Evolution of Structure and Function of Proteases

Amino acid sequences of proteolytic enzymes reflect phylogenetic relationships.

Hans Neurath, Kenneth A. Walsh, William P. Winter

Most of the information on the relation between the structure and function of proteolytic enzymes that has been obtained over the years is based on observations derived from a few well-defined enzymes believed to be representative of groups. Thus, bovine pancreatic trypsin and chymotrypsin have served as models for the "serine proteases," bovine carboxypeptidase A as a model for the "metallo peptidases," papain as a representative of the "sulfhydryl enzymes," and pepsin as a model of the "acid proteases" (*1*). The last few years have seen spectacular advances in the characterization of the structure and function of several of these proteolytic enzymes. We have seen the completion of the amino acid sequence of bovine trypsinogen (*2, 3*) chymotrypsinogen A (*4*) and B (*5*), of two strains of subtilisin (*6*) and of papain (*7*). Sequence analyses of other proteolytic enzymes are nearing completion. The development of "site-specific" reagents has led to the identification of amino acid residues of the active site and has provided a means for the chemical modulation of the enzymatic specificity or activity of several of these enzymes, such as the carboxypeptidases (*8*), the chymotrypsins (*9*), and the subtilisins (*10*). X-ray crystallography has recently yielded a three-dimensional model of the structure of one proteolytic enzyme, chymotrypsin (*11, 12*), and promises to give us a detailed understanding of the mode of the interaction of the active site with specific substrates and inhibitors.

A new parameter has recently been introduced by considerations of the evolution of enzymes and their phylogenetic variations. Thus enzymes, or for that matter proteins, which are believed to fulfill similar catalytic functions have been found to show certain similarities in structure. This approach is well documented (*13*) by the extensive investigations on cytochrome *c*, the hemoglobins, various dehydrogenases, phosphoglucomutases, and the aldolases. Conversely, it is possible to deduce the structural requirements for function from considerations of the evolutionary patterns and phylogenetic variations of classes of enzymes. This approach appears particularly promising in the case of the proteolytic enzymes, since, by and large, they contain no prosthetic groups and thus their function is directly coded in the amino acid sequence.

The phylogenetic relationships among enzymes have been considered in terms of "analogies" and "homologies." In applying these terms to proteolytic enzymes, it seems necessary to clarify their meaning.

Analogy

The term analogy is used to denote similarities in function without regard to structure. It is important, however, to recognize several independent types of analogy, based on the functional comparison that is being made. Within the context of the present discussion, three types of analogy could be considered: (i) analogy in enzymatic function (such as specificity); (ii) analogy in enzymatic mechanisms; and (iii) analogy in biological properties.

Enzymatic function relates to the nature of the bond being split and therefore is an expression of the specificity of the enzyme. Representative examples are given in Table 1. Analogy in enzymatic mechanism relates to those chemical features of the active site that are directly concerned with the bond-breaking mechanism. The serine-, metallo-, sulfhydryl-, and acid-proteases (*1*) represent the four most commonly recognized classes of proteolytic enzymes operating by analogous mechanisms (Table 1). Finally, it is possible to compare enzymes on the basis of their biological properties, such as tissue of origin, site of action, or physiological function. This definition cannot be formulated as precisely as the preceding ones because a more complex level of organization is involved (Table 1).

Enzymes that are analogous by one criterion need not be analogous by another. For instance, trypsin and cocoonase (*14*), an enzyme involved in the escape of certain moths from their cocoons, have analogous enzymatic functions and mechanisms, but obviously different biological functions. Conversely, pepsin and chymotrypsin have analogous biological and enzymatic functions, but operate by nonanalogous mechanisms. Thus, in one case nature has used enzymes that are analogous both in specificity and mechanism to fulfill different biological functions, whereas in another, different enzymatic mechanisms serve similar biological functions.

Homology

The term homology as applied to proteins refers to similarity in amino acid sequence. The only completely satisfactory criterion for homology is a comparison of the complete amino acid sequence of two proteins for areas of identity. As a first approximation, comparison of short elements of sequence is often used. Such restricted comparison becomes more meaningful if it can be demonstrated that the sequences being compared represent identical areas in the linear structure of these two proteins. In either event, the reliability of such a comparison must be interpreted on a statistical basis lest we misinterpret random similarities (*15*). Similarities in peptide maps was the first evidence for homologies among proteins (*16*) and is still a potent analytical criterion. Comparison of the

* Dr. Neurath is chairman and professor of biochemistry at the department of biochemistry, University of Washington, Seattle, 98105. Dr. Walsh is associate professor of biochemistry, University of Washington; Dr. Winter is investigator at the Howard Hughes Medical Institute department of biochemistry, University of Washington.

amino acid composition of two proteins is the most expedient experimental procedure, but it is the least reliable criterion of homology. Therefore, while amino acid composition as such is no proof of homology, it is becoming increasingly evident that enzymes or proteins which are believed to be homologous do have similar amino acid composition.

Whereas any of the three types of analogy already mentioned may be suggestive of homology, analogy by itself is not diagnostic of homology. For instance, while acetyltyrosine ethyl ester is usually considered a good and characteristic substrate for chymotrypsin and in fact has led to the search for and demonstration of homology between chymotrypsin A and B (*5*), the choice of a less specific substrate might lead to erroneous conclusions. Thus, the hydrolysis of *p*-nitrophenylacetate is catalyzed by chymotrypsin (*17*), carbonic anhydrase (*18*), as well as glyceraldehyde-3-phosphate dehydrogenase (*19*), but, in fact, these three enzymes have no other known structural or functional relation to each other.

Table 1. Analogies among proteases.

Representative groups	Examples
Enzymatic function	
Aminopeptidases	Leucine aminopeptidase
Carboxypeptidases	Carboxypeptidase A
Endopeptidases	Trypsin, papain, pepsin
Enzymatic mechanism	
Serine enzymes	Trypsin
Metallo enzymes	Carboxypeptidase A
Sulfhydryl enzymes	Papain
Acid proteases	Pepsin
Biological properties	
Digestive enzymes	Trypsin, carboxypeptidases
Enzymes for intracellular protein degradation	Cathepsins
Circulatory enzymes	Thrombin, plasmin, renin

Categories of Homologous Proteins

It is useful to divide observed homologies into two categories on the basis of the relationship between the individuals from which the proteins were derived. When these individuals are members of the same species, or in the special case where only one individual is involved, the homology is said to be *intraspecial*. When the individuals are members of distinct species, the homology is *interspecial*.

In intraspecial homology, it is possible to determine by conventional genetic studies whether the proteins are products of the same gene locus or of different loci, thus providing an additional criterion for classification. The equivalent subdivision is not possible in interspecial homologies because true species are noninterbreeding, precluding the possibility of genetic studies. However, a useful subdivision may be made on the basis of functional analogy. Representative examples of documented homologies are shown in this context in Table 2.

Homologous proteins obtained from the same species and derived from the same locus are well exemplified by the allotypic β-chains of hemoglobins A and S, which are the products of two allelic genes (*20*). No example is known of the second type in this category (Table 2), although homologies of this type doubtlessly occur. This situation could arise if a somatic mutation occurred in a developing organ. In that individual, two types of some protein would be produced by that organ, but the aberration would not be transmitted to the offspring. Somatic mutations may contribute to the apparent microheterogeneity that has been evident in certain sequence determinations.

Among homologous proteins from the same species, but produced at different loci, trypsin and chymotrypsin (*2*) are probably the best-documented examples. The recently characterized homology between chymotrypsins A and B (*5*) also belongs to this group.

The subclassification of interspecial homology is more equivocal than that of intraspecial homology because of the necessity for involving functional analogy as a basis. The best-known example of analogous interspecial homology is that of cytochrome *c* (*13*). In this case one has no difficulty in determining that the proteins from two species are analogous. Similarly, there is little question that porcine elastase and bovine trypsin (*4*) are nonanalogous. However, it is impossible, on this basis, to satisfactorily categorize the presumed homology between cocoonase (*14*) and trypsin, since the proteins appear to be chemically related in all ways but are biologically distinct, serving the animal in different ways and being elaborated by a different organ.

The value of a classification of this type is that it provides a frame of reference within which to discuss mechanisms of protein evolution and the significance of gene duplication.

Gene Duplication

The principal mechanism for increasing the size of the genome in the population is the process of gene duplication. Gene duplication provides the population with a certain "plasticity" to evolve new functions by subsequent gene modification at the new locus. This mechanism appears to have played a major role in the evolution of enzyme structure and function. The utility of this process lies in the fact that the original function of the duplicated gene is preserved, thus freeing the new gene from selection pressure. In all probability, the homologous enzymes trypsin and chymotrypsin have developed their characteristic specificities by this mechanism. It is remarkable that this has been accompanied by the replacement of approximately one-half of the amino acid residues in the protein (*2*). Gene

Table 2. Examples of homologies among proteins.

Product of the same gene locus	Product of different loci
Intraspecial homology	
1. Allelomorphs HbA_β vs Hbs_β Carboxypeptidase A_{val} vs A_{leu}	3. Products of duplicated genes Trypsin vs chymotrypsin
2. Somatic mutants No known examples	
Interspecial homology	
4. Products of equivalent loci Cyt c_{cow} vs Cyt c_{man} Porcine trypsin vs bovine trypsin	5. Products of unrelated or distantly related loci. $Elastase_{pig}$ vs $trypsin_{cow}$

duplication is definitely implicated in intraspecial homology involving independent loci, whereas this is not the case when products of the same locus in the same species are compared with each other. Thus carboxypeptidase A_{val} and carboxypeptidase A_{leu} appear to be the products of allelomorphism (*21*), whereas the two chymotrypsins, A and B, must have arisen by gene duplication.

It is not necessary to invoke gene duplication to account for phylogenetic variations of enzymes of identical biological and enzymatic functions, as in the case of bovine and porcine trypsin. Where homology is observed between biologically nonanalogous enzymes, such as cocoonase and trypsin, which are derived from different species, it is not possible to determine whether gene duplication has occurred. When the species yielding the enzymes are closely related, it is most probable that gene duplication took place, unless the two species have no counterpart of the other's enzymes.

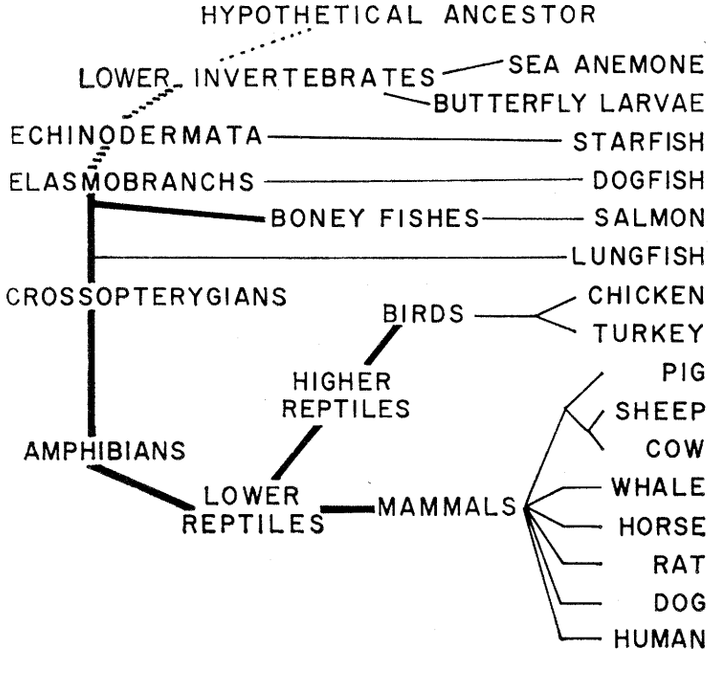
Since the survival of a duplicated gene is not influenced by negative selection pressure, it is possible that either "silent" genes or radically altered genes may develop, retaining nonetheless extensive homology. Several examples of this type may exist. Among these is muscle phosphorylase and glycogen synthetase which have identical phosphorylation sites (*22*). Another is the recently demonstrated homology between lysozyme and α-lactalbumin (*23*) where mutations in the region of the active site have permitted the divergence of radically contrasting activities. A third possibility is fraction III of the bovine procarboxypeptidase A complex which, in contrast to fraction II, is seemingly devoid of any biological function, yet like fraction II, contains the histidine loop found in serine proteases (*24*).

Homologies and Analogies among Proteolytic Enzymes

In the following discussion, we consider homologies among proteolytic enzymes, using analogous mechanisms as a basis for their classification.

The serine proteases appear to fall into two well-defined homologous systems, the microbial proteases on one hand and the proteases found among the higher animals on the other. The former seem to be characterized by the occurrence of the tetrapeptide sequence Thr-Ser-Met-Ala (*25*) at the active site; among the serine proteases of the higher animals, the characteristic sequence seems to be Gly-Asp-Ser-Gly. In both groups of proteases the serine is the functional component of the active site. In the case of the microbial enzymes, the involvement of a histidine residue in the catalytic mechanism rests only on kinetic evidence; whereas in higher animals, the participation of histidine has been confirmed by chemical evidence (*26*). On the basis of complete sequence data, the proteases from two strains of *Bacillus subtilis*, subtilisins BPN′ (Nagarse) and Carlsberg, are homologous proteins over 70 percent of the sequence (*6*). Considering the close phylogenetic proximity of the parent organisms, this degree of homology is surprisingly low. Subtilisin NOVO (*27*) and aspergillopeptidase B (*28*) are probably members of the same group.

There is one known exception to the generalization that microbial proteases are characterized by the Thr-Ser-Met sequence. The α-lytic protease from *Sorangium* (*29*) has been found to contain the active site sequence Asp-Ser-Gly-Gly (*30*). It remains to be seen whether the proteases from other microorganisms such as *Arthrobacter* (*31*), *Streptomyces moderatus* (*32*), and



	TRYPSIN		CHYMO-TRYPSIN		CARBOXY-PEPTIDASE		REF.
	An.	Cat.	An.	Cat.	A	B	
SEA ANEMONE	+		O	+	—	—	63
BUTTERFLY LARVAE	O	+	O	+	—	—	70
STARFISH	+	O	+		+	+	71
DOGFISH	+	O	+	+	+	+	54,72
SALMON	+	+	+	+	—	—	73
LUNGFISH	+	O	O	+	+	+	74
CHICKEN	+		O	+	—	—	60
TURKEY	O	+	+		—	—	60
PIG	O	+	+	+	+	+	75
SHEEP	O	+	—	—	—	—	52
COW	O	+	+	+	+	+	76
WHALE	O	+	+	+	—	—	77
HORSE	O	+	—	—	—	—	78
RAT	+	O	+	O	+	+	79
DOG	O	+	+	O	+	+	75
HUMAN	+	O	+	O	+	+	80

+ = Present O = Absent — = Unknown

Fig. 1. The distribution of animal trypsins, chymotrypsins, and carboxypeptidases with respect to their phylogenetic rank. The designations *An* and *Cat* stand, respectively, for anionic and cationic and refer to the net charge on the molecule at neutral *p*H. With regard to the invertebrates, the use of the terms trypsin, chymotrypsin, and carboxypeptidases A and B is based on analogous specificities.

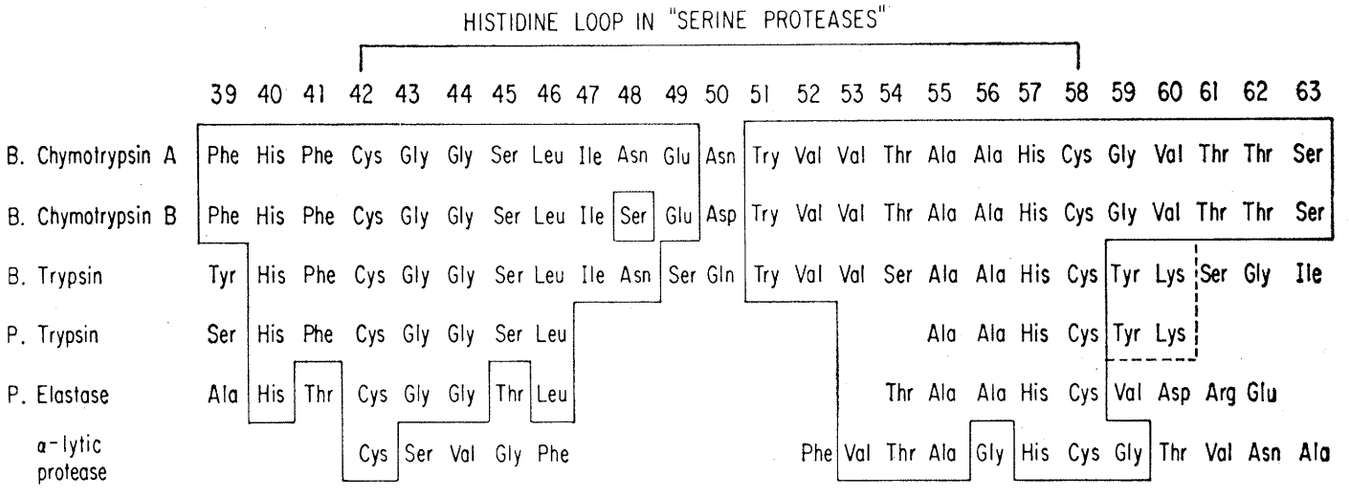


Fig. 2. The amino acid sequences about the active histidine residue (No. 57) in a variety of serine proteases. The solid lines enclose regions of general homology; the broken lines enclose a region of limited homology. In all these cases, a disulfide bridge between cysteines 42 and 58 form a loop. The residue numbering is that for bovine chymotrypsin A. The deletions in the last three have been made to maximize the homology.

others will prove to conform to established pattern or whether they, too, will resemble the proteases from higher animals, as indeed they now seem to.

A review of the literature shows that serine proteases have been found in or isolated from a large number of higher animals. These are listed with respect to the phylogenetic ranking of these species in Fig. 1. There are relatively few species that have been studied in a like manner. As a matter of fact, of the 16 living orders of mammals, only six are represented in past or present studies, and only one, Artiodactyla (cow, sheep, pig, and other), has been investigated extensively. No work on proteases of amphibia or reptilia has been recorded. As a group, the serine enzymes listed in Fig. 1 are nonhomologous to the microbial serine proteases and probably represent a case of convergent evolution (*33*). Within the group, certain homologies have been established, and others are suggested by preliminary evidence.

Bovine trypsinogen and chymotrypsinogen A were the first proteolytic enzymes shown to be homologous, to the extent of approximately 40 percent (*2*). The most prominent and seemingly invariable features of these two enzymes are the peptides surrounding the functional serine and histidine residues, respectively. This homology has since been extended to include bovine chymotrypsin B (*5*), porcine elastase (*4*), and bovine thrombin (*34*); and except for some conservative amino acid replacements, the generalization holds that homologous and invariant regions of the sequence reflect their involvement in analogous enzymatic mechanisms (*2*). The validity of this hypothesis is documented in Fig. 2. All but one of the serine proteases represented in this table contain two histidine residues in corresponding positions in a segment referred to as a "histidine loop." This finding led to the suggestion that perhaps, in analogy to ribonuclease (*35*), both of these histidines may be functionally important. Yet all site-specific reagents, particularly the chloromethyl ketone reagents of Ong, Shaw, and Shoellman (*26*) seem to react exclusively with one of the histidines only, namely that corresponding to position 57 in chymotrypsinogen, and to position 46 in trypsinogen (*26*). The α-lytic protease isolated from *Sorangium* contains only one histidine residue; and this one seems to be homologous in the sequence to histidine 57 in chymotrypsinogen, posing the possibility that the second histidine in the mammalian enzyme may be functionally inconsequential (*30*). This is not a compelling conclusion, however, since in the first place it remains to be established that the α-lytic protease is homologous to the serine proteases in higher animals in a manner suggesting a direct evolutionary relationship. Moreover, it is conceivable that the second histidine residue bestows upon the higher-animal enzymes a catalytic advantage but is not essential for enzymatic activity. The only other fully established case of homology is that of bovine chymotrypsinogen B and chymotrypsinogen A (*5*). Porcine trypsin (*37*), porcine chymotrypsin C (*38*), and porcine elastase (*4*) all contain the serine and histidine peptide sequences characteristic of the serine proteases, and hence are probably homologous to each other and to the bovine enzymes (Fig. 2). The other serine proteases listed in Fig. 1 are also potentially homologous to bovine trypsinogen (Table 3). To these might be added kallikrein (*39*) and bovine plasmin (*40*).

The homology between trypsin and chymotrypsin raises the general question whether enzymes that are homologous in primary structure also reveal extensive similarity in their three-dimensional structure or "conformational homology." Purely on the basis of the hypothesis that the primary sequence determines the three-dimensional structure (*41*), would one expect proteins like trypsin and chymotrypsin, which show only 40 percent homology, to have identical conformations? The fact that a histidine and a serine, separated by 138 residues in the primary structure, find themselves in juxtaposition at the active sites of both enzymes, indicates one common three-dimensional parameter. The findings of four homologous disulfides (*4*) strengthens the hypothesis of conformational homology. Trypsinogen is lacking one of the disulfide bonds found in chymotrypsinogen, but it has two other disulfide bonds which do not occur in chymotrypsinogen. When a model of trypsinogen was built, with the recently published structure of α-chymotrypsin (*11*, *12*) as a guide, it was found that the two nonhomologous disulfide bonds in trypsinogen smoothly fall into place without requiring any significant distortions. Similar observations suggest

Table 3. Comparative amino acid composition of certain serine proteases (residues per molecule); nd, not determined.

Source	Lys+ Arg	Tyr+ Phe Trp	Ser+ Thr	Ile+ Leu Val	Gly	Ala	Pro	His	Met	Cys/2
				Trypsinogen						
Bovine (*2*)	17	17	43	47	25	14	9	3	2	12
Porcine (*58*)	15	13	36	47	26	16	11	4	2	12
Ovine* (*59*)	12	18	31	36	19	12	8	3	2	nd
Turkey* (*60*)	10	16	35	36	19	11	8	3	2	8
				Cocoonase						
Antheraca (*14*)	19	17	39	44	22	16	9	4	1–2	4
				Elastase						
Porcine (*61*)	15	20	39	52	24	16	7	6	2	8
				Chymotrypsinogen						
A–Chicken* (*60*)	12–13	18	50–52	52	21	16–17	9–10	5	1	nd
A–Bovine (*2*)	18	18	52	52	20	22	9	2	2	10
A–Porcine (*62*)	15	16	41	48	19	20–21	12	2	2	10
A–Dogfish (*63*)	17	21	35	44	21	23	14	4	4	8
B–Bovine (*5*)	16	18	45	53	23	23	13	2	4	10
				PCP-fraction II						
Bovine (*64*)	17	20	31	54	21	15	12	5	1	8

* Active enzyme.

conformational homology between α-lactalbumin and lysozyme (*42*). While these findings are only suggestive, they are in keeping with the idea that disulfide bonds reinforce, rather than determine the three-dimensional characteristics of the protein molecule. They also permit the prediction that the three-dimensional conformation of trypsin may be strikingly similar to that of chymotrypsin, at least in the major aspects of its chain folding, in spite of differences in 60 percent of the amino acid sequence.

In the case of the metallo-peptidases, notably carboxypeptidases A and B, much evidence relating primarily to function (*8*) has accumulated, but only limited information relating to structure is available (*43*). The known distribution of carboxypeptidases A and B among higher animals is also given in Fig. 1. On the basis of composition (Table 4), carboxypeptidases A and B can be predicted to be homologous enzymes that have evolved from a common ancestor, much as chymotrypsin and trypsin have. These two enzymes differ primarily in substrate specificity, but they are cross-reactive to a certain degree (*44*). Thus carboxypeptidase B is active toward substrates for carboxypeptidase A, but not vice versa. This dual specificity has not yet been satisfactorily explained, except that it is not a product of enzyme contamination. Carboxypeptidases A and B have been isolated from bovine and porcine pancreas and more recently from pancreas of the Pacific spiny dogfish. These carboxypeptidases are all similar in amino acid composition (Table 4) except for differences in histidine and in the sulfur-containing amino acids, and hence they have probably evolved from a common archetype. Bovine carboxypeptidase A occurs in at least two allotypic forms which differ from each other in a leucine-valine dimorphism in the antipenultimate position of the polypeptide chain (*21*). The allelomorphic character of this dimorphism was strongly suggested by the distribution of the two forms among enzymes isolated from single bovine pancreas glands. Of 14 glands, approximately one-half contained both variants, and the remainder was evenly divided, showing either one or the other variant exclusively. This discovery was made possible by the fortunate fact that after degradation with cyanogen bromide, a carboxy-terminal hexapeptide could be easily isolated from carboxypeptidase. Preliminary evidence suggests the possibility of additional variants arising from amino acid replacements elsewhere in the polypeptide chain.

Among the sulfhydryl proteases only one, papain, has been subjected to complete or nearly complete sequence determination (*7*). However, sequences in the vicinity of the active cysteine residue are known for three other plant proteases, chymopapain B (*45*), stem bromelain (*46*), and ficin (*47*), and for one bacterial enzyme, the protease from group A *Streptococci* (*48*). Papain and ficin are clearly homologous at least to the extent that the sequence of ficin is known (Fig. 3). Bromelain, chymopapain B, and streptococcal protease all contain several features in common with papain and ficin, namely the occurrence of a glycine in position 9 and, with the exception of chymopapain B, the dipeptide Gln-Gly or Gln-Ala, in positions 5 and 6. However, it cannot be established from such short sequences containing so many differences whether homology truly exists among any or all of these enzymes.

Among the acid proteases, pepsin and rennin have been reported to bear evidence of homology to each other (*49*).

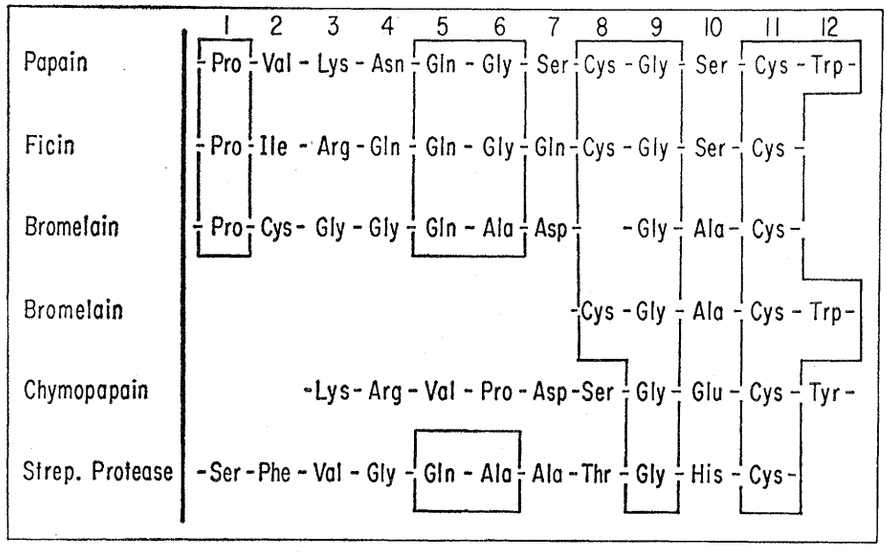


Fig. 3. The amino acid sequences adjacent to the active cysteine (Cys No. 11) in some sulfhydryl proteases. The significance of the solid lines is explained in Fig. 2. The deletion in the upper bromelain sequence was made in order to reconcile the two contradictory sequences which have been published (*46*). The residue numbering is arbitrary.

Zymogens

One of the unique features of proteolytic enzymes of higher animals is their occurrence as inactive precursors, the zymogens. Induction of enzymatic activity occurs through the primary event of peptide-bond cleavage in the amino-terminal region of the zymogen. This imposes certain limitations on structural variations in this region of the molecule, because in the process of activation the zymogens serve as substrates and hence must conform to the specificity requirements of the activating enzyme, usually trypsin. The selectivity of peptide-bond hydrolysis during the primary step of zymogen activation, by and large, may be the same for all homologous zymogens, as exemplified by trypsinogens and chymotrypsinogens, though the end products are different. It is worthy of note that in all known cases, activation involves peptide-bond cleavage in the amino- rather than the carboxy-terminal region of the zymogen. Desnuelle and co-workers have shown that bovine chymotrypsinogens A and B and in porcine chymotrypsinogen A, the amino-terminal regions of the zymogens are homologous (*50*, *51*) (Fig. 4). The same peptide bond, that between positions 15 and 16, is cleaved in the primary event, but secondary peptide-bond cleavage between positions 13 and 14 seemingly depends on the conformational details in this region of the zymogen, which appear to be different in the three proteins. The same kind of homology exists when bovine, porcine, and ovine trypsinogen are compared with each other, as shown in Fig. 5. Except for a single amino acid difference, ovine trypsinogen (*52*) appears to consist of a mixture of porcine and bovine types. While the structural significance of this dichotomy has not yet been established, it is consistent with the postulated phylogenetic relationship of these three artiodactyls.

The activation of procarboxypeptidases is less specific and of a somewhat different character. This has been demonstrated for the monomeric forms of bovine procarboxypeptidase B (*44*) and dogfish procarboxypeptidases A (*53*) and B (*54*). The peptide ultimately released is a large fragment having a molecular weight of approximately 8000. The same is probably true of the immediate precursor of bovine carboxypeptidase A, namely subunit I of the procarboxypeptidase aggregate (*55*). More importantly, primary peptide-bond cleavage is not a unique phenomenon, since activation can be induced not only by trypsin but also by other serine proteases which differ in specificity and hence probably cleave different bonds in the zymogen molecule. In the activation of the endopeptidases, the unique site of the primary bond cleavage is indicative of the involvement of the newly created α-amino group in the catalytic mechanism. This hypothesis (*56*) has been strengthened by the observation by Blow and his collaborators (*11*) that in α-chymotrypsin the α-amino group formed during activation forms a salt linkage with the β-carboxyl of the aspartic acid residue adjacent to the active serine. In contrast to the vertebrate enzymes, no conclusive evidence has yet been established for the occurrence of the digestive enzymes of invertebrates as zymogens.

Conclusions

One of the striking features of the proteolytic enzymes as a group is the immense variety of biological functions served by enzymes employing one of a few basic mechanisms. For example, in the higher animals, enzymes for activation of zymogens (trypsin), for digestion of dietary proteins (trypsin, chymotrypsin, elastase), for blood clotting (thrombin), for clot lysis (plasmin), and for sensing pain (kallikrein) all appear to use the same mechanism and to have evolved from the same ancestral gene by the process of gene duplication and subsequent divergent evolution. Equally striking is the variety of chemical solutions of the same functional problem, such as the peptide-bond cleavage by sulfhydryl proteases on the one hand and serine proteases on the other.

This is not simply a case of the

Table 4. Comparative amino acid composition of carboxypeptidases A and B (residues per molecule).

	Lys+ Arg	Tyr+Phe Trp	Ser+ Thr	Ile+Leu Val	Gly	Ala	Pro	His	Met	Cys/2
				Bovine						
A (*65*)	26	43	61	59	23	20	10	8	3	2
B (*66*)	30	44	52	50	21	22	12	7	6	7
				Porcine						
A (*67*)	24	41	55	52	26	23	14	9	3	2
B (*68*)	28	41	48	51	23	25	13	6	5	8
				Dogfish						
A (*69*)	28	43	47	55	29	21	17	7	9	4
B (*54*)	29	39	50	54	20	24	14	4	8	7

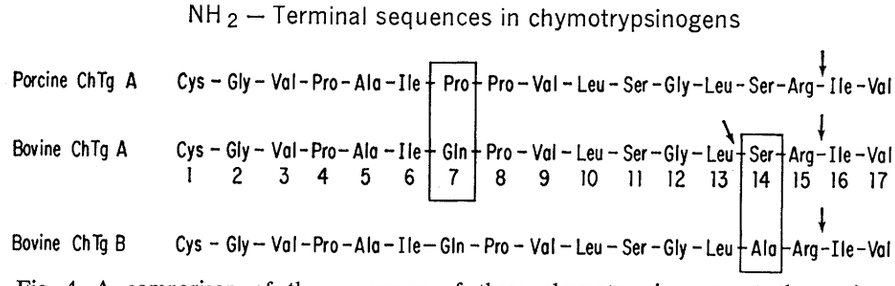


Fig. 4. A comparison of the sequences of three chymotrypsinogens at the amino-terminal end (*51*). The solid lines enclose the two areas of differences. The arrows indicate the site of the trypsin-catalyzed breaks upon activation.

Species	Activation peptides	Reference
Bovine	Val-Asp-Asp-Asp-Asp-Lys-	81
Porcine	Phe-Pro-Thr-Asp-Asp-Asp-Asp-Lys-	82
Ovine	Val-Asp-Asp-Asp-Asp-Lys- Phe-Pro-Val-Asp-Asp-Asp-Asp-Lys-	52

Fig. 5. The amino-terminal peptides released by tryptic activation of the trypsinogens from three closely related species. The occurrence of valine (underlined) in the phenylalanine activation peptide of ovine trypsinogen is the only factor which distinguishes that peptide from the single porcine trypsinogen activation peptides.

mutual replacement of two elements closely related on the periodic table in that recent experiments in which the serine hydroxyl of subtilisin has been converted to a sulfhydryl through "chemical mutation" failed to maintain enzymatic activity toward the normal specific substrates of this enzyme (*57*). Thus the effectiveness of the histidine-serine mechanism is not simply paralleled by an analogous activation of a sulfhydryl group. These experiments, though negative, amply illustrate the fact that the unique characteristics of these two different basic mechanisms reflect the aggregate contribution of the primary structure of the protein rather than a simple juxtaposition of two specific residues in the three-dimensional structure.

While in most examples cited in this discussion, gene duplication has led to the modification of enzyme specificity without affecting the enzymatic mechanism, this is by no means a compelling generalization. Thus it is quite possible that homologs exist which differ in amino acid residues at the active site and which may have acquired new functions. Such homologs would be difficult to detect by criteria other than complete sequence analysis and have not yet been encountered among proteolytic enzymes. The only documented case of that type is the homology between lysozyme and α-lactalbumin (*23*).

Finally, the minor variations in structures resulting from allelomorphism represent an important mechanism for divergent evolution and modulation of a particular enzyme species. In turn, it provides the raw material for the continuing evolution of new variants of a particular enzyme in newly emerging species. However, observation of the occurrence of these allotypes complicates the sequence analysis of a protein and raises the general question as to how many homogeneous enzymes are in fact heterogeneous in a molecular sense. To what extent do sequence determinations ignore the less-major species? Obviously the enzymes and proteins which are being investigated today present but one frame in the film of evolution covering a period of millions of years. In that kind of perspective, enzymes should not be expected to be functionally or structurally homogeneous but rather to display on the molecular level the residual traits of their ancestry.

References and Notes

1. B. S. Hartley, *Ann. Rev. Biochem.* **29**, 45 (1960).
2. K. A. Walsh and H. Neurath, *Proc. Nat. Acad. Sci. U.S.* **52**, 889 (1964).
3. O. Mikes, V. Tomasek, V. Holeysorsky, F. Sorm, *Biochim. Biophys. Acta* **117**, 281 (1966).
4. B. S. Hartley, J. R. Brown, D. L. Kauffman, L. B. Smillie, *Nature* **207**, 1157 (1965); B. S. Hartley, D. L. Kauffman, *Biochem. J.* **101**, 229 (1966); L. B. Smillie and B. S. Hartley, *ibid.*, p. 232.
5. L. B. Smillie, personal communication.
6. E. L. Smith, F. S. Markland, C. B Kaspar, R J. DeLange, M. Landon, W. H. Evans, *J. Biol. Chem.* **241**, 5974 (1967).
7. A. Light, R. Frater, J. R. Kimmel, E. L. Smith, *Proc. Nat. Acad. Sci. U.S.* **52**, 1276 (1964).
8. B. L. Vallee, *Fed. Proc.* **23**, 8 (1964).
9. H. Weiner, C. W. Batt, D. E. Koshland Jr., *J. Biol. Chem.* **241**, 2687 (1966).
10. J. T. Johansen, M. Ottesen, I. Svendsen, *Biochim. Biophys. Acta* **139**, 211 (1967).
11. B. W. Matthews, P. B. Sigler, R. Henderson, D. M. Blow, *Nature* **214**, 652 (1967).
12. J. Kraut, H. T. Wright, M. Kellerman, S. T. Freer, *Proc. Nat. Acad. Sci. U.S.* **58**, 304 (1967).
13. V. Bryson and H. J. Vogel, Eds., *Evolving Genes and Proteins* (Academic Press, New York, 1965), pp. 97, 167, 207, 221, 243, 279.
14. F. C. Kafatos, A. M. Tartakoff, J. H. Law, *J. Biol. Chem.* **242**, 1477, 1488 (1967).
15. W. M. Fitch, *J. Mol. Biol.* **16**, 9 (1966).
16. V. M. Ingram, *Biochim. Biophys. Acta* **28**, 539 (1958).
17. B. S. Hartley and B. A. Kilby, *Biochem. J.* **50**, 672 (1952).
18. Y. Pocker and Y. J. Stone, *Biochemistry* **6**, 668 (1967).
19. J. H. Park, B. P. Meriwether, P. Clodfelter, L. W. Cunningham, *J. Biol. Chem.* **236**, 136 (1961).
20. V. M. Ingram, *Biochim. Biophys. Acta* **36**, 402 (1959).
21. K. A. Walsh, L. H. Ericsson, H. Neurath, *Proc. Nat. Acad. Sci. U.S.* **56**, 1339 (1966).
22. J. Larner and F. Sanger, *J. Mol. Biol.* **11**, 491 (1965).
23. K. Brew, T. C. Vanaman, R. L. Hill, *Fed. Proc.* **26**, 724 (1967).
24. W. D. Behnke, unpublished data.
25. Abbreviations: Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys or Cys/2, cysteine or half-cystine; Cyt, cytochrome; Gln, glutamine; Glu, glutamic acid; Gly, glycine; Hb, hemoglobin; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; PCP, procarboxypeptidase; Phe, phenylananine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine.
26. E. B. Ong, E. Shaw, G. Shoellman, *J. Amer. Chem. Soc.* **86**, 1271 (1964).
27. F. Sanger and D. C. Shaw, *Nature* **187**, 872 (1960).
28. F. Sanger, *Proc. Chem. Soc. London*, **1963**, 76 (1963).
29. L. B. Smillie and D. R. Whitaker, *J. Amer. Chem. Soc.* **89**, 3350 (1967).
30. H. Kaplan and D. R. Whitaker, *Fed. Proc.* **26**, 448 (1967).
31. B. V. Hofsten, H. V. Kley, D. Eaker, *Biochim. Biophys. Acta* **110**, 585 (1965).
32. F. Rensser, *Arch. Biochem. Biophys.* **112**, 156 (1967).
33. G. H. Dixon in *Essays in Biochemistry*, P. N. Campbell and G. D. Greville, Eds. (Biochemical Society, London, 1966), vol. 2, p. 147.
34. J. A. Gladner and H. Laki, *J. Amer. Chem. Soc.* **80**, 1263 (1958).
35. A. M. Crestfield, W. H. Stein, S. Moore, *J. Biol. Chem.* **239**, 2421 (1963).
36. E. Shaw and S. Springhorn, *Biochem. Biophys. Res. Commun.* **27**, 391 (1967).
37. J. Travis and I. E. Liener, *J. Biol. Chem.* **241**, 1967 (1965).
38. T. Tobita and J. E. Folk, *Proc. Int. Congr. Biochem. 7th Tokyo, 1967, Abstracts*, p. 599.
39. N. Back, *Fed. Proc.* **25**, 77 (1966).
40. D. L. Kline, *ibid.*, p. 31.
41. D. Givol, F. DeLorenzo, R. F. Goldberger, C. B. Anfinsen, *Proc. Nat. Acad. Sci. U.S.* **53**, 676 (1965).
42. K. Brew, T. C. Vanaman, R. L. Hill, *Fed. Proc.* **26**, 724 (1967).
43. D. R. Babin, N. G. Srinivasan, M. Barton, L. H. Ericsson, unpublished experiments; W. N. Lipscomb, M. L. Ludwig, J. A. Hartsuck, T. A. Steitz, H. Muirheard, J. C. Coppola, G. N. Reeke, F. A. Quiocho, *Fed. Proc.* **26**, 385 Abstr. (1967); M. L. Ludwig, J. A. Hartsuck, T. A. Steitz, H. Muirheard, J. C. Coppola, G. N. Reeke, W. N. Lipscomb, *Proc. Nat. Acad. Sci. U.S.* **57**, 511 (1967).
44. D. J. Cox, E. Wintersberger, H. Neurath, *Biochemistry* **1**, 1069 (1962).
45. J. N. Tsunoda and K. T. Yasunobu, *J. Biol. Chem.* **241**, 4610 (1966).
46. L. Chao and I. E. Liener, *Biochem. Biophys. Res. Commun.* **27**, 100 (1967); T. Murachi, T. Miyake, M. Mizuno, *Int. Congr. Biochem. Tokyo, 7th 1967*, Abstracts, p. 765.
47. R. C. Wong and I. E. Liener, *Biochem. Biophys. Res. Commun.* **17**, 470 (1964).
48. T. Liu, W. H. Stein, S. Moore, S. D. Elliott, *J. Biol. Chem.* **240**, 1143 (1965).
49. B. S. Hartley, *Fed. Proc. Symp. 1967*, in press.
50. O. Guy, M. Rovery and P. Desnuelle, *Biochim. Biophys. Acta* **69**, 191 (1963).
51. M. Charles, D. Gratecos, M. Rovery, P. Desnuelle, *ibid.* **140**, 395 (1967).
52. S. Bricteux-Gregoire, R. Schyns, M. Florkin, *ibid.* **127**, 277 (1966).
53. A. G. Lacko and H. Neurath, *Biochem. Biophys. Res. Commun.* **26**, 272 (1967).
54. J. W. Prahl and H. Neurath, *Biochemistry* **5**, 4137 (1966).
55. J. H. Freisheim, K. A. Walsh, H. Neurath, *Biochemistry*, in press.
56. H. L. Oppenheimer, B. Labouesse, G. P. Hess, *J. Biol. Chem.* **241**, 2720 (1966).
57. K. W. Neet and D. E. Koshland, Jr., *Proc. Nat. Acad. Sci. U.S.* **56**, 1606 (1966); L. Polgar and M. L. Bender, *Biochemistry* **6**, 610 (1967).
58. P. Desnuelle and M. Rovery, *Advance. Protein Chem.* **16**, 139 (1961).
59. F. Buck, A. J. Vithayathil, M. Bier, F. F. Nord, *Arch. Biochem. Biophys.* **97**, 417 (1962).
60. C. A. Ryan, J. J. Clary, Y. Tomimatsu, *ibid.* **110**, 169, 175 (1965).
61. A. Gertler and T. Hofman, *J. Biol. Chem.* **242**, 2522 (1967).
62. M. Rovery, M. Charles, O. Guy, A. Guidoni, P. Desnuelle, *Bull. Soc. Chim. Biol.* **42**, 1235 (1960).
63. D. M. Gibson and G. H. Dixon, *Fed. Proc.* **25**, 650 Abstr. (1966).
64. J. R. Brown, R. N. Greenshields, M. Yamasaki, H. Neurath, *Biochemistry* **2**, 867 (1963).
65. J. P. Bargetzi, K. S. V. Sampath Kumar, D. J. Cox, K. A. Walsh, H. Neurath, *ibid.* **2**, 1468 (1963).
66. E. Wintersberger, D. J. Cox, H. Neurath, *ibid.* **1**, 1078 (1962).
67. J. E. Folk, *J. Biol. Chem.* **238**, 3895 (1963).
68. ———, K. A. Piez, W. R. Carroll, J. A. Gladner, *ibid.* **235**, 2272 (1960).
69. A. G. Lacko, unpublished data.
70. M. Lecadet and R. Dedonder, *Bull. Soc. Chim. Biol.* **48**, 631, 661 (1966).
71. W. P. Winter, unpublished data.
72. J. W. Prahl and H. Neurath, *Biochemistry* **5**, 2131 (1966).
73. C. B. Croston, *Arch. Biochem. Biophys.* **122**, 218 (1965).
74. W. P. Winter and G. R. Reeck, unpublished data.
75. G. Marchis-Mouren, M. Charles, A. Ben Abdeljlil, P. Desnuelle, *Biochim. Biophys. Acta* **50**, 186 (1961).
76. P. J. Keller, E. Cohen, H. Neurath, *J. Biol. Chem.* **233**, 344 (1958).
77. Y. Matsuoka and A. Koide, *Arch. Biochem. Biophys.* **114**, 422 (1966).
78. J. W. Prahl, thesis, Univ. of Washington (1964).
79. J. P. Reboud, L. Pasero, P. Desnuelle, *Biochem. Biophys. Res. Commun.* **17**, 347 (1964).
80. P. J. Keller and B. J. Allen, *J. Biol. Chem.* **242**, 287 (1967).
81. E. Davie and H. Neurath, *ibid.* **212**, 515 (1955).
82. M. Charles, M. Rovery, A. Guidoni, P. Desnuelle, *Biochim. Biophys. Acta* **69**, 115 (1963).