Molecular and Biotechnological Aspects of Microbial Proteases

Mala B. Rao, Aparna M. Tanksale, Mohini S. Ghatge and Vasanti V. Deshpande

Molecular and Biotechnological Aspects of Microbial Proteases†

MALA B. RAO, APARNA M. TANKSALE, MOHINI S. GHATGE, AND VASANTI V. DESHPANDE*

Division of Biochemical Sciences, National Chemical Laboratory, Pune 411008, India

INTRODUCTION........................................................................................................................................598
SCOPE OF THE REVIEW...........................................................................................................................599
SOURCES OF PROTEASES..........................................................................................................................599
Plant Proteases.........................................................................................................................................599
Papain.......................................................................................................................................................599
Bromelain................................................................................................................................................599
Keratinases.................................................................................................................................................599
Animal Proteases.....................................................................................................................................599
Trypsin......................................................................................................................................................599
Chymotrypsin.........................................................................................................................................600
Pepsin........................................................................................................................................................600
Rennin.......................................................................................................................................................600
Microbial Proteases.................................................................................................................................600
Bacteria....................................................................................................................................................600
Fungi.........................................................................................................................................................600
Viruses......................................................................................................................................................600
CLASSIFICATION OF PROTEASES........................................................................................................600
Exopeptidases.........................................................................................................................................601
Aminopeptidases....................................................................................................................................601
Carboxypeptidases.................................................................................................................................601
Endopeptidases.......................................................................................................................................601
Serine proteases....................................................................................................................................601
(i) Serine alkaline proteases ...................................................................................................................602
(ii) Subtilisins.........................................................................................................................................602
Aspartic proteases.................................................................................................................................602
Cysteine/thiol proteases .........................................................................................................................602
Metalloproteases...................................................................................................................................602
MECHANISM OF ACTION OF PROTEASES..........................................................................................603
Serine Proteases.....................................................................................................................................603
Aspartic Proteases.................................................................................................................................603
Metalloproteases...................................................................................................................................604
Cysteine Proteases.................................................................................................................................604
PHYSIOLOGICAL FUNCTIONS OF PROTEASES..................................................................................605
Protein Turnover.....................................................................................................................................605
Sporulation and Conidial Discharge.........................................................................................................605
Germination............................................................................................................................................605
Enzyme Modification.............................................................................................................................605
Nutrition..................................................................................................................................................606
Regulation of Gene Expression.............................................................................................................606
APPLICATIONS OF PROTEASES............................................................................................................606
Detergents...............................................................................................................................................606
Leather Industry.......................................................................................................................................606
Food Industry.........................................................................................................................................607
Dairy industry.........................................................................................................................................607
Baking industry......................................................................................................................................607
Manufacture of soy products..................................................................................................................607
Debittering of protein hydrolysates.......................................................................................................607
Synthesis of aspartame...........................................................................................................................607
Pharmaceutical Industry.........................................................................................................................607
Other Applications.................................................................................................................................607

* Corresponding author. Mailing address: Division of Biochemical Sciences, National Chemical Laboratory, Pune-411008, India. Phone: 091-212-338234, Fax: 091-212-338234.
† National Chemical Laboratory communication 6440.
INTRODUCTION

Proteases are the single class of enzymes which occupy a pivotal position with respect to their applications in both physiological and commercial fields. Proteolytic enzymes catalyze the cleavage of peptide bonds in other proteins. Proteases are degradative enzymes which catalyze the total hydrolysis of proteins. Advances in analytical techniques have demonstrated the cleavage of peptide bonds in other proteins. Proteases execute a large variety of functions, extending from cascade systems such as hemostasis and inflammation. They

enzymes by limited proteolysis, blood clotting and lysis of fibrin clots, and processing and transport of secretory proteins across the membranes. The current estimated value of the worldwide sales of industrial enzymes is $1 billion (72). Of the industrial enzymes, 75% are hydrolytic. Proteases represent one of the three largest groups of industrial enzymes and account for about 60% of the total worldwide sale of enzymes (Fig. 1). Proteases execute a large variety of functions, extending from the cellular level to the organ and organism level, to produce cascade systems such as hemostasis and inflammation. They
are responsible for the complex processes involved in the normal physiology of the cell as well as in abnormal pathophysiological conditions. Their involvement in the life cycle of disease-causing organisms has led them to become a potential target for developing therapeutic agents against fatal diseases such as cancer and AIDS. Proteases have a long history of application in the food and detergent industries. Their application in the leather industry for dehairing and bating of hides to substitute currently used toxic chemicals is a relatively new application in the food and detergent industries. Their application in the leather industry for dehairing and bating of hides to substitute currently used toxic chemicals is a relatively new application in the food and detergent industries.

**SCOPE OF THE REVIEW**

Since proteases are enzymes of metabolic as well as commercial importance, there is a vast literature on their biochemical and biotechnological aspects (64, 128, 192, 235, 309). However, the earlier reviews did not deal with the molecular biology of proteases, which offers new possibilities and potentials for their biotechnological applications. This review aims at analyzing the updated information on biochemical and genetic aspects of proteases, with special reference to some of the advances made in these areas. We also attempt to address some of the deficiencies in the earlier reviews and to identify problems, along with possible solutions, for the successful applications of proteases for the benefit of mankind. The genetic engineering approaches are also discussed, from the perspective of making better use of proteases. The reference to plant and animal proteases has been made to complete the overview.

**TABLE 1. Major protease producers**

<table>
<thead>
<tr>
<th>Company</th>
<th>Country</th>
<th>Market share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novo Industries</td>
<td>Denmark</td>
<td>40</td>
</tr>
<tr>
<td>Gist-Brocades</td>
<td>Netherlands</td>
<td>20</td>
</tr>
<tr>
<td>Genencor International</td>
<td>United States</td>
<td>10</td>
</tr>
<tr>
<td>Miles Laboratories</td>
<td>United States</td>
<td>10</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

**TABLE 2. Specificity of proteases**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Peptide bond cleaved*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>Lys (or Arg) ---------</td>
</tr>
<tr>
<td>Chymotrypsin, subtilisin</td>
<td>Trp (or Tyr, Phe, Leu)</td>
</tr>
<tr>
<td>Staphylococcus V8 protease</td>
<td>Asp (or Glu)</td>
</tr>
<tr>
<td>Papain</td>
<td>Phe (or Val, Leu) ----</td>
</tr>
<tr>
<td>Thermolysin</td>
<td>Leu (or Phe) ---------</td>
</tr>
<tr>
<td>Pepsin</td>
<td>Phe (or Tyr, Leu) ---- Trp (or Phe, Tyr)</td>
</tr>
</tbody>
</table>

* The arrow indicates the site of action of the protease. Xaa, any amino acid residue.

However, the major emphasis of the review is on the microbial proteases.

**SOURCES OF PROTEASES**

The use of plants as a source of proteases is governed by several factors such as the availability of land for cultivation and the suitability of climatic conditions for growth. Moreover, production of proteases from plants is a time-consuming process. Papain, bromelain, keratinases, and ficin represent some of the well-known proteases of plant origin. Papain. Papain is a traditional plant protease and has a long history of use (250). It is extracted from the latex of Carica papaya fruits, which are grown in subtropical areas of west and central Africa and India. The crude preparation of the enzyme has a broader specificity due to the presence of several proteinase and peptidase isozymes. The performance of the enzyme depends on the plant source, the climatic conditions for growth, and the methods used for its extraction and purification. The enzyme is active between pH 5 and 9 and is stable up to 80 or 90°C in the presence of substrates. It is extensively used in industry for the preparation of highly soluble and flavored protein hydrolysates.

Bromelain. Bromelain is prepared from the stem and juice of pineapples. The major supplier of the enzyme is Great Food Biochem., Bangkok, Thailand. The enzyme is characterized as a cysteine protease and is active at pH 5 to 9. Its inactivation temperature is 70°C, which is lower than that of papain.

Keratinases. Some of the botanical groups of plants produce proteases which degrade hair. Digestion of hair and wool is important for the production of essential amino acids such as lysine and for the prevention of clogging of wastewater systems.

**Animal Proteases**

The most familiar proteases of animal origin are pancreatic trypsin, chymotrypsin, pepsin, and rennins (23, 97). These are prepared in pure form in bulk quantities. However, their production depends on the availability of livestock for slaughter, which in turn is governed by political and agricultural policies.

Trypsin. Trypsin (Mr, 23,300) is the main intestinal digestive enzyme responsible for the hydrolysis of food proteins. It is a serine protease and hydrolyzes peptide bonds in which the carboxyl groups are contributed by the lysine and arginine residues (Table 2). Based on the ability of protease inhibitors to inhibit the enzyme from the insect gut, this enzyme has received attention as a target for biocontrol of insect pests.
Trypsin has limited applications in the food industry, since the protein hydrolysates generated by its action have a highly bitter taste. Trypsin is used in the preparation of bacterial media and in some specialized medical applications.

**Chymotrypsin.** Chymotrypsin (Mr 23,800) is found in animal pancreatic extract. Pure chymotrypsin is an expensive enzyme and is used only for diagnostic and analytical applications. It is specific for the hydrolysis of peptide bonds in which the carboxyl groups are provided by one of the three aromatic amino acids, i.e., phenylalanine, tyrosine, or tryptophan. It is used extensively in the deallergenizing of milk protein hydrolysates. It is stored in the pancreas in the form of a precursor, chymotrypsinogen, and is activated by trypsin in a multistep process.

**Pepsin.** Pepsin (Mr 34,500) is an acidic protease that is found in the stomachs of almost all vertebrates. The active enzyme is released from itszymogen, i.e., pepsinogen, by autocatalysis in the presence of hydrochloric acid. Pepsin is an aspartyl protease and resembles human immunodeficiency virus type 1 (HIV-1) protease, responsible for the maturation of HIV-1. It exhibits optimal activity between pH 1 and 2, while the optimal pH of the stomach is 2 to 4. Pepsin is inactivated above pH 6.0. The enzyme catalyzes the hydrolysis of peptide bonds between two hydrophobic amino acids.

**Rennin.** Rennet is a pepsin-like protease (rennin, chymosin; EC 3.4.23.4) that is produced as an inactive precursor, prorennin, in the stomachs of all nursing mammals. It is converted to active rennin (Mr 30,700) by the action of pepsin or by its autocatalysis. It is used extensively in the dairy industry to produce a stable curd with good flavor. The specialized nature of the enzyme is due to its specificity in cleaving a single peptide bond in κ-casein to generate insoluble para-κ-casein and C-terminal glycopeptide.

**Microbial Proteases**

The inability of the plant and animal proteases to meet current world demands has led to an increased interest in microbial proteases. Microorganisms represent an excellent source of enzymes owing to their broad biochemical diversity and their susceptibility to genetic manipulation. Microbial proteases account for approximately 40% of the total worldwide enzyme sales (72). Proteases from microbial sources are preferred to the enzymes from plant and animal sources since they possess almost all the characteristics desired for their biotechnological applications.

**Bacteria.** Most commercial proteases, mainly neutral and alkaline, are produced by organisms belonging to the genus *Bacillus*. Bacterial neutral proteases are active in a narrow pH range (pH 5 to 8) and have relatively low thermostability. Due to their intermediate rate of reaction, neutral proteases generate less bitterness in hydrolyzed food proteins than do the animal proteases and hence are valuable for use in the food industry. Neutrase, a neutral protease, is insensitive to the natural plant protease inhibitors and is therefore useful in the brewing industry. The bacterial neutral proteases are characterized by their high affinity for hydrophobic amino acid pairs. Their low thermostability is advantageous for controlling their reactivity during the production of food hydrolysates with a low degree of hydrolysis. Some of the neutral proteases belong to the metalloprotease type and require divalent metal ions for their activity, while others are serine proteases, which are not affected by chelating agents.

Bacterial alkaline proteases are characterized by their high activity at alkaline pH, e.g., pH 10, and their broad substrate specificity. Their optimal temperature is around 60°C. These properties of bacterial alkaline proteases make them suitable for use in the detergent industry.

**Fungi.** Fungi elaborate a wider variety of enzymes than do bacteria. For example, *Aspergillus oryzae* produces acid, neutral, and alkaline proteases. The fungal proteases are active over a wide pH range (pH 4 to 11) and exhibit broad substrate specificity. However, they have a lower reaction rate and worse heat tolerance than do the bacterial enzymes. Fungal enzymes can be conveniently produced in a solid-state fermentation process. Fungal acid proteases have an optimal pH between 4 and 4.5 and are stable between pH 2.5 and 6.0. They are particularly useful in the cheesemaking industry due to their narrow pH and temperature specificities. Fungal neutral proteases are metalloproteases that are active at pH 7.0 and are inhibited by chelating agents. In view of the accompanying peptidase activity and their specific function in hydrolyzing hydrophobic amino acid bonds, fungal neutral proteases supplement the action of plant, animal, and bacterial proteases in reducing the bitterness of food protein hydrolysates. Fungal alkaline proteases are also used in food protein modification.

**Viruses.** Viral proteases have gained importance due to their functional involvement in the processing of proteins of viruses that cause certain fatal diseases such as AIDS and cancer. Serine, aspartic, and cysteine proteases are found in various viruses (236). All of the virus-encoded peptides are endopeptidases; there are no metallopeptidases. Retroviral aspartyl proteases that are required for viral assembly and replication are homodimers and are expressed as a part of the polyprotein precursor. The mature protease is released by autolysis of the precursor. An extensive literature is available on the expression, purification, and enzymatic analysis of retroviral aspartic protease and its mutants (147). Extensive research has focused on the three-dimensional structure of viral proteases and their interaction with synthetic inhibitors with a view to designing potent inhibitors that can combat the relentlessly spreading and devastating epidemic of AIDS.

Thus, although proteases are widespread in nature, microbes serve as a preferred source of these enzymes because of their rapid growth, the limited space required for their cultivation, and the ease with which they can be genetically manipulated to generate new enzymes with altered properties that are desirable for their various applications.

**CLASSIFICATION OF PROTEASES**

According to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, proteases are classified in subgroup 4 of group 3 (hydrolyases) (114a). However, proteases do not comply easily with the general system of enzyme nomenclature due to their huge diversity of action and structure. Currently, proteases are classified on the basis of three major criteria: (i) type of reaction catalyzed, (ii) chemical nature of the catalytic site, and (iii) evolutionary relationship with reference to structure (12).

Proteases are grossly subdivided into two major groups, i.e., exopeptidases and endopeptidases, depending on their site of action. Exopeptidases cleave the peptide bond proximal to the amino or carboxyl termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate. Based on the functional group present at the active site, proteases are further classified into four prominent groups, i.e., serine proteases, aspartic proteases, cysteine proteases, and metalloproteases (85). There are a few miscellaneous proteases which do not precisely fit into the standard classification, e.g., ATP-dependent proteases which require ATP for activity (183). Based on their amino acid sequences,
Table 3. Classification of proteases

<table>
<thead>
<tr>
<th>Protease</th>
<th>Mode of action</th>
<th>EC no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exopeptidases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminopeptidases</td>
<td></td>
<td>3.4.11</td>
</tr>
<tr>
<td>Dipeptidyl peptidase</td>
<td></td>
<td>3.4.14</td>
</tr>
<tr>
<td>Tripeptidyl peptidase</td>
<td></td>
<td>3.4.14</td>
</tr>
<tr>
<td>Carboxypeptidase</td>
<td></td>
<td>3.4.16-3.4.18</td>
</tr>
<tr>
<td>Serine type peptase</td>
<td></td>
<td>3.4.16</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td></td>
<td>3.4.17</td>
</tr>
<tr>
<td>Cysteine type peptase</td>
<td></td>
<td>3.4.18</td>
</tr>
<tr>
<td>Peptidyl dipeptidase</td>
<td></td>
<td>3.4.15</td>
</tr>
<tr>
<td>Dipeptidas</td>
<td></td>
<td>3.4.13</td>
</tr>
<tr>
<td>Omega peptidases</td>
<td></td>
<td>3.4.19</td>
</tr>
<tr>
<td>Endopeptidases</td>
<td></td>
<td>3.4.21-3.4.34</td>
</tr>
<tr>
<td>Serine protease</td>
<td></td>
<td>3.4.21</td>
</tr>
<tr>
<td>Cysteine protease</td>
<td></td>
<td>3.4.22</td>
</tr>
<tr>
<td>Aspartic peptase</td>
<td></td>
<td>3.4.23</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td></td>
<td>3.4.24</td>
</tr>
<tr>
<td>Endopeptidases of unknown catalytic mechanism</td>
<td>3.4.99</td>
<td></td>
</tr>
</tbody>
</table>

* Open circles represent the amino acid residues in the polypeptide chain. Solid circles indicate the terminal amino acids, and stars signify the blocked termini. Arrows show the sites of action of the enzyme.

Endopeptidases

Endopeptidases are characterized by their preferential action at the peptide bonds in the inner regions of the polypeptide chain away from the N and C termini. The presence of the free amino or carboxyl group has a negative influence on enzyme activity. The endopeptidases are divided into four subgroups based on their catalytic mechanism. (i) Serine proteases, (ii) aspartic proteases, (iii) cysteine proteases, and (iv) metalloproteases. To facilitate quick and unambiguous reference to a particular family of peptidases, Rawlings and Barrett have assigned a code letter denoting the catalytic type, i.e., S, C, A, M, or U (see above) followed by an arbitrarily assigned number (236).

Serine proteases. Serine proteases are characterized by the presence of a serine group in their active site. They are numerous and widespread among viruses, bacteria, and eukaryotes, suggesting that they are vital to the organisms. Serine proteases are found in the exopeptidase, endopeptidase, oligopeptidase, and omega peptidase groups. Based on their structural similarities, serine proteases have been grouped into 20 families, which have been further subdivided into about six clans with common ancestors (12). The primary structures of the members of four clans, chymotrypsin (SA), subtilisin (SB), carboxypeptidase C (SC), and Escherichia coli-D-Ala-D-Ala peptidase A (SE) are totally unrelated, suggesting that there are at least four separate evolutionary origins for serine proteases. Clans SA, SB, and SC have a common reaction mechanism consisting of a common catalytic triad of the three amino acids, serine (nucleophile), aspartate (electrophile), and histidine (base). Although the geometric orientations of these residues are similar, the protein folds are quite different, forming a typical example of a convergent evolution. The catalytic mechanisms of clans SE and SF (repressor LexA) are distinctly different from those of clans SA, SB, and SC, since they lack the classical Ser-His-Asp triad. Another interesting feature of the serine proteases is the conservation of glycine residues in the vicinity of the catalytic serine residue to form the motif Gly-Xaa-Ser-Yaa-Gly (25).

Serine proteases are recognized by their irreversible inhibition by 3,4-dichloroisocoumarin (3,4-DCI), 1,3-carboxytrans-2,3-epoxypropyl-leucylamido (4-guanidine) butane (E.64), diisopropylfluorophosphate (DFP), phenylmethylsulfonyl fluoride (PMSF), and tosyl-l-lysine chloromethyl ketone (TLCK). Some of the serine proteases are inhibited by thiol reagents such as p-chloromercuribenzoate (PCMB) due to the presence of a cysteine residue near the active site. Serine proteases are generally active at neutral and alkaline pH, with an optimum between pH 7 and 11. They have broad substrate specificities including esterolytic and amides activity. Their molecular masses range between 18 and 35 kDa, for the serine protease from Blakeslea trispora, which has a molecular mass of 126 kDa (76). The isoelectric points of serine proteases are generally between pH 4 and 6. Serine alkaline proteases that are active at highly alkaline pH represent the largest subgroup of serine proteases.
(i) Serine alkaline proteases. Serine alkaline proteases are produced by several bacteria, molds, yeasts, and fungi. They are inhibited by DFP or a potato protease inhibitor but not by tosyl-L-phenylalanine chloromethyl ketone (TPCK) or TLCK. Their substrate specificity is similar to but less stringent than that of chymotrypsin. They hydrolyze a peptide bond which has tyrosine, phenylalanine, or leucine at the carboxyl side of the splitting bond. The optimal pH of alkaline proteases is around pH 10, and their isoelectric point is around pH 9. Their molecular masses are in the range of 15 to 30 kDa. Although alkaline serine proteases are produced by several bacteria such as Arthrobacter, Streptomyces, and Flavobacterium spp. (21), subtilisins produced by Bacillus spp. are the best known. Alkaline proteases are also produced by S. cerevisiae (189) and filamentous fungi such as Conidiobolus spp. (219) and Aspergillus and Neurospora spp. (165).

(ii) Subtilisins. Subtilisins of Bacillus origin represent the second largest family of serine proteases. Two different types of alkaline proteases, subtilisin Carlsberg and subtilisin Novo or bacterial protease Nagase (BPN’), have been identified. Subtilisin Carlsberg produced by Bacillus licheniformis was discovered in 1947 by Lindnerstrom, Lang, and Ottesen at the Carlsberg laboratory. Subtilisin Novo or BPN’ is produced by Bacillus amyloliquefaciens. Subtilisin Carlsberg is widely used in detergents. Its annual production amounts to about 500 tons of pure enzyme protein. Subtilisin BPN’ is less commercially important. Both subtilisins have a molecular mass of 27.5 kDa but differ from each other by 58 amino acids. They have similar properties such as an optimal temperature of 60°C and an optimal pH of 10. Both enzymes exhibit a broad substrate specificity and have an active-site triad made up of Ser221, His64 and Asp32. The Carlsberg enzyme has a broader substrate specificity and does not depend on Ca²⁺ for its stability. The active-site conformation of subtilisins is similar to that of trypsin and chymotrypsin despite the dissimilarity in their overall molecular arrangements. The serine alkaline protease from the fungus Conidiobolus coronatus was shown to possess a distinctly different structure from subtilisin Carlsberg in spite of their functional similarities (218).

Aspartic proteases. Aspartic acid proteases, commonly known as acidic proteases, are the endopeptidases that depend on aspartic acid residues for their catalytic activity. Acidic proteases have been grouped into three families, namely, pepsin (A1), retropepsin (A2), and enzymes from pararetroviruses (A3) (13), and have been placed in clan AA. The members of families A1 and A2 are known to be related to each other, while those of family A3 show some relatedness to A1 and A2. Most aspartic proteases show maximal activity at low pH (pH 3 to 4) and have isoelectric points in the range of pH 3 to 4.5. Their molecular masses are in the range of 30 to 45 kDa. The members of the pepsin family have a bilobal structure with the active-site cleft located between the lobes (259). The active-site aspartic acid residue is situated within the motif Asp-Xaa-Gly, in which Xaa can be Ser or Thr. The aspartic proteases are inhibited by pepstatin (63). They are also sensitive to diazoketone compounds such as diazocetyl-o-toluidine methyl ester (DAN) and 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP) in the presence of copper ions. Microbial acid proteases exhibit specificity against aromatic or bulky amino acid residues on both sides of the peptide bond, which is similar to pepsin, but their action is less stringent than that of pepsin. Micobial aspartic proteases can be broadly divided into two groups, (i) pepsin-like enzymes produced by Aspergillus, Penicillium, Rhizopus, and Neurospora and (ii) rennin-like enzymes produced by Endothia and Mucor spp.

Cysteine/thiol proteases. Cysteine proteases occur in both prokaryotes and eukaryotes. About 20 families of cysteine proteases have been recognized. The activity of all cysteine proteases depends on a catalytic dyad consisting of cysteine and histidine. The order of Cys and His (Cys-His or His-Cys) residues differs among the families (12). Generally, cysteine proteases are active only in the presence of reducing agents such as HCN or cysteine. Based on their side chain specificity, they are broadly divided into four groups: (i) papain-like, (ii) trypsin-like with preference for cleavage at the arginine residue, (iii) specific to glutamic acid, and (iv) others. Papain is the best-known cysteine protease. Cysteine proteases have neutral pH optima, although a few of them, e.g., lysosomal proteases, are maximally active at acidic pH. They are susceptible to sulphydryl agents such as PCMB but are unaffected by DFP and metal-chelating agents. Clostripain, produced by the anaerobic bacterium Clostridium histolyticum, exhibits a stringent specificity for arginyl residues at the carboxyl side of the splitting bond and differs from papain in its obligate requirement for calcium. Streptopain, the cysteine protease produced by Streptococcus spp., shows a broader specificity, including oxidized insulin B chain and other synthetic substrates. Clostripain has an isoelectric point of pH 4.9 and a molecular mass of 50 kDa, whereas the isoelectric point and molecular mass of streptopain are pH 8.4 and 32 kDa, respectively.

Metalloproteases. Metalloproteases are the most diverse of the catalytic types of proteases (13). They are characterized by the requirement for a divalent metal ion for their activity. They include enzymes from a variety of origins such as collagenases from higher organisms, hemorrhagic toxins from snake venoms, and thermolysin from bacteria (92, 210, 253, 311, 314). About 30 families of metalloproteases have been recognized, of which 17 contain only endopeptidases, 12 contain only exopeptidases, and 1 (M3) contains both endo- and exopeptidases. Families of metalloproteases have been grouped into different clans based on the nature of the amino acid that completes the metal-binding site; e.g., clan MA has the sequence HEXXH and clan MB corresponds to the motif HEXXH-H. In one of the groups, the metal atom binds at a motif other than the usual motif.

Based on the specificity of their action, metalloproteases can be divided into four groups, (i) neutral, (ii) alkaline, (iii) Mxobacter I, and (iv) Mxobacter II. The neutral proteases show specificity for hydrophobic amino acids, while the alkaline proteases possess a very broad specificity. Mxobacter protease I is specific for small amino acid residues on either side of the cleavage bond, whereas protease II is specific for lysine residue on the amino side of the peptide bond. All of them are inhibited by chelating agents such as EDTA but not by sulphydryl agents or DFP.

Thermolysin, a neutral protease, is the most thoroughly characterized member of clan MA. Histidine residues from the HEXXH motif serve as Zn ligands, and Glu has a catalytic function (311). Thermolysin produced by B. steatorrhophilus is a single peptide without disulfide bridges and has a molecular mass of 34 kDa. It contains an essential Zn atom embedded in a cleft formed between two folded lobes of the protein and four Ca atoms which impart thermostability to the protein. Thermolysin is a very stable protease, with a half-life of 1 h at 80°C.

Collagenase, another important metalloprotease, was first discovered in the broth of the anaerobic bacterium Clostridium histolyticum as a component of toxic products. Later, it was found to be produced by the aerobic bacterium Achromobacter tophagus and other microorganisms including fungi. The action of collagenase is very specific; i.e., it acts only on collagen and
The catalytic site of proteases is flanked on one or both sides by specificity subsites, each able to accommodate the side chain of the primary substrate-binding site in trypsin-like proteases. The primary specificity is affected only by the Pl residue; the residues at other positions affect the rate of cleavage. The subsite interactions are localized to specific amino acids around the Pl residue to a unique set of sequences on the enzyme. Some of the serine peptidases from *Achromobacter* spp. are lysine-specific enzymes (179), whereas those from *Clostridium* spp. are arginine specific (clostripain) (71) and those from *Flavobacterium* spp. are post proline-specific (329). Endopeptidases that are specific to glutamic acid and aspartic acid residues have also been found in *B. licheniformis* and *S. aureus* (52).

The recent studies based on the three-dimensional structures of proteases and comparisons of amino acid sequences near the primary substrate-binding site in trypsin-like proteases of viral and bacterial origin suggest a putative general substrate binding scheme for proteases with specificity towards glutamic acid involving a histidine residue and a hydroxyl function. However, a few other serine proteases such as peptidase A from *E. coli* and the repressor LexA show distinctly different mechanism of action without the classic Ser-His-Asp triad (12). Some of the glycine residues are conserved in the vicinity of the catalytic serine residue, but their exact positions are variable (25).

The chymotrypsin-like enzymes are confined almost entirely to animals, the exceptions being trypsin-like enzymes from actinomycetes and *Saccharopolyspora* spp. and from the fungus *Fusarium oxysporum*.

A few of the serine proteases belonging to the subtilisin family show a catalytic triad composed of the same residues as in the chymotrypsin family; however, the residues occur in a different order (Asp-His-Ser). Some members of the subtilisin family from the yeasts *Trichoderma* and *Methanolizium* spp. require thiol for their activity. The thiol dependance is attributable to Cys173 near the active-site histidine (122).

The carboxypeptidases are unusual among the serine-dependent enzymes in that they are maximally active at acidic pH. These enzymes are known to possess a Glu residue preceding the catalytic Ser, which is believed to be responsible for their acidic pH optimum. Although the majority of the serine proteases contain the catalytic triad Ser-His-Asp, a few use the Ser-base catalytic dyad. The Glu-specific proteases display a pronounced preference for Glu-Xaa bonds over Asp-Xaa bonds (8).

**Aspartic Proteases**

Aspartic endopeptidases depend on the aspartic acid residues for their catalytic activity. A general base catalytic mechanism has been proposed for the hydrolysis of proteins by aspartic proteases such as penicillopepsin (121) and endothiapepsin (215). Crystallographic studies have shown that the enzymes of the pepsin family are bilobed molecules with the active-site cleft located between the lobes and each lobe contributing one of the pair of aspartic acid residues that is essential for the catalytic activity (20, 259). The lobes are homologous to one another, having arisen by gene duplication. The retropepsin molecule has only one lobe, which carries only one aspartic residue, and the activity requires the formation of a noncovalent homodimer (184). In most of the enzymes from the pepsin family, the catalytic Asp residues are contained in
an Asp-Thr-Gly-Xaa motif in both the N- and C-terminal lobes of the enzyme, where Xaa is Ser or Thr, whose side chains can hydrogen bond to Asp. However, Xaa is Ala in most of the retropepsins. A marked conservation of cysteine residue is also evident in aspartic proteases. The pepsins and the majority of other members of the family show specificity for the cleavage of bonds in peptides of at least six residues with hydrophobic amino acids in both the Pl and Pl$^9$ positions (132).

The specificity of the catalysis has been explained on the basis of available crystal structures (166). The structural and kinetic studies also have suggested that the mechanism involves general acid-base catalysis with lytic water molecule that directly participates in the reaction (Fig. 3A). This is supported by the crystal structures of various aspartic protease-inhibitor complexes and by the thiol inhibitors mimicking a tetrahedral intermediate formed after the attack by the lytic water molecule (120).

**Metalloproteases**

The mechanism of action of metalloproteases is slightly different from that of the above-described proteases. These enzymes depend on the presence of bound divalent cations and can be inactivated by dialysis or by the addition of chelating agents. For thermolysin, based on the X-ray studies of the complex with a hydroxamic acid inhibitor, it has been proposed that Glu143 assists the nucleophilic attack of a water molecule on the carbonyl carbon of the scissile peptide bond, which is polarized by the Zn$^{2+}$ ion (98). Most of the metalloproteases are enzymes containing the His-Glu-Xaa-Xaa-His (HEXXH) motif, which has been shown by X-ray crystallography to form a part of the site for binding of the metal, usually zinc.

**Cysteine Proteases**

Cysteine proteases catalyze the hydrolysis of carboxylic acid derivatives through a double-displacement pathway involving general acid-base formation and hydrolysis of an acyl-thiol intermediate. The mechanism of action of cysteine proteases is thus very similar to that of serine proteases.

A striking similarity is also observed in the reaction mechanism for several peptidases of different evolutionary origins. The plant peptidase papain can be considered the archetype of cysteine peptidases and constitutes a good model for this family of enzymes. They catalyze the hydrolysis of peptide, amide ester, thiol ester, and thiono ester bonds (226). The initial step in the catalytic process (Fig. 3B) involves the noncovalent...
binding of the free enzyme (structure a) and the substrate to form the complex (structure b). This is followed by the acylation of the enzyme (structure c), with the formation and release of the first product, the amine R'\textbf{−}NH₂. In the next deacylation step, the acyl-enzyme reacts with a water molecule to release the second product, with the regeneration of free enzyme.

The enzyme papain consists of a single protein chain folded to form two domains containing a cleft for the substrate to bind. The crystal structure of papain confirmed the Cys25-His159 pairing (11). The presence of a conserved asparagine residue (Asn175) in the proximity of catalytic histidine (His159) creating a Cys-His-Asn triad in cysteine peptidases is considered analogous to the Ser-His-Asp arrangement found in serine proteases.

Studies of the mechanism of action of proteases have revealed that they exhibit different types of mechanism based on their active-site configuration. The serine proteases contain a Ser-His-Asp catalytic triad, and the hydrolysis of the peptide bond involves an acylation step followed by a deacylation step. Aspartic proteases are characterized by an Asp-Thr-Gly motif in their active site and by an acid-base catalysis as their mechanisms of action. The activity of metalloproteases depends on the binding of a divalent metal ion to a His-Glu-Xaa-Xaa-His motif. Cysteine proteases adopt a hydrolysis mechanism involving a general acid-base formation followed by hydrolysis of an acyl-thiol intermediate.

PHYSIOLOGICAL FUNCTIONS OF PROTEASES

Proteases execute a large variety of complex physiological functions. Their importance in conducting the essential metabolic and regulatory functions is evident from their occurrence in all forms of living organisms. Proteases play a critical role in many physiological and pathological processes such as protein catabolism, blood coagulation, cell growth and migration, tissue arrangement, morphogenesis in development, inflammation, tumor growth and metastasis, activation of zymogens, release of hormones and pharmacologically active peptides from precursor proteins, and transport of secretory proteins across membranes. In general, extracellular proteases catalyze the hydrolysis of large proteins to smaller molecules for subsequent absorption by the cell whereas intracellular proteases play a critical role in the regulation of metabolism. In contrast to the multitude of the roles contemplated for proteases, our knowledge about the mechanisms by which they perform these functions is very limited. Extensive research is being carried out to unravel the metabolic pathways in which proteases play an integral role; this research will continue to contribute significantly to our present state of information. Some of the major activities in which the proteases participate are described below.

Protein Turnover

All living cells maintain a particular rate of protein turnover by continuous, albeit balanced, degradation and synthesis of proteins. Catabolism of proteins provides a ready pool of amino acids as precursors of the synthesis of proteins. Intracellular proteases are known to participate in executing the proper protein turnover for the cell. In E. coli, ATP-dependent protease La, the lan gene product, is responsible for hydrolysis of abnormal proteins (38). The turnover of intracellular proteins in eukaryotes is also affected by a pathway involving ATP-dependent proteases (91). Evidence for the participation of proteolytic activity in controlling the protein turnover was demonstrated by the lack of proper turnover in protease-deficient mutants.

Sporulation and Conidial Discharge

The formation of spores in bacteria (142), ascospores in yeasts (58), fruiting bodies in slime molds (205) and conidial discharge in fungi (221) all involve intensive protein turnover. The requirement of a protease for sporulation has been demonstrated by the use of protease inhibitors (41). Ascospore formation in yeast diploids was shown to be related to the increase in protease A activity (58). Extensive protein degradation accompanied the formation of a fruiting body and its differentiation to a stalk in slime molds. The alkaline serine protease of Conidiobolus coronatus was shown to be involved in forcible conidial discharge by isolation of a mutant with less conidial formation (221). Formation of the less active protease by autoproteolysis represents a novel means of physiological regulation of protease activity in C. coronatus (219).

Germination

The dormant spores lack the amino acids required for germination. Degradation of proteins in dormant spores by serine endoproteinas makes amino acids and nitrogen available for the biosynthesis of new proteins and nucleotides. These proteases are specific only for storage proteins and do not affect other spore proteins. Their activity is rapidly lost on germination of the spores (227). Microconidal germination and hyphal fusion also involve the participation of a specific alkaline serine protease (159). Extracellular acid proteases are believed to be involved in the breakage of cell wall polypeptide linkages during germination of Dictyostelium discoideum spores (118) and Polysphondylium pallidum microcysts (206).

Enzyme Modification

Activation of the zymogenic precursor forms of enzymes and proteins by specific proteases represents an important step in the physiological regulation of many rate-controlling processes such as generation of protein hormones, assembly of fibrils and viruses, blood coagulation, and fertilization of ova by sperm. Activation of zymogenic forms of chitin synthase by limited proteolysis has been observed in Candida albicans, Mucor rouxii, and Aspergillus nidulans. Kex-2 protease (kexin; EC 3.4.21.61), originally discovered in yeast, has emerged as a prototype of a family of eukaryotic precursor processing enzymes. It catalyzes the hydrolysis of prohormones and of integral membrane proteins of the secretory pathway by specific cleavage at the carboxyl side of pairs of basic residues such as Lys-Arg or Arg-Arg (12). Furin (EC 3.4.21.5) is a mammalian homolog of the Kex-2 protease that was discovered serendipitously and has been shown to catalyze the hydrolysis of a wide variety of precursor proteins at Arg-X-Lys or Arg-Arg sites within the constitutive secretory pathway (266). Pepsin, trypsin, and chymotrypsin occur as their inactive zymogenic forms, which are activated by the action of proteases.

Proteolytic inactivation of enzymes, leading to irreversible loss of in vivo catalytic activity, is also a physiologically significant event. Several enzymes are known to be inactivated in response to physiological or developmental changes or after a metabolic shift. Proteinases A and B from yeast inactivate several enzymes in a two-step process involving covalent modification of proteins as a marking mechanism for proteolysis.

Proteolytic modification of enzymes is known to result in a protein with altered physiological function; e.g., leucyl-t-RNA synthetase from E. coli is converted into an enzyme which
catalyzes leucine-dependent pyrophosphate exchange by removal of a small peptide from the native enzyme.

**Nutrition**

Proteases assist the hydrolysis of large polypeptides into smaller peptides and amino acids, thus facilitating their absorption by the cell. The extracellular enzymes play a major role in nutrition due to their depolymerizing activity. The microbial enzymes and the mammalian extracellular enzymes such as those secreted by pancreas are primarily involved in keeping the cells alive by providing them with the necessary amino acid pool as nutrition.

**Regulation of Gene Expression**

Modulation of gene expression mediated by protease has been demonstrated (241). Proteolysis of a repressor by an ATP-requiring protease resulted in a derepression of the gene. A change in the transcriptional specificity of the B subunit of *Bacillus thuringiensis* RNA polymerase was correlated with its proteolytic modification (154). Modification of ribosomal proteins by proteases has been suggested to be responsible for the regulation of translation (128).

Besides the general functions described so far, the proteases also mediate the degradation of a variety of regulatory proteins that control the heat shock response, the SOS response to DNA damage, the life cycle of bacteriophage (75), and programmed bacterial cell death (303). Recently, a new physiological function has been attributed to the ATP-dependent proteases conserved between bacteria and eukaryotes. It is believed that they act as chaperones and mediate not only proteinolysis but also the insertion of proteins into membranes and the disassembly or oligomerization of protein complexes (275). In addition to the multitude of activities that are already assigned to proteases, many more new functions are likely to emerge in the near future.

**APPLICATIONS OF PROTEASES**

Proteases have a large variety of applications, mainly in the detergent and food industries. In view of the recent trend of developing environmentally friendly technologies, proteases are envisaged to have extensive applications in leather treatment and in several bioremediation processes. The worldwide requirement for enzymes for individual applications varies considerably. Proteases are used extensively in the pharmaceutical industry for preparation of medicines such as ointments for debriding of wounds, etc. Proteases that are used in the food and detergent industries are prepared in bulk quantities and used as crude preparations, whereas those that are used in medicine are produced in small amounts but require extensive purification before they can be used.

**Detergents**

Proteases are one of the standard ingredients of all kinds of detergents ranging from those used for household laundering to reagents used for cleaning contact lenses or dentures. The use of proteases in laundry detergents accounts for approximately 25% of the total worldwide sales of enzymes. The preparation of the first enzymatic detergent, “Burnus,” dates back to 1913; it consisted of sodium carbonate and a crude pancreatic extract. The first detergent containing the bacterial enzyme was introduced in 1956 under the trade name BIO-40. In 1960, Novo Industry A/S introduced alcalase, produced by *Bacillus licheniformis*; its commercial name was BIOTEX. This was followed by Maxataze, a detergent made by Gist-Brocades. The biggest market for detergents is in the laundry industry, amounting to a worldwide production of 13 billion tons per year. The ideal detergent protease should possess broad substrate specificity to facilitate the removal of a large variety of stains due to food, blood, and other body secretions. Activity and stability at high pH and temperature and compatibility with other chelating and oxidizing agents added to the detergents are among the major prerequisites for the use of proteases in detergents. The key parameter for the best performance of a protease in a detergent is its pI. It is known that a protease is most suitable for this application if its pI coincides with the pH of the detergent solution. Esperase and Savinase T (Novo Industry), produced by alkalophilic *Bacillus* spp., are two commercial preparations with very high isoelectric points (pI 11); hence, they can withstand higher pH ranges. Due to the present energy crisis and the awareness for energy conservation, it is desirable to use proteases that are active at lower temperatures. A combination of lipase, amylase, and cellulase is expected to enhance the performance of protease in laundry detergents.

All detergent proteases currently used in the market are serine proteases produced by *Bacillus* strains. Fungal alkaline proteases are advantageous due to the ease of downstream processing to prepare a microbe-free enzyme. An alkaline protease from *Comidiotholbus coronatus* was found to be compatible with commercial detergents used in India (219) and retained 43% of its activity at 50°C for 50 min in the presence of Ca²⁺ (25 mM) and glycine (1 M) (16).

**Leather Industry**

Leather processing involves several steps such as soaking, dehairing, bating, and tanning. The major building blocks of skin and hair are proteinaceous. The conventional methods of leather processing involve hazardous chemicals such as sodium sulfide, which create problems of pollution and effluent disposal. The use of enzymes as alternatives to chemicals has proved successful in improving leather quality and in reducing environmental pollution. Proteases are used for selective hydrolysis of noncollagenous constituents of the skin and for removal of nonfibrillar proteins such as albumins and globulins. The purpose of soaking is to swell the hide. Traditionally, this step was performed with alkali. Currently, microbial alkaline proteases are used to ensure faster absorption of water and to reduce the time required for soaking. The use of nonionic and, to some extent, anionic surfactants is compatible with the use of enzymes. The conventional method of dehairing and dewooling consists of development of an extremely alkaline condition followed by treatment with sulfide to solubilize the proteins of the hair root. At present, alkaline proteases with hydrated lime and sodium chloride are used for dehairing, resulting in a significant reduction in the amount of wastewater generated. Earlier methods of bating were based on the use of animal feces as the source of proteases; these methods were unpleasant and unreliable and were replaced by methods involving pancreatic trypsin. Currently, trypsin is used in combination with other *Bacillus* and *Aspergillus* proteases for bating. The selection of the enzyme depends on its specificity for matrix proteins such as elastin and keratin, and the amount of enzyme needed depends on the type of leather (soft or hard) to be produced. Increased usage of enzymes for dehairing and bating not only prevents pollution problems but also is effective in saving energy. Novo Nordisk manufactures three different proteases, Aquaderm, NUE, and Pyrase, for use in soaking, dehairing, and bating, respectively.
**Food Industry**

The use of proteases in the food industry dates back to antiquity. They have been routinely used for various purposes such as cheesemaking, baking, preparation of soya hydrolysates, and meat tenderization.

**Dairy industry.** The major application of proteases in the dairy industry is in the manufacture of cheese. The milk-coagulating enzymes fall into three main categories, (i) animal rennets, (ii) microbial milk coagulants, and (iii) genetically engineered chymosin. Both animal and microbial milk-coagulating proteases belong to a class of acid aspartate proteases and have molecular weights between 30,000 to 40,000. Rennet extracted from the fourth stomach of unweaned calves contains the highest ratio of chymosin (EC 3.4.23.4) to pepsin activity. A world shortage of calf rennet due to the increased demand for cheese production has intensified the search for alternative microbial milk coagulants. The microbial enzymes exhibited two major drawbacks, i.e., (i) the presence of high levels of nonspecific and heat-stable proteases, which led to the development of bitterness in cheese after storage; and (ii) a poor yield. Extensive research in this area has resulted in the production of enzymes that are completely inactivated at normal pasteurization temperatures and contain very low levels of nonspecific proteases. In cheesemaking, the primary function of proteases is to hydrolyze the specific peptide bond (the Phe105-Met106 bond) to generate para-c-casein and macropetides. Chymosin is preferred due to its high specificity for casein, which is responsible for its excellent performance in cheesemaking. The proteases produced by GRAS (genetically regarded as safe) cleared microbes such as *Mucor miehei*, *Bacillus subtilis*, and *Endothia parasitica* are gradually replacing chymosin in cheesemaking. In 1988, chymosin produced through recombinant DNA technology was first introduced to cheesemakers for evaluation. Genencor International increased the production of chymosin in *Aspergillus niger* var. *avamori* to commercial levels. At present, their three recombinant chymosin products are available and are awaiting legislative approval for their use in cheesemaking (72).

Whey is a by-product of cheese manufacture. It contains lactose, proteins, minerals, and lactic acid. The insoluble heat-denatured whey protein is solubilized by treatment with immobilized trypsin.

**Baking industry.** Wheat flour is a major component of baking processes. It contains an insoluble protein called gluten, which determines the properties of the bakery doughs. Endo- and exopeptidases from *Aspergillus oryzae* have been used to modify wheat gluten by limited proteolysis. Enzymatic treatment of the dough facilitates its handling and machining and permits the production of a wider range of products. The addition of proteases reduces the mixing time and results in increased loaf volumes. Bacterial proteases are used to improve the extensibility and strength of the dough.

**Manufacture of soy products.** Soybeans serve as a rich source of food, due to their high content of good-quality protein. Proteases have been used from ancient times to prepare soy sauce and other soy products. The alkaline and neutral proteases play an important role in the processing of soy sauce. Proteolytic modification of soy proteins helps to improve their functional properties. Treatment of soy proteins with alcalase at pH 8 results in soluble hydrolysates with high solubility, good protein yield, and low bitterness. The hydrolysate is used in protein-fortified soft drinks and in the formulation of dietetic feeds.

**Debittering of protein hydrolysates.** Protein hydrolysates have several applications, e.g., as constituents of dietetic and health products, in infant formulae and clinical nutrition supplements, and as flavoring agents. The bitter taste of protein hydrolysates is a major barrier to their use in food and health care products. The intensity of the bitterness is proportional to the number of hydrophobic amino acids in the hydrolysate. The presence of a proline residue in the center of the peptide also contributes to the bitterness. The peptidases that can cleave hydrophobic amino acids and proline are valuable in debittering protein hydrolysates. Aminopeptidases from lactic acid bacteria are available under the trade name Debitrase. Carboxypeptidase A has a high specificity for hydrophobic amino acids and hence has a great potential for debittering. A careful combination of an endoprotease for the primary hydrolysis and an aminopeptidase for the secondary hydrolysis is required for the production of a functional hydrolysate with reduced bitterness.

**Synthesis of aspartame.** The use of aspartame as a noncaloric artificial sweetener has been approved by the Food and Drug Administration. Aspartame is a dipeptide composed of L-aspartic acid and the methyl ester of L-phenylalanine. The L-configuration of the two amino acids is responsible for the sweet taste of aspartame. Maintenance of the stereospecificity is crucial, but it adds to the cost of production by chemical methods. Enzymatic synthesis of aspartame is therefore preferred. Although proteases are generally regarded as hydrolytic enzymes, they catalyze the reverse reaction under certain kinetically controlled conditions. An immobilized preparation of thermolysin from *Bacillus thermoproteolyticus* is used for the enzymatic synthesis of aspartame. Toya Soda (Japan) and DSM (The Netherlands) are the major industrial producers of aspartame.

**Pharmaceutical Industry**

The wide diversity and specificity of proteases are used to great advantage in developing effective therapeutic agents. Oral administration of proteases from *Aspergillus oryzae* (Luzyme and Nortase) has been used as a digestive aid to correct certain lytic enzyme deficiency syndromes. Clostridial collagenase or subtilisin is used in combination with broad-spectrum antibiotics in the treatment of burns and wounds. An asparagine isolated from *E. coli* is used to eliminate asparagine from the bloodstream in the various forms of lymphocytic leukemia. Alkaline protease from *Conidiobolus coronatus* was found to be able to replace trypsin in animal cell cultures (36).

**Other Applications**

Besides their industrial and medicinal applications, proteases play an important role in basic research. Their selective peptide bond cleavage is used in the elucidation of structure-function relationship, in the synthesis of peptides, and in the sequencing of proteins.

In essence, the wide specificity of the hydrolytic action of proteases finds an extensive application in the food, detergent, leather, and pharmaceutical industries, as well as in the structural elucidation of proteins, whereas their synthetic capacities are used for the synthesis of proteins.

**GENETIC ENGINEERING OF MICROBIAL PROTEASES**

Gene cloning is a rapidly progressing technology that has been instrumental in improving our understanding of the structure-function relationship of genetic systems. It provides an excellent method for the manipulation and control of genes. More than 50% of the industrially important enzymes are now produced from genetically engineered microorganisms (96).
TABLE 4. Cloning, sequencing, and/or expression of protease genes or cDNAs from microbial sources

<table>
<thead>
<tr>
<th>Source of protease gene</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
</tr>
<tr>
<td>Bacilli</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> 168</td>
<td></td>
</tr>
<tr>
<td>apr</td>
<td>270</td>
</tr>
<tr>
<td>npr</td>
<td>90, 295, 297, 323</td>
</tr>
<tr>
<td>epr</td>
<td>27, 263</td>
</tr>
<tr>
<td>bpr</td>
<td>265</td>
</tr>
<tr>
<td>mpr</td>
<td>264</td>
</tr>
<tr>
<td>1sp-1</td>
<td>138</td>
</tr>
<tr>
<td><em>B. subtilis</em> (Natto) 16</td>
<td>319</td>
</tr>
<tr>
<td><em>B. subtilis</em> N 515-N (nprX)</td>
<td>157</td>
</tr>
<tr>
<td>Alkalophilic <em>Bacillus</em> strain</td>
<td>129</td>
</tr>
<tr>
<td><em>B. alkalophilus</em> PB92</td>
<td>300</td>
</tr>
<tr>
<td><em>Bacillus</em> sp. strain Y</td>
<td>293</td>
</tr>
<tr>
<td>Alkalophilic <em>Bacillus</em> sp. NKS-21</td>
<td>318</td>
</tr>
<tr>
<td>Alkalophilic <em>Bacillus</em> sp. LG-12</td>
<td>251</td>
</tr>
<tr>
<td><em>Bacillus</em> sp. EA (Npr)</td>
<td>249</td>
</tr>
<tr>
<td>Lactococci</td>
<td></td>
</tr>
<tr>
<td><em>Streptococcus cremoris</em> Wg2</td>
<td>139–141</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> subsp. cremoris H2</td>
<td>317</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em> N PDO 763</td>
<td>137</td>
</tr>
<tr>
<td><em>L. lactis</em> subsp. cremoris SK11</td>
<td>50</td>
</tr>
<tr>
<td><em>L. lactis</em> subsp. lactis VC 317</td>
<td>153</td>
</tr>
<tr>
<td><em>L. lactis</em> subsp. cremoris Wg2</td>
<td>58</td>
</tr>
<tr>
<td><em>Lactobacillus delbrueckii</em> subsp. bulgaricus</td>
<td>69</td>
</tr>
<tr>
<td>Streptomycetes</td>
<td></td>
</tr>
<tr>
<td><em>S. griseus</em></td>
<td>89</td>
</tr>
<tr>
<td><em>S. griseus</em> ATCC 10137</td>
<td>107</td>
</tr>
<tr>
<td><em>S. canaori</em> YMI5</td>
<td>32</td>
</tr>
<tr>
<td><em>S. fradiae</em> ATCC 14544</td>
<td>135</td>
</tr>
<tr>
<td><em>S. lividans</em> 66</td>
<td>17, 18, 163</td>
</tr>
<tr>
<td>Serratia</td>
<td></td>
</tr>
<tr>
<td><em>Serratia</em> sp. strain E-15</td>
<td>198</td>
</tr>
<tr>
<td><em>S. marcescens</em> SM6</td>
<td>24</td>
</tr>
<tr>
<td><em>S. marcescens</em></td>
<td>187</td>
</tr>
<tr>
<td><em>S. marcescens</em> ATCC 27117</td>
<td>134</td>
</tr>
<tr>
<td>Pseudomonas</td>
<td></td>
</tr>
<tr>
<td><em>P. aeruginosa</em> IFO 3455</td>
<td>7, 254</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PAO1</td>
<td>83</td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td>82</td>
</tr>
<tr>
<td><em>P. halvojense</em></td>
<td>68</td>
</tr>
<tr>
<td>Aeromonas</td>
<td></td>
</tr>
<tr>
<td><em>A. hydrophila</em> SO 212</td>
<td>238, 239</td>
</tr>
<tr>
<td><em>A. hydrophila</em> D13</td>
<td>238</td>
</tr>
<tr>
<td>Vibrio</td>
<td></td>
</tr>
<tr>
<td><em>V. anguillarum</em> NB10</td>
<td>185</td>
</tr>
<tr>
<td><em>V. parahaemolyticus</em></td>
<td>155</td>
</tr>
<tr>
<td><em>V. vulnificus</em></td>
<td>34</td>
</tr>
<tr>
<td><em>V. proteolyticus</em></td>
<td>43</td>
</tr>
<tr>
<td><em>V. angionolyticus</em></td>
<td>45</td>
</tr>
<tr>
<td><em>V. cholerae</em></td>
<td>86</td>
</tr>
<tr>
<td>E. coli</td>
<td></td>
</tr>
<tr>
<td>Membrane proteases</td>
<td></td>
</tr>
<tr>
<td>lbpA, lep</td>
<td>42, 333</td>
</tr>
<tr>
<td>sppA</td>
<td>108, 276</td>
</tr>
<tr>
<td>ompT</td>
<td>80</td>
</tr>
<tr>
<td>ATP-dependent proteases</td>
<td></td>
</tr>
<tr>
<td>La/Lon</td>
<td>3, 35, 334</td>
</tr>
<tr>
<td>Clp/Ti</td>
<td>181</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
</tr>
<tr>
<td>Lysobacter enzymogenes 495</td>
<td>260</td>
</tr>
</tbody>
</table>

**Fungi**

Filamentous fungi

Acidic proteases

*Mucor* *pasiulus* rennin (MPR) ........................................... 94, 296
*Mucor miehei* aspartyl protease (MMAP) .................. 51, 79
*R. niveus* aspartic protease (RNAP) .................. 100, 101
*A. awamori* aspergilopepsin A ..................................... 15
*A. oryzae* aspergilopepsin A ..................................... 74
*A. fumigatus* aspergilopepsin F .................................. 156, 237
*A. oryzae* M-9 .................................................. 284
*A. niger* ATCC 14332 ........................................ 257
*A. niger* var. *macropsorus* Proctase A ............................ 114, 125, 283
Proctase B .............................................................. 113, 175

Alkaline proteases (Alp)

*Apergillus*  
*A. oryzae* ATCC 20386 ........................................... 195, 207, 286, 288
*A. oryzae* Thailand industrial strain .................................... 33
*A. soya* ...................................................................... 211
*A. fumigatus* .................................................................. 123, 237
*A. flavus* ..................................................................... 233
*A. nidulans* .................................................................. 131

*Acremonium*  
*A. chrysogenum* ATCC 11550 ............................................. 115
*Aspergillus*  
*A. oryzae* ATCC 14332 ........................................ 257
*Alcaligenes*  
*A. niger* var. *macropsorus* Proctase A ............................ 114, 125, 283
Proctase B .............................................................. 113, 175

Serine proteases

*Trichoderma album* Limber Protease K ............................. 81
Protease T .............................................................. 247

Metalloproteases

*A. fumigatus* MEP .............................................................. 124, 262
*A. flavus* MEP-20 .............................................................. 234
*A. fumigatus* MEP-20 .............................................................. 234

Yeasts

Acidic proteases

*S. fibuligera* (PEP1) ...................................................... 95, 320
*S. cerevisiae* (PEP4) ...................................................... 315
*S. cerevisiae* (BAR1) ...................................................... 177
*S. cerevisiae* (YAP3) ...................................................... 53
*C. albicans* (SAP) ...................................................... 106, 170, 196
*C. tropicalis* (ACP) ...................................................... 294
*Y. lipolytica* 146 (AXP) ...................................................... 351
Wild-type yeast ...................................................... 332

Alkaline proteases

*Y. lipolytica* (AEP) XRP2 ...................................................... 44, 201

Serine proteases

*Kluveromyces*  
*L. enzymogenes* .............................................................. 57
*A. lyticus* .............................................................. 169
*Erwinia* sp. .............................................................. 2, 307
*Rhodococcus glutinosa* APR 3-2 ............................................. 116
*Bacteroides nodosus* ...................................................... 194
*Xanthomonas* campestris pv. campestris .................................. 168
*Treponema denticola* ATCC 33520 ...................................... 176
*Staphylococcus aureus* V8 .............................................. 29
*Thermus aquaticus* YT-1 .................................................... 148
*Thermomonospora fusca* YX ................................................ 152
*Alteromonas* sp. strain O-7 .................................................. 298, 299

IgA family of proteases

*N. gonorrhoeae* .............................................................. 62, 224, 232
*N. meningitidis* .............................................................. 169
*H. influenzae* .............................................................. 228
*Streptococcus sanguis* ATCC 10556 ...................................... 70
*S. pneumoniae* .............................................................. 229, 308

Continued on following page
Several reports have been published in the past decade (Table 4) on the isolation and manipulation of microbial protease genes with the aim of (i) enzyme overproduction by the gene dosage effect, (ii) studying the primary structure of the protein and its role in the pathogenicity of the secretory microorganism, and (iii) protein engineering to locate the active-site residues and/or to alter the enzyme properties to suit its commercial applications. Protease genes from bacteria, fungi, and viruses have been cloned and sequenced (Table 4).

Bacteria

The objective of cloning bacterial protease genes has been mainly the overproduction of enzymes for various commercial applications in the food, detergent and pharmaceutical industries. The virulence of several bacteria is related to the secretion of several extracellular proteases. Gene cloning in these microbes was studied to understand the basis of their pathogenicity and to develop therapeutics against them. Proteases play an important role in cell physiology, and protease gene cloning, especially in E. coli, has been attempted to study the regulatory aspects of proteases.

Bacilli. (i) B. subtilis as a host for cloning of protease genes from Bacillus spp. The ability of B. subtilis to secrete various proteins into the culture medium and its lack of pathogenicity make it a potential host for the production of foreign polypeptides by recombinant DNA technology. Several Bacillus spp. secrete two major types of protease, a subtilisin or alkaline protease and a metalloprotease or neutral protease, which are of industrial importance. Studies of these extracellular proteases are significant not only from the point of view of overproduction but also for understanding their mechanism of secretion. Table 5 describes the cloning of genes for several neutral (npr) and alkaline (apr) proteases from various bacilli into B. subtilis.

(ii) B. subtilis. B. subtilis 168 secretes at least six extracellular proteases into the culture medium at the end of the exponential phase. The structural genes encoding the alkaline protease (apr) or subtilisin (270), neutral protease A and B (aprA and aprB) (90, 297, 323), minor extracellular protease (epr) (27, 263), bacillopeptidase E (bpr) (265), and metalloprotease (mpr) (264) have been cloned and characterized. These proteases are synthesized in the form of a “pro” enzyme. To increase the expression of subtilisin and neutral proteases, Henner et al. replaced the natural promoters of apr and npr genes with the amylase promoter from B. amyloquefaciens and the neutral protease promoter from B. subtilis, respectively (90). To understand the regulation of npr A gene expression, Toma et al. cloned the genes from B. subtilis 168 (normal producer) and Basc 1A341 (overproducer) (295). The two genes were found to be highly homologous except for a stretch of 66 bp close to the promoter region, which is absent in the Basc 1A341 gene. The epr gene shows partial homology to the apr gene and to the major intracellular serine protease (Isp-1) gene of B. subtilis (138). The apr gene was mapped at a locus different from the apr and npr loci on the B. subtilis chromosome and was shown not to be required for growth or sporulation, similar to apr or npr genes. Deletion of 240 amino acids (aa) from the C-terminal region of the epr gene product did not abolish the enzyme activity (27, 263). The deduced amino acid sequence of the mature bpr gene product is similar to those of other serine proteases of B. subtilis, i.e., subtilisin, Isp-1, and Epr. B. subtilis strains containing mutations in five extracellular protease genes (apr, npr, epr, mpr, and bpr) have been constructed (264) with the aim of expressing heterologous gene products in B. subtilis. The total amino acid sequence of B. subtilis Isp-1 deduced from the nucleotide sequence showed considerable homology (45%) to subtilisin. Highly conserved sequences are present around the essential amino acids, Ser, His, and Asp, indicating that the genes for both the intra- and extracellular serine proteases have a common ancestor.

In 1995, Yamagata et al. cloned and sequenced a 90-kDa serine protease gene (hspK) from B. subtilis (Natto) 16 (319). The large size of the enzyme may represent an ancient form of bacterial serine protease.

Analysis of DNA sequences of subtilisin BPN’ from B. amyloliquefaciens (304, 313) and subtilisin Carlsberg from B. licheniformis (119) revealed that the two sequences are highly conserved in the coding region for the mature protein and must therefore have a common ancestral precursor. Yoshimoto et al. characterized the gene encoding subtilisin amylosacchariticus from B. subtilis subsp. amylosacchariticus (327, 328). The sequence was highly homologous to that of subtilisin E from B. subtilis 168 (269). The gene was expressed in B. subtilis ISW 1214 by using the vector pHY300PLK, with 20-
(iii) Alkalophilic Bacillus spp. Bacillus proteases with an extremely alkaline pH optimum are generally used in detergent powders and are preferred over the subtilisins (optimal pH, 8.5 to 10.0). The information on these enzymes is helpful in designing new subtilisins. Kaneko et al. cloned and sequenced the \( \text{ale} \) gene, encoding alkaline elastase YaB, a new subtilisin from an alkalophilic \( \text{Bacillus} \) strain (129). The deduced amino acid sequence showed 55% homology to subtilisin BPN\(_{\text{e}}\). Almost all the positively charged residues have been predicted to be present on the surface of the alkaline elastase YaB molecule, facilitating its binding to elastin. The deduced amino acid sequence of the highly alkaline serine protease from another alkalophilic strain, \( \text{B. alcalophilus} \) PB92, showed considerable homology to YaB (300). The cloned gene was further used to increase the production level of the protease by gene amplification through chromosomal integration. Increased enzyme production and gene stabilization was observed when non tandem duplication occurred.

A gene encoding ISP-1 was characterized from alkalophilic \( \text{Bacillus} \) sp. strain NKS-21 (318). The nucleotide sequence was 50% homologous to genes encoding ISP-1 from \( \text{B. subtilis} \), \( \text{B. polymyxa} \), and the alkalophilic \( \text{Bacillus} \) sp. strain 221.

(iv) Other bacilli. A gene encoding the highly thermostable neutral proteinase (Npr) from \( \text{Bacillus} \) sp. strain EA1 was shown to be closely related to an \( \text{npr} \) gene from \( \text{B. caldolyticus} \) YP-T, except for a single-amino-acid change in the gene product (249). The enzyme from \( \text{Bacillus} \) sp. strain EA1 was more thermostable than the enzyme from \( \text{B. caldolyticus} \) YP-T; this can be attributed to the single-amino-acid change.

### Table 5. Cloning of protease genes in \( \text{B. subtilis} \)

<table>
<thead>
<tr>
<th>Source of proteases</th>
<th>Type of protease</th>
<th>Expression (fold)</th>
<th>Characterization of gene</th>
<th>Reference(s)</th>
</tr>
</thead>
</table>
| \( \text{B. amylo NFC} \) | Neutral | 50 | —
| \( \text{B. stearothermophilus} \) | Neutral | 5 | \( \text{nprS} \) sequenced; homologous to \( \text{nprM} \)
| \( \text{B. amylo NFC} \) | Neutral | 15 | Partially sequenced |
| \( \text{B. cereus} \) | Neutral (metalloprotease) | 1 | 99, 178 |
| \( \text{B. stearothermophilus} \) | Thermostable, neutral | 29 | — |
| \( \text{B. caldolyticus} \) | Thermostable, neutral | 258 | — |
| \( \text{B. amylo NFC} \) | Thermostable, metalloprotease | 304 | — |
| \( \text{B. brevis} \) | Metalloprotease | 324 | — |
| \( \text{B. licheniformis} \) | Alkaline and neutral | 325 | — |
| \( \text{B. amylo NFC} \) | Alkaline | 388 | — |
| \( \text{B. licheniformis} \) | Alkaline | 289 | — |
| \( \text{B. polymyxa} \) | Subtilisin | 313 | — |
| \( \text{B. natto} \) | Alkaline | 197 | — |
| \( \text{B. licheniformis} \) | C-terminal glutamic acid specific (BLase) | 127 | — |

*—, no data is available.

**Lactococci.** Lactococci (\( \text{Lactococcus lactis} \) subsp. \( \text{lactis} \) and \( \text{cremoris} \), previously \( \text{Streptococcus lactis} \) and \( \text{Streptococcus cremoris} \), respectively), the dairy starter cultures, have a complex proteolytic system which enables them to grow in milk by degrading casein into small peptides and free amino acids. This leads to the development of the texture and flavor of various dairy products. The importance of the cell envelope-located proteolytic system for dairy product quality has resulted in an increased fundamental research of the involved enzymes and their genes. On the basis of differences in caseinolytic specificity, the lactococcal proteases have been classified into two main groups: the PI-type protease, which degrades predominantly \( \beta \)-casein, and the PIII-type protease which degrades \( \alpha \)-S1-, \( \beta \)-, and \( \kappa \)-casein (305). Most of the genetic studies have focused on the PI-type protease genes. Lactococcal protease genes are located mostly on plasmids, which differ considerably in size and genetic organization in different strains (49). Curing experiments have suggested that plasmid pWV05 of \( \text{S. cremoris} \) Wg2 specifies proteolytic activity. The entire plasmid was subcloned in \( \text{E. coli} \) (140). A 4.3-MDa HindIII fragment of the plasmid, specifying the proteolytic activity, was cloned in \( \text{B. subtilis} \) and in a protease-deficient \( \text{S. lactis} \) strain. In \( \text{S. lactis} \), the recombinant plasmid enabled the cells to grow normally in milk with rapid acid production. The HindIII fragment specifying the proteolytic activity of \( \text{S. cremoris} \) Wg2 was fully sequenced (141). The nucleotide sequence revealed two open reading frames (ORFs), ORF-1, a small ORF containing 295 codons, and ORF-2, a large ORF containing 1,772 codons. The protein specified by ORF-2 contained regions of extensive homology to subtilisins. The amino acids Asp32, His64, and Ser221, involved in the formation of the active site, were well
conserved. Deletion analysis of the protease gene of *S. cremoris* Wg2 showed that deletion of the C-terminal 343 aa did not influence the enzyme specificity of β-casein degradation (139). *L. lactis* subsp. *cremoris* H2 carries plasmid pDI21, containing the gene for the protease-positve phenotype (Prt⁺). The 6.5-kbp *HindIII* DNA fragment of pDI21 encoding the protease was cloned in *E. coli* as well as in *L. lactis* subsp. *lactis* 4125 (317). Protease that specifically degrades β-casein was expressed in both the transformed organisms. *S. lactis* NCDO 763 harbors plasmid pLP763, containing the gene for Prt⁺, which enables it to grow to a higher density in milk. The deduced amino acid sequence (1,902 aa) of the Prt⁺ phenotype was homologous to that of the serine protease from *S. cremoris* Wg2, suggesting that the genes encoding both products must have been derived from a common ancestral gene (137).

The PriI-type protease is found only in *L. lactis* subsp. *cremoris* AM1 and SK11. These strains are related, and they both contain the proteases encoded by the 78-kbp plasmid psk111. The *L. lactis* subsp. *cremoris* SK11 *prtP* gene was cloned and expressed in *E. coli* as well as in other subspecies of *L. lactis* (50). The location and orientation of the *prtP* gene on psk111 was determined by deletion analysis. A region at the C terminal of the *prtP* product, which is involved in cell envelope attachment, was identified. A deletion derivative of *prtP* specifying a C-terminally truncated protease was able to express and fully secrete the protease in the medium and showed the capacity to degrade αS1-, β-, and κ-casein. The N-terminal catalytic domain of the matrix enzyme shows significant sequence homology to the serine proteases of the subtilisin family (subtilases). Comparison with the known sequences of proteolytic activity and an active site, as predicted by comparing the deduced amino acid sequence with that of *B. subtilis* 168, a chromotypin-like serine protease (SAM-P20) (17), and SlpD and SlpE (homologs of the Tp [major tripeptidyl aminopeptidase] mycelium-associated proteases) (18) were cloned from *S. lividans* 66.

*Serratia*. The gram-negative bacteria belonging to the family *Enterobacteriaceae* are known to secrete large amounts of extracellular proteases into the surrounding medium. *Serratia* sp. strain E-15 produces a potent extracellular metalloprotease, which is widely used as an anti-inflammatory agent. The gene encoding the protease from *Serratia* sp. strain E-15 was expressed both in *E. coli* and in *S. marcescens* (198). Nucleotide sequence analysis revealed three zinc ligands (essential for proteolytic activity) and an active site, as predicted by comparing the deduced amino acid sequence with that of *B. thermoproteolyticus* thermolysin and *B. subtilis* neutral protease.

In another study, the extracellular serine protease (SSP) of *S. marcescens* was excreted through the outer membrane of *E. coli*. The nucleotide sequence of the cloned SSP gene, together with the determination of the N and C termini of the excreted enzymes, suggested that this protease is produced as a 112-kDa preproenzyme composed of an N-terminal signal sequence, the mature protease, and a large C-terminal domain (187).

**Pseudomonas.** *Pseudomonas aeruginosa* is an opportunistic pathogen and can cause fatal infections in compromised hosts. This virulence is related to the secretion of several extracellular proteins (167). *P. aeruginosa* secretes two proteases, an alkaline protease and an elastase. The alkaline protease genes (*apr*) from *P. aeruginosa* IFO 3455 and PAO1 were cloned in *E. coli* (7, 83, 254). The DNA fragment (8.8 kbp) coding for the alkaline protease from strain PAO1 was expressed in *E. coli* under the control of a tac promoter. Active enzyme was found to be synthesized and secreted into the medium in the absence of cell lysis.

The LasA protease (elastin degrading) of *P. aeruginosa* is also an important contributor to the pathogenesis of this bacterium. The enzyme shows a high level of staphylocoytic activity. The lasA gene from strain FRD1 was overexpressed in *E. coli* (82). It encodes a precursor, prepro-LasA, of about 45 kDa. N-terminal sequence analysis allowed the identification of a 31-aa signal peptide. pro-LasA (42 kDa) does not undergo autoproteolytic processing and possesses little anti-staphylo-
coccal activity. The digestion of pro-LasA either by trypsin or by culture filtrate of the \textit{P. aeruginosa lasA} deletion mutant yielded the active (20-kDa) staphyloptic protease.

\textit{Aeromonas}. \textit{Aeromonas hydrophila} and the related aero-
monads are opportunistic pathogens of humans and fish. The pathogenicity of the microbe may involve several extracellular enzymes, and it has been suggested that the proteases excrated by \textit{Aeromonas} spp. play an important role in invasiveness and in establishment of the infection. Two distinct types of extra-
cellular proteases, a temperature-stable metalloprotease and a temperature-labile serine protease, are found in various strains of \textit{A. hydrophila} and other aeromonads (160). Structural genes encoding extracellular proteases from two different \textit{A. hy-
drophila} strains, SO2/2 and D13, were cloned in \textit{E. coli} C600-1 by using pBR322 (238). A temperature-stable protease is se-
creted into the periplasm of \textit{E. coli} and exhibits properties identical to those of the protease purified from \textit{A. hydrophila} SO2/2 culture supernatant. A gene for the temperature-labile serine protease was also expressed from \textit{A. hydrophila} SO2/2 into \textit{E. coli} C600-1 and \textit{S. lividans} 1326 (239).

\textit{Vibrio}. To facilitate genetic analyses of the role of proteases in the pathogenesis of various \textit{Vibrio} species, the genes encoding the Zn$^{2+}$-metalloprotease from \textit{V. anguillarum} NB 10 (185), \textit{V. para
ahaemolyticus} (155), and \textit{V. vulnificus} (34) were cloned and sequenced. The conserved Zn$^{2+}$-binding domains were identified by measuring homology to other metallopro-
teases. The nucleotide sequence of the \textit{nprV} gene encoding the extracellular neutral protease, vibriolysin (NprV), of \textit{V. proteo-
lyticus} revealed an ORF encoding 609 aa including a putative signal peptide sequence followed by a long prospeque of 172 aa (43). Comparative analysis of the mature NprV with the sequences of the neutral proteases from bacilli revealed extensive regions of conserved amino acid homology with respect to the active site and zinc- and calcium-binding residues. NprV was overproduced in \textit{B. subtilis} by placing the DNA encoding the pro-NprV and the mature NprV downstream of the \textit{Bacillus} promoter and signal sequences.

In one of the studies, the nucleotide sequence analysis of the structural gene, \textit{hap}, for the extracellular haemagglutinin protease of \textit{V. cholerae} revealed that the enzyme is produced as a large precursor, with the amino-terminal signal sequence following a propeptide (86). The deduced amino acid sequence of the mature enzyme showed 61.5% identity to the \textit{P. aeruginosa} elastase.

\textit{E. coli}. (i) Membrane proteases. In a bacterium, a protein that is to be exported across the cytoplasmic membrane is synthesized as a large precursor with a signal peptide at its amino terminus (19). The processing of this precursor involves two sequential events: (i) removal of the signal peptide from the precursor through an endo-type cleavage and (ii) digestion of the cleaved signal peptide. The membrane proteases involved are (i) signal peptidases (lipoprotein signal peptidase [Lsp] and leader peptidase [Lep]) and (ii) signal peptide pepti-
dase (protease IV). The genes \textit{lspA} (333), \textit{lep} (42), and \textit{sppA} (108, 276) for protease IV of \textit{E. coli} have been characterized and mapped on \textit{E. coli} chromosomal DNA. Protease IV was shown to be a tetramer of the \textit{sppA} gene product.

(ii) ATP-dependent proteases. ATP-dependent proteolysis plays a major role in the turnover of both abnormal proteins and a variety of regulatory proteins in both prokaryotic and eukaryotic cells. Three families of ATP-dependent proteases are found in \textit{E. coli}: La (or Lon), Clp (or Ti), and FtsH (or HflB) proteases. Lon and Clp are soluble proteins, whereas FtsH is a membrane-anchored protein.

In vitro studies on ATP-dependent proteolysis have shown that the major ATP-dependent activity in the extracts of \textit{E. coli} cells is the Lon protease (73). The \textit{lon} gene of \textit{E. coli} K-12 has been cloned (334), sequenced (3, 35), and shown to be dispensable by insertional mutagenesis of the gene (180). Extracts from Lon-deficient \textit{E. coli} cells still catalyze ATP-dependent proteolysis mediated by a soluble two-component protease, Clp. Two dissimilar components of Clp are (i) the ClpA reg-
ulatory polypeptide, with two ATP-binding sites and an intrin-
sic ATPase activity, and (ii) the ClpP subunit, with a proteo-
lytic active site. Clp is a serine protease, and its nucleotide sequence (181) showed little homology to the known classes of serine proteases representing a unique family of serine prote-
tases (182).

The cleavage of proteins such as casein and albumin by Clp proteases requires both ClpP and the regulatory subunit ClpA and ATP. However, it has been observed that ClpP can inde-
dependently catalyze endopeptidolytic cleavage of short peptides at a lower rate than in the presence of ClpA and ATP. The gene encoding ClpP is, at 10 min on the \textit{E. coli} map, nearer to the gene encoding the ATP-dependent \textit{Lon} protease of \textit{E. coli} and farther from the gene encoding ClpA. Primer extension experiments indicate that the transcription initiates immediately upstream of the coding region for ClpP, with a major transcription start at 120 bases in front of the start of translation. ClpP insertion mutants have been isolated, and strains devoid of ClpP are viable in the presence as well as the absence of Lon protease. Genetic evidence is available demonstrating that ClpA and ClpP act together in vivo (181). Processing of ClpP appears to involve an intermolecular autocatalytic cleavage reaction which is shown to be independent of ClpA (182). A speculative model for the chaperone-like function of ATP-
dependent proteases has been postulated by Suzuki et al. (275). The dual function of the ATP-dependent protease is determined by the affinity of the protein for the subunit or domain. Based on this, the ATP-dependent protease may regulate the subunit stoichiometry of protein complexes.

\textit{Miscellaneous}. Among the bacterial representatives of the trypsin family, \textit{\alpha}-lytic protease, an extracellular enzyme of the gram-negative soil bacterium \textit{Lysobacter enzymogenes} 495, is of particular interest. Nucleotide sequence analysis and S1 mapping of the structural gene for the \textit{\alpha}-lytic protease from \textit{L. enzymogenes} 495 indicated that the enzyme is synthesized as a prepro-protein (41 kDa) that is subsequently processed to its mature extracellular form (20 kDa) (260). The gene was further expressed in \textit{E. coli} by fusing the promoter and signal sequence of the \textit{E. coli phoA} gene to the proenzyme portion of the \textit{\alpha}-lytic protease gene (261). Following induction, an active enzyme was produced both intra- and extracellularly. Fusion of the mature protein domain alone resulted in the production of an inactive enzyme, indicating that the large N-terminal pro-
protein region is necessary for activity. Epstein and Wensink also cloned and sequenced the gene for \textit{\alpha}-lytic protease, a 19.8-kDa serine protease secreted by \textit{L. enzymogenes} (57). The nucleotide sequence contains an ORF which codes for the 198-residue mature enzyme and a potential prepro-peptide, also of 198 residues.

\textit{Achromobacter} protease 1 (API) is a mammalian-type, lys-
ine-specific serine protease that specifically hydrolyzes the lysyl peptide bond. The nucleotide sequence analysis of API from \textit{Achromobacter lyticus} M497-1 revealed that the gene codes for a single polypeptide chain of 653 aa (208). The 263-aa mature protein, which was identified by protein se-
quencing, was found to be flanked N-terminally by 205 aa including a signal peptide and C-terminally by 180 aa. \textit{E. coli} carrying a recombinant plasmid containing the API gene over-
produced and secreted the protein (API) into the periplasm. The N-terminal amino acid sequence of API was the same as
that of mature API, whereas the enzyme retained the C-terminal extended polypeptide chain. The structural gene for β-lytic protease was cloned from \textit{A. lyticus}, and the nucleotide sequence analysis of the gene revealed a mature enzyme of 179 aa, with additional 195 aa at the N-terminal end of the enzyme, which includes the signal peptide (161).

Characterization of a serine protease gene that cleaves specifically on the carbonyl side of acidic residues from \textit{Staphylococcus aureus} V8 revealed a 68-residue N-terminal extension which includes a 19- to 29-residue signal peptide, the mature protein, and the C-terminal region with several repeated acidic amino acid-rich tripeptides (29). The C-terminus may function as a competitive inhibitor of the pro-protein form of the enzyme, perhaps to prevent activity prior to secretion.

Aqualysin I, an alkaline serine protease, is secreted into the culture medium by an extreme thermophile, \textit{Thermus aquaticus} YT-1. Aqualysin I shows high DNA sequence homology to the subtilisin-type serine proteases, especially in the regions containing the active-site residues (Asp32, His64, and Ser221) of subtilisin BPN’ (148). The nucleotide sequence also revealed that the enzyme is produced as a large precursor, containing the N-terminal portion, the protease, and the C-terminal portion.

The gene (\textit{fgfA}) for the major extracellular protease of \textit{Thermomonospora fusca} YX was isolated, sequenced, and expressed in \textit{Streptomyces lividans} (152). The ORF encoded 375 residues including a 31-residue potential signal sequence, an N-terminal 150-residue prosequence, and the 194-residue mature protease belonging to chymotrypsin family. \textit{Alteromonas} sp. strain O-7, a marine bacterium, excretes alkaline serine proteases or subtilases (AprI and AprII) into the growth medium. The results of the deduced amino acid sequence analysis of genes for both AprI and AprII indicated that both the enzymes are produced as large precursors consisting of four domains: the signal sequence, the N-terminal pro-region, the mature Apr or AprII, and the C-terminal extension (298, 299). The amino acid sequence of mature AprI shows high sequence homology to that of class I subtilase, while the sequence of AprII shows high sequence homology to that of class II subtilase. Repeated sequences were observed in the C-terminal pro-region, showing high homology to sequences from the C-terminal pro-region of other known gram-negative bacteria (\textit{V. ang溜ticum}, \textit{Xanthomonas campestris}, and \textit{V. proteolyticus}).

\textbf{IgA family of proteases.} Immunoglobulin A1 (IgA1) proteases form a very heterogenous group of extracellular endopeptidases produced by a number of bacterial pathogens that colonize human mucosal surfaces. The enzymes specifically cleave human IgA1, which participates in the immune system surveillance in the human mucosa. A number of reports (62, 224, 232) on the cloning of the \textit{iga} gene, encoding the IgA1 protease from \textit{Neisseria gonorrhoeae}, are available. Nucleotide sequence analysis revealed that the enzyme is produced as a large precursor with three functional domains, i.e., the N-terminal leader peptide, the protease, and the carboxy-terminal "helper" domain. An overall structural similarity to the \textit{iga} gene from \textit{N. meningitidis} was also demonstrated (169).

Comparison of the deduced amino acid sequence of the gene of \textit{Haemophilus influenzae} serotype b with that of a similar protease from \textit{N. gonorrhoeae} revealed several domains with a high degree of homology (228). An enzyme secretion mechanism analogous to that for \textit{N. gonorrhoeae} IgA1 protease was proposed for \textit{H. influenzae} IgA1 protease. Limited diversity has been found among the IgA1 protease genes of \textit{H. influenzae}, serotype b strains (230), information that is useful from the point of view of vaccine preparation.

Cloning of streptococcal IgA1 genes from \textit{Streptococcus sanguis} ATCC 10556 (70) and \textit{S. pneumoniae} (229, 308) has been reported. Hybridization experiments with an \textit{S. sanguis} IgA1 protease gene probe showed no detectable homology to chromosomal DNA of gram-negative bacteria secreting IgA1 proteases. The gene encoding IgA1 protease from \textit{S. pneumoniae} was identified by using the \textit{S. sanguis} protease probe. However, the \textit{iga} gene was found to be highly heterogenous among streptococcal species.

From the foregoing, it can be seen that subtilisins (270) and neutral proteases (279, 323) of various \textit{Bacillus} species, the α-lytic protease from \textit{L. enzymogenes} (57, 260), and proteases A and B from \textit{S. griseus} (89) have long polypeptide extensions at their N termini. The IgA protease of \textit{N. gonorrhoeae} (224) and the protease of \textit{S. marcescens} (322) have C-terminal extensions. \textit{Achromobacter} protease I (208), aqualysin I from \textit{T. aquaticus} (148), and AprI and AprII from \textit{Alteromonas} sp. strain O-7 (298, 299) bear long peptide chains at both the N and C termini. The function of the pre-peptide portion (signal peptide) in these precursors is possibly to assist in the transport of the secretory protein across the cytoplasmic membrane. The exact role of the pro-peptide region is not known; possibly the long peptide serves to inhibit the mature protease to which it is connected (29, 57). It is also possible that the pro-peptide helps the protease to fold into its active form (111, 261).

\textbf{Fungi}

As in bacteria, cloning of the protease genes of fungi has been attempted from both the commercial and pathogenicity points of view.

\textbf{Filamentous fungi.} (i) Acidic proteases. (a) \textit{Mucor}. Two closely related species of zygomycete fungus, \textit{Mucor pusillus} and \textit{Mucor miehei}, secrete aspartate proteases, also known as mucor rennins, into the medium. The enzymes possess high milk-clotting activity and low proteolytic activity, enabling them to be used as substitutes for calf chymosin in the cheese industry.

Sequencing of the cloned gene encoding \textit{M. pusillus} rennin (MPR) revealed an ORF without introns, encoding possible pre-pro-sequences (66 aa) upstream of the mature MPR sequence (296). The deduced amino acid sequence showed a high degree of homology to that of \textit{M. miehei} rennin (MMR). The gene encoding \textit{M. miehei} aspartyl protease (MMP) has also been cloned and sequenced (79). The deduced primary translation product showed an N-terminal extension which appears to comprise a signal peptide of 22 aa and a propeptide of 47 aa. Fungal aspartyl proteases are structurally related to each other and to the gastric aspartyl proteases chymosin and pepsin; therefore, they may be activated in a manner similar to their gastric counterparts. When the gene encoding the pre-pro-form of MPR was cloned in \textit{S. cerevisiae} under the control of the yeast GAL7 promoter, an inactive zymogen of the enzyme with the 44-aa pro-sequence was identified in the medium during the initial stage of cultivation (94). In vitro conversion of the zymogen to mature MRP was shown to proceed autocatalytically under the acidic conditions.

(b) \textit{Rhizopus}. \textit{Rhizopus niveus}, belonging to the zygomycete class, also secretes aspartyl protease abundantly. The gene encoding \textit{R. niveus} aspartic protease (RNP) was cloned and sequenced (100). Comparison of the deduced amino acid sequence with the amino acid sequence of rhizopuspepsin of \textit{R. chinensis} (282) revealed that the RNP gene has an intron within its coding region. A prepro-sequence of 66 aa upstream of the mature enzyme was also revealed. High-level secretion of RNP-I was achieved by subcloning the RNP-I gene into
Saccharomyces cerevisiae (101). Yeast cells carrying the intact RAP-I gene under the control of the glyceraldehyde-3-phosphate dehydrogenase gene promoter of S. cerevisiae were unable to synthesize RAP-I. On removal of the intron of the RAP-I gene, the cell secreted the enzyme with high efficiency.

(c) Aspergillus. The pepA gene encoding the aspartic protease, aspergillopepsin A, from Aspergillus awamori (15), the pepA gene from A. oryzae (74), and the cDNA coding for an elastinolytic aspartic protease, aspergillopepsin F, from A. fumigatus (156, 237) were cloned and sequenced. The nucleotide sequence data revealed that the ORFs encoding aspartic proteases in these aspergilli are composed of four exons. Prepro-peptides of 69, 78, and 70 aa were found to precede 395-, 326-, and 332-aa mature proteins of A. awamori, A. oryzae, and A. fumigatus, respectively. The amino acid sequence of aspergillopepsin F shows 70, 66, and 67% homology to the sequences of those from A. oryzae, A. awamori, and A. saitoi, respectively. The primary structure of aspergillopepsin I from A. saitoi ATCC 14332 (now designated A. phoenicis) was deduced from the nucleotide sequence of the gene (257). The cDNA of the gene was also cloned and expressed in yeast cells.

Two types of acid endopeptidases, acid proteases A and B (commercially named proctase A and B), are known to be secreted into the medium by A. niger var. macrosporus. Pro tease B is a typical aspartic protease, inhibited by pepstatin, whereas protease A is not inhibited by pepstatin. Sequencing of the protease A gene revealed an 846-bp structural gene without any introns encoding the precursor form of the enzyme (114, 125, 285). The precursor, of 282 residues, includes an N-terminal prepro-sequence of 59 residues, the L chain of 39 residues, an intervening sequence of 11 residues, and the H chain of 173 residues linked in that order. The deduced amino acid sequence (394 residues) of the prepro-form of protease B showed 98% homology to the sequences of aspergillopepsin I from A. awamori and A. saitoi and 68% homology to that of aspergillopepsin I from A. oryzae (113, 175). The cDNA was expressed in E. coli, and the purified pro-protease B showed protease activity under acidic conditions (pH 2 to 4).

(ii) Alkaline proteases. (a) Aspergillus. Alkaline protease (Alp) produced by A. oryzae, a filamentous ascomycete used in the manufacture of soy sauce, is considered to play an important role in producing the flavor of soy sauce by hydrolyzing the raw materials. Tatsumi et al. constructed the cDNA library of A. oryzae ATCC 20386 in pUC119 and isolated a cDNA (1,100 bp) encoding the mature region of Alp (286). The nucleotide sequence of the cDNA lacked most of the DNA sequences corresponding to the prepro-region. The entire cDNA coding for prepro-Alp was cloned and expressed in S. cerevisiae (288). The character of the Alp secreted from S. cerevisiae was shown to be identical to that of the native Alp. The predicted mature Alp consists of 282 aa and shows homology to other serine proteases of subtilisin families from bacteria as well as from fungi. Alp has a 121-aa prepro-region wherein the N-terminal 21 residues show the characteristics of a signal peptide. Alp expressed in S. cerevisiae was secreted with the N terminus processed correctly, analogous to the expression in S. cerevisiae of aspartic protease from M. pusillus (321). The prepro-Alp cDNA of A. oryzae was further cloned into an osmophilic yeast, Zygosacharomyces rouxii (207). The recombinant Z. rouxii secreted a large amount of Alp (about 300 mg/liter) into the culture medium. The Alp gene is 1,374 nucleotides long and contains three introns, one in the pre-region and two in the mature protein region (195).

A gene encoding Alp from the A. oryzae Thailand industrial strain was isolated from the genomic library by using oligode-oxiribonucleotide probes based on the A. oryzae ATCC 20386 cDNA sequence (33). By comparison with the published cDNA sequence (286), Alp from A. oryzae Thailand was found to be encoded by four exons. Transformation of the alpA4 gene in the high-level-Alp-producing A. oryzae strain U212, obtained by UV mutagenesis, resulted in the production of up to five times as much Alp as in the parental strain. A. fumigatus and A. flavus, the agents of invasive aspergillosis, secrete highly homologous serine proteases. The genomic as well as cDNA clones encoding elastinolytic Alp from both A. fumigatus (123, 237) and A. flavus (233) were sequenced. The A. nidulans prtA gene coding for Alp was isolated by using the gene encoding A. oryzae Alp (131). The nucleotide sequence of prtA was determined, and the deduced amino acid sequence showed a high degree of similarity to Alp from A. fumigatus, A. flavus, and A. oryzae. prtA transcription was shown to be dependent on the medium composition.

(b) Acremonium. Acremonium chrysogenum ATCC 11550 (Cephalosporium acremonium) produces a considerable amount of extracellular Alp. The cDNA and genomic DNA encoding Alp were isolated from the A. chrysogenum cDNA and genomic DNA libraries, respectively (115). The nucleotide sequence of the gene was determined. The deduced amino acid sequence showed 57% homology to that of A. oryzae Alp. Cloning of the entire cDNA encoding A. chrysogenum Alp into S. cerevisiae directed the secretion of enzymatically active Alp into the culture medium.

(c) Fusarium. The transfer of the Fusarium alkaline protease gene (136) into A. chrysogenum resulted in transformants producing large amount of Alp (193). Southern hybridization analysis, as well as PCR of genomic DNAs from these transformants, showed chromosomal integration of the full-length alp gene. The enzyme secreted by A. chrysogenum had properties identical to that of the native Fusarium Alp, indicating that the Alp promoter, signal sequence, and introns functioned correctly in A. chrysogenum.

(iii) Serine proteases. (a) Tritirachium. Proteinase K is a serine endoprotease excreted by the fungus Tritirachium album Limber. The enzyme is able to hydrolyze native proteins rapidly and is active in the presence of detergents (urea, sodium dodecyl sulfate, etc.), making the proteinase K one of the most useful tools in molecular biology. The enzyme exhibits strong similarity to the bacterial subtilisins. The genomic DNA as well as the cDNA encoding proteinase K from T. album Limber have been cloned in E. coli, and the entire nucleotide sequence of the coding region, including the 5′- and 3′-flanking regions, has been determined (81). The nucleotide sequence analysis revealed that the primary secreted product is azymogen containing a 15-aa signal sequence and a 90-aa pro-peptide. The pro-peptide is presumably removed in the later steps of the secretion process or upon secretion into the medium. The proteinase K gene was shown to be composed of two exons and one 63-bp intron located in the proregion. The pro-proteinase K gene was expressed in E. coli under the control of the tac promoter.

The coding sequence of proteinase T from T. album Limber (247) was shown to be interrupted by two introns. The deduced amino acid sequence showed 53% identity to that of proteinase K. The presence of four cysteines in the mature proteinase, probably in the form of two disulfide bonds, explains the thermal stability of proteinase T. The proteinase T cDNA was expressed in E. coli, and the authenticity of the product was confirmed by Western blotting and N-terminal analysis of the recombinant product.

(iv) Metalloproteases. (a) Aspergillus. Jaton-Ogay et al. (124) and Sirakova et al. (262) cloned and sequenced the gene...
as well as the cDNA encoding the 42-kDa elastinolytic metalloproteinase (MEP) of \textit{A. fumigatus}. Comparison of the nucleotide sequences revealed that the genomic and the cDNA sequences are analogous except for four introns interrupting the ORF. The enzyme was shown to be produced in a preproform, with a 384-aa mature protease region. In another study, no intron was found in the ORF of \textit{A. flavus} mep20 (encoding a 23-kDa MEP) whereas a 59-bp intron was present in the gene from \textit{A. fumigatus} (a homolog of mep20) (234). The MEP20 proteins of \textit{A. flavus} and \textit{A. fumigatus} have 68% identity.

\textbf{Yeasts. (i) Acidic proteases.} The yeast \textit{Saccharomyces fibuligera} produces an extracellular acid protease (PEP1). DNA coding for the secretable acid protease gene of \textit{S. fibuligera} was isolated (95, 320). The enzyme produced by \textit{Saccharomyces cerevisiae} species such as \textit{S. fibuligera} is a multigene family. Three putative new members, \textit{S. cerevisiae} SAP1, SAP2, and SAP3, were also isolated and sequenced. Evidence was obtained for the existence of SAP multigene families in other \textit{Saccharomyces} species (332). The amino acid sequence of an acid extracellular protease (AXP) from \textit{Yarrowia lipolytica} (AXP) from \textit{C. guilliermondii} (191).

\textbf{Yeast carboxypeptidase (CPY) is a glycosylated yeast vacuolar protease that is used commercially in peptide synthesis. CPY is encoded by the \textit{PRC1} gene. To increase the production of CPY in \textit{S. cerevisiae}, \textit{PRC1} was placed under the control of the strongly regulated yeast GAL1 promoter on the multicopy plasmids and introduced into \textit{wpl1} mutant strains (202). About a 200-fold increase in the level of secreted CPY (40 mg/liter) was obtained compared to the level in a \textit{wpl1} mutant carrying a single copy of the wild-type \textit{PRC-1} gene. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed two forms of secreted active CPY, probably due to the different levels of glycosylation. The structural gene \textit{PRB1}, encoding the vacuolar protease \textit{B} of \textit{S. cerevisiae}, was cloned by complementation of the \textit{prb1}-\textit{1122} mutation (190).

\textbf{PRG1}, a yeast gene encoding the 32-kDa proteasome, which shows 55.6% sequence homology to 80% of the RING10 gene product (human proteasome), was identified (65). Genomic disruption of \textit{PRG1} revealed that it is essential for yeast growth. The results strongly indicate that the antigen-processing system present in vertebrates has evolved from a basic cellular process present in all organisms.

\textbf{Viruses}

Gene cloning of viral proteases has been undertaken for the isolation and overexpression of the gene and for subsequent screening of inhibitory compounds that may be used in the development of chemotherapeutic agents. Viral protease is responsible for processing of polyprotein precursors into the structural proteins of the mature virion. Among viruses, reports on cloning of protease genes are limited mainly to animal viruses (Table 5).

\textbf{Animal viruses. (i) Herpesviruses.} Each member of the herpesvirus family encodes a unique serine protease in association with a capsid assembly protein, with the associated ORFs being designated \textit{U}1 and \textit{U}2, respectively. The \textit{U}1 gene encodes a protease responsible for the C-terminal cleavage of the nucleocapsid-associated proteins (ICP35c and ICP35d) to their posttranslationally modified...
counterparts (ICP35e and ICP35f). The protease expressed in E. coli exhibited autoprocessing and specifically cleaved the ICP35 protein assembly (47). Similarly, genes encoding proteases from HSV-2, murine cytomegalovirus (MCMV), and human herpesvirus 6 (HHV-6) have been studied (172, 271, 292). Such studies assist in the investigation of the role of proteolytic processing in the virus.

(ii) Adenoviruses. Adenoviruses code for a serine-centered, neutral protease specific for selected Gly-Ala bonds in several virus-encoded precursor proteins that are required for virion maturation and infectivity. To determine the functional domains of this key enzyme, protease genes from various types of adenoviruses have also been cloned and sequenced (102–104, 306).

(iii) Retroviruses. The genomic organization of retroviruses is 5′-LTR-gag-pro-pol-env-LTR-3′ (where LTR is a long terminal repeat). The pro/prt gene product is an aspartyl protease, which is responsible for processing the gag and pol polyprotein precursors into the structural proteins of the mature virion. Comparison of the genomic organization of certain retroviruses revealed that prt lies in the carboxyl terminus of gag in Rous sarcoma virus (RSV) (252) and avian sarcoma-leukosis virus (ASLV) (144); in the amino terminus of pol in AIDS-associated retrovirus type 2 (ARV-2) (248); in the same reading frame as both gag and pol in Moloney murine leukemia virus (M-MuLV) (256); and as a separate reading frame in simian AIDS retrovirus type 1 (SRV-I) (231), human T-cell leukemia virus type 2 (HTLV-2) (255), bovine leukaemia virus (BLV) (245), and Mason-Pfizer monkey virus (MPMV) (268). Besides cloning and sequencing of the prt gene, there are a few reports on expression of the gene in E. coli (40, 105, 144). Significant inhibition of the expressed protease activity by pepstatin A confirmed that HTLV-1 protease is a member of the aspartyl protease group (246).

Human immunodeficiency virus (HIV), a causative agent of AIDS, is also a member of the family Retroviridae. The virus exhibits the same overall gag-pol-env genome organization as that of other retroviruses. The genome-size mRNA of HIV-1 is translated into two polyproteins: Pr55 (gag gene product) and Pr160 (gag-pol gene product). Cleavage of these polyproteins by the viral protease into smaller structural proteins and replication enzymes such as reverse transcriptase and integrase is necessary to produce infectious progeny virions from immature virus particles. The enzyme, a part of the polyprotein, has a highly conserved sequence, Asp-Thr-Gly, which is homologous to the active site of the aspartic proteases and is thought to belong to this enzyme family (216). The protease is essential for the retroviral life cycle, as indicated by the production of noninfectious, replication-deficient virions by Moloney murine leukemia virus variants mutated in the protease-encoding region (130). This suggests that HIV protease is a good target for chemotherapy and that specific inhibitors of this enzyme may have a significant function in the treatment of AIDS without interfering with the host cell physiology. To obtain sufficient quantities of the HIV protease for biochemical and structural analyses, several groups have described expression of the recombinant HIV-1 protease in E. coli (46, 78, 171). Pichuang et al. have reported extracellular expression of HIV-1 aspartic protease in S. cerevisiae (222). The expressed enzyme was shown to exhibit a proteolytic activity, as has been shown to be associated with the purified HIV-1 virions (164). Debouck et al. expressed the HIV protease gene product in E. coli (46). The product was shown to autocatalyze its maturation from a larger precursor and to process an HIV Pr55 gag protein when coexpressed in E. coli. This allowed a structure-function analysis of the HIV protease and provided a simple assay for the development of potential therapeutic agents directed against the critical viral enzyme.

(iv) Picornaviruses. Human rhinovirus is a member of the picornavirus (small RNA) family. Rhinovirus has commercial importance since it is the causative agent of about 15% of cases of the common cold. A cDNA encoding the viral protease from the 3C region of human rhinovirus type 14 was expressed in E. coli through the use of a periplasmic secretion vector (162). A comparison of the 3C protease regions of all the available picornaviruses (foot-and-mouth disease virus, encephalomyocarditis virus, and poliovirus) sequences revealed two completely conserved residues, Cys147 and His161, which may be the reactive residues of the active sites of these cysteine proteases (69).

Plant viruses. Potyviruses are a cause of serious losses of several major crop plants. In plants infected with the potyviruses, inclusions consisting of viral proteins are found in the cell nucleus. One of them, the nuclear inclusion protease (NIA), is the major viral protein responsible for the proteolytic maturation of the polyprotein encoded by the virus. The elucidation of the structure of such virus-encoded proteins could potentially facilitate the design of novel polypeptides which bind to them and inhibit their functions. With this objective, cDNAs for NIA proteases were cloned and sequenced from bean yellow mosaic virus (22) and zucchini yellow mosaic virus (Singapore isolate) (316).

The potential contributions of genetic engineering to mankind are enormous and will benefit agriculture, animal husbandry, environmental protection, food production and processing, human health care, manufacture of biochemicals and biofuels, etc. In general, the application of genetic engineering to proteases will facilitate their use in industry and enable the development of therapeutic agents against the proteases that are important in the life cycle of organisms which cause serious diseases.

PROTEIN ENGINEERING

Many industrial applications of proteases require enzymes with properties that are nonphysiological. Protein engineering allows the introduction of predesigned changes into the gene for the synthesis of a protein with an altered function that is desired for the application. Recent advances in recombinant DNA technology and the ability to selectively exchange amino acids by site-directed mutagenesis (SDM) have been responsible for the rapid progress of protein engineering. Identification of the gene and knowledge of the three-dimensional structure of the protein in question are the two main prerequisites for protein engineering. The X-ray crystallographic structures of several proteases have been determined (39, 143, 223, 267). Proteases from bacteria, fungi, and viruses have been engineered to improve their properties to suit their particular applications.

Bacteria

Subtilisin has been chosen as a model system for protein engineering since a lot of basic information about this commercially important enzyme is available. Its pH dependance (290), catalytic activity (278, 281), stability to heat or denaturing agents (112, 199), and substrate specificity (10, 14, 30, 59, 243) have been altered through SDM. A slightly reduced rate of thermal inactivation was observed for a subtilisin BPN’ variant containing two cysteine residues (Cys22, Cys87) (186, 214). Oxidation of Met222 adjacent to the Ser221 in the active site of subtilisin reduces the catalytic activity of subtilisin. The
effect of substitution of Met222 with different amino acids revealed that small side chains yield the highest activity. The mutant enzymes Ser222, Ala222, and Leu222 were active and stable to peroxide for 1 h. Probing of the specificity of the S1 binding site of Met222 Cys/Ser mutants of subtilisin from Bacillus licheniformis with boronic acid inhibitors revealed similar binding trends for the mutant and the parent (269). The disulfide bonds introduced into subtilisin away from its catalytic center were shown to possess increased autolytic stability (312). Higher thermostability of subtilisin E as a result of introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease has been reported (280). Strausberg et al. have created the environment for stabilization of subtilisin by deleting the calcium-binding loop from the protein (273). Analysis of the structure and stability of the prototype with the loop deleted followed by SDM resulted in a mutant with native proteolytic activity and 1,000-fold greater stability under strongly chelating conditions. SDM-mediated substitution of Asn241 buried in the neutral protease of B. steatorrhophilus by leucine resulted in an increase in thermostability of 0.7 ± 0.1°C (55). The thermostability of the neutral protease from B. subtilis was increased by 0.3 and 1.0°C by replacing Lys with Ser at positions 249 and 290, respectively, whereas the Asp249 and Asp290 mutants exhibited an increased stabilization by 0.6 and 1.2°C, respectively (54).

A protein engineering study was undertaken by Brünenberg et al. to determine the functions of one of the largest loop insertions (residues 205 to 219), predicted to be spatially close to the substrate-binding region of the SK11 protease from L. delbrueckii and susceptible to autolysis (28). Deletion or modification of this loop was shown to affect the activity and autolysis of the protease. Graham et al. showed that random mutagenesis of the substrate-binding site of α-lytic protease, a serine protease secreted by the soil bacterium Lysobacter enzymogenes, generated enzymes with increased activities and altered primary structures (77). Substitution of His120 by Ala in the LasA protease of P. aeruginosa yielded an enzyme devoid of staphylolytic activity. Thus, His120 was shown to be essential for LasA activity (82).

**Fungi**

Fungal aspartic proteases are able to cleave substrate with “Lys” in the P1 position. Sequencing and structural comparison suggest that two aspartic acid residues (Asp30 and Asp77) may be responsible for conferring this unique specificity. Lowther et al. engineered the substrate specificity of rhizopuspepsin from Rhizopus niveus and demonstrated the role of Asp77 in the hydrolysis of the substrates with lysine in the P-1 position (173).

The primary structure of aspergillopepsin I from Aspergillus saitoi ATCC 14332 (now designated A. phoenicis) was deduced from the nucleotide sequence of the gene (257). To identify the residue responsible for determining the specificity of aspergillopepsin I, we performed the substrate-binding pocket, Asp76 was replaced with a Ser residue by SDM. The striking feature of this mutation was that only the tryptophan-activating activity of the enzyme was destroyed, suggesting the importance of the Asp76 residue in binding to basic substrates. To elucidate whether the processing of the pro-region occurs by autolysis or by involving a processing enzyme, Tatsumi et al. changed Ser228 to Ala by SDM (287). S. cerevisiae cells harboring a recombinant plasmid with mutant Alp did not secrete active Alp into the culture medium. The yeast cells accumulated a protein of 44 kDa, probably a precursor of Alp (the 34-kDa mature Alp plus the 10-kDa pro-peptide), suggesting that autoproteolytic processing of the pro-region was occurring.

Introduction of a disulfide bond by SDM is known to enhance the thermostability of a cysteine-free enzyme. Aqualysin I, a thermostable subtilisin-type protease from Thermus aquaticus YT-1, contains four Cys residues forming two disulfide bonds (149). The primary structure of Alp showed 44% homology to that of aqualysin I, and sites for Cys substitutions to form a disulfide bond were chosen in the Alp based on this homology. Ser69, Gly101, Gly169, and Val200 were replaced by Cys in the mutant Alp. Both Cys69-Cys101 and Cys169-Cys200 mutant Alps were expressed in S. cerevisiae, and the enzymes were purified to homogeneity. The Cys69-Cys200 disulfide bond was shown to increase the thermostability as well as the thermostolerance of Aspergillus oryzae Alp (110).

In vitro mutation of an aspartic acid residue predicted to be in the active site abolished the barrier activity of S. cerevisiae (177). BAR1 possesses a carboxyl-terminal domain of an unknown function, and deletion of 166 of 191 aa of this region had no significant effect on the barrier activity.

**Viruses**

The protease of HCMV was rendered stable by conversion of one of the three Val141, Val207, or Val254 residues to Gly by SDM (151). The resulting stable proteases are useful as screening tools for HCMV antiviral agents and as diagnostic tools for diseases resulting from HCMV infection.

Replacement of Asp64, a residue from the catalytic core of the enzyme present among aspartyl proteases, with Gly was shown to abolish the correct processing of the 53K gag precursor by HTLV-I gag protease (87).

In poliovirus, the mutation of highly conserved residues, e.g., Cys147 or His161, produced an inactive enzyme while mutation of a nonconserved residue, Cys154, had only a negligible effect on the proteolytic activity (117).

The protein-engineering technique has been exploited successfully for obtaining proteases which show unique specificity and/or stability at high temperature and pH. It has also contributed substantially to our understanding of the structure-function relationship of proteases. In future, protein engineering will offer possibilities of generating proteases possessing entirely new functions.

**SEQUENCE HOMOLOGY**

Studies of DNA and protein sequence homology are important for a variety of purposes and have therefore become routine in computational molecular biology. They serve as a prelude to phylogenetic analysis of proteins and assist in predicting the secondary structure of DNA and proteins. Proteases are a complex group of enzymes and vary enormously in their physicochemical and catalytic properties. The nucleotide and amino acid sequences of a number of proteases have been determined, and their comparison is useful for elucidating the structure-function relationship (5). The homology of proteases with respect to the nature of the catalytic site has been studied (12, 13). Accordingly, the enzymes have been allocated to the three main evolutionary families and clans. It has been suggested that there may be as many as 60 evolutionary lines of peptidases with separate origins. Some of these contain members with quite diverse peptidase activities, and yet there are some striking examples of convergence (236).

A number of reports on the homology of proteases are available. Takagi et al. found that the thermostable proteases
of Bacillus steatorrhophilus and B. thermoproteolyticus are 85% homologous and the thermolabile proteases of B. subtilis and B. amyloliquifaciens are 82% homologous, whereas the thermolabile protease of B. steatorrhophilus is only 30% homologous to the thermolabile protease of B. subtilis (279). However, an amino acid sequence of 17 residues, which also includes the active-site histidine residue, was found to be highly conserved in all four neutral proteases, suggesting that they have the same three-dimensional structure around the active site despite the difference in their source and physicochemical properties such as thermostability.

Koide et al. compared the amino acid sequences of intracellular serine proteases from B. subtilis with those of subtilisin Carlsberg and subtilisin BPN' and showed that they were 45% homologous (138). The sequence around the catalytic triad of serine, aspartate, and histidine is highly conserved, suggesting that the genes for both the intracellular and extracellular proteases have evolved from a common ancestor by divergent evolution (200).

The amino acid sequence of an extracellular alkaline protease, subtilisin J, is highly homologous to that of subtilisin E and shows 69% identity to that of subtilisin Carlsberg, 89% identity to that of subtilisin BPN', and 70% identity to that of subtilisin DY. The amino acid sequence of subtilisin J is completely identical to that of the protease from B. amyloliquifaciens except for two amino acid substitutions, Thr130 to Ser130 and Thr162 to Ser162, in addition to one amino acid substitution in the signal peptide and two in the propeptide region. The probable active-site residues of subtilisin J, i.e., Asp32, His64, and Ser221, are identical to those of other subtilisins from Bacillus. Therefore, it was concluded that the alkaline protease from B. steatorrhophilus is a subtilisin. Similarly, the various Bacillus serine alkaline proteases, such as bacillopeptidase F, subtilisin, Epr, and ISP-1, show considerable homology and conserved amino acids around the active-site residues, i.e., Ser, Asp, and His (265).

The extracellular proteases of B. subtilis are synthesized as prepro-enzymes. Four neutral proteases from bacilli with known pro-sequences were compared, and considerable homology within the pro-peptide region was observed (297). Since the pro-peptide region mediates the folding of the protease, it would be interesting to learn about the residues essential for folding and to determine whether the mechanism of folding is similar in these proteases. Sequences corresponding to the mature form of these enzymes were compared by using thermolysin sequence as a reference. The zinc-binding site residues, i.e., Ser, Asp, and His (265).

The extracellular proteases of B. subtilis are synthesized as prepro-enzymes. Four neutral proteases from bacilli with known pro-sequences were compared, and considerable homology within the pro-peptide region was observed (297). Since the pro-peptide region mediates the folding of the protease, it would be interesting to learn about the residues essential for folding and to determine whether the mechanism of folding is similar in these proteases. Sequences corresponding to the mature form of these enzymes were compared by using thermolysin sequence as a reference. The zinc-binding site residues, i.e., Ser, Asp, and His (265).

Alkaline proteases from various species of Aspergillus also show a high degree of homology (131). Alp from A. oryzae shows considerable homology (29 to 44%) to the members of the subtilisin family with conserved active-site residues (288). However, Alp exhibits little homology to mammalian serine proteases such as trypsin and chymotrypsin. The deduced structure of the KEX-1 protein, required for the production of the killer toxin of Kluyveromyces lactis contains an internal domain with a striking homology to the sequences of subtilisin-type proteases (242). Therefore it was deduced that the product of the KEX-1 gene of K. lactis is a protease involved in the processing of the toxin precursor.

The characteristic of trypsin-related enzymes is the presence of disulfide bonds, which are absent in all known subtilisins. Proteinase K from Trichtrichium album Limber is a single chain protein of 277 aa with two disulfide bonds at positions 34-124 and positions 179-248 and a free -SH group at position 73. Sequences around the active-site residues correspond to those around the active-site residues of subtilisins. Comparison of the proteinase K sequence with known subtilisins shows 35% homology and 44% sequence identity to thermolysin, which is indicative of a relationship between proteinase K and the subtilisin family. It is likely that these enzymes have evolved from a common ancestral precursor serine protease (122). However, there is a distinct difference between the typical subtilisins and proteinase K, since the latter has two disulfide bonds, which are lacking in subtilisins. Therefore, it has been assumed that the two progenitors diverged from an ancestral protease, separating the subtilisin-related enzymes into two subclasses: (i) cysteine-containing subtilisins e.g., proteinase K and thermolysin, and (ii) cysteine-free subtilisins, e.g., subtilisin Novo, Carlsberg, or DY.

The proteasome or multicatalytic endopeptidase complex is a highly-molecular-mass multisubunit complex that is ubiquitous in eukaryotes and also found in the archaeabacterium Thermoplasma acidophilum (336). While eukaryotic proteasomes contain 15 to 20 different subunits, the archaeabacterial proteasome is made of only two different subunits (α and β), yet the complexes are almost identical in size and shape. The α (233-aa) and β (211-aa) subunits of T. acidophilum have a sequence identity of 24% and an overall similarity of 47%, indicating that the genes encoding the two subunits arose from a common ancestor. All the sequences of proteasomal subunits from eukaryotes available to date can be related to either the α or β subunit of the T. acidophilum “urproteasome,” and they can be distinguished by the presence or absence of a highly conserved N-terminal extension which is characteristic of α-type subunits. In terms of evolution, the genes for these α and β subunits can be considered paralogous (genes resulting from duplication and divergence of one gene within one genome) and therefore are able to acquire different functions. The α subunit of the T. acidophilum proteasome shows sequence similarity to the S. cerevisiae wild-type suppressor gene self-encoded polypeptide, which is probably identical to the subunit YC7-α of the yeast proteasome. This lends support to a putative role of proteasomes in the regulation of gene expression (337). The amino acid sequence of Xenopus proteasome subunit XT3 is highly homologous (95.3%) to those of the rat RC3 and human HC3 subunits (66). The presence of an accessible nuclear targeting signal at the C terminus of the subunits suggests that it is probably involved in the regulation of the cellular distribution of the proteasome.

The secretable acid protease of the yeast Saccharomyces fibuligera carries a hydrophobic amino-terminal segment of about 20 aa which resembles signal sequences found in a wide variety of secretory protein precursors (95). Alignment of this sequence with the aspartyl protease family showed significant homologies, especially in the regions surrounding the two active-site aspartate residues. These results suggest that the PEPI gene is a structural gene for the secretable acid protease of S. fibuligera. The aspartic protease from Rhizopus niveus (RNAP) shows 76% homology to rhizopuspepsin, 42% homology to penicillopepsin, and 41% homology to human pepsin (100, 101). The homology between RNAP and rhizopuspepsin is found throughout their structures. Based on this homology,
an intron within the coding region and a prepro-enzyme sequence of 66 aa upstream of the mature enzyme were detected in RNAP. Studies of the homology of proteases have shown that the residues involved in the substrate and metal ion binding, catalysis, disulfide bond formation and active-site formation are conserved. Analysis of sequence homology is used in deciphering the structure-function relationship of proteases.

**EVOLUTIONARY RELATIONSHIP OF PROTEASES**

Proteases are present in all living organisms and are considered to have arisen in the earliest phases of biological evolution, some 1 billion years ago. Comparisons of amino acid sequences, three-dimensional structures, and mechanism of action of proteases assist in deciphering their course of evolution. Changes in molecular structure have accompanied the demands for altered functions of proteases during evolution. We have compiled the amino acid sequences of proteases from diverse origins such as microbes, plants, and animals and have arranged them in three different groups based on the pH of their action. These sequences, which have been selected from SWISS-PROT and PIR entries, are of comparable length and have been aligned with CLUSTAL W software for multiple alignments (291) (Fig. 4).

**Acidic Proteases**

The proteases selected here for comparison of amino acid sequences are active between pH 2 and 6. They include mostly aspartic proteases and also some of the cysteine proteases and metalloproteases. They are about 380 to 420 aa long and have different amino acid residues constituting the active site, as shown in Table 6. The homology between these acidic proteases is shown in Fig. 4A. The sequences belonging to pepsin family (A1) are grouped and are aligned below the other sequences. As expected, there is considerable homology among these five acidic proteases. The sequences around the two aspartic residues (D97 and D258, residues numbered according to the Bajra protease) constituting the active site are conserved. Among these five proteases, the rat and monkey proteases show maximum homology (68.4%) and are related to the mosquito lysosomal aspartic protease. When four monkey pepsinogens which show development-dependent expression were compared, a very high homology was observed (126). Pepsinogens A-1 and A-2/3 differed in seven amino acids and only in five amino acids when the pepsin moiety alone was examined. The mosquito lysosomal protease is very closely related to human cathepsin D, exhibiting 92% homology (37).

The amino acid sequences of *C. tropicalis* and *Saccharomyces fibuligera* show considerable homology (42.6%). High similarity scores were obtained when the acid protease from *C. tropicalis* was compared with *Rhizopus* aspartic proteases, human pepsinogen A precursor, protease A from yeast, the barrier protein from *S. cerevisiae*, and an acidic protease from *S. fibuligera* (294).

The cysteine protease from *Hordeum vulgare* shows some homology to the snake venom metalloprotease from *Crotalus atrox*, which is not statistically significant, whereas the Gpr protease from *Bacillus megaterium*, which plays a vital role in spore germination, shows least homology to all other acidic proteases but shares one of the active-site aspartate residues (D258) with them. The Gpr acidic proteases of *B. subtilis* and *B. megaterium* showed 68% identity in their sequences, but comparison of the *B. subtilis* Gpr amino acid sequence with that of its serine protease or metalloprotease revealed no significant homologies (274), which supports our observations. This suggests that the genes encoding these proteases do not have a common ancestor or that if they do so, they have undergone much divergence. The lack of homology between the spore protease and other *B. subtilis* proteases can be explained by differences in their properties such as the number of subunits and sequence specificity for the substrate. Thus, our results, in agreement with previous reports, indicate that the extent of homology is greater if the proteases belong to the same family and that in the same family the homology is greater if the phylogenetic distance is shorter.

A pairwise computer comparison also provides more information about the evolutionary relationships between the members of the different families. The dendrograms generated by this analysis, using the TreeView package (213), demonstrate the relationship among the proteins based on the similarity of the amino acid sequences (Fig. 5a).

**Neutral Proteases**

The neutral proteases, which are active at neutral or weakly alkaline or weakly acidic pH, include cysteine proteases, metalloproteases, and some of the serine proteases. Brenner (25) has pointed out that the two codons for serine TCN and AGY cannot be interconverted by single nucleotide mutations but can be connected by two other codons, ACN for threonine and TGY for cysteine. Thus, there can be at least two different lines of descent for the active-site sequences of the serine proteases. The simplest pathway for this convergent evolution is by the divergence of each line from a precursor which was itself catalytically active and had much the same sequence. It was further demonstrated that modern serine enzymes are likely to have arisen from cysteine precursors. These findings encourage the search for evidence to connect the presumed and existing cysteine sequences with their postulated metalloenzyme predecessors. For this search and construction of phylogenetic trees, gene structure is important. Thus, multiple lines of descent can be realistically considered in situations with sequence similarity but with differences in gene structure.

The neutral proteases selected for sequence analysis in the present study are in the size range of 225 to 275 aa (Table 6). The homology between them is shown in Fig. 4B. Of 14 proteases, 9 belong to the T1A or proteasome A family of the multicatalytic endopeptidase complex. The sequences of the proteasomal subunits aligned here can be related to the α subunit of the Thermoplasma proteasome and show considerable homology. It is not clear to which family of proteases the proteasomes belong (93). As in the cysteine family of proteases, all nine proteasome subunits show a conserved proline residue (P-17), which may serve to prevent unwanted N-terminal proteolysis (12). Many residues at the N terminus are highly conserved, which is a characteristic of the α-type subunits. The similarity decreases toward the C terminus which appears to be rather variable (337). Although the β subunit shows no sequence motif characteristic of serine proteases, it contains all the essential amino acid residues forming the catalytic triad or the “charge relay system” (Ser, Asp, and His). These residues are found to be conserved (Ser16, His73, and Asp84), except for the histidine in the α subunits of *Thermoplasma*, yeast (*S. cerevisiae*), and *Caenorhabditis elegans* (residues are numbered according to the *Thermoplasma* α-subunit...
FIG. 4.
FIG. 4—Continued.

Vol. 62, 1998 MICROBIAL PROTEASES 621

Downloaded from http://mmbr.asm.org/ on July 27, 2012 by USP/REITORIA/SIBI
sequence). Therefore, it is possible that the active site is shared between the α and β subunits (336). The tyrosine autophosphorylation site at Tyr123 is conserved in six of the nine sequences. The cAMP/cGMP-dependent phosphorylation sites between aa 31 and 37 are found only in Thermoplasma and Drosophila spp. (84), as reported by Zwickl et al. (337). A consensus nuclear localization signal between aa 49 and 56 (240) and a region complementary to the nuclear localization signal consensus sequence (326) between aa 201 and 212 can be identified in these sequences. Thus, the sequence comparison of various α proteasome subunits from archaea to mammals shows high homology.

The bovine and porcine proteases which belong to the calpain or C2 family of cysteine proteases differ from each other in only six amino acid residues and thus show almost 99% homology to highly conserved calcium-binding domains and the N-terminal glycine-rich hydrophobic region. The region rich in proline residues (aa 76 to 81, numbered as in the Thermoplasma protease) is also conserved except at position 79, where proline is replaced by valine.

Tryptase precursors from humans and dogs (301), which belong to the S1 or trypsin family of serine proteases, show 76% sequence identity. The signal sequence from residues 1 to 30 is 60% identical; the difference is only in the site of glycosylation, which is Asn132 in the canine sequence and Asn233 in the human sequence. The sequences for active-site and disulfide bond formation are highly conserved and correspond to those of chymotrypsinogen (302).
The relationship between these neutral proteases is evident from the dendrogram shown in Fig. 5b.

**Alkaline Proteases**

The alkaline proteases selected here are active in the pH range of 8 to 13 and are about 420 to 480 residues in length. Six of them belong to the S8 or subtilase family of serine proteases (Table 6). They are aligned in their phylogenetic order, as shown in Fig. 4C. Considerable homology within the same genus is observed for *Bacillus* and *Aspergillus* proteases and three other fungal proteases. However, these proteases show comparatively lower homology among themselves. The active-site residues, as well as the residues around the active site, are highly conserved, suggesting that they may have evolved from a common ancestor. The sequences of *E. coli* and *Cyprinus carpio* seem to be homologous to some extent, but they do not have common active-site residues and therefore do not have a common ancestor. These two, in turn, show no significant homology to the other seven alkaline proteases. The overall homology among all these sequences can be represented by the dendrogram in Fig. 5c.

The results of our analysis of the amino acid sequences of the acidic, neutral, and alkaline proteases indicate that the members of the pepsin family of acidic proteases may have evolved from a common ancestor by convergent evolution. High homology between the sequences of the α subunits of proteasomes provides evidence for the presence of an evolutionarily conserved gene family. No amino acid residues are conserved in all the acidic or neutral proteases, except glycine. The alkaline serine proteases seem to have evolved from a
common ancestor by divergent evolution. In general, the sequences belonging to the same family show more homology or are more closely related. This criterion is currently used to assign a particular sequence to a particular family, i.e., the serine protease, cysteine protease, aspartate protease, or metalloprotease family. Within a family, the extent of homology is inversely proportional to the phylogenetic distance. The proteases from distantly related organisms show less homology or more diversity. However, this needs extensive sequence analysis of proteases, since the homology depends on many parameters or factors such as structure, function, source, and nature of the catalytic or active site. Thus, proteases are highly diverse enzymes having different active sites and metal-binding regions. The residues involved in disulfide bond formation and their positions, which vary in different proteases, can be detected by multiple alignments.

FIG. 4—Continued.
FIG. 4—Continued.
CURRENT PROBLEMS AND POTENTIAL SOLUTIONS

Proteases are a complex group of enzymes which differ in their properties such as substrate specificity, active site, and catalytic mechanism. Their exquisite specificities provide a basis for their numerous physiological and commercial applications. Despite the extensive research on several aspects of proteases from ancient times, there are several gaps in our knowledge of these enzymes and there is tremendous scope for improving their properties to suit projected applications. The future lines of development would include (i) genetic approaches to generate microbial strains for hyperproduction of the enzymes, (ii) application of SDM to design proteases with unique specificity and increased resistance to heat and alkaline pH, (iii) synthesis of peptides (synzymes) to mimic proteases, (iv) production of abzymes (catalytic antibodies) with specific protease activity, and (v) understanding of the structure-function relationship of the enzymes. Although the section on protein engineering describes in detail how SDM has been used to alter the properties and functions of proteases of bacterial, fungal, and viral origins, some of the important problems faced in their desired usages and the possible solutions to overcome these hurdles are discussed below.

Enhancement of Thermostability

The industrial use of proteases in detergents or for leather processing requires that the enzyme be stable at higher temperatures. One of the common strategies to enhance the thermostability of the enzyme is to introduce disulfide bonds into the protease by SDM. Introduction of a disulfide bond into subtilisin E from *Bacillus subtilis* resulted in an increase of 4.5°C in the $T_{m}$ of the mutant enzyme without causing any change in its catalytic efficiency (280). However, the properties of the mutant enzyme were found to revert to those of the wild-type enzyme. Enhanced stability of subtilisin was observed as a result of mutations of Asn109 and Asn218 to Ser. The analog containing both the mutations showed an additive effect on thermal stability. Thermostability of the alkaline protease from *Aspergillus oryzae* is important because of its extensive use in the manufacture of soy sauce. The optimal temperature of the wild-type enzyme was enhanced from 51 to 56°C by the introduction of a disulfide (Cys 169-Cys 200) bond. Another strategy for improving the stability of the protease was by replacing the polar amino acid groups by hydrophobic groups. The presence of positively charged amino acids in the N-terminal turn of an $\alpha$-helix is undesirable in view of the possibility of an occurrence of the repulsive interactions with the helix dipole. Replacement of Lys by Ser or Asp resulted in an increase in the thermostability of the neutral protease from *B. subtilis* in the range of 0.3 to 1.2°C (54). Although these approaches result in an increased stability of proteases, the difference in the thermostabilities of the parent and the mutant enzymes is only marginal, and further research involving cassette mutagenesis, etc., is necessary to yield an enzyme with substantially enhanced thermostability.
**Prevention of Autoproteolytic Inactivation**

Subtilisin, an extensively studied protease, is widely used in detergent formulations due to its stability at alkaline pH. However, its autolytic digestion presents a major problem for its use in industry. It was deduced that there is a correlation between the autolytic and conformational stabilities. Computer modeling followed by introduction of a Cys24 or Cys87 mutation resulted in destabilization of subtilisin from *Bacillus amyloliquefaciens* (312). Introduction of a disulfide bond increased the stability of the mutant to a level less than or equal to that of the wild-type enzyme. It appears logical that mutations in the gene coding for the enzyme with an alkaline pH optimum, whereas the use in the cheesemaking industry requires an acidic protease. Protein engineering enables us to tailor the pH dependence of the enzyme catalysis to optimize the industrial processes. Modifications in the overall surface charge of the proteins are known by systematic multiple mutations on the surface of the protein to alter the optimal pH of the enzyme. A change of Asp99 to Lys in subtilisin from *B. subtilis* resulted in destabilization of the enzyme. The properties needed for industrial applications of proteases differ from their physiological properties. The natural substrates of the enzyme are usually different from those desired for their industrial applications. Despite extensive research on proteases, relatively little is known about the factors that control their specificities toward nonphysiological sub-
strates. Strategies involving SDM are being explored to tailor these specificities at will. A combinatorial random-mutagenesis approach has been used to generate mutants that secrete proteases with functional properties different from those of the parent enzyme (77). Introduction of several point mutations into the substrate-binding site of a lytic protease was shown to affect its specificity in a predictable manner. The protease preferentially cleaves on the C-terminal side of small uncharged residues such as Ala, mainly because the pocket that accommodates the substrate P-1 residue is shallow due to the presence of two bulky methionine residues (Met190 and Met213) at the subsite. Replacement of Met213 with a His residue had a beneficial effect on its substrate specificity.

Improvement of Yield

The cost of enzyme production is a major obstacle in the successful application of proteases in industry. Protease yields have been improved by screening for hyperproducing strains and/or by optimization of the fermentation medium. Strain improvement by either conventional mutagenesis or recombinant-DNA technology have been useful in improving the production of proteases. Hyperexpression by genetic manipulation of microbes is described in the section on genetic engineering. Increases in the yield of viral proteases are particularly important for developing therapeutic agents against devastating diseases such as cancer, malaria, and AIDS. Advances in genetic manipulation of microorganisms by SDM of the cloned gene opens new possibilities for the introduction of predesigned changes, resulting in the production of tailor-made proteases with novel and desirable properties. The development of recombinant rennin and its commercialization by Pfizer and Genencor is an excellent example of the successful application of modern biology to biotechnology. The advent of techniques for rapid sequencing of cloned DNA has yielded an explosive increase in protease sequence information. Analysis of sequences for acidic, alkaline, and neutral proteases has provided new insights into the evolutionary relationships of proteases.

Despite the systematic application of recombinant technology and protein engineering to alter the properties of proteases, it has not been possible to obtain microbial proteases that are ideal for their biotechnological applications. Industrial applications of proteases have posed several problems and challenges for their further improvements. The biodiversity represents an invaluable resource for biotechnological innovations and plays an important role in the search for improved strains of microorganisms used in the industry. A recent trend has involved conducting industrial reactions with enzymes reaped from exotic microorganisms that inhabit hot waters, freezing Arctic waters, saline waters, or extremely acidic or alkaline habitats. The proteases isolated from extremophilic organisms are likely to mimic some of the unnatural properties of the enzymes that are desirable for their commercial applications. Exploitation of biodiversity to provide microorganisms that produce proteases well suited for their diverse applications is considered to be one of the most promising future alternatives. Introduction of extremophilic proteases into industrial processes is hampered by the difficulties encountered in growing the extremophiles as laboratory cultures. Revolutionary robotic approaches such as DNA shuffling are being
ACKNOWLEDGMENTS

We thank M. C. Srivivasan, U. S. Phadtare, K. R. Bandivadekar, S. H. Bhosale, and D. Nath for providing some of the literature information. We are grateful to A. S. Kolaskar, P. B. Vidyasagar, and S. Jagtap, Bioinformatics Centre, University of Pune, for their help in analyzing the protease sequences. Financial support to M. S. Ghatge and A. M. Tanksale from the Council of Scientific and Industrial Research is gratefully acknowledged.

REFERENCES


