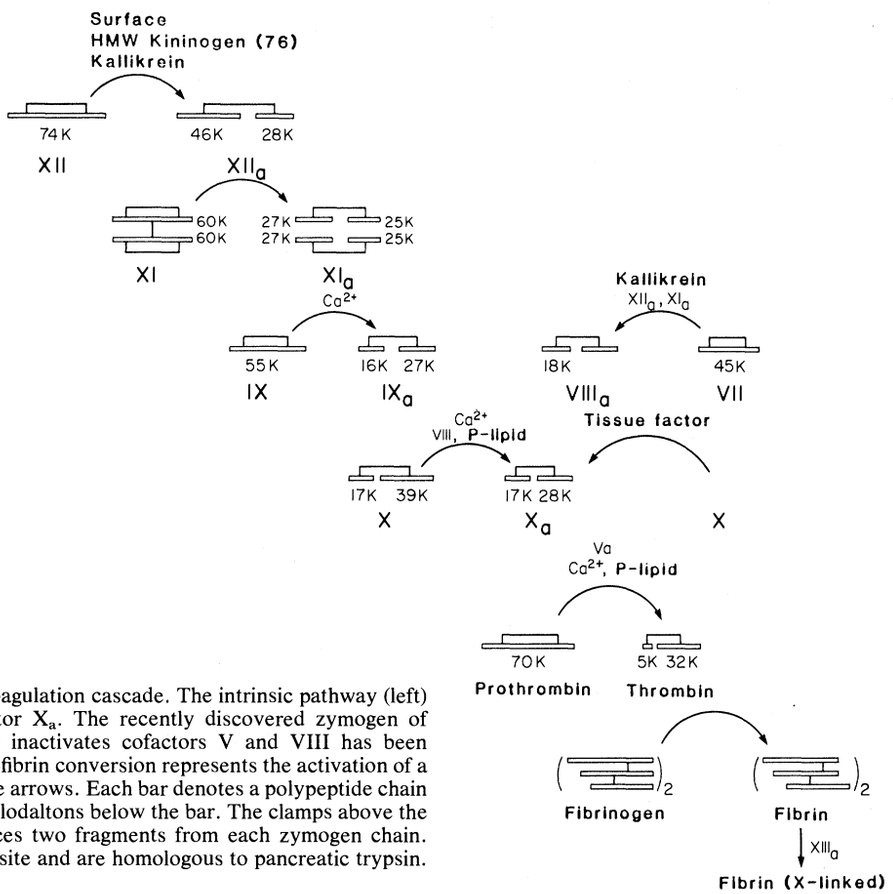


Fig. 4. Schematic representation of the blood coagulation cascade. The intrinsic pathway (left) and extrinsic pathway (right) converge at factor X_a. The recently discovered zymogen of protease C which upon activation specifically inactivates cofactors V and VIII has been omitted. Each reaction preceding the fibrinogen-fibrin conversion represents the activation of a zymogen of a serine protease, as indicated by the arrows. Each bar denotes a polypeptide chain proportional in length to its mass, indicated in kilodaltons below the bar. The clamps above the bars denote disulfide bonds. Activation produces two fragments from each zymogen chain. Those on the right-hand side contain the active site and are homologous to pancreatic trypsin. [Courtesy of Elsevier/North-Holland (*65*)]

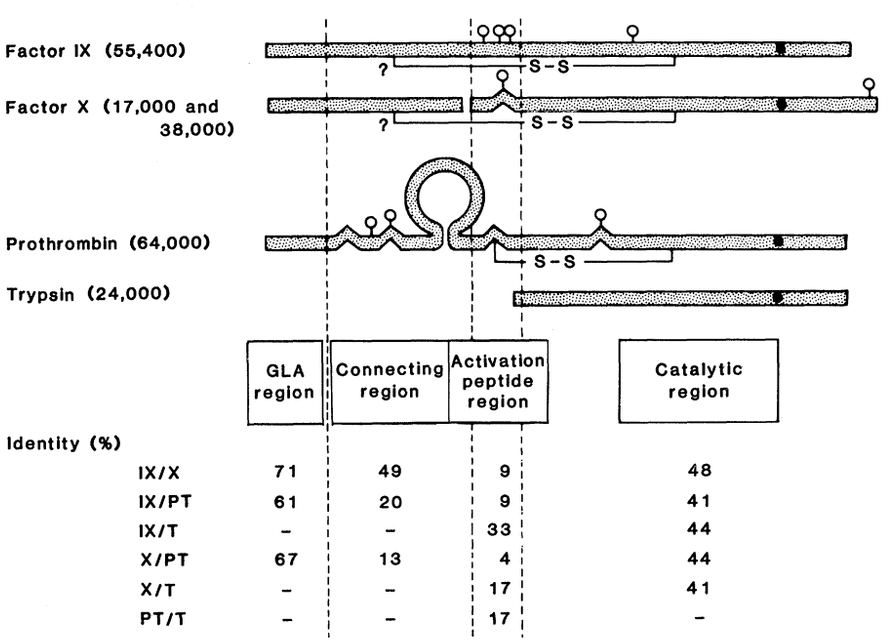
Cascades

Zymogen activation is of particular interest and physiological importance when it occurs in a series of consecutive reactions, or cascades. Such reactions occur with exquisite specificity in that the protease generated in the first step activates the zymogen in the next (*3*). Cascades constitute enzymatic amplification systems that can be further controlled by protease inhibitors that combine with the newly generated protease to assure that its action is limited to a pulse long enough to activate the next zymogen but short enough to avoid generalized proteolysis (Fig. 3). A typical example of a cascade is the process of blood coagulation (*30*). The intrinsic pathway, designated in this way because all the proteins and cofactors are found in the blood plasma, includes 11 zymogen activation reactions; the extrinsic pathway requires the participation of tissue factors and involves three consecutive zymogen activation reactions, plus the recently discovered zymogen of protease C, which upon activation, specifically inactivates cofactors V and VIII (*46*) (Fig. 4). The cascade is initiated in a complex reaction at solid surfaces. For example, the vascular basement membrane is the site of conversion of prekallikrein to kallikrein, which, in combination with the same complex, activates factor XII, the first zymogen in the cascade. The cascade terminates when fibrinogen is converted to fibrin, with subsequent formation of the fibrin clot. The specificity of the proteases that are generated is such that each protease can activate the next zymogen in the cascade but not the one following it. Were it otherwise, the succeeding step could be by-passed. It thus appears that the blood coagulation cascade has evolved from the bottom up because the insertion of each additional zymogen enhances the amplification and the fine-tuning of regulation.

Each of the proteases generated in the blood coagulation cascade is a serine protease related to pancreatic trypsin, except that, as previously noted, these coagulation proteases contain two covalently linked polypeptide chains, instead of one (Fig. 4). A comparison of the amino acid sequences of homologous zymogens, such as factors IX and X and prothrombin with that of pancreatic trypsinogen shows how functionally important regions of the polypeptide chains have preserved their identity (*47*) (Fig. 5). The highest degree of identity among these coagulation zymogens is observed in the amino-terminal region which contains the γ-carboxylglutamic acid residues that are essential for the binding of calcium ions (*48*), and in the carboxyl-terminal regions which include the components of the catalytic site. The sequences of the activation peptide and the "connecting region" have been preserved the least. If it is assumed that each homologous segment corresponds to a domain, that is, an independent folding segment of the polypeptide chain, and if these domains are integrated into the three-dimensional configuration of the active enzyme, it is possible to construct by computer graphics models of these three coagulation proteases and to compare them to that of trypsin. This has actually been done and the results suggest that homology in the linear sequence is accompanied by a close similarity of three-dimensional structure (*49*). The independent functional role of each of the two chains of regulatory proteases has been demonstrated for human plasma kallikrein (*50*) and for the light chain of activated human blood coagulation factor XI (*51*). In the former case, the chemically isolated light chain was as effective as the whole molecule in hydrolyzing synthetic substrates and in activating factor XI in solution. However, under physiological conditions, requiring solid surfaces, the whole molecule was 180 times more effective. In the case of activated human factor XI, the light chain was also as effective toward synthetic substrates as was the whole molecule, whereas under physiological conditions, requiring the presence of cal-

Table 2. Proteinase inhibitors in human plasma (*60*).

Inhibitor	Concentration (mg/100 ml)	Molecular mass (kD)	Polypeptide chains (No.)	Reactive sites (No.)
α_1-Proteinase	290	52	1	1
α_1-Antichymotrypsin	50	69	1	1
Inter-α_1-trypsin	50	160	1	2
α_2-Antiplasmin	7	70	1	1
Antithrombin III	24	65	1	1
C1-Inactivator	24	70	1	1
α_2-Macroglobulin	260	720	4	2



Identity (%)	GLA region	Connecting region	Activation peptide region	Catalytic region
IX/X	71	49	9	48
IX/PT	61	20	9	41
IX/T	-	-	33	44
X/PT	67	13	4	44
X/T	-	-	17	41
PT/T	-	-	17	-

Fig. 5. Schematic comparison of the zymogens of four mammalian (bovine) serine proteases, that is, factor IX, factor X, prothrombin, and trypsinogen. The length of each bar is proportional to the actual length of each polypeptide chain. Wrinkles and loops are included to indicate adjustments required to place homologous amino- and carboxyl-terminal regions in register. The circles above each bar indicate the sites of carbohydrate attachments and the filled circle near the right end, the position of the serine residue of the active site. Activation splits a peptide bond at the position marked by the rightmost vertical dotted line. Prior to activation, factor X contains two polypeptide chains connected by a disulfide bond; the other three zymogens are each single-chain proteins. The active enzymes are each disulfide-bonded two-chain proteins except for trypsin which in the native form contains a single polypeptide chain. Homologies in the various regions denoted by squared boxes are expressed as percent identities. For further interpretations, see the text. [Courtesy of the *Proceedings of the National Academy of Sciences U.S.A.* (*47*)]

cium ions, the light chain activated factor IX only 1/2000 as fast as the whole molecule and failed to interact with kininogen, an attribute entirely due to the heavy chain.

The cleavage of an amino-terminal extension peptide is not restricted to the activation of zymogens. Mammalian proteases and secretory proteins in general contain another amino-terminal extension, the so-called signal peptide or leader sequence that is required for the transport of nascent proteins across the membranes of the endoplasmic reticulum (*52*). The signal peptide specifies the destination, acting as a "zip code" for the nascent protein, and is removed on emergence of the polypeptide chain in the lumen by the action of a membrane-bound "signal peptidase" (*53*) of as yet unknown mechanism of action. The signal peptide appears to prevent the proper folding of the polypeptide chain and hence the "pre-protein" is physiologically inactive (*54*). There is no discernible sequence homology among the 100 or so signal peptides of known structure, suggesting that the signal peptide, like the activation peptide, has been acquired by the incorporation of extraneous DNA segments into the genome. An obvious common feature of signal peptides is the predominance of hydrophobic amino acid residues, which conceivably form a hydrophobic core or interact with the phospholipid bilayer of the membrane.

Protease Inhibitors

As indicated at the outset, protease inhibitors are essential for the protection of the organism against tissue damage by active proteases. Hence it is not surprising that protease inhibitors have been found in animal, plant, and microbial organisms (*5*). Less is known about the evolution of protease inhibitors than of the proteases themselves. This is so because the molecular structure of many of these inhibitors is as yet unknown and their specificity is less pronounced than that of the proteases with which they combine. The basic mechanisms of the protease inhibition by these proteins has been derived from studies of a few well-characterized inhibitors isolated from the pancreas (*5*), from bird eggs (*55*), and from certain legumes such as the soybean and the lima bean (*56*). In these instances, the inhibitors act as a pseudosubstrate by combining, essentially irreversibly, with the active site of the enzyme, thereby being converted into a modified form in which a single peptide bond, corresponding to the specificity requirements of the protease, is split

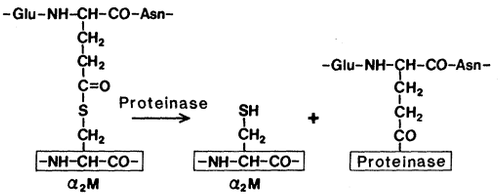


Fig. 6. Schematic representation of the reaction path of the thiol ester implicated in the trapping of a protease by α_2-macroglobulin. Shown on the left is a segment of one of the four polypeptide chains of α-macroglobulin, which forms a thiol ester between the γ-carboxyl group of glutamic acid with the thiol group of a cysteinyl residue. When the protease is trapped in the "bait" region, it undergoes a transacylation reaction and becomes covalently linked to the inhibitor that undergoes a conformational change which tightens the grip over the protease (*62*).

(*57*). It is likely, although not proved, that other protease inhibitors operate by a similar mechanism. Proteins that specifically inhibit chymotrypsin and carboxypeptidase, respectively, have been isolated from potatoes and tomatoes. They are expressed in response to a systemic signal, a polysaccharide, which these plants release when they are wounded by insects. The insects, in turn, become starved as a result of the inhibition of their own digestive proteases by the ingested plant protease inhibitor (*58*).

Bird eggs contain a strange assortment of single-headed and multiheaded inhibitors (inhibitors that can combine with one or more protease molecules at the same time). It is remarkable that, in contrast to the proteases which they inhibit, some of these inhibitors display an extraordinary variability in the amino acid residues at their "active site," that is, the region cleaved by the protease when it combines with the inhibitor (*55*). The reason for this high mutability at the active site is not understood. Yeast contains at least six different intracellular protease inhibitors of yeast proteases (*59*). They have been implicated as a safeguard against proteolytic damage that might result from leaky or broken vacuoles and as a regulatory control element of protease activity. Human and other mammalian plasmas contain a number of potent serum protease inhibitors (*60*) (Table 2). α_1-Proteinase inhibitor (also referred to as α_1-antitrypsin), already alluded to, is particularly effective against leukocyte elastase but also inactivates other serine proteases by forming stoichiometric complexes. This inhibitor is a glycoprotein having a molecular weight of 50,000 and containing a single polypeptide chain of known sequence. The complementary DNA (cDNA) of the protein has been cloned (*61*). The amino acid sequence of the protein shows striking homology to that of α_1-antichymotrypsin, and antithrombin III, and strangely enough also to that of ovalbumin which has no known relationship to proteolytic enzymes (*61, 62*). These homologies suggest that the plasma inhibitors, like the proteases which they inhibit, had a common ancestor. In fact, the evolutionary distance between α_1-protease inhibitor and antithrombin III—that is, 160 accepted point muta-

Table 3. Representative physiological reactions triggered by proteolytic enzymes (*3, 8, 64*).

Type of reaction	Example
Multiple cleavage of precursor "side chains"	Adrenocorticotropin
	Oxytocin-neurophysin
Transmembrane processes	Secretory pre-proteins
	Bacterial outer membrane and periplasmic proteins
	Bacterial toxins
Hormone precursors	Proinsulin
	Proglucagon
	Proparathyroid
Macromolecular assembly	Picornaviruses
	Collagen
	Fibrin
	Phage head protein
Zymogen activation	Pancreatic protease zymogens
	Blood coagulation (see Fig. 4)
	Complement (C1r, C1s, C3 convertase)
Fibrinolysis	Plasminogen proactivator
	Plasminogen
Release of physiologically active peptides	Angiotensinogen
	Kininogen
Development	Proacrosin
	Prococoonase
	Prochitin synthetase
	Procollagenase

tions per 100 residues—is the same as that which separates the respective enzymes elastase and thrombin. Similarly, the degree of sequence identity of α_1-protease inhibitor and α_1-antichymotrypsin is similar to that between the cognate enzymes trypsin and chymotrypsin (about 42 percent). This parallelism suggests that the specialization of the inhibitors has occurred in response to, and simultaneously with, that of the corresponding proteases.

Perhaps the most remarkable plasma protease inhibitor is α_2-macroglobulin, a protein of molecular weight 720,000 and composed of four nearly identical polypeptide chains (*63*). This inhibitor displays an almost universal specificity and a unique mode of action. When the protease is trapped in the so-called "bait region" it splits two of the four peptide chains near the carboxyl-third of the chain, near the amino side of a glutamic acid residue whose gamma carboxyl group forms a thiol ester with the SH group of a cysteinyl residue (*64*) (Fig. 6). The trapped enzyme then becomes covalently anchored to the inhibitor by a process of transacylation. The cleavage of the thioester bond is accompanied by a conformational change which tightens the grip of the inhibitor over the trapped protease. α_2-Macroglobulin thus acts as a scavenger of active plasma proteases.

Conclusions

The evolution of proteases from those with digestive functions to those with highly specialized, regulatory functions is essentially a restriction and refinement of the process of enzymatic peptide bond cleavage. Restricted or "limited" proteolysis can best be understood by considering the mutual fitting of the protein substrate to the configuration of the active site of the protease (*65*). The more stringent the requirements for productive interaction, the greater the selectivity of enzyme and the fewer the peptide bonds that will be cleaved. Zymogen activation and the cleavage of signal peptides are only two of the many kinds of physiological reactions that are triggered by limited proteolysis. Some of the more important ones are listed in Table 3, and in all known cases they produce a conformational change in the protein molecule, generating new functions. Peptide bond cleavage is only one of various posttranslational modifications that proteins undergo, but in many respects it is the most important one since it affects the delicate balance of interactions that determine the conformation of a protein molecule. Protein phosphorylation (*17*) is another widespread posttranslational modification that regulates physiological processes, but in contrast to proteolytic regulation, phosphorylation is a physiologically reversible process. The kinases participating in such cascades, like the proteases involved in the blood coagulation cascade, show sequence homology (*66*), a possible indication that, in part at least, they too have evolved from a common ancestor. The common denominator in each of these groups of proteins is the functional domain, for example, the catalytic site in the case of the serine proteases, or the cyclic nucleotide-binding site of the catalytic subunits in the case of the protein kinases. If a protein molecule is considered an assembly of domains, it is easy to conceive how proteins may be formed from combinations of structural DNA sequences, some of which code for the same functional domain such as the NAD-binding fold (NAD, nicotinamide adenine dinucleotide) in certain dehydrogenases (*67*) and thus display partial structural homology. Proteases, particularly those endowed with regulatory functions, usually recognize interdomain regions and direct their action on peptide bonds located between rather than within the domains themselves (*65*).

The processes of gene assembly, translation, and transcription are certainly far more complex than originally conceived, and the rules determining the distribution between sequences coding for domains and intervening sequences are not clearly understood (*42, 68*). The evolutionary forces that give rise to point mutations, gene duplication, and recombinations of gene segments, and thus to new proteins, are complex, but the end effect can well be likened to that of a tinkerer who, quoting again from Jacob (*1*), "does not produce innovations from scratch. It works on what already exists, either transforming a system to give it a new function or combining several systems to produce a more complex one" (*69*). Most of these rearrangements may not result in expression of a gene product but some may result in the formation of a protein sequence endowed with new cellular function.

References and Notes

1. F. Jacob, *The Possible and the Actual* (Pantheon, New York, 1982), pp. 25–46.
2. M. N. G. James, *Can. J. Biochem.* **58**, 251 (1980).
3. H. Neurath and K. A. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3825 (1976).
4. H. Neurath, *Adv. Protein Chem.* **12**, 320 (1957); *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **23**, 1 (1964).
5. M. Laskowski, Jr., and I. Kato, *Annu. Rev. Biochem.* **49**, 593 (1980).
6. O. D. Ratnoff, *Progr. Hemostasis Thromb.* **1**, 39 (1972).
7. R. W. Carell *et al.*, *Nature (London)* **298**, 329 (1982); V. J. Kidd, R. B. Wallace, K. Itakura, S. L. C. Woo, *ibid.* **304**, 230 (1983).
8. K. A. Walsh, in *Proteases and Biological Control*, E. Reich, D. B. Rifkin, E. Shaw, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1975), pp. 1–12.
9. B. W. Matthews, in *The Proteins*, H. Neurath and R. L. Hill, Eds. (Academic Press, New York, ed. 3, 1977), vol. 3, pp. 404–590.
10. M. K. Pangburn and K. A. Walsh, *Biochemistry* **14**, 4050 (1975); W. R. Kester and B. W. Matthews, *ibid.* **16**, 2506 (1977).
11. V. K. Hopsu-Havu and G. G. Glenner, *Histochemie* **3**, 197 (1966); V. K. Hopsu-Havu, P. Rintola, G. G. Glenner, *Acta Chem. Scand.* **22**, 299 (1968).
12. T. E. Cawston and G. Murphy, *Methods Enzymol.* **80**, 711 (1981).
13. L. Waxman, *ibid.*, p. 664.
14. For a review of sequence comparison of proteins and protein evolution, see R. F. Doolittle, *Science* **214**, 149 (1981); C. de Haen, H. Neurath, D. C. Teller, *J. Mol. Biol.* **92**, 225 (1975).
15. R. E. Dickerson and R. Timkovich, in *The Enzymes*, D. D. Boyer, Ed. (Academic Press, New York, 1976), vol. 11, p. 377.
16. M. O. Dayhoff, L. T. Hunt, P. J. McLaughlin, D. D. Jones, in *Atlas of Protein Sequence and Structure 1972*, M. O. Dayhoff, Ed. (National Medical Research Foundation, Washington, D.C., 1972), pp. 17–30.
17. E. G. Krebs, *Philos. Trans. R. Soc. London Ser. B* **302**, 3 (1983).
18. B. W. Matthews, D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, *Trends Biochem. Sci.* **8**, 25 (1983); M. Watson *ibid.*, p. 229; Y. Takeda, D. H. Ohlendorf, W. F. Anderson, B. W. Matthews, *Science* **221**, 1020 (1983).
19. D. B. McKay and T. A. Steitz, *Nature (London)* **290**, 744 (1981).
20. M. G. Gruetter, L. H. Weaver, B. W. Matthews, *ibid.* **303**, 828 (1983).
21. K. Brew, T. C. Vanaman, R. L. Hill, *J. Biol. Chem.* **242**, 3747 (1967).
22. K. Takio, T. Towatari, N. Katunuma, D. C. Teller, K. Titani, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3666 (1983).
23. R. F. Doolittle, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 2145 (1976).
24. H. Neurath, K. A. Walsh, W. P. Winter, *Science* **158**, 1638 (1967).
25. K. Titani, T. Sasagawa, R. G. Woodbury, L. H. Ericsson, H. Doersam, M. Kraemer, H. Neurath, R. Zwilling, *Biochemistry* **22**, 1459 (1983).
26. K. Titani, L. H. Ericsson, S. Kumar, F. Jacob, H. Neurath, R. Zwilling, in preparation.
27. B. S. Hartley, *Philos. Trans. R. Soc. London Ser. B* **257**, 77 (1970); D. M. Shotton and H. C. Watson, *ibid.*, p. 111; J. Kraut, *Annu. Rev. Biochem.* **46**, 331 (1977).
28. W. Bode, Z. Chen, K. Bartels, C. Kutzbach, G. Schmidt-Kastner, H. Bartunik, *J. Mol. Biol.* **164**, 237 (1983).
29. S. Maroux, J. Baratti, P. Desnuelle, *J. Biol. Chem.* **246**, 5031 (1971); J. W. Brodrick, C. Largman, M. W. Hsiang, J. H. Johnson, M. C. Geokas, *ibid.* **253**, 2737 (1978).
30. E. W. Davie and K. Fujikawa, *Annu. Rev. Biochem.* **46**, 799 (1977); _____, K. Kurachi, W. Kisiel, *Adv. Enzymol.* **48**, 277 (1979).
31. R. G. Woodbury and H. Neurath, *FEBS Lett.* **114**, 189 (1980).
32. W. Mueller-Esterl and H. Fritz, *Methods Enzymol.* **80**, 621 (1980).
33. K. B. M. Reid and R. R. Porter, *Annu. Rev. Biochem.* **50**, 433 (1981).
34. E. W. Davie and H. Neurath, *J. Biol. Chem.* **212**, 515 (1955); W. J. Dreyer and H. Neurath, *ibid.* **217**, 527 (1955).
35. D. A. Matthews, R. A. Alden, J. J. Birktoft, S. T. Freer, J. Kraut, *J. Biol. Chem.* **252**, 8875 (1977); M. A. Kerr, K. A. Walsh, H. Neurath, *Biochemistry* **15**, 5566 (1976).
36. H. Neurath, R. A. Bradshaw, R. Arnon, in *Structure-Function Relationships of Proteolytic Enzymes*, P. Desnuelle, H. Neurath, M. Ottsesen, Eds. (Munksgaard, Copenhagen, 1970), pp. 113–134.
37. R. J. MacDonald, G. H. Swift, C. Quinto, W. Swain, R. L. Pictet, W. J. Rutter, *Biochemistry* **21**, 1453 (1983).
38. S. Magnusson, T. E. Petersen, L. Sottrup-Jensen, H. Claeys, in *Proteases and Biological Control*, E. Reich, D. B. Rifkin, E. Shaw, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1975), pp. 123–150; C. M. Jackson and Y. Nemerson, *Annu. Rev. Biochem.* **49**, 765 (1980); D. Pennica *et al.*, *Nature (London)* **301**, 214 (1983).
39. S. Magnusson, L. Sottrup-Jensen, T. E. Peder-

sen, G. Dudek-Wojciechowska, H. Claeys, in *Proteolysis and Physiological Regulation*, D. W. Ribbons and K. Brew, Eds. (Academic Press, New York, 1976), pp. 189–202; L. Sottrup-Jensen, H. Claeys, M. Zajdel, T. E. Petersen, S. Magnusson, in *Progress in Chemical Fibrinolysis and Thrombosis*, J. F. Davidson, R. M. Rowman, M. Samama, P. C. Desnoyers, Eds. (Raven, New York, 1978), vol. 3, pp. 191–209; F. J. Castellino, *Trends Biochem. Sci.* **4**, 1 (1979).
40. B. San Segundo, M. C. Martinez, C. M. Cuchillo, F. X. Aviles, *Biochim. Biophys. Acta* **707**, 194 (1982); F. X. Aviles, B. San Segundo, M. Vilanova, C. M. Cuchillo, C. Turner, *FEBS Lett.* **149**, 257 (1982).
41. M. Trexler and L. Patthy, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2457 (1983); M. Trexler, L. Banyai, L. Patthy, N. D. Pluck, R. J. P. Williams, *FEBS Lett.* **154**, 311 (1983).
42. W. Gilbert, *Nature (London)* **271**, 501 (1978); C. C. F. Blake, *ibid.* **273**, 267 (1978); *ibid.* **277**, 598 (1979); C. S. Craik, S. Sprang, R. Fletterick, W. J. Rutter, *ibid.* **299**, 180 (1982).
43. S. J. Friezner Degen, R. T. A. MacGillivray, E. W. Davie, *Biochemistry* **22**, 37 (1983).
44. C. Quinto *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 31 (1982).
45. K. W. Jackson and J. Tang, *Biochemistry* **21**, 6620 (1982).
46. W. Kisiel, W. M. Canfield, L. H. Ericsson, E. W. Davie, *ibid.* **16**, 5824 (1977); G. A. Vehar and E. W. Davie, *ibid.* **19**, 401 (1980).
47. K. Katayama *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4994 (1979).
48. P. Fernlund, J. Stenflo, A. Tufvesson, *ibid.* **75**, 5889 (1978).
49. B. Furie, D. H. Bing, R. J. Feldmann, D. J. Robison, J. P. Burnier, B. C. Furie, *J. Biol. Chem.* **257**, 3875 (1982).
50. F. van der Graaf, G. Tans, B. N. Bouma, J. H. Griffin, *ibid.*, p. 14300.
51. F. van der Graaf, J. S. Greengard, B. N. Bouma, D. M. Kerbiriou, J. H. Griffin, *ibid.* **258**, 9669 (1983).
52. G. Kreil, *Annu. Rev. Biochem.* **48**, 317 (1981); M. Zimmerman, R. A. Mumford, D. F. Steiner, Eds., *Precursor Processing in the Biosynthesis of Proteins* (vol. 343, *Annals of the New York Academy of Sciences*, New York, 1980).
53. M. O. Lively and K. A. Walsh, *J. Biol. Chem.* **258**, 9488 (1983).
54. S. N. Thibodeau, R. D. Palmiter, K. A. Walsh, *ibid.* **253**, 9018 (1978); G. Scheele and R. Jacoby, *ibid.* **258**, 2005 (1983).
55. M. Laskowski, Jr., I. Kato, T. R. Leary, J. Schrode, R. W. Sealock, in *Proteinase Inhibitors*, H. Fritz, H. Tschesche, L. J. Greene, E. Truscheit, Eds. (Springer-Verlag, New York, 1974), pp. 597–614.
56. C. A. Ryan, *Annu. Rev. Plant Physiol.* **24**, 173 (1973); Y. Birk, *Methods Enzymol.* **45**, 695 (1976).
57. D. Kowalski, T. R. Leary, R. E. McKee, R. W. Sealock, D. Wang, M. Laskowski, Jr., in *Proteinase Inhibitors*, H. Fritz, H. Tschesche, L. J. Greene, E. Truscheit, Eds. (Springer-Verlag, New York, 1974), pp. 311–324.
58. C. Nelson, M. Walker-Simmons, D. Makus, G. Zuroske, J. Graham, C. A. Ryan, in *Mechanism of Plant Resistance to Insects*, P. Hedin, Ed. (ACS Monograph, 1983), p. 103; P. Bishop, D. J. Markus, G. Pearce, C. A. Ryan, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3536 (1981).
59. H. Holzer, in *Metabolic Interconversion of Enzymes 1975*, S. Shaltiel, Ed. (Springer-Verlag, New York, 1975), pp. 168–174.
60. J. Travis and G. S. Salvesen, *Annu. Rev. Biochem.* **52**, 655 (1983).
61. K. Kurachi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6826 (1981).
62. T. Chandra, R. Stackhouse, V. J. Kidd, K. J. H. Robson, S. L. C. Woo, *Biochemistry*, in press.
63. L. Sottrup-Jensen, T. E. Petersen, S. Magnusson, *FEBS Lett.* **121**, 275 (1980); S. R. Mortensen, L. Sottrup-Jensen, H. F. Hansen, S. Magnusson, *ibid.* **135**, 295 (1981).
64. A similar thiol ester occurs in the C3, C4, and C5 components of complement [J. B. Howard, *J. Biol. Chem.* **255**, 7082 and 7084 (1980)].
65. H. Neurath, in *Protein Folding*, R. Jaenicke, Ed. (Elsevier/North-Holland, New York, 1980), pp. 501.
66. W. C. Barker and M. O. Dayhoff, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2836 (1982); K. Titani, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **42**, 2256 (1983); S. Shoji, L. H. Ericsson, K. A. Walsh, E. H. Fischer, K. Titani, *Biochemistry* **22**, 3702 (1983).
67. M. G. Rossman, in *The Enzymes of Glycolysis: Structure, Activity and Evolution*, C. C. F. Blake, D. C. Phillips, H. C. Watson, Eds. (Royal Society of London, London, 1981), p. 191.
68. W. F. Bodmer, in *ibid.*, p. 173; J. S. Richardson, *Adv. Protein Chem.* **34**, 168 (1981).
69. M. Eigen, *Naturwissenschaften* **58**, 465 (1971).
70. J. D. Robertus, R. A. Alden, J. J. Birktoft, J. Kraut, J. C. Powers, P. E. Wilcox, *Biochemistry* **11**, 2439 (1972); J. D. Robertus, J. Kraut, R. A. Alden, J. J. Birktoft, *ibid.*, p. 4293; D. A. Matthews, R. A. Alden, J. J. Birktoft, S. T. Freer, J. Kraut, *J. Biol. Chem.* **256**, 8875 (1977).
71. J. Drenth, J. N. Jansonius, R. Koekoek, B. G. Walthers, *Adv. Protein Chem.* **25**, 79 (1971).
72. E. N. Baker, *J. Mol. Biol.* **115**, 263 (1977).
73. M. N. G. James and A. R. Sieleck, *ibid.* **163**, 299 (1983).
74. J. Tang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3437 (1973); N. S. Andreeva, A. A. Fedoroc, A. E. Gushchina, R. R. Riskulov, N. E. Shutskever, M. G. Safro, *Mol. Biol.* **12**, 922 and 936 (1978).
75. E. Subramanian *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 556 (1977).
76. R. Bott, E. Subramanian, D. R. Davies, *Biochemistry* **21**, 6596 (1982).
77. T. Blundell, B. L. Sibanda, L. Pearl, *Nature (London)* **304**, 273 (1983).
78. R. G. Bradshaw, L. H. Ericsson, K. A. Walsh, H. Neurath, *Proc. Natl. Acad. Sci. U.S.A.* **63**, 1389 (1969); B. L. Vallee, J. F. Riordan, D. S. Auld, S. A. Latt, *Philos. Trans. R. Soc. London Ser. B* **257**, 215 (1970); J. A. Hartsuck and W. N. Lipscomb, in *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, ed. 3, 1971), p. 1; B. L. Vallee, A. Glades, D. S. Auld, J. F. Riordan, in *Zinc Enzymes*, T. G. Spiro, Ed. (Wiley, New York, 1983), p. 26; W. N. Lipscomb, *Annu. Rev. Biochem.* **52**, 17 (1983).
79. K. Titani, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1666 (1975); M. F. Schmid and J. R. Herriott, *J. Mol. Biol.* **103**, 175 (1976).
80. K. Titani, M. A. Hermodson, L. H. Ericsson, K. A. Walsh, H. Neurath, *Nature (London) New Biol.* **80**, 35 (1972); B. W. Matthews, J. N. Jansonius, P. M. Colman, B. P. Schoenborn, D. Dupourque, *Nature (London)* **238**, 37 (1972); W. R. Kester and B. W. Matthews, *J. Biol. Chem.* **252**, 7704 (1977); *Biochemistry* **16**, 2506 (1977); M. G. Holmes and B. W. Matthews, *ibid.* **20**, 6912 (1982).
81. Supported by NIH grant GM-15731. I thank Drs. K. A. Walsh and E. W. Davie for valuable suggestions and critical reading of the manuscript.

Biotechnology as an Intellectual Property

Reid G. Adler

One of the most significant issues created by the emergence of modern biotechnology has been the legal characterization and treatment of biotechnological industrial products. Advances in most other technologies have been readily assimilated by the patent system and routinely licensed and marketed. Because of the tremendous potential impact of biotechnology on many diverse areas, however, it has received an unusual amount of attention and generated a variety of public policy issues and legal uncertainties. This article focuses on biotechnology as an intellectual property.

The term property is generally associated with physical objects, such as household goods or land for which ownership and associated rights are guaranteed and protected by the government. Intellectual property, on the other hand, is intangible. It includes patents, trade secrets, copyrights, and trademarks—rights (which can be bought, sold, or licensed) to exclude others from making, copying, or in some instances using or selling tangible embodiments of the proprietary subject matter.

Although microorganisms have been used for industrial purposes such as baking or fermenting for millennia (*1*), the recent use of restriction enzymes to create recombinant DNA has fueled interest in developing genetic engineering techniques and encouraged the creation of a host of new processes and products. The characterization of these research results as intellectual properties encourages industry to allocate labor, research and development, and funding to facilitate the production of commercially marketable items. As is similarly becoming evident in several other areas, including gene therapy (*2*) and environmental dissemination of organisms (*3*), biotechnology as an intellectual property has also challenged legal and public policies and will continue to catalyze change for several years.

Reid G. Adler is an attorney and is a technical law clerk to the Honorable Giles S. Rich, U.S. Court of Appeals for the Federal Circuit, Washington, D.C. He also teaches "Science and the Law: Biotechnology" at the Graduate School at the National Institutes of Health.