Plant Proteolytic Enzymes

Volume I

Editor

Michael J. Dalling, Ph.D.
Senior Lecturer
School of Agriculture and Forestry
The University of Melbourne
Parkville, Victoria, Australia

CRC Press, Inc.
Boca Raton, Florida

Library of Congress Cataloging-in-Publication Data
Main entry under title:

plant proteolytic enzymes
Bibliography: p.
Includes index.
1. Proteolytic enzymes. 2. Botanical chemistry.
I. Dalling, Michael J.
QK898.P82P55 1986 581.19'256 85-24319
ISBN 0-8493-5682-2 (v. 1)
ISBN 0-8493-5683-0 (v. 2)

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431.

International Standard Book Number 0-8493-5682-2 (v.1)
International Standard Book Number 0-8493-5683-0 (v.2)

Library of Congress Card Number 85-24319
Printed in the United States

PREFACE

Knowledge on protein degradation and on the proteolytic enzymes involved appears to be equally important for the understanding of cell metabolism and development and knowledge on protein synthesis. This is not an overstatement. It is simply concluded from the fact that protein degradation and protein synthesis depend on each other. Both processes contribute equally to the vitally important protein metabolism, both of them are equally involved in protein turnover and in all of the changes in quantity and quality of cellular protein as associated with cell differentiations and, hence, with development. Leaf senescence provides just one example demonstrating that even the marked loss of protein depends on protein synthesis. Seed germination may be mentioned as an example for the opposite case in which protein synthesis is intimately associated and even dependent on proteolysis.

Biologists have been much more interested and successful in the elucidation of protein synthesis as compared with protein degradation which has been, and to some extent still is, a rather neglected field. This is certainly due to the fact, that protein synthesis was appreciated as the key to the understanding of the connection between genome and metabolism. However, the elucidation of protein synthesis may have been easier than the work toward the understanding of protein degradation although the latter represents, in biochemical terms, nothing but simple hydrolysis.

A considerable number of plant proteolytic enzymes has been described so far, but only in rare cases have the catalysts been associated with a clear cut function. It is a truism that the number of proteases (including the number of proteases which will be discovered in the future) is much smaller than the number of protein species, i.e., the potential substrates present in plant cells. The million dollar question concerns, therefore, the explanation of specificity of protein degradation which has been observed in various instances, e.g., when turnover rates were determined for individual protein species or losses of certain proteins were followed in senescent leaves. It appears to be impossible to explain such phenomena with activities of highly specific proteolytic enzymes responsible for selective degradation of certain proteins. It rather seems that a comparatively low number of proteolytic enzymes is responsible for unspecific degradations of proteins with which they get into contact. Hence, the mechanisms responsible for the contact between proteolytic enzymes and those proteins which are destined for degradation appear to be an important aspect of protein degradation.

A hypothetical solution of this problem is subcellular compartmentation. The concept of lysosomes, originally developed for rat liver cells, was adopted for plant cells. In the past few years it has been documented convincingly that plant cells, indeed, contain a lysosomal extraplasmatic compartment, the vacuole, in which proteolytic enzymes together with other hydrolases are located. The concept of lysosomes is intelligible because proteolytic enzymes appear to be separated in a specific compartment, the cytoplasm, the truly living entity of the cell, being protected against the uncontrolled attack by digestive enzymes. Yet, the concept certainly does not explain the mechanism of selective and controlled protein degradation. It merely switches over from the specificity of proteolytic enzymes to the specificity of transport of cytoplasmic proteins into the vacuoles. In fact, the role of vacuolar porteolytic enzymes has so far been demonstrated unambiguously only in the very special case of protein bodies, in which the compartmentalization of the substrates, the reserve proteins, coincides with that of the proteolytic enzymes responsible for intracellular protein digestion during seed germination.

It is quite possible and even probable that only a fraction of proteolytic enzymes which can be assessed by using the conventional substrates has been discovered so far. It is not unlikely that proteolytic enzymes which are more directly related to the degradation of cytoplasmic proteins than are the vacuolar enzymes remained undiscovered because they

have comparatively low activities and unusual properties or specificities that do not allow the determination with the common substrates. Recent findings of proteolytic activities in chloroplasts and in mitochondria support this view. The example of the proteolytic system of yeast (not covered in the present volume) may show the prerequisites for discovering minor proteolytic enzymes. At the same time the example demonstrates the limitations of work with higher plants.

Bakers yeast cells contain two principal endopeptidases (A and B), two carboxypeptidases (Y—formerly protease C and S), and several aminopeptidases. This proteolytic system is located in the vacuoles (A, B, Y, S, and two aminopeptidases). It is particularly interesting that yeast cells also contain several proteins which specifically inhibit the proteases A, B, and C. These inhibitor proteins are located in the cytosol. Hence, the proteolytic machinery in the vacuoles is fully active with the cytoplasm being protected, not only through compartmentation, but also by virtue of protease inhibitor proteins. Assuming that this system is responsible for protein degradation, one would expect that proteinase deficient mutants are not viable. Yet, mutants lacking the two vacuolar endopeptidases are able to grow, differentiate, and even sporulate, although the rate of protein turnover is markedly lower than in normal strains. Working with normal strains, Wiemken was able to demonstrate unambiguously that vacuoles are "the sole compartments" of endopeptidases.[1]

Working with mutants deficient in the vacuolar proteases, Wolf has not only discovered several novel proteolytic enzymes the activity of which in normal strains is completely masked; he also showed that "vacuoles are not the sole compartment of proteolytic enzymes in yeast";[2] the newly discovered proteinases D and E were found to be located outside the vacuoles. These extravacuolar enzymes may play essential roles in proteolysis and it is even feasible that the functions will be elucidated should selection for corresponding mutants be practicable. In any case, the researcher dealing with yeast is in a much more promising situation that his colleague working with higher plants who will have to wait for the availability of corresponding genetial research tools.

The present volume is undoubtedly a most valuable source of information about plant proteolytic enzymes. It covers not only the enzymological aspects, but also the various functions including those which are hypthetical at the moment and probably also difficult to prove in the future. Hopefully, the book will stimulate plant physiologists to step into the fascinating field of protein degradation and help to overcome the difficulties in understanding how the proteases are integrated in the metabolism of the living plant cells.

Michael J. Dalling

REFERENCES

1. **Wienken, A., Schellberg, M., and Urech, K.,** Vacuoles: The sole compartments of digestive enzymes in yeast *(Saccharonyces cerevisiae)* ?, *Arch. Microbiol.*, 123, 23, 1979.
2. **Emter, O. and Wolf, D. H.,** Vacuoles are not the sole compartments of proteolytic enzymes in yeast, *FEBS Lett.*, 166, 321, 1984.

THE EDITOR

Michael J. Dalling, Ph.D., is a Reader in Crop Physiology in the School of Agriculture and Forestry at the University of Melbourne. A native of Australia, he received his B.Agr.Sc. from the University of Melbourne in 1967 and his M.Agr.Sc. from the same university in 1969. Dr. Dalling received his Ph.D. in Agronomy in 1971 from the University of Illinois. In 1974, he was appointed a Lecturer in Agronomy at the University of Melbourne and in 1985 he was appointed to the position of Reader in Crop Physiology. In 1981 he was a Senior Scholar of the Australian-American Education Foundation at the University of California at Davis and in 1981/82 he was a Visiting Professor in the Department of General Botany at the Swiss Federal Institute of Technology. Dr. Dalling is the author of numerous journal articles and is a member of the Australian Society of Plant Physiologists, Agronomy Society of Australia, The Wheat Breeding Society of Australia, and the American Society of Plant Physiologists. His current research includes an attempt to understand the process of plant senescence and how it determines or influences grain yield.

CONTRIBUTORS

Alan J. Barrett, Ph.D., Sc.D.
Department of Biochemistry
Strangeways Laboratory
Worts Causeway
Cambridge, England

Prem Lata Bhalla, Ph.D.
Research Fellow
School of Agriculture and Forestry
The University of Melbourne
Parkville, Victoria, Australia

Thomas Boller, Ph.D.
Lecturer
Botanisches Institut
Abteilung Pflanzenphysiologic
University of Basel
Basel, Switzerland

William R. Ellis, Ph.D.
Director of Inoculation Research
Research Seed, Inc.
St. Joseph, Missouri

Urs Feller, Ph.D.
Privatdozent
Institute of Plant Physiology
University of Bern
Bern, Switzerland

Arthur W. Galston, Ph.D.
Professor
Department of Biology
Yale University
New Haven, Connecticut

Ravindar Kaur-Sawhney, Ph.D.
Research Associate
Department of Biology
Yale University
New Haven, Connecticut

Heinrich Kauss, Ph.D.
Professor
Department of Biology
University of Kaiserslautern
Kaiserslautern, West Germany

R. Bruce Knox, Ph.,D., D.Sc.
Professor
Department of Botany
Plant and Cell Biology Research Center
The University of Melbourne
Parkville, Victoria, Australia

James E. Kruger, Ph.D.
Research Scientist
Grain Research Laboratory
Canadian Grain Commission
Winnipeg, Manitoba, Canada

John Michael Lord, Ph.D.
Reader
Department of Biological Sciences
University of Warwick
Coventry, England

Juhani Mikola, Ph.D.
Associate Professor
Department of Biology
University of Jyvškylä
Jyväskylä, Finland

Leena Mikola, Ph.D.
Research Associate
Department of Biology
University of Jyväskylä
Jyväskylä, Finland

Angela M. Nettleton
School of Agriculture and Forestry
The University of Melbourne
Parkville, Victoria, Australia

Ann Oaks, PhD.
Professor
Department of Biology
McMaster University
Hamilton, Ontario, Canada

K. R. Preston, Ph.D.
Grain Research Laboratory
Canadian Grain Commission
Winnipeg, Manitoba, Canada

Paul Reibach, Ph.D.
Plant Physiologist
Rohm and Haas Research Laboratories
Spring House, Pennsylvania

Colin Robinson, Ph.D.
Lecturer
Department of Biological Sciences
University of Warwick
Coventry, England

Mohan B. Singh, Ph.D.
Research Fellow
Department of Botany
Plant Cell Biology Research Center
The University of Melbourne
Parkville, Victoria, Australia

Richard Storey, Ph.D.
Associate Professor of Biology
Department of Biology
The Colorado College
Colorado Springs, Colorado

Carroll P. Vance, Ph.D.
Professor and Research Plant Physiologist
U.S. Department of Agriculture
Agricultural Research Service
Department of Agronomy and Plant Genetics
University of Minnesota
St. Paul, Minnesota

Fred W. Wagner, Ph.D.
Professor
Department of Agricultural Biochemistry
University of Nebraska-Lincoln
Lincoln, Nebraska

William Wallace, Ph.D.
Senior Lecturer
Department of Agricultural Biochemistry
Waite Agricultural Research Institute
University of Adelaide
Glen Osmond, South Australia

Karl A. Wilson, Ph.D.
Associate Professor
Department of Biological Sciences
State University of New York at Binghamton
Binghamton, New York

TABLE OF CONTENTS

Chapter 1

THE CLASSES OF PROTEOLYTIC ENZYMES

Alan J. Barrett

TABLE OF CONTENTS

I. PROTEASES

For about a century, the enzymes that play the central role in the degradation of proteins by hydrolyzing peptide bonds have been known as "proteases". The term protease is therefore equivalent to "peptide hydrolase". A little over 50 years ago, the German physiological chemists realized that there was an important distinction to be made between the hydrolases that act directly on proteins, and those that act on the peptides that are intermediates in proteolysis. The proteases that act on *proteins* were called *proteinases*, and those that act on *peptides* were called *peptidases*.[1,2] Thus, the footnote to the paper of Grassmann and Dyckerhoff[1] may be freely translated as follows:

> *Remarks on the nomenclature of proteolytic enzymes:* . . . the word "protease" . . . is used as a general name for all those enzymes that participate in protein degradation. Therefore . . . we propose to designate as "proteinases" enzymes that hydrolyze natural, intact proteins, and which have been called "proteases in the strict sense", "primary proteases", and "actual proteases" . . . Although most probably all enzymes that take part in protein degradation hydrolyze . . . –CO–NH– bonds, the name "peptidases" should be restricted . . . to those proteolytic enzymes that have been described as hydrolyzing peptide substrates of established structure . . .

Later, the underlying reasons for the preference of proteinases for action on intact proteins or long polypeptide chains, and that of "peptidases" for action on small peptides, became clear. These depended on the acceptability of the amino and carboxyl termini of polypeptides in the specificity sites of the enzymes. Thus, the end groups were not required, and typically were not tolerated, in the specificity sites of the proteinases, so that the enzymes acted well on long chains, away from the ends, but acted less well on the peptide products. In contrast, the specificity of the "peptidases" required at least one terminus to occupy a specificity site, so that these enzymes had little action on intact proteins, in which few of the peptide bonds are close to a terminus.

In the course of discussion of the mechanism of proteolysis (e.g., Reference 3), the understanding of the general characteristics of the action of the two subdivisions of proteases was embodied in the alternative names *endopeptidase* for those acting in the interior of polypeptide chains, and *exopeptidase* for those acting at the termini.[4] These names are, of course, analogous to those used for hydrolases acting on other polymers such as polysaccharides and polynucleotides. They are undoubtedly valuable, and one should note that they use the root "-peptidase" in the broad sense of "peptide hydrolase" or "polypeptidase", and not as Grassmann had used the word.*

Most proteases are unmistakably endopeptidases or exopeptidases, but a few have marked activities of both types. By convention, any enzyme that has distinct proteinase activity is treated as a proteinase, even though it may have peptidase activity too.

In the physiological context, an inevitable consequence of the specificities of the endopeptidases and exopeptidases is that the former are responsible for the early stages of protein breakdown, and the latter enzymes take over at an intermediate stage and complete the generation of free amino acids. As is generally found with biochemical pathways, the endopeptidase-catalyzed initial stages are typically the rate-limiting ones. Once the degradation of a protein molecule has started, it proceeds rapidly, and there is little or no accumulation of partial degradation products. Most of the known examples of limited proteolysis, such as those in the post-translational processing of proteins (Chapter 11), and the activation of enzyme precursors, are attributable to endopeptidases, but there is no doubt that exopeptidases also participate in these important aspects of biological control.

* *Note added in proof:* It has now been decided that "peptidase" should, in the future, be used only in the broad sense, meaning "peptide (bond) hydrolase", and the term therefore is synonymous with "protease" (Barrett, A. J., Nomenclature: a possible solution to the "peptidase" anomaly, *Biochem. J.*, 231, 807, 1985).

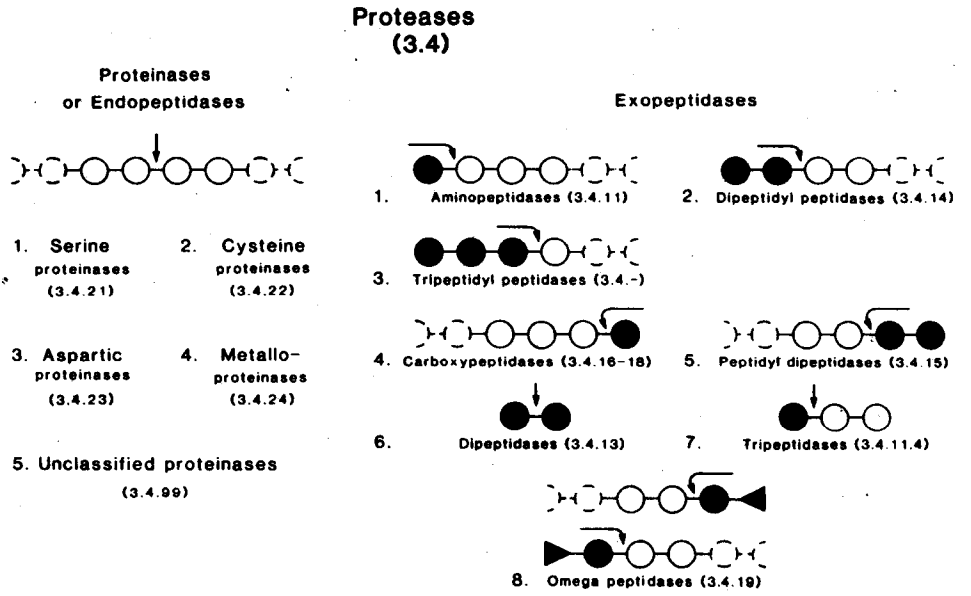

FIGURE 1. The classification of proteolytic enzymes. As is explained more fully in the text, the enzymes that hydrolyze peptide bonds, all of which are termed proteases, are divided into endopeptidases and exopeptidases. The endopeptidases are classified on the basis of their catalytic mechanism, which tends to reflect their evolutionary relationships, whereas the exopeptidases are classified on the basis of their substrate specificity, like almost all other enzymes. The numbers given in parentheses indicate the divisions into which the enzymes have been placed in the enzyme nomenclature scheme of the International Union of Biochemistry.[5,6] In the diagrammatic representation of the types of activity of the exopeptidases, the open circles represent amino acid residues, and the filled circles are those of the fragment released.

The way in which the proteases are classified, as described in this and the following sections, is summarized in Figure 1.

II. PROTEINASES OR ENDOPEPTIDASES

A. The Active Sites

Proteinases differ from almost all other enzymes in that their substrate specificities are invariably extremely difficult to define, and certainly do not represent an acceptable basis on which to name and classify the enzymes. A solution to this problem was put forward by Hartley,[7] who showed that the proteinases could be seen to act through four distinct catalytic mechanisms, and hence could be regarded as *serine, thiol, acid,* or *metal* proteinases. Now that much more is known about the chemistry of the catalytic sites, three of the four names have been amended, so that we now speak of *serine, cysteine, aspartic,* and *metallo-*proteinases,[8] but the concept of distinguishing these groups of enzymes remains completely valid.

An important development in the study of the catalytic mechanism of proteinases stemmed from the observation that the hydrolysis of 4-nitrophenyl acetate by chymotrypsin appeared to be a two-step process.[9] This has now been found to be true for the action of proteases generally, and the sequence of reactions may be illustrated thus:

$$E + S \underset{}{\overset{K_s}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} ES + P_1 \overset{k_3}{\rightarrow} E + P_2$$

As with all enzyme reactions, the first step is the reversible formation of a noncovalent enzyme-substrate complex, with equilibrium constant K_s. This involves interactions between specificity subsites of the enzyme, and amino acid sidechains and the peptide backbone of the substrate, and dictates the selectivity of the enzyme. The second stage of the reaction involves an attack on the scissile peptide bond on the substrate by an activated chemical grouping in the catalytic site. This results in the liberation of the first product of the peptide bond cleavage, e.g., the C-terminal portion of the peptide, and leaves the other part of the substrate still associated with the enzyme, e.g., as an acyl (peptidyl)-enzyme. The reaction is then completed by release of the second product, and regeneration of the free enzyme.

A special type of substrate that illustrates the two-stage action of proteases in a very clear and useful way is the "burst" substrate, of which the best example is *p*-nitrophenyl-*p'*-guanidinobenzoate with trypsin.[10] The reaction proceeds rapidly to the release of free *p*-nitrophenol (which is readily followed spectrophotometrically), and the formation of the guanidinobenzoyl-trypsin. The release of the second product is very slow, however, so that what is observed is a "burst" of nitrophenol, equimolar to the functional active sites of trypsin, and then very little further turnover of the substrate. This provides a valuable method to determine the concentration of active trypsin. There are a few other useful burst titrants for serine and cysteine proteinases.

The various types of protease that we shall survey briefly in the remainder of this chapter can be distinguished in terms of (1) the requirements of the substrate binding sites that first bring the substrate into the correct orientation for the activity of the catalytic machinery, (2) the nature of the group in the enzyme that forms the temporary E-P_2 bond in ES, and (3) the identity of the groups that catalyze the decay of this intermediate.

The serine, cysteine, aspartic, and metallo-proteinases are most clearly distinguished in the laboratory by use of tests with appropriate active site directed inhibitors. Practical suggestions for the use of some of these were made by Barrett,[11] and some of the more recently discovered reagents are also discussed in Chapter 2.

Precise molecular structures have been determined for representatives of all four classes of proteinase by X-ray crystallography, and in each case it has been found that the catalytic site lies in a cleft on the surface of the enzyme molecule. The substrate polypeptide chain lies along the active site cleft, and on either side of the catalytic site are specificity subsites adapted to binding amino acid sidechains, or the polypeptide backbone.

A study of the specificity of papain led Berger and Schechter[12] to put forward a terminology for the specificity subsites and for the complementary parts of the substrate structure that has become accepted for endopeptidases generally (Figure 2).

B. Evolution of the Endopeptidases

It is one of the valuable aspects of the system of classifying proteinases by catalytic mechanism that the groups thus created do not cut across evolutionary relationships in the way that classification by specificity would do. Nevertheless, the four major groups of proteinases are not all monophyletic. Evolutionary homology of the proteinases is demonstrated by similarities in amino acid sequence and molecular structure as revealed by X-ray crystallography.[13,14] Similarities of amino acid sequence are susceptible to statistical analysis, and Dayhoff et al.[14] has introduced the valuable term "superfamily" to describe the largest groups of proteins that show statistical evidence of evolutionary homology with each other. As is explained more fully later, the modern aspartic proteinases seem to have arisen from a single primitive enzyme, whereas the serine proteinases are derived from two stocks, the cysteine proteinases probably from at least four, and the evolution of the metallo-proteinases is unclear. The aspartic proteinases appeared most recently.

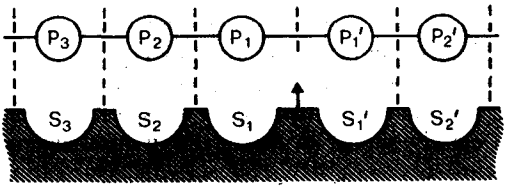

FIGURE 2. Scheme for the terminology of specificity subsites of proteases, and the complementary features of the substrate. This scheme, for the active site of an endopeptidase, shows how substrate-binding subsites are considered to be located on either side of the catalytic group, in the active site cleft. The subsites are numbered S_1, S_2, etc. away from the catalytic site toward the amino terminus of the substrate, and S_1', S_2', etc. toward the carboxyl terminus. The subsites are usually thought of as binding amino acid sidechains, but there are also important interactions with the polypeptide backbone and with terminal blocking groups of synthetic substrates. In exopeptidases, the cleft may be "blind" on one side, so as to prevent the binding of extended polypeptides. (Adapted from Berger and Schechter.[12])

C. Serine Proteinases

1. Evolution and Occurrence

The serine-dependent mechanism of peptide bond hydrolysis may be the most successful that has evolved for endopeptidases, in that the serine proteinases (EC number 3.4.21) seem to be the most numerous group, and are extremely widespread and diverse.

There are two superfamilies of serine proteinases: the *trypsin* superfamily, and the *subtilisin* superfamily.[15,16] The subtilisin-related enzymes have been found only in bacteria, whereas the trypsin-related proteinases are found in both prokaryotic and eukaryotic microorganisms, plants, and both invertebrate and vertebrate animals. Young et al.[17] have reviewed the evolution of the trypsin superfamily. The primitive proteinases of the myxobacteria and streptomycetes contain 200 to 220 amino acid residues in a single chain. Most of the related proteinases in higher animals have extensions of the N-terminus that form activation peptides such that proteolytic cleavage requires 200 to 260 residues back from the C-terminus to generate the enzymically active molecule. These extensions are short in trypsinogen and proelastase, and the small activation peptides are lost after cleavage. By contrast, the activation peptides of the precursors of the plasma proteinases, factor Xa, thrombin, and plasmin are very complex, and confer new properties on the zymogens and enzymes. For example, a specialized Ca^{2+}-binding site exists at the C-terminus of factor X and prothrombin (and several of the other coagulation factors). Also, prothrombin and plasminogen have acquired multiple copies of a complex, disulfide-crosslinked structure known as the kringle. In each of these proteins, the activation process yields a chain of very roughly 250 residues that contains the residues of the "catalytic triad", serine, histidine, and aspartic acid. In a few proteins, this chain is further fragmented, but much greater complexity arises from continued association of parts of the extended N-terminal sequences with the active chain.

2. Active Sites

a. Catalytic Sites

The hydroxyl group of the active site serine residue attacks the carbonyl carbon atom of the substrate, with general base catalysis by histidine. This leads to the formation of a tetrahedral intermediate and an imidazolium ion. The intermediate breaks down by general acid catalysis to an acyl-enzyme, an imidazole base, and alcohol or amine. The acyl enzyme is hydrolyzed through the reverse reaction pathway, with the hydroxyl group of water acting

as the nucleophile instead of the hydroxyl group of serine. This mechanism is described, and details of the mechanism are critically discussed by Polgar and Halasz.[18] The "charge-relay system", once widely accepted, now seems unlikely to contribute, but the "oxyanion hole" seems important. It is remarkable that the enzymes of the subtilisin superfamily have acquired a mechanism closely similar to that of the trypsin superfamily by convergent evolution.

Activity of the serine proteinases is most commonly maximal at slightly alkaline pH. There are no absolute activator requirements, but Ca^{2+} ions are in the activation reaction of some of the proenzymes, and to stabilize some of the enzymes.

b. Specificity Sites

The specificities of serine proteinases are exceedingly diverse. Fitting of the S rather than the S′ sites (see Figure 2) is important, and perhaps the simplest example is that of trypsin, in which the overriding requirement is the occupation of subsite S_1 by a sidechain of arginine or lysine. This type of specificity has been refined in numerous enzymes evolved for the limited proteolysis of particular substrates, especially in the coagulation[19] and complement[20] systems. Such refinement has been achieved by making subsites S_2 and S_3 more discriminating, while maintaining the specificity of S_1 for basic residues. Generally, synthetic substrates and inhibitors designed to match the sequences of the natural substrates have been found to show good specificity.[21-22]

Other serine proteinases show primary specificity for large hydrophobic sidechains (chymotrypsin, subtilisin) or small aliphatic sidechains (elastases) in S_1. There are innumerable other variations, many of which are described in References 23 and 24.

c. Inhibitors

Diisopropyl fluorophosphate is by far the most useful inhibitor for identification of serine proteinases; it has very little action on nonserine proteinases, but it must be noted that rates of reaction with serine proteinases differ by many orders of magnitude. Phenylmethane sulfonylfluoride is somewhat more convenient to obtain and use, but reacts only very slowly with some serine proteinases. Other important irreversible inhibitors of serine proteinases are the blocked aminoacyl and peptidyl chloromethanes, which can be made highly selective.[25] When using any of these irreversible inhibitors, one should expect to obtain complete inhibition, or to show that partial inhibition is the result of an incomplete, time-dependent process with simple kinetics. Data often presented in the literature for partial inhibition are difficult to evaluate.

Important reversible inhibitors include natural and synthetic peptide aldehydes.[26,27]

None of the many known protein inhibitors of serine proteinases reacts with all of these enzymes, but nevertheless the protein inhibitors can be useful since they have very little reactivity with other types of proteinase. Examples are soya bean trypsin inhibitor, lima bean trypsin inhibitor, and aprotinin.[28,29] These are all more or less tight-binding reversible inhibitors, so the degree of inhibition should be expected to be dependent on concentration but not on time (in normal assays), and to comply with a definite K_i (equilibrium constant for the dissociation of the enzyme-inhibitor complex).

For experiments with living materials, in which inhibitors are used to characterize the proteinase responsible for a particular biological process, the broad-specificity, irreversible inhibitors of serine proteinases are generally too unstable and too toxic. Irreversible inhibitors with structure optimized for specific enzymes can be used in vivo, however.[25] The protein inhibitors are also applicable, although again somewhat narrow in specificity. Low molecular weight reversible inhibitors such as the aldehydes, leupeptin, chymostatin, and elastatinal have been used with success,[26,30] but these too are selective for only certain serine proteinases, and moreover, commonly inhibit cysteine proteinases.

D. Cysteine Proteinases

1. Evolution and Occurrence

Cysteine proteinases (EC number 3.4.22) are found in bacteria,[31] in eukaryotic microorganisms,[32] and in plants[33] and animals.[23] There may be four or more distinct superfamilies of cysteine proteinases, however.

The most studied cysteine proteinase is papain, from the latex of *Carica papaya L.*, and it is now clear that the papain superfamily is a large one.[34] Amino acid sequence data show that cysteine proteinases of other higher plants are members of the papain superfamily, as are cathepsins B, H, and L of mammalian lysosomes.[34] Many protozoa contain cysteine proteinases that have M_r about 25,000, and other properties consistent with their being related to papain.[32] It is conceivable that the first of the cysteine proteinases of the papain superfamily functioned in the digestive vacuoles of a protozoan. The enzymes have continued in evolution to tend to be associated with the intracellular vacuoles of plant cells (and thence to latex), and of animal cells.

Two bacterial cysteine proteinases that currently seem to represent independent superfamilies not occurring in higher plants and animals are clostripain from *Clostridium histolyticum*,[35] and the proteinase from a *Streptococcus* species.[36] Clostripain is highly specific for substrates with an arginyl residue at P_1, like some serine proteinases,[35] and is calcium activated. The amino acid sequence around the essential cysteine is unlike that of other known cysteine proteinases.[37] The streptococcal proteinase has an inactive zymogen that is activated by limited proteolysis.[38] Zymogens of serine-, aspartic, and metallo-proteinases are common in higher animals, but are uncommon among the cysteine proteinases, and in primitive organisms.

There are calcium-dependent cysteine proteinases in invertebrate and vertebrate animals that are widely distributed, generally in solution in the cytoplasmic fraction of the cells. These enzymes are forms of *calpain*, which differ in requiring micromolar or millimolar concentrations of Ca^{2+} for activity, but have a highly conserved molecular structure comprising two subunits of M_r 80,000 and 30,000, respectively.[39,40] The essential cysteine residue is in the M_r 80,000 subunit, and there is, at present, no evidence that these cysteine proteinases are related to papain.

2. Active Sites

a. Catalytic Sites

The sensitivity of papain, ficin, and bromelain to inactivation by thiol-blocking reagents, and their activation by reagents expected to regenerate thiol groups from disulfides, led Hartley[7] to classify them as "thiol proteinases". Of course, cysteine is the only generally occurring amino acid that has a thiol group, and the essential thiol group of papain was shown to be on the sidechain of Cys-25 by Light et al.[41] By analogy with the term "serine proteinase", it was clear that these enzymes should be called "cysteine proteinases". The term was used,[8,24] and eventually adopted officially.[5]

Polgar and Halasz[18] have reviewed what is known of the catalytic mechanism of the (papain superfamily) cysteine proteinases. The thiolate-imidazolium ion pair between Cys-25 and His-159 (papain numbering) is the reactive nucleophile. An acyl enzyme is formed with the thiol of Cys-25, and hydrolyzed as much as for the serine proteinases. A role proposed for Asp-158 in the catalytic mechanism seems to be precluded on mechanistic grounds,[18] and by the fact that this residue is replaced by other amino acids in the sequences of cathepsins B and H, and bromelain.[34] The role of the "oxyanion hole" is more doubtful in the cysteine proteinases than in the serine proteinases.[18]

Conditions required for the activity of cysteine proteinases vary widely with enzymes and substrates. As regards pH, at the low end of the range, the lysosomal cysteine proteinases all show acid pH optima. This is due to several factors, and is especially marked with

proteins, e.g., the optimum for action on collagen can be as low as pH 3.5. The ionic interaction of cysteine and histidine in the catalytic site produces a structure with pK_a about 4,[43] as contrasted with a value close to pH 8 for a typical thiol group. This greatly increases reactivity at acid pH, but other factors are the denaturation of the enzymes above pH 7, which causes a skewing of the pH dependence curve in most assays, and the unfolding of some protein substrates at acid pH. Papain has a broad pH optimum, with activity extending well into the alkaline range, and the calpains are most active at about pH 7.5.

Of course, a low molecular weight thiol compound is usually necessary as an activator of cysteine proteinases, and in the test-tube dithiothreitol is convenient because of its resistance to autoxidation. The activation of the enzymes can be a slow reaction, especially at acidic pH, and care is necessary to ensure that it is complete in a quantitative assay. A second-order rate constant of 4.1 $M^{-1}s^{-1}$ has been reported for the activation of cathepsin B by dithiothreitol at pH 6; this implies that complete activation by 2 m*M* dithiothreitol takes about 7 min.[44] High concentrations of dithiothreitol are undesirable, however, since they may reduce disulfides in protein substrates, with uncontrolled results, and also destabilize the enzymes.[45]

For the papain-like cysteine proteinases, it is usual to include a chelating agent, since some metal ions, e.g., Zn^{2+}, inhibit even in the presence of a thiol, but of course calcium ions and no chelating agent are appropriate for the calpains and clostripain.

b. Specificity Sites

Some progress has been made in mapping the specificity sites of papain and related cysteine proteinases with synthetic substrates and inhibitors. Papain has an extended specificity site, containing about 5 to 7 significant subsites,[12] but the dominant one is S_2, which is a hydrophobic pocket that binds a phenylalanyl sidechain very well.[46]

Other enzymes of the superfamily tend to retain this characteristic, but there are some striking variations. For example, although cathepsin B accepts a phenylalanyl sidechain in S_2, it also binds an arginyl sidechain well. One possible explanation for this is that Val-133, which forms part of the end of the P_2 pocket,[46] is replaced by glycine, which may allow the sidechain in P_2 to form an ionic interaction with residue 205 (glutamic acid in cathepsin B, serine in papain).[34] Shaw et al.[47] has shown how low M_r inhibitors can be used to explore the S_1 subsite of cathepsin B.

The detailed studies of the preferences of the specificity subsites of the papain superfamily proteinases for binding of low M_r compounds have not yet made it possible to predict with any confidence what bonds will be cleaved in proteins.[48] Indeed, there may be fundamental differences between the reactions of the enzymes with low M_r compounds and with proteins, for although the enzymes tend to have similar specific activities as proteinases (within about an order of magnitude), some have very low or negligible activities with synthetic substrates tested so far.[49]

Although, on balance, it seems that the streptococcal cysteine proteinase is unrelated to papain,[36] this enzyme has an S_2 subsite with a preference for sidechains resembling that of phenylalanine, like papain. In contrast, the cysteine proteinase of *Clostridium histolyticum* is highly specific for the cleavage of arginyl bonds in both synthetic and protein substrates.[35] Very little is known of the specificity of the calpains.

c. Inhibitors

The identification of cysteine proteinase activity in the test-tube is not usually difficult, since the activity is generally stimulated by low M_r thiol compounds, as well as being inhibited by thiol-blocking reagents such as iodoacetate. The reactivity of the essential thiol with iodoacetate is generally much greater than that of a low M_r thiol compound, so that iodoacetate will readily inactivate the enzyme, even in the presence of an excess of the thiol

activator, and at low pH. The covalent reactions with iodoacetatamide and N-ethylmaleimide tend to be slower than that with iodoacetate, and the mercurial reagents also are less satisfactory, reacting reversibly.

Peptidyl chloromethanes, and aldehydes such as leupeptin and antipain, inactivate serine proteinases as well as cysteine proteinases, so they are of little use as diagnostic reagents. Much more valuable are the diazomethanes such as Z-Phe-Ala-CHN_2,[50] and the epoxides such as compound E-64 (3-carboxyl-2,3-L-*trans*-epoxypropyl-leucylamido(4-guanidino)butane).[51] Not only are these specific for cysteine proteinases and unreactive with low M_r thiol compounds, but several of them react rapidly enough with the papain-like proteinases to allow stoichiometric titration of their activity in very dilute solutions.[51,52]

In addition to the low M_r inhibitors, the cysteine proteinases are sensitive to protein inhibitors. The papain-like enzymes are mostly inhibited by a group of proteins called cystatins, of which chicken egg-white cystatin is not too difficult to obtain as a reagent.[53] Cystatins do not inhibit the calpains, but these enzymes are sensitive to another protein, calpastatin,[39,40] and to a plasma protein α-cysteine proteinase inhibitor.

For in vivo inhibition of cysteine proteinases, both the diazomethanes and the epoxides seem to have promise.

E. Aspartic Proteinases

1. Evolution and Occurrence

The aspartic proteinases seem to comprise a single evolutionary superfamily, and to be confined to the eukaryotes, which suggests that they are one of the younger superfamilies of proteinases. The extensive literature on the aspartic proteinases of eukaryotic microorganisms has been reviewed by North,[32] who notes that the fungal aspartic proteinases generally have M_r values in the range 30,000 to 45,000 and isoelectric points below pH 5.1. The mammalian aspartic proteinases have similar properties,[23] although forms of cathepsin D occur with pI up to 6.5.

The most thoroughly studied of the aspartic proteinases is undoubtedly pepsin, the proteinase responsible for the digestion of food proteins in the stomach in higher animals. Foltmann and Pedersen[54] have used sequencing and immunological methods to show that the mammalian gastric aspartic proteinases are all evolutionary relatives of pepsin, as is cathepsin D, the intracellular aspartic proteinase of the lysosomal system.[55] A combination of amino acid sequence analysis and X-ray crystallography has shown that the acid proteinases of the fungi, *Rhizopus chinensis*, *Penicillium ianthinellum*, and *Endothia parasitica* also belong to the same superfamily (reviewed by North[32]). Nepenthesin is the aspartic proteinase in the digestive fluid of the pitcher plant.

The gastric aspartic proteinases of higher animals are secreted from the gastric lining cells as zymogens — inactive precursors that are converted to the active enzymes by proteolytic cleavage of an N-terminal extension of the polypeptide chain. This limited proteolysis is mediated by the zymogens themselves, at acid pH. Pepsin A is irreversibly inactivated at and above neutral pH, which is probably a significant mechanism by which its activity is kept localized in the body. Pepsinogen, on the other hand, is stable to neutral pH.

2. Active Sites

a. Catalytic Sites

Hartley[7] termed the enzymes that we now know as aspartic proteinases "acid proteinases" because it was not possible at that time to identify the catalytic residues. The very low pH optima of some of the acid proteinases indirectly implicated carboxyl groups in the catalytic mechanism, but more direct evidence came from the inactivation of pepsin by carboxyl-blocking reagents (reviewed by Fruton[56]).