ACCOUNTS OF CHEMICAL RESEARCH

VOLUME 3 NUMBER 8 AUGUST, 1970

Evolution of Proteolytic Function

HANS NEURATH AND RALPH A. BRADSHAW

Department of Biochemistry, University of Washington, Seattle, Washington 98105
and Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110

Received December 24, 1969

One of the most fundamental biochemical reactions is the hydrolytic cleavage of peptide bonds. The overall reaction brings about scission of the –CO–NH– linkage so as to form two fragments, one with a newly formed carboxyl terminal and one with a newly formed amino terminal, as illustrated in Figure 1.

In biological systems, enzymes (known as "proteolytic" enzymes or "proteases") are required to catalyze the hydrolytic cleavage of peptides. The mechanism of action of these enzymes and the inherent specificity in the events that they catalyze pose a complex problem whose solution is the key to understanding such important biological processes as digestion, fertilization, sporulation, blood clotting, zymogen activation, and the release of protein hormones.

Proteolytic enzymes tend to be specific as to the peptide bond whose hydrolysis they catalyze. For example, trypsin and chymotrypsin, both serine proteases, attack different sorts of peptide linkages, either in proteins or in model substrates. Trypsin preferentially hydrolyzes peptide linkages that involve the carbonyl group of a lysine or arginine residue; chymotrypsin preferentially hydrolyzes peptide bonds involving the carbonyl group of aromatic and some branched aliphatic amino acids. Similar differences in specificity exist between carboxypeptidases B and A, which hydrolyze sequentially the carboxyl-terminal peptide bond in proteins or peptides. In this instance specificity is determined primarily by the side chain providing the amino function of the peptide bond. An additional degree of specificity is imparted by the steric accessibility of the susceptible peptide bond in the substrates. This restriction prevents random proteolysis and directs the attack of the various proteases toward the key bonds whose cleavages control many physiological processes. Thus proteolytic specificity is regulated by the inherent structural features in the active site of the enzymes and by the structural environment of the susceptible peptide bond in the substrate.

Detailed examination of several proteolytic enzymes has provided growing evidence that the seemingly enormous diversity of nature can be reduced to a limited number of structurally and functionally related classes. At the current level of our knowledge, it appears that four basic mechanisms exist by which enzymatic cleavage of peptide bonds is accomplished.[1] These are represented by the four classes of proteolytic enzymes,[2] *i.e.*, the serine proteases, the metallopeptidases, the sulfhydryl proteases, and the acid proteases, as summarized in Table I. Although these enzymatic classes differ from one another in the detailed features of their mechanism, the same underlying chemical principles characterizing hydrolysis of amides are involved in each case.

The simplicity of such a classification system, compared to earlier notions of structural and functional complexity, has provided a new route for the investigation of the molecular basis of enzyme function. The rationale of this approach can be briefly stated as follows: the chemical relatedness of enzymes, or proteins in general, is determined by the similarity in the amino acid sequence. Proteins with similar sequences are said to be *homologous*.[3] In view of the developing concept that the three-dimensional structure of proteins is a result of the amino acid sequence,[4] it is a logical hypothesis that homologous proteins will possess similar three-dimensional structures or exhibit *conformational homology*.[5] Homologous enzymes of related function may be expected to possess catalytic centers with the same array of functional groups and hence to exhibit common mechanistic traits. A detailed examination of the chemical structure of enzymes

(1) "Symposium on the Structure, Function, and Evolution of Proteolytic Enzymes," P. Desnuelle, H. Neurath, and M. Ottesen, Ed., Munksgaard Press, Copenhagen, Denmark, in press.

(2) B. S. Hartley, *Annu. Rev. Biochem.*, **29**, 45 (1960).

(3) (a) Throughout this discussion, the terms *homology* and *analogy* will be used in the sense defined by Neurath, *et al.*[3b] Because of the fact that these terms are defined differently by other investigators,[3c] including classical biologists, it is important for clarity to note the different usages of these terms. Whereas *homology*, as defined by classical biologists, is reserved for proteins that have evolved from a common ancestral gene,[3c] the definition used here[3b] is more empirical and simply relates to similarities in amino acid sequences that are greater than would be expected by chance alone. This latter definition avoids the difficult and perhaps insolvable problem of proving a common hereditary origin. The term *analogy* has been used here[3b] to denote similarities in function without regard to structure and thus differs from the biologists' definition which requires distinct evolutionary origins which have produced functional convergence.[3c] (b) H. Neurath, K. A. Walsh, and W. P. Winter, *Science*, **158**, 1638 (1967); W. P. Winter, K. A. Walsh, and H. Neurath, *ibid.*, **162**, 1433 (1968). (c) C. Nolan and E. Margoliash, *Annu. Rev. Biochem.*, **37**, 727 (1968).

(4) C. J. Epstein, R. F. Goldberger, and C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 439 (1963).

(5) H. Neurath, R. A. Bradshaw, and R. Arnon in ref 1.

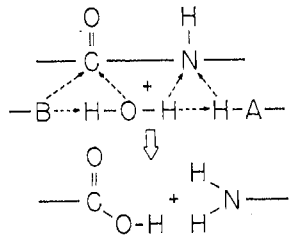

Figure 1. Schematic representation of the hydrolysis of a peptide bond as mediated by a nucleophile ("B") and a proton donor ("AH"). During enzyme-catalyzed hydrolysis, these moieties are contributed by amino acid side chains of the protein.

operating by the same mechanism should provide evidence for their relationships.

The origin of homologous enzymes is a very interesting problem. Evolutionary changes in protein structure occur principally by the mechanism of gene duplication, a process whereby a section of the chromosome containing the structural information for the synthesis of the polypeptide chain is duplicated during gametogenesis.[6] As a result, the new gene could mutate independently of the original one. The importance of gene duplication as an evolutionary mechanism lies in the increase in genetic information which allows the development of new enzymatic functions by alterations in protein structure while the original function is retained. In all probability, the homologous enzymes trypsin, chymotrypsin, elastase, and other pancreatic serine proteases have developed from a common precursor by this mechanism.[7]

An additional route of evolutionary change is allelomorphism, *i.e.*, mutations which occur in the same structural gene. In contrast to gene duplication, allelomorphism will be segregated throughout the species according to the laws of chance as modulated by the forces of natural selection. A striking example of allelomorphism is the large number of abnormal human hemoglobins which differ from each other in certain amino acid replacements with concomitant alterations in function.[8] The two genetic variants of bovine carboxypeptidase A, to be discussed below, also exemplify this phenomenon.[9]

The examination of proteins or enzymes for evolutionary relationships may be conducted in two different ways. On the one hand, systematic examination of proteins of similar function isolated from many different species provides evolutionary information about the development of the species. Such a study has been made on the electron carrier cytochrome *c*, isolated from species ranging from yeast to man.[10] More important for the study of differentiation are investigations of the evolutionary events which mark the appearance of new functions. Such studies involve comparisons of structures serving related functions in the same species such as bovine trypsin, chymotrypsin, and thrombin or porcine chymotrypsin and elastase, all of which have proven to be homologous[3] and appear to have arisen from an evolutionary precursor by a process of gene duplication.[1,7] The features of evolutionary development as applied to proteolytic enzymes are best described by an examination of specific examples.

Table I
Classes of Proteolytic Enzymes

Class	Representative enzyme	Postulated functional groups of active site
Serine proteases	Chymotrypsin	Serine, histidine, aspartic acid[a]
Metallopeptidases	Carboxypeptidase A	Zn, glutamic acid, tyrosine, arginine[b-d]
Sulfhydryl proteases	Papain	Cysteine, histidine[e-g]
Acid proteases	Pepsin	Aspartic acid[h]

[a] D. M. Blow, J. J. Birktoft, and B. S. Hartley, *Nature*, **221**, 337 (1969). [b] B. L. Vallee and J. F. Riordan, *Brookhaven Symp. Biol.*, **21**, 91 (1969). [c] W. N. Lipscomb, J. A. Hartsuck, G. N. Reeke, Jr., F. A. Quiocho, P. H. Bethge, M. L. Ludwig, T. A. Steitz, H. Muirhead, and J. C. Coppola, *ibid.*, **21**, 24 (1969). [d] R. A. Bradshaw, L. H. Ericsson, K. A. Walsh, and H. Neurath, *Proc. Nat. Acad. Sci. U. S.*, **63**, 1389 (1970). [e] J. Drenth, J. N. Jansonius, R. Kolkolk, H. M. Swen, and B. G. Wolthers, *Nature*, **218**, 929 (1968). [f] A. Light, R. Frater, J. R. Kimmel, and E. L. Smith, *Proc. Nat. Acad. Sci. U. S.*, **52**, 1267 (1964). [g] G. Lowe and A. Williams, *Biochem. J.*, **96**, 194 (1965). [h] J. R. Knowles and G. B. Wybrandt, *FEBS Lett.*, **1**, 211 (1968).

Specific Examples

Acid and Sulfhydryl Processes. Of the four principal classes of proteases, the least information is available about the evolution of the acid and sulfhydryl proteases. As judged by limited sequence data and gross molecular properties, the acid proteases pepsin and rennin appear to be closely related.[11] In the case of pepsin, an aspartic acid residue has been implicated as a component of the active site.[12] However, the lack of conclusive information regarding the functional and mechanistic characteristics of this group of enzymes precludes detailed conclusions concerning their evolutionary origin.

Somewhat more information is known about the sulfhydryl proteases where at least two different families appear to exist. These are exemplified on the one hand by the plant proteases papain, ficin, and bromelain and on the other by the proteases found in certain bacterial organisms such as the streptococcal proteases isolated by Liu, *et al.*[13,14] Preliminary sequence data indicate that structural homology does not exist between the two families.[15] In both instances, however, a single sulfhydryl residue appears to function as a nucleophilic

(6) G. H. Dixon, "Essays in Biochemistry," Vol. 2, P. N. Campbell and G. D. Greville, Ed., Academic, New York, N. Y., 1966, p 147.
(7) B. S. Hartley, *Phil. Trans.*, in press.
(8) M. F. Perutz and H. Lehmann, *Nature*, **219**, 902 (1968).
(9) P. H. Pétra, R. A. Bradshaw, K. A. Walsh, and H. Neurath, *Biochemistry*, **8**, 2762 (1969).
(10) W. M. Fitch and E. Margoliash, *Brookhaven Symp. Biol.*, **21**, 217 (1969).
(11) B. Foltmann, *Compt. Rend. Trav. Lab. Carlsberg*, **34**, 295 (1964); **35**, 143 (1966).
(12) See Table I, footnote *h*.
(13) T.-Y. Liu and S. D. Elliott, *J. Biol. Chem.*, **240**, 1138 (1965).
(14) T.-Y. Liu, N. P. Neumann, S. D. Elliott, S. Moore, and W. H. Stein, *ibid.*, **238**, 251 (1963).
(15) S. Moore in ref 1.

moiety in a manner analogous to the hydroxyl group of the serine residue in the active site of the serine proteases (*vide infra*). In the best studied example, papain, mechanistic, sequence, and X-ray diffraction analyses indicate that, in addition to the active thiol group, a histidyl residue also is important for activity.[16] Although the mechanism of action of this enzyme is still highly conjectural,[1] it is likely that this latter residue acts primarily as a proton donor. Studies on the related enzymes ficin and bromelain are still too fragmentary to ascertain the extent of structural homology and functional analogy. If both classes of sulfhydryl proteases function by a similar mechanism, then it would appear that these subclasses exemplify convergent evolution from different ancestral proteins. Such a distinction may prove to be difficult to ascertain since the wide phylogenetic gulf between bacteria and higher green plants might obscure the similarities in amino acid sequence which indicate homology in the classical sense. Clearly, structural information for both of these classes of enzymes must be forthcoming before distinct conclusions regarding their evolutionary origin can be reached.

Serine Proteases. Enzymes belonging to this group —such as bovine trypsin and chymotrypsin—were the first proteolytic enzymes to be characterized both as proteins and as enzymes.[17] They are also the first for which the complete amino acid sequence has been elucidated. The common functional feature of the serine proteases is the participation of a unique serine and histidine residue at the active site, *i.e.*, serine-195 and histidine-57 in the sequence of chymotrypsinogen[18] and serine-183 and histidine-46 in the sequence of trypsinogen.[19] Examination of the location of these residues in the three-dimensional structure of α-chymotrypsin has brought to light their interaction with yet a third residue, aspartic acid-102[20] (corresponding to aspartic acid-90 in trypsin) and has led to the proposal that these three residues function catalytically as a charge relay system in a manner indicated in Figure 2.

Early observations demonstrated a unique chemical reactivity of the serine residue in the active site of both chymotrypsin and trypsin. This residue is the site of acylation in the enzymatic reaction with certain ester substrates and the site of phosphorylation during inactivation with diisopropyl fluorophosphate.[21] In a like manner, specific modification of the active histidine residue with certain methyl chloro ketones possessing characteristics of specific substrates leads to complete enzyme inactivation.[22] The introduction of an acetonyl group into histidine-46 of trypsin (by reaction

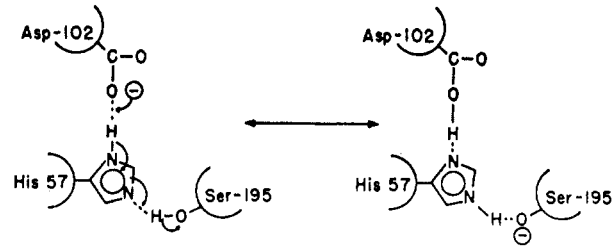

Figure 2. Schematic representation of a portion of the active-site structure of α-chymotrypsin as deduced by X-ray diffraction analysis of the crystalline enzyme. This geometry is probably characteristic of all "serine" proteases. See Blow, *et al.*, Table I, footnote *a*.

with bromoacetone) abolishes the usual reactivity of serine-183, thus providing chemical evidence in support of the hypothesis that interaction between these residues is essential for the catalytic function of the enzyme.[23] No chemical evidence has yet been obtained to prove the participation of the aspartic acid residue in the catalytic mechanism. However, a fourth residue has been implicated as a component of the active site, *i.e.*, the newly formed α-amino group of the enzyme which arises from activation of the zymogen. The formation of a salt linkage between the new α-amino group with aspartic acid-194 (adjacent to the reactive serine) appears to be both a structural[24] and a functional[25] characteristic of chymotrypsin, but not necessarily a common structural feature of all serine proteases.[26]

A detailed comparison of the amino acid sequence of bovine chymotrypsinogen and trypsinogen in 1964[19] has revealed 40% homology, and over 50% when structurally similar amino acids are compared in groups.[18,19,27] As is to be expected, the regions of most extensive homology surround the residues of functional importance, *e.g.*, the serine, histidine, and aspartic acid residues of the active site. In addition, eight of the ten half-cystinyl residues in chymotrypsinogen form corresponding disulfide bonds in trypsinogen. This homology in chemical sequence appears to be fully compatible with the assumption that the two proteins, trypsinogen and chymotrypsinogen, have homologous three-dimensional structures.[28-30] A model for trypsin has therefore been proposed which incorporates the amino acid sequence of trypsinogen in the three-dimensional structure of α-chymotrypsin. Such a model, adapted from Sigler, *et al.*,[30] is shown in Figure 3 (Walsh *et al.*[31]). While no chemical evidence is as yet available, comparison of the model with the amino acid sequence suggests that the major difference in specificity between

(16) See Table I, footnotes *e–g*.
(17) J. H. Northrop, M. Kunitz, and R. H. Herriott, "Crystalline Enzymes," Columbia University Press, New York, N. Y., 1948.
(18) B. S. Hartley, J. R. Brown, D. L. Kauffman, and L. B. Smillie, *Nature*, **207**, 1157 (1965).
(19) K. A. Walsh and H. Neurath, *Proc. Nat. Acad. Sci. U. S.*, **52**, 884 (1964).
(20) See Table I, footnote *a*.
(21) L. W. Cunningham, Jr., *Compr. Biochem.*, **16**, 85 (1965).
(22) E. Shaw, M. Mares-Guia, and W. Cohen, *Biochemistry*, **4**, 2219 (1965).
(23) J. G. Beeley and H. Neurath, *ibid.*, **7**, 1239 (1968).
(24) D. M. Blow, *Biochem. J.*, **112**, 261 (1969).
(25) G. P. Hess, *Brookhaven Symp. Biol.*, **21**, 155 (1969).
(26) H. Kaplan and H. Dugas, *Biochem. Biophys. Res. Commun.*, **34**, 681 (1969).
(27) F. Šörm, V. Holeysovskey, O. Mikes, and V. Tomašek, *Collect. Czech. Chem. Commun.*, **30**, 2103 (1965).
(28) See ref 3c.
(29) B. W. Matthews, P. B. Sigler, R. Henderson, and D. M. Blow, *Nature*, **214**, 652 (1967).
(30) P. B. Sigler, D. M. Blow, B. W. Matthews, and R. Henderson, *J. Mol. Biol.*, **35**, 143 (1968).
(31) K. A. Walsh, L. L. Houston, and R. A. Kenner in ref 1.

chymotrypsin and trypsin relates to the presence of a negatively charged aspartic acid residue (177) in trypsin.[20,32a]

This evidence, briefly summarized in the preceding paragraph, suggests that bovine trypsin and chymotrypsin have arisen from a common evolutionary precursor by a process of gene duplication and subsequent mutation. Additional evidence for this thesis has recently been provided by the observation that trypsin and chymotrypsin cross-react immunologically and that antibodies to trypsin bind a peptide fragment of chymotrypsin which includes prolyl residues which are homologous in both enzymes and the serine residue of the active site.[32b] This evolutionary relationship among serine proteases has been recently extended, by similar reasoning, to other serine proteases of different biological functions. These include pancreatic elastase, a protease which has strong lytic activity toward the elastic fibrous protein of connective tissue, chymotrypsin B, fraction II of the bovine procarboxypeptidase A complex (two pancreatic enzymes similar in specificity to chymotrypsin but differing in amino acid composition), and thrombin, plasmin, factor X, and other enzymes involved in the blood-clotting process.[1] The most complete evidence in this regard relates to the homology between porcine elastase and bovine chymotrypsin A.[33] These two proteins show nearly 40% homology in amino acid sequence, which is considerably higher than that of other closely related proteins such as sperm whale myoglobin and the α chain of human hemoglobin (18%).[34] The most strongly conserved regions are concentrated in areas which are of functional importance and, in addition, are more prevalent in segments which occupy internal positions in the three-dimensional structure. The most conclusive proof for conformational homology has come from the coincidence of the detailed three-dimensional structure predicted on the one hand from a model constructed from the actual sequence of elastase along the three-dimensional model of chymotrypsin and on the other as obtained by direct solution of the X-ray structure of porcine elastase.[35] Where deletions or additions occur in the primary sequence of elastase, they appear as shortening or lengthening of loops on the surface of the molecule, but not as major differences in the overall conformation when compared to that of chymotrypsin. The conclusion is therefore inescapable that these serine proteases are homologous in structure and analogous in mechanism and that they have evolved as the product of a common ancestral gene. Differences in substrate specificities among these enzymes can be traced, at least in part, to the substitution or deletion of amino acid residues in regions of the molecule which are believed to bind the substrate.

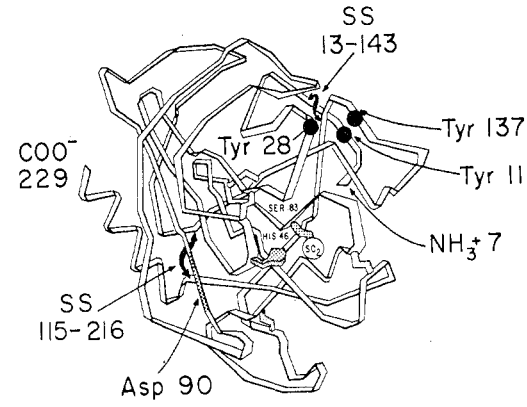

Figure 3. Hypothetical three-dimensional structure of trypsin, deduced from the structure of α-chymotrypsin as elucidated by Sigler, *et al.*,[30] on the basis of assumed conformational homology. See also Walsh, *et al.*[31]

These serine enzymes are derived from species of relatively recent evolutionary origin. To trace the pattern of their evolution, it would be important therefore to compare some of these enzymes to analogous serine proteases of more primitive species. Trypsin-like enzymes have indeed been found in a variety of species ranging from the higher mammals (such as bovine, procine, or equine) to lower vertebrates, invertebrates, and bacteria.[28] Many of these show considerable similarities in amino acid composition, and in some instances peptide sequences surrounding the functional residues of the active site have been found to show a considerable degree of homology.

Such investigations have been directed in our own laboratory toward a trypsin-like enzyme isolated from the spiny Pacific dogfish, a species which is believed to have arisen some 200 million years preceding the mammal and seems to have undergone little evolutionary change since that time. Preliminary sequence analysis has indeed proven this enzyme to show a high degree of identity with that of bovine origin (in the order of 70%).[36] Another trypsin-like enzyme has been isolated from the secum of the starfish *Evasterias trochelii* and, as shown in Table II, is similar in amino acid composition to trypsin from higher organisms.[37] Recently, Jurasek, *et al.*,[38] have reported that a bacterial enzyme which is a component of pronase, a mixture of enzymes isolated from *Streptomycis griseus*, is approximately 20% homologous to bovine trypsin. It seems clear, therefore, that the serine proteases exemplified by bovine trypsin and chymotrypsin are the products of an extensive process of evolution, but to date no information has been derived from the structure of any of the serine proteases that would reveal the steps leading to the present enzyme.

(32) (a) E. Shaw in ref 1; (b) M. M. Sanders, K. A. Walsh, and R. Arnon, *Biochemistry*, **9**, 2356 (1970).

(33) B. S. Hartley and D. M. Shotton, presented at the Symposium on Structure, Function, and Evolution of Proteolytic Enzymes, Copenhagen, Denmark, June 1969; D. M. Shotton and B. S. Hartley, *Nature*, **225**, 802 (1970).

(34) M. F. Perutz, J. C. Kendrew, and H. C. Watson, *J. Mol. Biol.*, **13**, 669 (1965).

(35) H. C. Watson and D. M. Shotton, *Phil. Trans.*, in press; *Nature*, **225**, 811 (1970).

(36) R. A. Bradshaw, H. Neurath, R. W. Tye, K. A. Walsh, and W. P. Winter, *ibid.*, in press.

(37) W. P. Winter and H. Neurath, in preparation.

(38) L. Jurasek, D. Fackre, and L. B. Smillie, *Biochem. Biophys. Res. Commun.*, **37**, 99 (1969).

Table II

Amino Acid Composition of Trypsinogens (or Trypsins)[a]

Protein	Lys + Arg	Tyr + Phe + Trp	Ser + Thr	Ile + Leu + Val	Gly + Ala	Met	Pro	His	Cys/2
Dogfish trypsinogen[b]	13	18	24	45–46	45	8	11	7–8	12
Bovine trypsinogen[c]	17	17	43	47	39	2	9	3	12
Porcine trypsinogen[d]	15	17	36	47	39–40	2	11	4	12
Ovine trypsinogen[e]	17	20	42	46–47	42	2	10	3	11
Starfish trypsin[f]	13	18	33	43	44	2	13	4	8
Human trypsin[g]	17	14	34	40	33	1	9	3	8

[a] Exclusive of glutamic and aspartic acids since their amide distribution is not known in all proteins listed herein. [b] R. A. Bradshaw, H. Neurath, R. W. Tye, K. A. Walsh, and W. P. Winter, *Nature*, in press. [c] K. A. Walsh and H. Neurath, *Proc. Nat. Acad. Sci. U. S.*, **52**, 884 (1964). [d] M. Charles, M. Rovery, A. Guidoni, and P. Desnuelle, *Biochim. Biophys. Acta*, **69**, 115 (1963). [e] R. Schyns, S. Bricteux-Grégoire, and M. Florkin, *ibid.*, **175**, 97 (1969). [f] W. P. Winter and H. Neurath, in preparation. [g] J. Travis and R. C. Roberts, *Biochemistry*, **8**, 2884 (1969).

It would be erroneous to conclude, however, that all serine enzymes have in fact arisen by a single event. The bacterial enzyme subtilisin is a telling example of a "serine" enzyme operating by a mechanism analogous to if not identical with that of the mammalian counterparts but yet being so different in sequence[39] and in three-dimensional structure[40] that it is highly improbable that it is the product of the same ancestral gene. Like trypsin or chymotrypsin, this bacterial enzyme contains the active-site configuration of aspartic acid–histidine–serine,[40] but their linear sequence in the polypeptide chain starting from the N-terminus is different, *i.e.*, histidine–aspartic acid–serine in the pancreatic enzymes and aspartic acid–histidine–serine in subtilisin.[41] It is obvious, therefore, that the catalytic apparatus has arisen in subtilisin by an evolutionary process independent from that of the pancreatic enzymes.

Metallocarboxypeptidases. Of the class of proteolytic enzymes which sequentially remove amino acid residues from the carboxyl terminus of peptides and proteins, pancreatic carboxypeptidases A and B are the best studied examples. Bovine carboxypeptidase A in particular has been examined in detail, and from these studies a picture of the structure–function relationship of this enzyme has evolved. Although several facets of this relationship remain to be clarified, a consideration of a model of this enzyme is essential for an understanding of its evolution.

The basic mechanism for the hydrolysis of peptide bonds by carboxypeptidase A is illustrated in Figure 4.[42] In this mechanism, the cleavage of the peptide bond is brought about by the interaction of four components of the active site. The elements of a molecule of water are added to the peptide bond by means of the interaction of a nucleophile (indicated in Figure 4 as B) and a proton donor (indicated in Figure 4 as AH). The action of the nucleophile, either on the carbonyl carbon directly or through a molecule of water, is aided by the interaction of the carbonyl oxygen with the zinc

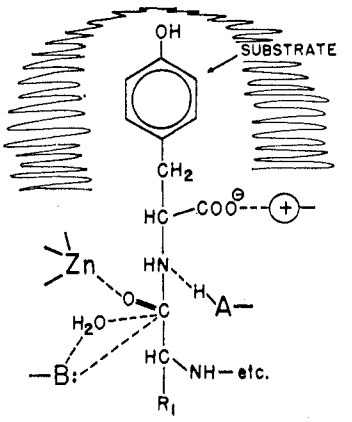

Figure 4. Schematic representation of a portion of the active site of bovine carboxypeotidase A; see Vallee and Riordan.[42]

atom, thus increasing the electron deficiency of the carbon atom. Another principal feature is the presence of a positively charged group which supplies the binding site for the required carboxylate group of peptide substrates. The relation of these components of the active site to the structure of the enzyme will now be considered.

In 1954, Vallee and Neurath[43] established that a single zinc atom was structurally and functionally an integral part of the active site. Subsequent studies have shown that certain other transition metals could replace zinc, producing derivatives of various levels of activity toward peptide and ester substrates.[44] However, the nature of the protein ligands to zinc (and presumably the ligands to the other metal moieties) eluded identification. Early experiments indicated a nitrogenous and sulfhydryl ligand in a bidentate structure.[45] While these conclusions were consistent with the behavior of model compounds, subsequent identification of the actual tridentate ligand structure was not in accord with these model systems. The correct ligand structure has now been ascertained by a combination of

(39) E. L. Smith, F. S. Markland, C. B. Kasper, R. J. Delange, M. Landon, and W. H. Evans, *J. Biol. Chem.*, **241**, 5974 (1966).
(40) J. Kraut, *Phil. Trans.*, in press; C. S. Wright, R. A. Alden, and J. Kraut, *Nature*, **221**, 235 (1969).
(41) E. L. Smith in ref 1.
(42) See Table I, footnote *b*.
(43) B. L. Vallee and H. Neurath, *J. Amer. Chem. Soc.*, **76**, 5006 (1954).
(44) J. E. Coleman and B. L. Vallee, *J. Biol. Chem.*, **236**, 2244 (1961).
(45) T. L. Coombs, Y. Omote, and B. L. Vallee, *Biochemistry*, **3**, 653 (1964).

X-ray diffraction analysis[46] and amino acid sequence determination.[47] The complete amino acid sequence of bovine carboxypeptidase A, as deduced by chemical methods, is shown in Figure 5. Thus it is now known that the zinc ligands are histidine-69, glutamic acid-72, and residue 196, which was believed to be lysine (or glutamic acid or glutamine) according to X-ray data,[46] but has been identified as histidine by sequence analysis.[47] This is the first instance where the ligand structure of a metalloenzyme has been established.

The identification of the residue contributing the positive charge (see Figure 4) to the binding of the carboxyl group in peptides was ascertained as arginine-145 from X-ray diffraction analysis of peptide substrate–protein complexes.[46] In a like manner, the nucleophilic moiety (B in Figure 4) was identified as glutamic acid-270. Both identifications were confirmed by sequence analysis.[47] The proton donor (AH in Figure 4) has been identified as tyrosine-248. This identification is consistent with a number of different chemical modifications of carboxypeptidase A, such as acetylation,[48] iodination,[49,50] and nitration,[51] which cause the esterase activity to be enhanced and peptidase activity to be abolished. In view of the fact that no X-ray data are available for complexes of enzyme and ester substrates, it is not known whether these same residues also participate in the same manner in ester hydrolysis, an activity first noted by Snoke, Schwert, and Neurath in 1948.[52] In fact, several lines of evidence—especially kinetic studies of the chemically modified enzymes—suggest that ester hydrolysis may involve different residues of the enzyme. It is clear that the mechanism of action of carboxypeptidase A can be fully understood only when the mode of hydrolysis of both the peptide and ester substrates is known.

The formation of the active-site structure is the result of zymogen activation. The actual proteolytic events accompanying activation are known in some detail. Specifically, it has been established that three forms of bovine carboxypeptidase A, designated α, β, and γ, may be obtained as the result of different conditions of activation.[53] The enzyme may thus be isolated containing either 307, 305, or 300 amino acid residues. These differences are manifested in the length of the amino-terminal portion of the enzyme.[54] None of the conditions of activation produces a single form of the enzyme, and hence homogeneous preparations of carboxypeptidase A can only be obtained by separation using ion-exchange chromatography.[55] It is worthy of note that neither X-ray analysis nor chemical modification studies have been carried out with homogeneous enzyme preparations, thus underscoring the problem of obtaining pure protein preparations, particularly in proteolytic systems.

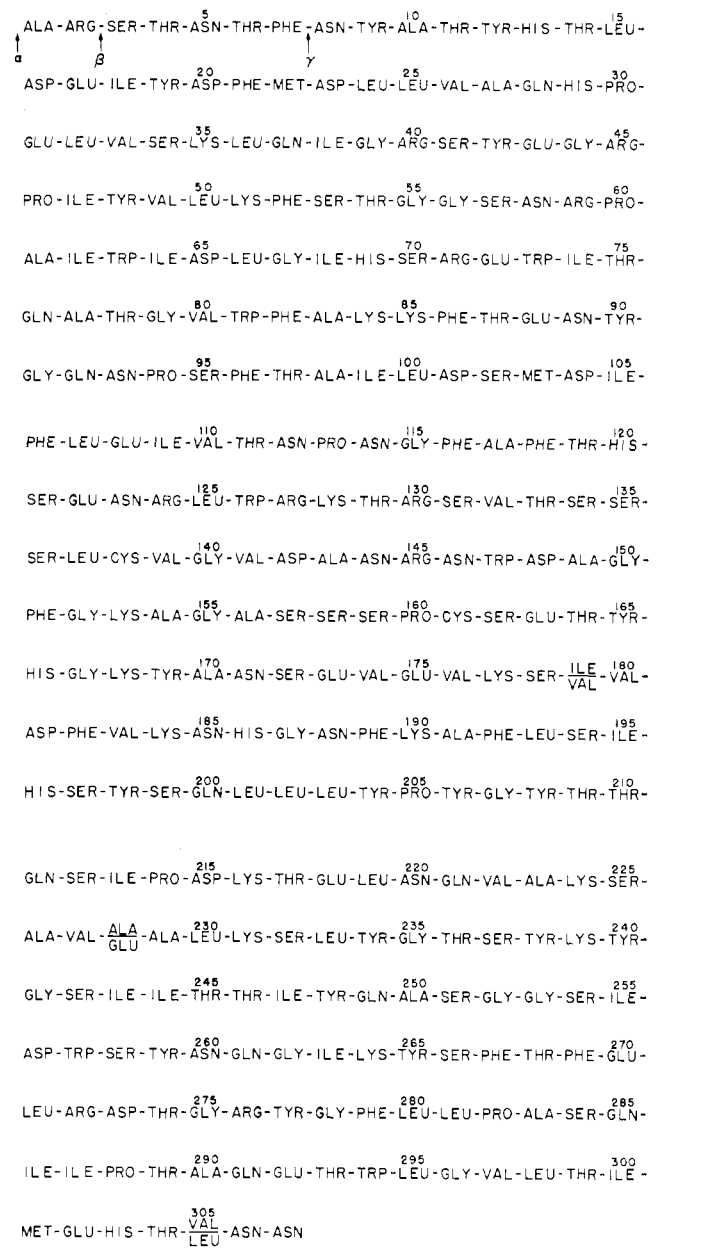

Figure 5. The amino acid sequence of bovine carboxypeptidase A.[47]

An additional element of heterogeneity is introduced by genetic variation which gives rise to two allotypic forms of carboxypeptidase differing from each other at three sites in the amino acid sequence. The identification of these sites, by chemical means, has shown that positions 179, 228, and 305 in the amino acid sequence of bovine carboxypeptidase Aα are occupied by either isoleucine, alanine, and valine or by valine, glutamic acid,

(46) See Table I, footnote *c*.
(47) See Table I, footnote *d*.
(48) R. T. Simpson, J. F. Riordan, and B. L. Vallee, *Biochemistry*, **2**, 616 (1963).
(49) O. A. Roholt and D. Pressman, *Proc. Nat. Acad. Sci. U. S.*, **58**, 280 (1967).
(50) R. T. Simpson and B. L. Vallee, *Biochemistry*, **5**, 1760 (1966).
(51) J. F. Riordan, M. Sokolovsky, and B. L. Vallee, *ibid.*, **6**, 358 (1967).
(52) J. E. Snoke, G. W. Schwert, and H. Neurath, *J. Biol. Chem.*, **175**, 7 (1948).
(53) H. Neurath, R. A. Bradshaw, L. H. Ericsson, D. R. Babin, P. H. Pétra, and K. A. Walsh, *Brookhaven Symp. Biol.*, **21**, 1 (1969).
(54) K. S. V. Sampath Kumar, J. B. Clegg, and K. A. Walsh, *Biochemistry*, **3**, 1728 (1964).
(55) P. H. Pétra and H. Neurath, *ibid.*, **8**, 2466 (1969).

and leucine.[9] Thus, as a consequence of activation and genetic heterogeneity, there are six forms of bovine carboxypeptidase A, and these have been obtained in a pure form by chromatographic means. No significant catalytic differences exist among the six forms.[53]

Bovine carboxypeptidase A contains two half-cystinyl residues located at positions 138 and 161.[47] The oxidation state of the sulfur of these residues has been a matter of conjecture,[42,46,53] particularly since it appeared that one of them in the form of a sulfhydryl group served as zinc ligand. From an examination of the electron density map, Lipscomb, *et al.*,[46] concluded that in carboxypeptidase Aα the cystinyl residues were present in disulfide linkage. The disulfide nature has recently been established chemically by isolation in over 50% yield of a single cystinyl peptide from peptic digests of carboxypeptidase Aγ.[56] Supporting evidence has also been obtained more recently by use of a chloride ion nmr probe of the behavior of reversibly enzyme bound mercuric ions.[57] The apparent importance of this disulfide loop in the amino acid sequence to the evolutionary development of this enzyme will be discussed below.

The site of binding of the α-carboxyl group of the polypeptide chain of carboxypeptidase A was identified by Lipscomb, *et al.*,[46] as residue 265. This residue has been correctly identified by sequence analysis as tyrosine instead of arginine[47] as proposed on the basis of the X-ray data. Thus the nature of the interaction may be considered as the hydrogen-bond type rather than the salt interaction previously proposed. This internal binding may be of physiological significance as a mechanism for protecting the enzyme from autocatalytic degradation.

Secondary sites of catalytic and binding activity proposed on the basis of the X-ray analyses have been confirmed or corrected on the basis of the amino acid sequence. In particular, residue 279 has been identified as phenylalanine instead of histidine, thus eliminating any catalytic role for this residue. In addition, residue 127, which has been identified as an arginyl residue rather than a Glx residue (deduced from the electron density map), introduces an additional positive charge into the active-site pocket. In view of the other arginyl residues identified in the sequence and positioned in the active site by the X-ray data, this particular residue may play an important role in enhancing the overall positive charge associated with the active site, thus increasing the number of possible interactions with the negatively charged substrate.

Homologies

In analogy with the serine proteases (*vide supra*), it might be expected that bovine carboxypeptidase A will show similarities to other carboxypeptidases. Six pancreatic carboxypeptidases from three different species have thus far been isolated and characterized (Table III). They resemble each other in molecular weight, amino acid composition, metal content, and functional response to chemical and physical modification as well.[58] One of the enzymes most likely to be homologous to bovine carboxypeptidase A is bovine carboxypeptidase B. Rigorous identification of homology between these two enzymes was, however, hampered by the lack of sequence data for the B enzyme. A comparison of the available fragments of the B enzyme[59,60] with the complete sequence of carboxypeptidase A led to the prediction that carboxypeptidases A and B have homologous amino acid sequences.[61] This comparison, based on 46 residues, has subsequently been extended by isolation of a 14-amino acid residue peptide containing the apparently functional tyrosyl residue of the B enzyme[62] (Figure 6). It may be concluded on the basis of this limited comparison that carboxypeptidases A and B are homologous with regard to amino acid sequence and probably possess very similar three-dimensional structures.

Table III
Amino Acid Composition of Carboxypeptidases[a]

Enzyme	Lys + Arg	Tyr + Phe + Trp	Ser + Thr	Lu + Val	Gly + Ala	Pro	His	Met + Cys
Bovine								
A	26	43	61	59	43	10	8	5
B	30	44	52	50	43	12	7	13
Porcine								
A	24	41	55	52	48	14	9	5
B	28	41	48	51	48	13	6	13
Dogfish								
A	28	43	47	55	50	17	7	13
B	29	39	50	54	44	14	4	13

[a] Exclusive of aspartic and glutamic acid residues since their state of amidation has not been determined for all enzymes listed in this table. Data are expressed as amino acid residues per molecule.

It is of interest to note that the relative specificities of carboxypeptidases A and B are directed on the one hand toward aromatic and branched aliphatic amino acid side chains and on the other toward the positively charged side chains of lysine and arginine. These specificities are strikingly parallel to those of chymotrypsin and trypsin, respectively. By extending the analogy, it may be proposed that carboxypeptidase B will also contain a negatively charged residue in the active site to bind the principal side-chain residue which governs the specificity of the enzyme.

Evolution

The completion of the amino acid sequence of carboxypeptidase A has afforded the opportunity to ex-

(56) K. A. Walsh, L. H. Ericsson, R. A. Bradshaw, and H. Neurath, *Biochemistry*, **9**, 219 (1970).
(57) R. G. Bryant, H. J. C. Yeh, and T. R. Stengle, *Biochem. Biophys. Res. Commun.*, **37**, 603 (1969).
(58) H. Neurath, R. A. Bradshaw, P. H. Pétra, and K. A. Walsh, *Phil. Trans.*, in press.
(59) M. Elzinga and C. H. W. Hirs, *Arch. Biochem. Biophys.*, **123**, 361 (1968).
(60) E. Wintersberger, *Biochemistry*, **4**, 1533 (1965).
(61) R. A. Bradshaw, H. Neurath, and K. A. Walsh, *Proc. Nat. Acad. Sci. U. S.*, **63**, 406 (1969).
(62) T. H. Plummer, Jr., *J. Biol. Chem.*, **244**, 5246 (1969).

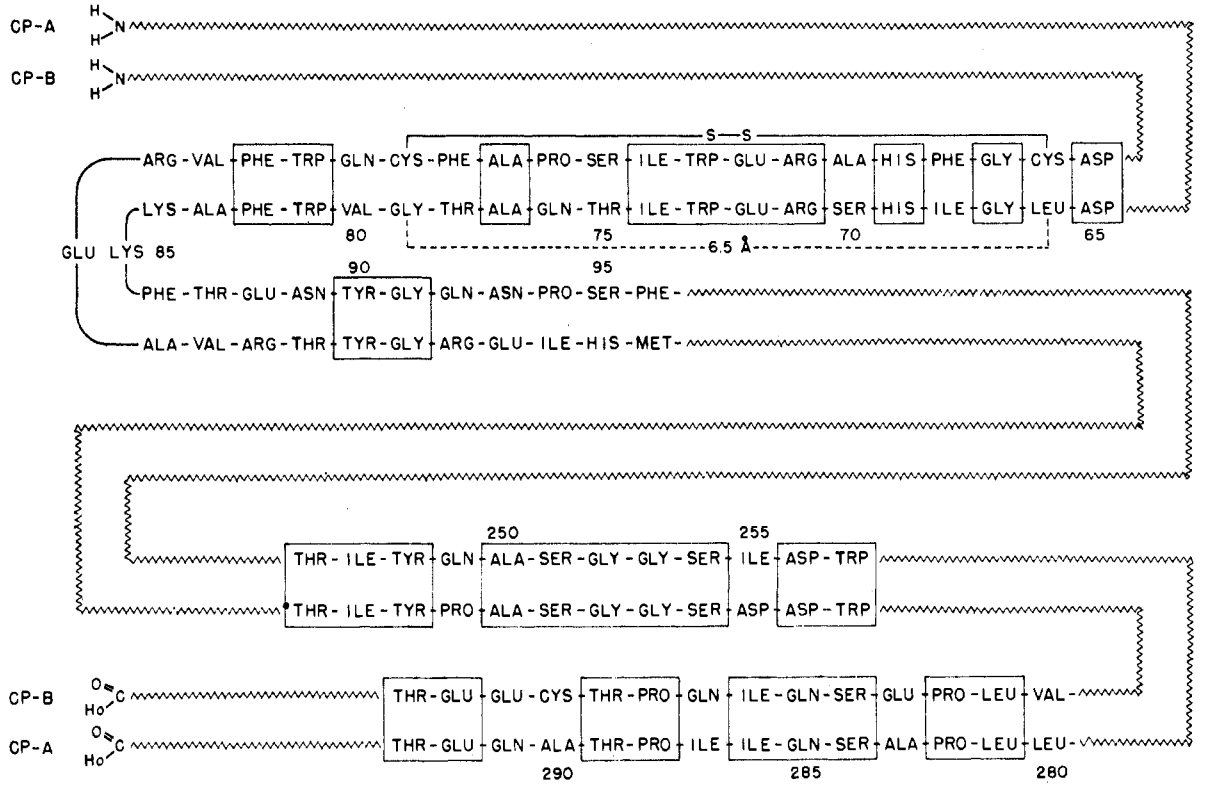

Figure 6. The structural homology of bovine carboxypeptidases A and B. The amino acid sequence data of bovine carboxypeptidase A are taken from Bradshaw, *et al.*,[47] and the amino and sequence data of bovine carboxypeptidase B from Elzinga and Hirs,[59] Wintersberger,[60] and Plummer,[62] as described in the text.

amine this class of enzymes for information concerning its evolution. If it be assumed that carboxypeptidase A evolved by a steady, systematic series of events from an unknown protein of unspecified size with undefined enzymatic function, then a possible series of events for the appearance of carboxypeptidase activity can be postulated. The hypothetical scheme presented in Figure 7 has been purposely kept simple and is only meant to reflect major changes in the developmental process. No evidence to date has been available to support such an hypothesis with the exception of the last step, *i.e.*, the diversion of carboxypeptidase activity between the A and B enzymes[61] (*vide supra*). However, two pieces of information provided by the sequence of the enzyme are consistent with such a scheme and add credence to the concept that the activity of this enzyme has developed by a series of events, either similar to or identical with those proposed.

In the first case, a comparison of the segments of polypeptide chains surrounding the half-cystinyl residues, as shown in the upper part of Figure 8, suggests that an internal partial gene duplication was responsible for the appearance of the disulfide loop in carboxypeptidase A. Including the half-cystinyl residues and the distinctive -Ser-Ser-Ser- sequences, 7 of the 16 residues are identical. A consequence of this event, in addition to providing two half-cystinyl residues, was the duplication of arginine-124. The new arginine, which eventually became residue 145, is of marked significance in that it supplies the binding site for the required α-carboxyl group of the peptide substrate (*vide supra*). In addition, the incorporation of the new segment into the protein was accomplished by creating a loop which was locked in place by joining the half-cystinyl residues in a disulfide bond. The formation of the 22-residue loop effectively would have closed the

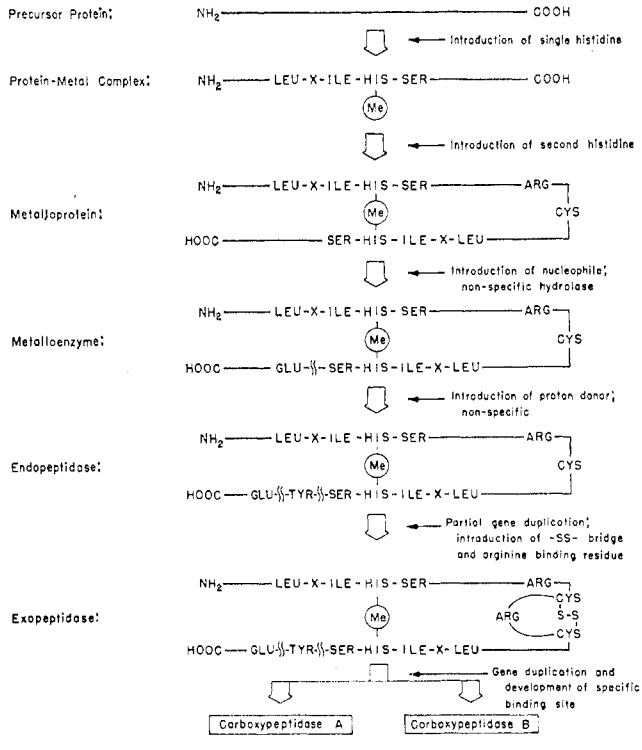

Figure 7. Hypothetical scheme for the evolution of carboxypeptidase function.[5] See text for details.

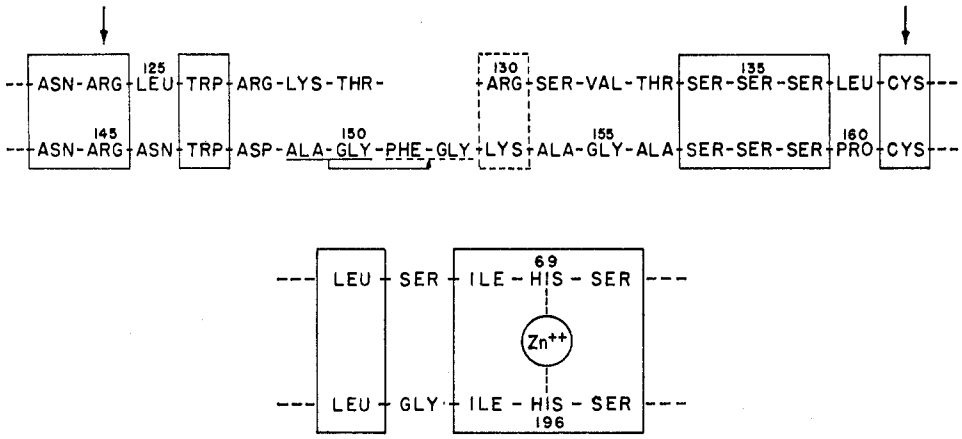

Figure 8. Comparison of the amino acid sequences surrounding the half-cystinyl residues in bovine carboxypeptidase A (upper) and the zinc-binding histidyl residues in bovine carboxypeptidase A (lower).

crevice containing the active-site residues, thus converting the endopeptidase enzyme into one with exopeptidase activity.

The second observation deals with the proposed binding site of the enzyme as deduced by chemical and X-ray analyses. As described above, residues 69 and 196 have been shown to be the histidyl residues binding the zinc. An examination of each of the sequences containing these histidyl residues, as shown in the lower part of Figure 8, shows pronounced similarity in that four of the five residues are identical. Thus it may be proposed that an extremely early genetic event may have resulted in partial or even entire gene duplication of the histidyl peptide sequence, resulting in a protein with two histidines of proper environment to bind the zinc metal. The preservation of this short sequence may be indicative of the requirements creating the appropriate orientation for the metal binding site. It should be noted that no apparent homology has been detected in the segments of polypeptide chain extending beyond that shown.

The observations described above give information about the evolutionary scheme proposed in Figure 6 only with regard to the appearance of endopeptidase activity going to exopeptidase activity and to the appearance of the metalloprotein from the protein–metal complex. The tenuous nature of the second observation is such that no detailed conclusions can be drawn regarding the early steps of the scheme. Further substantiation must be obtained by examination of proteins which would fit into the proposed scheme. For example, examination of other zinc metalloenzymes (such as carbonic anhydrase or thermolysin) would add substantial proof to the proposed scheme if they were found to contain similar structures. In this regard, it is significant that a histidine residue may be a zinc ligand in the carbonic anhydrase molecule, as judged by chemical modification experiments that indicate that one or possibly two histidines play an important role in the active site of this enzyme.[63,64] Solution of the primary and three-dimensional structures of these and other zinc metalloenzymes should supply the evidence needed to test the proposed hypotheses for the evolution of exopeptidase function.

(63) P. L. Whitney, G. Fölsch, P. O. Nyman, and B. G. Malmström, *J. Biol. Chem.*, **242**, 4206 (1967).
(64) S. L. Bradbury, *ibid.*, **244**, 2002 (1969).

Hydrolysis of Orthophosphate Esters

CLIFFORD A. BUNTON

Department of Chemistry, University of California, Santa Barbara, California 93106

Received February 11, 1970

In many biological systems phosphate esters, *e.g.*, nucleotides, nucleic acids, and sugar phosphates, are important intermediates whose *in vivo* reactions are of considerable interest to biochemists; an understanding of the mechanisms of phosphate ester hydrolysis is of obvious biochemical significance. These mechanisms have intrinsic interest to the physical organic chemist. Orthophosphoric acid is tribasic, and its esterification can give not only trisubstituted esters but di- and monosubstituted esters whose acidic dissociation is governed by the pH of the medium (for mono- and disubstituted phosphates, $ROPO(OH)_2$ and