8
Production of Active Mammalian and Viral Proteases in Bacterial Expression Systems

LILIA M. BABÉ[1]*, CHRISTOPHER J. LINNEVERS[1] AND BRIAN F. SCHMIDT[2]

[1]*Axys Pharmaceuticals Inc., 180 Kimball Way, South San Francisco, CA 94080, U.S.A. and* [2]*Coulter Pharmaceuticals Inc., South San Francisco, CA, U.S.A.*

Introduction

Mammalian endopeptidases and exopeptidases participate in a wide variety of cellular processes. They are responsible for the relatively non-specific degradation of proteins targeted for digestion or recycling, and they also perform highly specific single-site cleavages necessary for the activation or inactivation of functional proteins and peptides. Likewise, numerous viruses that infect mammalian cells utilize virus-encoded proteases to regulate their replication cycle. Mammalian proteases are expressed as enzymatically inactive zymogens requiring specific co- or post-translational processing by self or other proteases. Virus-encoded proteases are expressed as part of viral polyproteins that also require specific autoprocessing to release the fully active protease. Thus, the same theme is used, where structural motifs prevent the enzyme from being active before the appropriate time and place, and catalytic proficiency is regulated by the formation of the active protease (Babé and Craik, 1997). This theme must be kept in mind when designing heterologous expression systems for mammalian and viral proteases to ensure the production of active or activatable enzymes.

Advances in the study of proteases in the past two decades have been largely dependent on the ability of researchers to produce significant quantities of pure enzymes. Generally, recombinant gene expression systems have been used to accomplish this task, especially for proteases that are naturally produced in very limited amounts. Heterologous expression systems also have the advantage of being able to produce variant proteases at will, allowing the study of structure-function relationships and modifications of their properties.

In addition to basic research, the production of recombinant proteases has been crucial to the development of commercial products. For example, recombinant bovine chymosin, an aspartic protease, is used in the manufacture of cheese, while

*To whom correspondence may be addressed (lilia_babe@axyspharm.com)

Biotechnology and Genetic Engineering Reviews – Vol. 17, August 2000
0264–8725/00/17/213–252 $20.00 + $0.00 © Intercept Ltd, P.O. Box 716, Andover, Hampshire SP10 1YG, U.K.

human tissue-type plasminogen activator, a serine protease, is sold as a therapeutic agent for the dissolution of blood clots following a heart attack. The production of human and viral recombinant proteases has also been critical for the development of new therapeutic protease inhibitors. The ability to screen extremely large numbers of low molecular weight compounds for potent and selective protease inhibitors is dependent on the ability to produce large quantities of pure enzyme. Two examples of commercially successful protease inhibitor therapeutics are angiotensin-converting enzyme (ACE) inhibitors for proper blood pressure regulation, and the now familiar human immunodeficiency virus (HIV) protease inhibitors for the control of the acquired immune deficiency syndrome (AIDS).

There are many different eukaryotic and prokaryotic expression systems that can be applied to the production of enzymatically active proteases. The list includes: mammalian cell based systems (Chinese hamster ovary cells, COS monkey cells, 293 human embryonic kidney cells, or 3T3-NIH mouse fibroblast cells), insect cell based systems (baculovirus or Drosophila S2 cells), yeast systems (*Saccharomyces cerevisiae* or *Pichia pastoris*), and numerous bacterial systems. Many of the reagents and strains needed to use these systems are now commercially available (for a practical guide to protease production see Brömme and Schmidt, 1999).

This review will focus only on the use of bacterial systems for the expression of mammalian and viral protease genes. *Table 8.1* provides examples of *E. coli* expression systems successfully applied to express members of all four major classes of mammalian protease. *Table 8.2* provides a list of viral proteases from numerous virus families, which have also been successfully expressed in *E. coli*. Generally, bacterial expression systems are preferred, due to their ease of use and inexpensive fermentation and scale-up costs. The molecular genetics and physiology of *Escherichia coli*, the bacteria most often used for heterologous gene expression, are arguably better understood than any other organism. There is also an abundance of different *E. coli* expression systems that are commercially available. Of the eukaryotic systems, yeast can be grown relatively easily and inexpensively, but it still cannot typically compete with bacteria in terms of rapid doubling times and robust growth. However, bacteria usually cannot be used for the production of recombinant proteases that require post-translational modifications such as glycosylation and phosphorylation for activity. In general, large complex proteases with multiple sub-units or domains, especially ones that have a large number of disulphide cross-links, are not good candidates for production in bacteria. Since expression is still as much an art as a science, one cannot predict *a priori* what system may or may not work for any given protease. But, as listed in this review, mammalian and viral proteases of nearly every type have been successfully produced in bacteria. Since bacterial systems are quick, easy, and inexpensive to use, they are usually the first system to be tried for the production of a novel protease.

Expression systems

GENERAL PROPERTIES

The general components of a bacterial expression cassette for the production of a 'typical' protease are shown in *Figure 8.1*. Expression vectors with selectable

Table 8.1. Examples of mammalian proteases expressed in *E. coli*

Class	Protease	Promoter and Gene	Localization and Form	Yield	Activation
aspartic	bovine chymosin[1]	*tac* 12 residue linker-prochymosin	cytoplasm insoluble	12 mg/L	refolded acid activation
	porcine pepsin[2]	*T7* pepsinogen	cytoplasm insoluble	50 mg/L	refolded acid activation
	human cathepsin D[3]	*T7* procathepsin D-MBP- $(His)_6$	insoluble inclusion bodies	10 mg/L	acid activation
metallo	porcine collagenase 1 (MMP-1)[4]	*tac* GST-proMMP1	insoluble inclusion bodies	0.8 mg/L	organomercurial or trypsin
	human gelatinase B (MMP-9)[5]	*T7* OmpT leader-proMMP9	periplasm soluble	5 mg/L	trypsin
	human stromelysin-1 (MMP-3)[6]	*T7* proMMP3	cytoplasm soluble	8 mg/L	organomercurial or trypsin
serine	rat trypsin[7]	*tac* hisJ leader-mature trypsin	periplasm soluble	56 mg/L[8]	signal peptidase
	human t-plasminogen activator[9]	*BAD* stII leader-DsbC-protPA	periplasm soluble	0.2 mg/L	auto-activation
	human tissue kallikrein[10]	λP_L synthetic prokallikrein	cytoplasm insoluble	20 mg/L (purified)	thermolysin
	granzyme K[11]	*aprE* aprE leader-mature granzyme K	medium soluble	0.2 mg/L (purified)	signal peptidase
cysteine	human cathepsin S[12]	*T7* procathepsin S	cytoplasm insoluble	100 mg/L	auto-activation at low pH
	human caspace-1[13]	λP_L $(His)_6$-truncated procaspase 1	cytoplasm soluble	3 mg/L	auto-activation

References: (1) Kapralek *et al.*, 1990; (2) Lin *et al.*, 1989; (3) Sachdev and Chirgwin, 1998; (4) O'Hare *et al.*, 1995; (5) Pourmotabbed *et al.*, 1994; (6) Marcy *et al.*, 1991; (7) Higaki *et al.*, 1989; (8) Yee and Blanch, 1993; (9) Qiu *et al.*, 1998; (10) Lu *et al.*, 1996; (11) Babé *et al.*, 1998; (12) Kopitar *et al.*, 1996; (13) Dang *et al.*, 1996.

Table 8.2. Examples of viral proteases successfully expressed in *E. coli*.

Family	Virus name	Protease	Structure and mechanism	Special attributes
herpesviridae	herpesvirus simplex-1[1]	HSV-1 PR	novel fold, Ser nucleophile	allosteric homodimer
	cytomegalovirus[2]	CMV PR	novel fold, Ser nucleophile	allosteric homodimer
	human herpesvirus-8[3]	HHV-8 PR	novel fold, Ser nucleophile	allosteric homodimer
flaviviridae	dengue[4]	NS3	chymotrypsin-like, Ser nucleophile	peptide co-factors, tethered to helicase
	hepatitis C[5]	NS3	chymotrypsin-like, Ser nucleophile	peptide co-factors, tethered to helicase
picornaviridae	rhinovirus[6]	3C	chymotrypsin-like, Cys nucleophile	tethered to RNA polymerase
	hepatitis A[7]	3C	chymotrypsin-like, Cys nucleophile	tethered to RNA polymerase
	poliovirus[8]	3C	chymotrypsin-like, Cys nucleophile	tethered to RNA polymerase
	mengo[9]	3C	chymotrypsin-like, Cys nucleophile	tethered to RNA polymerase
	rhinovirus[10]	2A	unknown structure, Cys nucleophile	–
	polio virus[11]	2A	unknown structure, Cys nucleophile	–
	foot and mouth virus[12]	L PR	unknown structure, Cys nucleophile	–
retroviridae	HIV-1[13]	HIV-1 PR	aspartyl protease	obligate homodimer
	HIV-2[14]	HIV-2 PR	aspartyl protease	obligate homodimer
	SIV[15]	SIV PR	aspartyl protease	obligate homodimer
	HTLV-1[16]	HTLV-1 PR	aspartyl protease	obligate homodimer
	human foamy virus[17]	HFV PR	aspartyl protease	obligate homodimer
adenoviridae	adenovirus-2[18]	Ad2 PR	novel fold, Cys nucleophile	peptide co-factor

References: (1) Darke *et al.*, 1994; (2) LaFemina *et al.*, 1996; (3) Unal *et al.*, 1997; (4) Murthy *et al.*, 1999; (5) Vishnuvardhan *et al.*, 1997; (6) Birch *et al.*, 1995; (7) Malcolm *et al.*, 1992; (8) Baum *et al.*, 1991; (9) Hall and Palmenberg, 1996; (10) Wang *et al.*, 1998; (11) Alvey *et al.*, 1991; (12) Piconne *et al*, 1995; (13) Debouck *et al.*, 1987; (14) Salto *et al.*, 1994; (15) Grant *et al.*, 1991; (16) Ding *et al.*, 1998; (17) Pfrepper *et al.*, 1997; (18) Rancourt *et al.*, 1994.

markers, and transformation or transfection protocols for the introduction of DNA into the bacterium, must be readily available for the expression host of choice. In cases where plasmid instability or the cost of the antibiotic for selection is prohibitive for large-scale fermentations, gene stability can be achieved by integrating the entire expression cassette into the chromosome. Otherwise, simple plasmid-based systems are usually chosen for expression studies.

For any expression system to be successful, the following requisites must be met: large amounts of mRNA need to be synthesized, the mRNA should be stable and translated at high efficiency, the protein should fold correctly (or be amenable to refolding), and the product must be resistant to proteolysis (or produced in a protease-deficient host). Synthesizing large amounts of mRNA can be accomplished by placing a strong promoter upstream of the gene of interest and a strong terminator at the end of the gene. The promoter should carry an operator so that induction of expression can be controlled. High expression levels can typically slow the growth of the bacteria, or even be toxic, so it is desirable to induce expression only after the cells have grown to a high density. Strong inducible promoters have been characterized for commonly used expression hosts and are readily available from commercial molecular biology vendors.

Generally, the mRNA that is synthesized consists of a 5' untranslated region followed by a ribosomal binding site, a methionine initiation codon, the structural protease gene of interest, a stop codon, and a 3' untranslated region (*Figure 8.1*). The mRNA stability and the translational efficiency both generally depend on the sequence encoded in these regions. For example, a high translational rate may increase mRNA stability since the ribosomes may block nucleases from degrading the mRNA. Alternatively, changing the sequence to improve the ribosomal binding site may alter the mRNA secondary structure in such a way to make it more sensitive to degradation. It is not easy to predict how changes in the mRNA sequence will affect its stability. However, in general, secondary structures such as hairpin loops located at the 5' and 3' ends of the mRNA will increase stability.

Translational efficiency can be improved by using a strong ribosomal binding site that is not buried in mRNA secondary structure, using the best initiator codon (usually AUG rather than GUG), a gene relatively free of contiguous rare codons, and a good stop codon sequence (UAAU in *E. coli*, for example). In practice, one usually starts implementing the factors outlined above for efficient translation and then uses the native sequence for the gene of interest. If this fails, introduction of optimal codons or attempts to improve mRNA stability should be considered. Other effective ways to improve mRNA stability and/or translational efficiency include fusing the protein of interest to a highly expressed protein, or performing translational coupling of the gene of interest to a highly transcribed gene. The latter system involves making a two-gene cistron that delivers a steady stream of ribosomes to the translational start of the gene of interest.

An additional consideration is whether to produce soluble or insoluble material. Mammalian proteases can be produced in the bacterial cytoplasm, secreted into the culture medium or, in the case of gram-negative bacteria, exported into the periplasm. In most cases, recombinant proteins are targeted to the cytoplasm when the production of insoluble inclusion bodies is desired. In general, inclusion body formation simplifies purification and reduces the proteolytic degradation of the recombinant

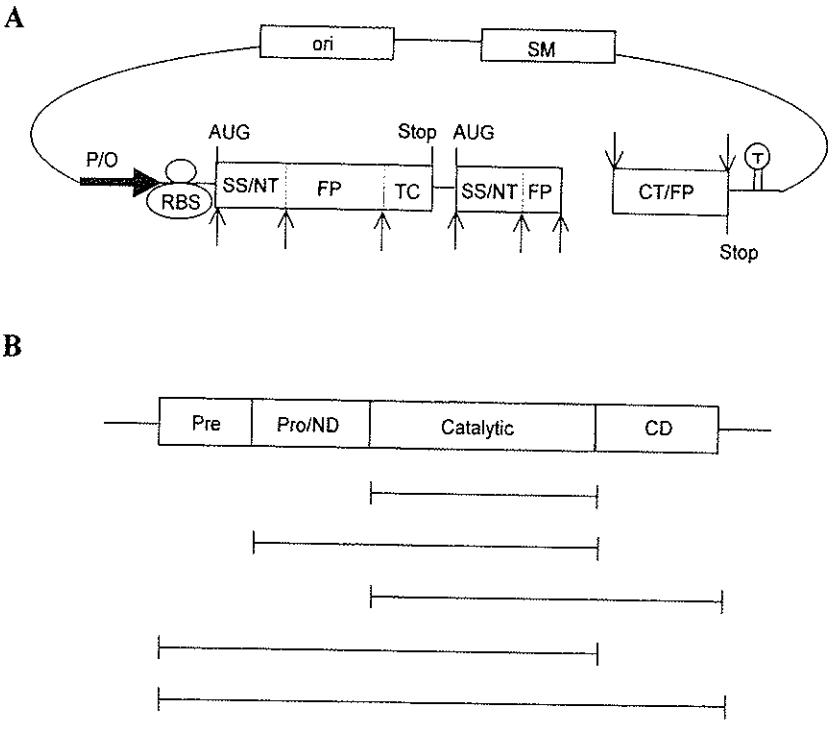


Figure 8.1. Schematic diagrams of the general components needed for the expression of a mammalian protease in bacteria. A) A schematic representation of a typical expression plasmid used to produce recombinant mammalian proteases. An origin of plasmid replication (ori) and a selectable marker (SM) are needed for maintenance of the plasmid in the bacterium. A strong promoter and operator (P/O) with a terminator (T) ensures high inducible levels of mRNA, and a consensus ribosomal binding site (RBS) promotes translational initiation. Upward pointing arrows denote locations where the 5' end of the protease gene may be fused to the expression plasmid, and downward pointing arrows are possible sites for fusion of the 3' end of the protease gene to the vector. The expression vector may be designed so that the produced protease is fused to a bacterial signal sequence or N-terminal tag (SS/NT) for secretion/export or ease of purification/identification, respectively. The protease may also be fused to a highly expressed or solubilizing protein (FP), or translationally coupled (TC) to a highly expressed gene. A fusion protein (FP) or C-terminal tag (CT) can be attached to the C-terminal end of the protease to aid purification or increase solubility. B) Schematic representation of a typical protease gene. The protease may have a pre-region (Pre) to direct secretion into the endoplasmic reticulum, a pro-peptide and/or N-terminal domain(s) (Pro/ND), the catalytic domain, and a C-terminal domain(s) (CD). Depending on the requirements of the experimentalist, various domains may be eliminated from the catalytic domain when the gene is inserted in the expression vector (bars shown below the schematic protease gene represent regions that may be inserted into the expression plasmid).

protease; however, renaturation of the protease to achieve native specific activity may be difficult. Soluble production of recombinant protease in the cytoplasm can be enhanced by using mutant hosts that allow disulphide bond formation (eg *trx*B in *E. coli*), co-expression of chaperones, or by fusing the protease of interest to a soluble protein. In most cases, soluble proteases may be more easily produced by exporting them to the periplasm or by secreting them into the culture medium. In these cases, it is usually desirable to use bacterial signal sequences to target the mammalian protease for export and secretion. No matter what compartment is chosen for production, the use of protease-deficient bacterial strains can be used to reduce proteolytic degradation of the recombinant protein.

A schematic diagram of a generalized protease gene is shown in *Figure 8.1*. Most proteases of interest have a native signal sequence, an N-terminal pro-peptide that must be cleaved to activate the enzyme, and a mature catalytic domain. For some mammalian serine proteases, the N-terminal pro-region has evolved into a large domain(s) that remains linked to the catalytic domain even after activation. Other proteases have one or more large C-terminal domain(s). Since it is often difficult to express large multi-domain proteins in bacteria, one may choose to express truncated versions of the protease gene that lack some or all of the N- or C-terminal domains. Also, the pro-region may not need to be included in the expression cassette if it is not important for the proper folding of the protease.

GRAM-NEGATIVE BACTERIAL SYSTEMS

As mentioned above, the gram-negative bacteria, *Escherichia coli*, is by far the most utilized vehicle for the production of mammalian proteases since the physiology and genetics of this organism are better characterized than any other bacterium. A detailed dissertation about expressing heterologous genes in *E.coli* will not be presented here since a number of excellent reviews have been published on this topic in the past several years (Georgiou, 1996; Hanning and Makrides, 1998; Makrides, 1996; Rosé and Craik, 1996; Weickert *et al.*, 1996). However, specific *E. coli* expression systems that have been used successfully for the production of each class of mammalian and viral protease will be outlined in this review. To help the researcher get started producing these proteases in *E. coli*, many different expression systems are now sold commercially and, in some cases, vectors, hosts, purification products, and detailed protocols are provided in a single kit.

Initially, it is wise to choose an *E.coli* system that has worked for the production of the most closely related proteases to the one of interest. The construction of the expression system and the selection of the *E. coli* host will depend on whether the protease will be targeted to the cytoplasm, exported to the periplasm, or secreted into the culture medium. Keep in mind that production in the periplasm or culture medium is usually desirable if a soluble product is needed. However, as mentioned earlier, several approaches can be tried to produce soluble proteins in the cytoplasm. Examples include: fusing the protease to another protein (LaVallie and McCoy, 1995; Ståhl *et al.*, 1997), co-expressing with a chaperone (Wall and Plückthun, 1995; Guise *et al*, 1996), using a thioredoxin deficient strain (to allow disulphide bond formation in the cytoplasm), or altering the bacterial growth conditions (Georgiou and Valax, 1996; Hockney, 1994; Zhang *et al*, 1998).

Ubiquitin fusion technology offers a new method for producing large quantities of good quality recombinant proteins. The expression systems consist of a suitable *E. coli* host strain paired with a plasmid that encodes the ubiquitin fusion, and an ubiquitin-specific protease, UCH-L3, which cleaves only C-terminal extensions from ubiquitin (Pilon *et al.*, 1997). Not only is overexpression achieved (multigram quantities of recombinant product from a 10 litre fermentor are common), but the chaperone-like function of the ubiquitin moiety ensures proper folding and bioactivity of the products.

For soluble expression, the best choices are titratable promoters, such as the *lac* or *lac* hybrid promoters, that are inducible by IPTG (Schweizer and Karkhoff-Schweizer,

1997) and can be induced at different temperatures. Recombinant proteins can be easily targeted to the periplasm by fusing a standard gram-negative signal sequence to the protease of choice (Pines and Inouye, 1997). The recombinant protease may leak into the medium once produced in the periplasm, or will be directly secreted into the medium if one uses a cell wall-less L-form of *E. coli* (Gumpert and Hoischen, 1998). Strategies for the active secretion of proteins through the periplasm have been developed (Blight and Holland, 1994; Sandkvist and Bagdasarian, 1996) but, as yet, not effectively utilized for recombinant mammalian or viral protease production. Proteolytic degradation by endogenous *E.coli* proteases can be a problem when producing soluble recombinant proteins in the cytoplasm or periplasm. Strategies such as the use of protease deficient strains (Meerman and Georgiou, 1994), or the production of proteins as insoluble inclusion bodies in the cytoplasm, have been helpful in reducing recombinant protein degradation (reviewed by Murby *et al.*, 1996). Refolding of material extracted from inclusion bodies has successfully produced active mammalian proteases (see below), and various renaturation protocols have been developed (Guise *et al.*, 1996; Lilie *et al.*, 1998). For the production of large amounts of insoluble protein, the use of a strong promoter system, such as the bacteriophage T7 promoter (Studier *et al.*, 1990), is usually advisable.

Although other gram-negative bacteria have been used as hosts for recombinant protein production (Billman-Jacobe, 1996), to our knowledge, only the cell wall-less L-form of *Proteus mirabilis* has been used successfully to produce a mammalian protease (Klessen *et al.*, 1989).

GRAM-POSITIVE BACTERIAL SYSTEMS

Gram-positive bacteria, most notably *Bacillus subtilis*, have been used very successfully for the production of recombinant enzymes used in industrial applications (Ferrari *et al.*, 1993; Harwood, 1992). However, to this date, the majority of the recombinant proteins produced commercially by gram-positive bacteria are of bacterial origin. Thus, these organisms have had little impact in the commercial (or even research) production of mammalian proteins when compared to the gram-negative bacteria, *E. coli*. Still, gram-positive bacteria are attractive hosts, since recombinant proteins can be secreted directly into the culture medium by virtue of having a single cell membrane, instead of having both a cytoplasmic and an outer membrane like gram-negative bacteria. They can also produce high titres of extracellular products in fermentors, where yields of > 1 gram/litre are not uncommon. In the case of *B. subtilis*, many reagents are available for heterologous gene expression (Wang and Doi, 1992). These include: a range of plasmid vectors (Perego, 1993; Jannière *et al.*, 1993), promoters (Conrad *et al.*, 1996; Henner, 1990; Lam *et al.*, 1998, Le Grice, 1990; Nagarajan, 1990), signal sequences (Nagarajan, 1993), and transformation/transduction protocols (Hoch, 1991). Unfortunately, *B. subtilis* produces many intracellular and extracellular proteases (Pero and Sloma, 1993) which can degrade recombinant products, and also the bacterium has difficulty secreting large multi-domain eukaryotic proteins (Bron *et al.*, 1998), especially ones with multiple disulphide bonds. Thus, the construction of protease deficient strains, and studies characterizing the secretion apparatus of *B. subtilis*, have been the focus of extensive research groups (Wong, 1995).

Other members of the *Bacillus* genus, especially ones such as *Bacillus brevis* (Udaka and Yamagata, 1993), that are naturally deficient in extracellular proteases, have been tested for their ability to produce heterologous proteins. *Corynebacterium glutamicum* was used to express and secrete the bacterial protease subtilisin (Billman-Jacobe *et al.*, 1995). Other gram-positive bacteria like *Streptomyces lividans* (Binnie *et al.*, 1997) and lactic acid bacteria (Chassy and Murphy, 1993) have been used to produce recombinant mammalian proteins. But to date, *B. subtilis* has been the major gram-positive bacteria used to express heterologous genes (Billman-Jacobe, 1996).

Specific expression systems for mammalian proteases

ASPARTIC PROTEASES

Aspartic proteases are characterized by having two aspartate residues in the active site cleft of the enzyme that aid in the nucleophilic attack of the peptide bond by an activated water molecule. Mammalian proteases of this type generally belong to the pepsin family of aspartic proteolytic enzymes, and include the gastric proteases such as pepsin, the lysosomal protease cathepsin D, and processing enzymes like renin. Production of aspartic proteases in bacterial systems has been useful both for the development of commercial products and for structural studies of aspartic proteases in general.

The first mammalian protease gene expressed in bacteria, and even one of the first active mammalian enzymes purified from *E. coli*, was bovine chymosin (renin). Pro-chymosin is secreted into the fourth stomach of the non-weaned calf where it is acid-activated to chymosin by the removal of the 42 amino acid N-terminal pro-peptide. Chymosin specifically cleaves the milk protein β-casein, inducing the coagulation of milk that apparently aids milk digestion in the calf (Foltman and Szecsi, 1998). For commercial applications, chymosin is used as a milk-clotting agent in the manufacture of cheese. Due to the limited supply of chymosin from natural sources, this protease became an early favourite of the biotechnology industry for commercial recombinant production (Yu, 1994). Since the early 1980s, many groups have worked on improving chymosin production in bacteria. Initial studies used the *lac*UV5 or the *trp* promoter to produce insoluble chymosin in inclusion bodies (Nishimori *et al.*, 1982; Emtage *et al.*, 1983). The expressed pro-chymosin was extracted from the inclusion bodies by urea, refolded by dialysis, and activated by acid treatment (Emtage *et al.*, 1983; McCaman *et al.*, 1985; Nishimori *et al.*, 1984). In another system, insoluble chymosin was produced in fermentors to a level of 5% of the total protein (Marston *et al.*, 1984). Eventually, by using the *tac* promoter, enzymatically active chymosin could be produced at 12 mg per litre of culture broth (Kaprálek *et al.*, 1990). The expression of chymosin has also been used as a model system in the development of more efficient ribosomal binding sites (Emtage *et al.*, 1983; Kawaguchi *et al.*, 1986; Wang *et al.*, 1995a). In addition, active chymosin has been produced in the culture medium by secretion from protoplast type L-forms of *Proteus mirabilis* at 17 mg/litre by using the streptococcal pyrogenic exotoxin type A promoter and signal sequence fused to the pro-chymosin gene (Klessen *et al.*, 1989). Soluble pro-chymosin has also been secreted into the medium from *B. subtilis* by using the neutral protease promoter and subtilisin signal sequence; however, the

yields were quite low (100 μg/litre of culture) (Parente *et al.*, 1991). It should be noted that pro-chymosin was produced in the lactic acid bacteria, *Lactococcus lactis*, but could not be activated (Simons *et al.*, 1991).

The zymogen form of pepsin, pepsinogen, is secreted into the gastric lumen. The acidic gastric environment promotes autocatalysis to yield the active pepsin, which then aids in the degradation of proteins during digestion. In humans, several different pepsinogen genes give rise to a number of different isozymes (such as pepsinogen A or C-type aspartic proteases) (Tang, 1998). Similar to pro-chymosin, pepsinogen has been produced in *E. coli* as insoluble material that can be refolded and activated to produce pepsin. Activatable porcine pepsin was efficiently expressed without a signal sequence. Levels of up to 50 mg/litre were achieved by using the *lac* or *tac* promoters (Lin *et al.*, 1989; Tsukagoshi *et al.*, 1988) and up to 500 mg/litre with a T7 promoter (~ 10% could be refolded and activated) (Lin *et al.*, 1994). An active site mutant of pepsinogen, D32A, produced in *E. coli* could not be activated to pepsin (Lin *et al.*, 1989). Acid-activatable soluble porcine pepsinogen was produced at 6 mg/litre as a thioredoxin fusion protein when the cultures were grown at 22°C, while mostly insoluble material was produced at 30°C (Tanaka and Yada, 1996). However, human pepsinogens were produced as insoluble aggregates which could be activated after refolding when expressed fused to either the N- or C-termini of the maltose-binding protein (Aoki *et al.*, 1998; Sachdev and Chirgwin, 1998). Soluble and activatable porcine pepsinogen was secreted by *Bacillus brevis* into the culture medium at levels of about 5 mg/litre by using the middle wall protein promoter and a signal sequence (Takao *et al.*, 1989).

Cathepsin D is a major lysosomal protease that is involved in the general intracellular degradation of proteins (Conner, 1998). It is present in lysosomes in its enzymatically active form. Human procathepsin D was produced in *E. coli* as insoluble inclusion bodies, but the enzyme could be extracted, refolded, activated (Conner and Udey, 1990), and further purified using an inhibitor affinity column (Conner and Richo, 1992). Unfortunately, unlike pepsin and chymosin, only about 1% of the pepsin could be refolded into its native state, and the majority remained as insoluble aggregates (Scarborough and Dunn, 1994). Further attempts to increase the soluble fraction by fusing maltose-binding protein or thioredoxin to the C-terminus of procathepsin D did increase the solubility of the refolded material, but less than 2% could be activated (Sachdev and Chirgwin, 1998).

Unlike the ubiquitous cathepsin D, cathepsin E is present only in a limited number of cells and tissues (Kay and Tatnell, 1998). It is found in non-lysosomal endosomal compartments that are generally associated with the endoplasmic reticulum, and the active enzyme is isolated as a homodimer linked by a disulphide bridge near the N-terminus. Using a T7 promoter system, insoluble procathepsin E was produced in *E. coli* (Hill *et al.*, 1993). The material could be extracted with urea, refolded by dilution, and autoactivated by acidification to produce the native dimer, even though several different N-termini are formed (Tatnell *et al.*, 1997). A monomeric mutant, C4A, was produced in *E. coli*, refolded, and found to have a similar specific activity as the native dimer (Fowler *et al.*, 1995).

The medically relevant protease renin cleaves angiotensinogen to produce angiotensin, which is a vasoactive peptide involved in blood pressure regulation and electrolyte balance (Suzuki *et al.*, 1998a). The pro-renin gene was expressed in *E.*

coli, fused to the *trp*E gene and produced in insoluble inclusion bodies (Imai *et al.*, 1986). This protease could be extracted, refolded and activated with trypsin but less than 1% of the material could be recovered.

CYSTEINE PROTEASES

The first cysteine proteases isolated from mammals were the cathepsins, which are closely related to the plant protease papain (Rawlings, 1998). These include the major lysosomal proteases such as cathepsin B, L, and S. The cysteine-type peptidases use the sulphydryl group of a Cys residue as the nucleophile. Their catalytic mechanism is similar to serine-type peptidases, where a nucleophile and a proton donor/general base are required. All cysteine proteases use a His as the proton donor (Storer and Ménard, 1994). Although there is evidence that a third residue (Asp or Glu) is required to orient the imidazolium ring of the histidine, His, there are a number of proteases in which only a catalytic dyad is necessary. The cathepsins are secreted into the endoplasmic reticulum as zymogens, modified, and transported to the lysosomes. The pro-peptide blocks the active site cleft until activation occurs, then the catalytic domain may be further processed by proteolytic cleavage. The cathepsins have been implicated in diseases related to abnormal matrix degradation such as arthritis, cancer and osteoporosis.

Using a T7 promoter system, procathepsin B was expressed in *E. coli*, the protein was solubilized in urea, refolded by dialysis, then activated with pepsin. Approximately 3 mg of active enzyme was recovered from cells grown in 1 litre of broth (Kuhelj *et al.*, 1995). Active cathepsin L was produced in *E. coli* using a similar system (Smith and Gottesman, 1989), and later with an improved guanidinium denaturation and refolding protocol (Dolinar *et al.*, 1995). Unfortunately, yields of the final active product were low (less than 0.2 mg per litre), even when the initial insoluble procathepsin L represented 5–10% of the total cellular protein. A similar *E. coli* system was used to generate the more novel cathepsin O (Velasco *et al.*, 1994), resulting in active protein after renaturation. Refolding and autoactivation at low pH of insoluble procathepsin S, produced in *E. coli,* was also accomplished by expressing the gene using a T7 promoter system (Kopitar *et al.*, 1996). Earlier systems used a glutathione S-transferase fusion to the mature cathepsin S gene, and purification on a glutathione affinity column produced soluble active cathepsin S (Petanceska and Devi, 1992). Recently, glutathione S-transferase fusions have also been used to produce soluble active cathepsin L2 (Santamaría *et al.*, 1998a), cathepsin Z (Santamaría *et al.*, 1998b), and cathepsin F (Santamaría *et al.*, 1999) in *E. coli*.

Caspases ('C' for cysteine protease, 'aspase' to denote specificity for aspartic acid residues) are the other major cysteine proteases found in mammals (Thornberry, 1998). These enzymes are not homologous to cathepsins, but do have an active site dyad of histidine and cysteine. Caspases are expressed as single chain polypeptides, and are converted to heterodimers by cleavage at four aspartyl bonds. This cleavage which yields the 22 kDa heavy chain and the 10 kDa light chain may be autocatalytic, or carried out by another caspase, or by granzyme B. Caspase-1 (interleukin-1β converting enzyme, ICE) cleaves pro-IL-1β to produce the active cytokine. Other caspases have been implicated in inducing apoptosis, most likely through a proteolytic cascade involving numerous caspases (Thornberry, 1998). Thus, due to their potential therapeu-

tic value and fairly recent discovery, intensive research has resulted in the identification of numerous caspase genes (ten human ones, thus far) and expression systems have followed. Bacterial systems have been very useful for the production of caspases since their overexpression in eukaryotic cells can induce apoptosis. Active soluble murine caspase-1 was produced in *E. coli* using a pro-region truncated version (Molineaux *et al.*, 1993). An autolysis-resistant mutant was expressed by replacement of the caspase-1 pro-region with an N-terminal poly-histidine to facilitate purification (Dang *et al.*, 1996). Soluble human caspase-1 was produced fused to TrxA, and some insoluble material was also present that could be refolded to active enzyme for a total yield > 1 mg per litre (Malinowski *et al.*, 1995). During refolding, caspase-1 is autocatalytically processed to the active heterodimer (Ramage *et al.*, 1995). Milligram quantities of caspase-1 were also prepared by producing each chain separately using the λp_L promoter. The two insoluble sub-units can be solubilized and then refolded together to yield active enzyme suitable for crystallization (Walker *et al.*, 1994a) and stability studies (Talanian *et al.*, 1996). A similar approach of expressing the sub-unit genes separately was successfully used for the production of caspase-3 (apopain/CPP32) (Rotonda *et al.*, 1996). Other caspases, such as: caspase-4 (ICH-2) (Kamens *et al.*, 1995), caspase-6 (Mch2) (Fernandes-Alnemri *et al.*, 1995a), caspase-7 (Mch3) (Fernandes-Alnemri *et al.*, 1995b); caspase-8 (Mch5) (Srinivasula *et al.*, 1996), caspase-9 (Mch6) (Stennicke *et al.*, 1999), and caspase-10 (Mch4) (Fernandes-Alnemri *et al.*, 1996; Srinivasula *et al.*, 1996), have been produced in *E. coli* as well. Some have been produced as soluble material either as a fusion to glutathione S-transferase, or with the pro-region deleted, while others have been produced as insoluble full-length or sub-unit aggregates that could be extracted and refolded to produce active enzyme. Fortunately, most caspases can be autoactivated after refolding.

Another cysteine protease not related to cathepsins or caspases, microsomal ER-60 protease, has also been successfully produced in *E. coli* (Urade *et al.*, 1997).

METALLOPROTEASES

Metalloproteases are characterized by having a metal ligand (usually zinc) coordinated to usually three amino acid side chains (such as His) on the protein. An additional coordinated water molecule is involved in the nucleophilic attack of the peptide bond, and there is normally another side chain (from a residue such as Glu) that is important for catalysis. Of the mammalian metalloproteases, the matrix metalloproteases (MMPs, matraxins) of the interstitial collagenase family have been the most intensively studied due to their potential involvement in diseases such as cancer and arthritis. In these proteases, the catalytic zinc is coordinated to three histidines that are present in the motif HEXXHXXGXXH (X can be any amino acid). In general, these MMPs consist of three domains, a pro-domain with a cysteine residue responsible for maintaining the latency of the enzyme by interacting with the catalytic zinc, a catalytic domain, and a hemopexin-like C-terminal domain that is most likely involved in substrate specificity. Expression of the pro-enzymes is tightly regulated by cytokines and growth factors, followed by secretion of the inactive pro-MMP in most instances. Activation ensues, and regulation of the mature enzymes is generally balanced by the locally produced TIPMs (tissue inhibitors of metalloproteases) (Nagase and Woessner, 1999).

The mammalian collagenases, interstitial collagenase (collagenase-1, fibroblast collagenase, MMP-1) (Cawston, 1998), and neutrophil collagenase (collagenase-2, MMP-8) (Tschesche and Pieper, 1998) cleave each strand of triple helical collagen Types I, II, and III at a specific site, to yield one-quarter and three-quarter collagen fragments. The substrate specificity for triple helical collagen is lost when the C-terminal domain of these proteases is removed. Interstitial and neutrophil collagenase have been produced in *E. coli* in insoluble inclusion bodies, extracted with urea or guanidine hydrochloride, then refolded by dilution or dialysis to produce active enzyme. Full-length mature interstitial collagenase (human or porcine) was made by expressing the pro-enzyme (Windsor *et al.*, 1991). N-terminal fusions of the mature gene (via a collagen hinge region) have also been made using β-galactosidase (O'Hare *et al.*, 1992; Clark *et al.*, 1995) or glutathione-S-transferase (O'Hare *et al.*, 1995). The pro-enzyme can be slowly autoactivated, or activated by treatment with organomercurials, while the fusion proteins were activated by cleaving at the introduced Factor Xa site after purification on antibody (Windsor *et al.*, 1991) or hydroxamic acid affinity columns (Clark *et al.*, 1995). Autoactivation occurs rapidly when the coordinating cysteine residue in the propeptide is replaced by serine (Windsor *et al.*, 1991). The additional, and undesired, autocatalytic cleavage within the C-terminal domain can be prevented by mutating the cleavage site (O'Hare *et al.*, 1995). In addition, the catalytic domain of human neutrophil collagenase was also produced in *E. coli* as insoluble aggregates that could be refolded to an active conformation (Schnierer *et al.*, 1993; Ho *et al.*, 1994).

The mammalian gelatinases, gelatinase A (Type IV collagenase, MMP-2) (Murphy, 1998) and gelatinase B (Type V collagenase, 92 kDa collagenase, MMP-9) (Collier and Goldberg, 1998), can cleave certain types of triple helical collagen such as Type V collagen and other matrix proteins like elastin, in addition to gelatin. Gelatinase A is expressed in a wide variety of tissues, while gelatinase B is restricted mainly to leukocytes and osteoclasts. The gelatinase domain structure and activation mechanism is similar to that of the collagenases described above, except that a fibronectin-like domain is inserted within the catalytic domain, and appears to be responsible for substrate specificity. The catalytic domain of gelatinase A was expressed in *E. coli* from a synthetic gene lacking the fibronectin-like domain. The protein was then refolded to generate an active protease (Ye *et al.*, 1995).

The catalytic domain of gelatinase B was expressed using a T7 promoter system to produce insoluble material that was also refolded to attain enzymatic activity (Kröger and Tschesche, 1997). In addition, soluble gelatinase B was produced in the periplasm of *E. coli* by fusion to the *ompT* gene, and the enzyme was activated by treatment with trypsin (Pourmotabbed *et al.*, 1994). More recently, a fibronectin-like domain deletion of gelatinase B fused to glutathione S-transferase was isolated from the *E. coli* cytoplasm, purified as a soluble fusion protein, and activated by an organomercurial compound (O'Farrell and Pourmotabbed, 1998).

Stromelysin-1 (MMP-3) (Matrisian, 1998), matrilysin (MMP-7) (Woessner, 1998), and macrophage elastase (macrophage metalloelastase, MMP-12) (Senior and Shapiro, 1998) cleave a wide variety of matrix proteins. Stromelysin-1 and matrilysin can also activate pro-collagenases and pro-gelatinases. Stromelysin-1 contains a structural zinc molecule in addition to the normal catalytic zinc. Unlike the other MMPs,

matrilysin does not have a C-terminal domain, and this domain is also rapidly removed from macrophage elastase during purification. Stromelysin-1, matrilysin, and macrophage elastase are not widely distributed in tissues but primarily found in stromal cells, glandular epithelial cells, and macrophages, respectively. Active catalytic domains of stromelysin-1 expressed by T7 promoter systems could be isolated from the soluble fraction (Marcy *et al.*, 1991; Ye *et al.*, 1992) of *E. coli,* or extracted from inclusion bodies (Suzuki *et al.*, 1998b; Ye *et al.*, 1992), with yields of up to about 10 mg per litre of culture broth. Also, growing the cultures at 27°C rather than 37°C increased the amount of soluble protein (Ye *et al.*, 1992). Full-length pro-stromelysin-1 (Rosenfeld *et al.*, 1994), as well as various pro-domain mutants (Freimark *et al.,* 1994), were refolded from inclusion bodies and activated with chymotrypsin or organomercurials.

Pro-matrilysin has also been produced in *E. coli*, either fused to ubiquitin (Welch *et al.*, 1995) or to a C-terminal $(His)_6$-tag (Itoh *et al.*, 1996) which aided purification from the extracted insoluble material. Human (Shapiro *et al.*, 1993; Gronski *et al.*, 1997) and mouse (Jeng *et al.*, 1995) macrophage elastases have been extracted from inclusion bodies produced in *E. coli*. During refolding and purification, the pro- and C-terminal domains are autocatalytically cleaved generating the active catalytic domain.

Another group of metalloproteases includes the membrane-type matrix metalloproteinases such as MT1-MMP. They can activate pro-gelatinase A and pro-collagenase 3 but not pro-gelatinase B, and can also degrade matrix proteins (Seiki, 1998). In addition to the typical MMP domains, MT1-MMP has a furin cleavage site between the pro- and mature domains, and has a C-terminal transmembrane domain with a short cytoplasmic tail. The catalytic domain, when refolded from inclusion bodies produced in *E. coli,* is autoactivated during purification (Lichte *et al.*, 1996; Will *et al.*, 1996), but the recombinant enzymes did not require TIMP-2 for pro-gelA activation, as observed during mammalian cell-mediated activation. The role of TIMP-2 binding regions was explored by expressing various forms of recombinant MT1-MMP fragments as fusion proteins with a FLAG tag at the C-termini (Kinoshita *et al.*, 1998). The protease can also be expressed as a fusion protein with glutathione S-transferase at the N-terminus and activated by exogenous furin (Sato *et al.*, 1996).

Recently, two new MMPs, MMP-19 (cloned from a human liver cDNA library) (Pendás *et al.*, 1997) and enamelysin (MMP-20, from odontoblastic cells) (Llano *et al.*, 1997), have been cloned and expressed in *E. coli*. These MMPs were expressed using a T7 promoter system, solubilized from inclusion bodies, and refolded to produce active enzyme.

Other metallopeptidases that have been successfully produced in *E. coli* include: a mitochondrial processing peptidase (Saavedra-Alanis *et al.*, 1994), thimet oligopeptidase (McKie *et al.*, 1995), pancreatic carboxypeptidase B (Edge *et al.*, 1998) and aspartyl aminopeptidase (Wilk *et al.*, 1998).

SERINE PROTEASES

Serine proteases generally have a catalytic triad of histidine, aspartate, and serine residues. The hydroxyl moiety of the serine residue directly attacks the scissile peptide bond to form an acyl-enzyme intermediate. The trypsin and subtilisin families

are the two major types of serine proteases, and they differ both in structure and sequence (Perona and Craik, 1995). Since there has been little success in expressing members of the mammalian subtilisin family, such as furin and pro-hormone convertases in bacteria, only the mammalian proteases of the trypsin family will be discussed in detail below. Unlike most other mammalian proteases that contain fairly long pro-regions and large C-terminal domains, members of the trypsin family tend to have short pro-peptides, or large N-terminal domains that remain covalently attached to the catalytic domain after activation.

Trypsin is produced in acinar cells of the pancreas as the inactive zymogen, trypsinogen (Halfon and Craik, 1998). Trypsinogen is activated by the cleavage of a short pro-peptide by enteropeptidase or by trypsin itself. This activated enzyme is then involved in the digestion of proteins in the intestine. In humans, five different trypsin genes are actively transcribed. Initially, trypsinogen (rat) was expressed in *E. coli* fused to the alkaline phosphatase (*phoA*) promoter and the signal sequence, then the soluble protein was isolated from the periplasm, purified, and activated by enteropeptidase cleavage of the pro-peptide (Graf *et al.*, 1987). By deleting the DNA coding for the activation hexapeptide of the zymogen in the above vector, active trypsin was expressed in the periplasm (Vasquez *et al*, 1989). Further improvements to this system were made by using the *tac* promoter and fusing the mature trypsin sequence directly to the *hisJ* signal sequence. In this instance, the protease was activated during export to the periplasm by the signal peptidases of *E. coli* (Higaki *et al.*, 1989). This expression system provided enough trypsin and various mutants for activity and structural determinations that helped characterize the catalytic mechanism of the enzyme. Using the above mentioned vector in a high cell density fermentation, induction at late or early log phase with IPTG yielded 56 and 13 mg/litre of active trypsin, respectively (Yee and Blanch, 1993). In addition, the gene III and VIII encoded coat proteins of the bacteriophage M13 have been fused to the C-terminus of trypsin in order to display active trypsin on the surface of bacteriophage (Corey *et al.*, 1993). Active human trypsin has also been produced using a *Bacillus subtilis* expression/secretion system similar to the one described in this section for chymase and granzyme K (Yoast and Schmidt, unpublished results). For the production of enteropeptidase (the enzyme that converts trypsinogen to trypsin), the catalytic domain of the enzyme was exported into the periplasm of *E. coli* using the DsbA protein as an N-terminal fusion partner. In this case, enteropeptidase was autoactivated by an engineered cleavage site at the fusion junction (Collins-Racie *et al.*, 1995).

The plasminogen activators, t-plasminogen activator (tPA, tissue-type plasminogen activator) (Lijnen and Collen, 1998) and u-plasminogen activator (uPA, urinary-type plasminogen activator, urokinase) (Ellis and Dano, 1998), are serine proteases that convert plasminogen to plasmin by cleavage at a specific position. Activated plasmin is a serine protease that then degrades fibrin in blood clots. Thus, the plasminogen activators have been studied for their potential clinical use as therapeutic thrombolytic agents. In fact, tPA is currently sold commercially as a thrombolytic agent for acute myocardial infarction. Both enzymes are produced as zymogens with multiple N-terminal domains (such as fibronectin-like, epidermal growth factor-like, or kringles) fused to the C-terminal serine protease catalytic domain. These zymogens are activated by plasmin cleavage, which converts the enzymes into two chains held together by a single disulphide bond. The binding of

tPA and uPA to their respective ligands (fibrin or uPA-receptor) greatly increases their activity towards plasminogen. Thus, tPA is a good acute thrombolytic agent when given systemically, since it is a very poor plasminogen activator in the absence of fibrin. Active human tPA was initially produced at very low levels in *E. coli* by refolding sonicated cell pellets extracted with guanidine hydrochloride (Pennica *et al.*, 1983). In a later study, about 3 mg/litre of tPA was recovered from inclusion bodies using an improved extraction and refolding protocol (Sarmientos *et al.*, 1989). Insoluble tPA was produced using either the *trp* or bacteriophage λ p_L promoters, and with or without the native signal sequence. Increased yields (up to 100 mg/litre) of refolded purified tPA could be produced when expressed with only one kringle domain (the fibronectin-like, epidermal growth factor-like, and first kringle domains were deleted) (Saito *et al.*, 1994). Also, some soluble material was recovered from the cytoplasm when tPA was expressed without the first N-terminal (fibronectin-like finger) domain (Kagitani *et al.*, 1985). In addition, soluble active tPA has been produced in the *E. coli* periplasm using the *tac* promoter and *phoA* signal sequence fused to the first kringle domain of tPA (Obukowicz *et al.*, 1990). Recently, full-length tPA was produced in the periplasm (~ 0.2 mg purified protein per litre) using the arabinose promoter and enterotoxin II signal sequence when co-expressed with an *E. coli* disulphide isomerase gene (*dsbC*) in a host deficient in periplasmic proteases (Qiu *et al.*, 1998). Thus, the production of a soluble, easily purified tPA in *E. coli* may eventually be economically feasible (Datar *et al.*, 1993). Insoluble tPA has also been produced in the cytoplasm of *Bacillus subtilis* (Wang *et al.*, 1989), but this system has not been optimized. The human pro-urokinase tissue plasminogen activator (pro-uPA, pro-urokinase, single-chain uPA, scuPA) gene was cloned, and low amounts of the active enzyme could be extracted from insoluble *E. coli* cell lysates (Holmes *et al.*, 1985). This system, utilizing the *trp* promoter induced with indole acetic acid, was scaled for a 10 litre fermentation that produced over 100 mg of pro-uPA. This protein could then be solubilized, refolded, and activated to the two-chain form by plasmin (Winkler and Blaber, 1986). Enough of the protein was produced to enable researchers to study the kinetics of plasminogen activation by recombinant uPA (Collen *et al.*, 1986; Lijnen *et al.*, 1986). Improvements in the expression system were made by using alternative ribosomal binding sites and transcriptional terminators (Brigelius-Flohé *et al.*, 1992). Recently, further improvements have been made in the production of uPA by simplifying the extraction and purification procedure. Examples include: taking advantage of a C-terminal $(His)_6$-tag (Tang *et al.*, 1997), or using a T7 promoter system with a synthetic gene constructed to have optimal *E. coli* codons (Hua *et al.*, 1996; Xue *et al.*, 1997). In addition, using the gram-positive bacterium *Lactococcus lactis* as a host, the catalytic domain of bovine plasmin was produced in the cytoplasm or secreted into the medium using the *prtP* promoter or *usp45* promoter and signal sequence, respectively; however, the yields were quite low (Arnau *et al.*, 1997).

There are a number of serine proteases that are necessary for blood coagulation. Thrombin (coagulation factor IIa, fibrinogenase) converts soluble fibrinogen into fibrin monomers that polymerize to form a blood clot (Stone and LeBonniec, 1998). Factor XIII, activated by thrombin, then cross-links fibrin to stabilize the clot. Thrombin also triggers the coagulation cascade by its involvement in platelet activation, but when bound to endothelial cells, it can also shut the coagulation cascade down by activating protein C. Coagulation factor Xa, when bound to calcium, Factor

Va, and phospholipids, activates pro-thrombin by the cleavage of two peptide bonds (Stenflo, 1998). Coagulation factor IXa (Bajaj, 1998) deficient in people with haemophilia B, can activate coagulation factor X (the zymogen of coagulation factor Xa). Thrombin consists of a short N-terminal A chain (36 amino acids in human thrombin, produced during activation) covalently attached by a disulphide bond to the serine protease B chain. Coagulation factors Xa and IXa have an N-terminal Gla domain (γ-carboxyglutamic acid domain) followed by two epidermal growth factor-like domains, and then the catalytic domain.

Enzymatically active ζ-thrombin naturally arises from a single chymotryptic cleavage at Trp-148 of the catalytic domain that yields two noncovalent polypeptide segments. The recombinant ζ2-thrombin segment (B chain residues 149–259) was produced in *E. coli* using the *trp* promoter and a *trp* E fusion, and refolded from inclusion bodies. The enzyme was reconstituted by adding purified ζ1-thrombin (A chain linked via disulphide to B chain residues 1–148) to the recombinant ζ2-thrombin (Gan *et al.*, 1991). More recently, prethrombin-2 (A chain fused to B chain as a single polypeptide without the coagulation factor Xa cleavage between chains) was produced in *E. coli* using a T7 promoter system. Inclusion bodies yielded approximately 0.6 mg per litre after refolding, and the prethrombin-2 produced could be processed to active thrombin (DiBella *et al.*, 1995). Similarly, truncated versions of coagulation factors X and IX (missing the Gla and first EGF-like domains) were produced in *E. coli*, extracted, refolded and activated for kinetic studies (Hopfner *et al.*, 1997).

Kallikreins are a family of serine proteases that release the vasoactive peptides, kinins (Lys-bradykinin or bradykinin) from kininogen. Kallikreins may also be involved in the processing of other hormones and growth factors. Semenogelase (prostate specific antigen, or insulin-like growth factor binding protein-3 proteinase) is closely related to kallikreins, is present in seminal plasma, and has been used as a diagnostic marker for prostate cancer (Chao, 1998a). Human tissue kallikrein (also glandular, pancreatic, renal, or urinary kallikrein) is produced as a single chain protein with a short (7 amino acids) pro-peptide (Chao, 1998b). Human tissue kallikrein was produced in the periplasm of *E. coli* using the *tac* promoter and *phoA* signal sequence; unfortunately, only about 1% of the protein was active (Angermann *et al.*, 1989). Active and soluble tissue kallikrein could also be extracted with detergents from insoluble aggregates produced from a T7 promoter system in *E. coli* (Wang *et al.*, 1991). A ten litre fermentation yielded about 3 mg of active tissue kallikrein following extraction, refolding from insoluble aggregates, and activation by thermolysin (Lu *et al.*, 1996). A fusion of the arabinose promoter to mouse kallikrein produced an insoluble fusion protein in *E. coli* that could be solubilized with guanidine hydrochloride, refolded by dialysis, and activated by limited proteolysis with trypsin (Blaber *et al.*, 1990). Similarly, insoluble pro-semenogelase was produced using a T7 promoter system, extracted, and refolded from *E. coli* refractile bodies, and activated by trypsin or by recombinant human tissue kallikrein (Takayama *et al.*, 1997).

Serine proteases can also be found in the granules of mast cells and cytotoxic lymphocytes. Mast cell chymase is released along with histamine and mast cell tryptase during the degranulation of mast cells triggered by an allergic response (Caughey, 1998). Granzymes (T-cell specific proteinases, or cytotoxic cell proteinases)

are released from granules into the intracellular space upon binding of a cytotoxic lymphocyte to its target cell, resulting in the apoptotic death of the target cell (Berke, 1995; Henkart, 1994). Active granzyme A (a disulphide linked homodimer) was produced in the periplasm of *E. coli* using a T7 promoter system by attaching a pelB signal sequence-enteropeptidase cleavage site fusion to the mature granzyme A sequence, and a C-terminal $(His)_6$-tag (Beresford *et al.*, 1997). Insoluble human chymase was expressed from the p_L promoter in *E. coli* fused to ubiquitin and an enteropeptidase cleavage site, was solubilized with urea, and refolded to yield about 125 μg of active enzyme per litre (Wang *et al.*, 1995b).

Soluble active chymase (McGrath *et al.*, 1997) and granzyme K (Babé *et al.*, 1998) were secreted directly into the culture medium using a *Bacillus subtilis* expression system (yields of purified enzyme were over 0.2 mg per litre of culture broth). In these studies, the subtilisin (*apr*E) promoter and signal sequence was fused to the mature mammalian protease that was activated by *B. subtilis* signal peptidases during secretion. Proteolytic degradation of the products was minimized by using extracellular protease deficient bacterial strains (Wu *et al.*, 1991).

Other serine protease types, such as prolyl-endopeptidase (Sommer, 1993), acylamino acid releasing peptidase (Mitta *et al.*, 1998), and a human homologue of the *E. coli* ClpP protease (Corydon *et al.*, 1998), have also been produced in *E. coli*.

Specific expression systems for viral proteases

PICORNAVIRIDAE PROTEASES

The picornavirus ssRNA genome is translated into a single polyprotein precursor. The subsequent proteolytic cleavages required to release mature proteins can be separated into three categories: (1) the precursor to the capsid proteins is cleaved co-translationally by the virus-encoded 2A protease (one cleavage event at its own N-terminus); (2) the N-terminal polyprotein region is autoprocessed (for members of the aphthovirus and cardiovirus genera, this is performed by the virus-encoded L-protease); and (3) the remaining cleavages within the viral polyprotein are carried-out by the virus-encoded 3C protease (Lawson and Semler, 1990). Members of the picornavirus family include numerous human pathogens such as polio virus, hepatitis A, rhinovirus, and coxsackievirus (Rueckert, 1996). The 2A and 3C proteases both contain a cysteine as the active site nucleophile, but the protein folds are different. The crystal structures of several 3C proteases reveal a chymotrypsin-like fold and confirm the catalytic triad including the active-site cysteine (Matthews *et al.*, 1994; Bergmann, *et al.*, 1997). The recently elucidated 2A protease structure is also chymotrypsin-related, but unusual (Petersen *et al.*, 1999). Both proteases have received considerable attention as potential targets for antiviral therapies, and numerous bacterial expression schemes have been used (*Figure 8.2* illustrates several examples for rhinovirus 3C protease).

The 2A protease of rhinovirus was expressed as a fusion protein in *E. coli* under the control of the λ promoter. The fusion consisted of the sequence for the first 98 amino acids of the MS2 polymerase and a portion of the viral genome that encompassed the 2A protease and extensive viral sequences at both ends (Sommergruber *et al.*, 1989). These early results showed that the protease could be expressed in *E. coli* as a fusion,

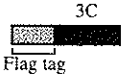
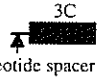
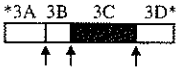
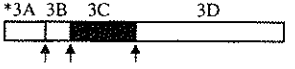
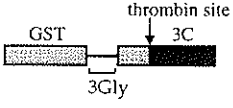
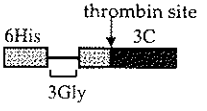
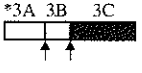
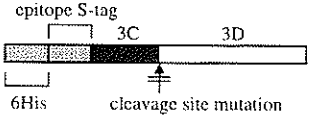
Vector/ Promoter	Expression Cassette	Comments
pIN-III-OmpA3 derived *lpp-lac* tandem		Periplasmic, partially soluble, ~ 1mg of chimera/L of culture after purification.[1]
pAT153 derived *tac*		At least 50% of 3C is soluble, ~ 0.5-1.0 mg/L of culture after purification.[2]
pKK322-2 *trp-lac*	(pR202)  (pR207)  (pR209) 	pR207 yielded highest level of soluble, active 3C protease, 3-5% of total cell protein.[3]
pGEX-1N *tac*		Efficient purification of 3C by affinity chromatography, thrombin cleavage and gel filtration.[4]
pQE-31 T5		Purified mature 3C protease used as reagent to cleave other fusions proteins.[5]
pET-3a *lacUV5*		Majority of 3C found in cytoplasmic granules, 10-20% total cell protein.[6]
pET-30a *lacUV5*		Two step purification yielded 3CI 0.02-0.04 mg/L culture, ~70% pure.[7]

Figure 8.2. Examples of diverse *E. coli* expression systems designed for the production of the rhinovirus 3C protease. The 3C catalytic domain (corresponding to bases 5240–5780 of HRV-14 genome) is shown as black bars. Viral capsid proteins 3A, 3B and 3D shown as white bars. Fusion partners are shown as gray bars. Asterix (*) denotes a truncated gene. Arrows beneath the 3A/3B, 3B/3C and 3C/3D junctions indicate cleavage sites for protease 3C within the HRV polyprotein. For simplicity, signal peptides are not shown in the cassette schemes. Figures are not to scale. References are: (1) Libby *et al.*, 1988; (2) Knott *et al.*, 1989; (3) Cordingley *et al.*, 1989; (4) Leong *et al.*, 1992; (5) Walker *et al.*, 1994b; (6) Birch *et al.*, 1995; (7) Davis *et al.*, 1997.

with autoprocessing taking place at its own N-terminus at predicted sites. The resulting material was identified by SDS/PAGE, but was not purified. It is worth noting that the picornavirus 2A proteases are very toxic to eukaryotic cells. As part of their replication cycle, the picornaviruses are known to block host protein synthesis, while allowing translation of the uncapped viral RNA. Biochemical and genetic data indicate that the 2A protease is responsible for this phenomenon by virtue of cleaving the eukaryotic initiation factor 4γ component eIF-4 (Wyckoff *et al.*, 1990). Thus, heterologous expression in bacteria overcomes the limitation of 2A expression in eukaryotic cells. The mature 2A proteases of rhinovirus 2 and coxsackievirus B4 were expressed in soluble form in *E. coli* (Liebig *et al.*, 1993). A pET8c vector was used to clone the entire 2A domain flanked by a short segment of the VP1 viral protein at the N-terminus. Both proteases migrated according to the size expected for the properly autoprocessed form, and approximately 1 mg of purified protein was obtained per gram of cell paste. Likewise, the 2A protease of polio virus was expressed in *E. coli* as a fusion protein (Alvey *et al.*, 1991). The N-terminal portion of the bacterial TrpE protein was linked to the 2A domain. This fusion protein underwent efficient autocatalytic cleavage at the N-terminus of 2A generating the mature protease. Soluble enzyme was produced and, following purification, it showed activity against various protein and peptide substrates. More recently, the 2A protease of rhinovirus 14 was expressed in *E. coli* in inclusion bodies, and recovered by refolding (Wang *et al.*, 1998). The coding sequence for the 2A protease was inserted into a heat-inducible expression vector, and optimal activity was obtained when refolding was performed in the presence of zinc. This metal is known to bind and stabilize the protein structure (Voss *et al.*, 1995). Previous knowledge of the zinc requirement facilitated the maximal recovery of enzymatic activity by supplementing the refolding buffer with this metal.

The 3C protease of rhinovirus has received considerable attention as a potential antiviral target; a representative list of expression systems is presented in *Figure 8.2*. A cDNA encoding the widely studied rhinovirus 14 (HRV-14) sequence was expressed in *E. coli* using a periplasmic secretion vector containing the OmpA signal region (Libby *et al.*, 1988). An octapeptide sequence Asp-Tyr-Lys-$(Asp)_4$-Lys, also known as the FLAG sequence, was fused to the N-terminus of the protease gene and was helpful in tracking the presence of the expression product using anti-FLAG antibodies. Yields of approximately 1 mg/litre of *E. coli* cultures were achieved. Despite the presence of a periplasmic signal sequence, the protease was localized in the membrane fraction of lysed cells, precluding efficient purification. This same HRV-14 3C sequence was cloned in a *tac* promoter-driven expression plasmid with an N-terminal methionyl residue (Knott *et al.*, 1989). This time, the protease was expressed as a soluble and active protein in *E. coli*. When the HRV-14 3C protease was expressed by fusing the protein to the C-terminus of glutathione-S-transferase (GST), efficient purification was achieved in a single step (Leong *et al.*, 1992). A cleavage recognition sequence for thrombin was inserted between the GST and the 3C domain using a $(Gly)_3$ linker. This greatly facilitated the production of mature 3C with the correct N-terminus, and the purification of a soluble protein. For the product of HRV-14 3C protease for crystallography, the protein was expressed as a precursor, including viral sequences of the 3A and 3B proteins. The protease correctly autoprocessed in *E. coli*, yet it had to be recovered from inclusion bodies by guanidinium

denaturation and slow-dialysis refolding (Matthews *et al.*, 1994).

Biochemical and genetic studies of HRV had suggested that the protein hybrid 3CD, containing both the protease and polymerase enzyme domains, is responsible for some of the 3C-like proteolytic activity in infected cells. To confirm this, the 3CD protein was expressed using the *E.coli* pET-30A (Novagen) system (Davis *et al.*, 1997). The recombinant protein was fused to two N-terminal affinity tags: $(His)_6$-tag and epitope S-tag, resulting in a soluble and enzymatically active product, following purification.

The mature domain of the 3C protease of the hepatitis A virus has also been studied. Using a multicopy vector with a *tac* promoter, soluble and active enzyme was produced as 10% of total cellular protein in *E. coli* (Malcolm *et al.*, 1992). This same protocol was employed to express mutants of the enzyme for x-ray crystallography (Bergmann *et al.*, 1997).

Mature 3C protease from polio virus was expressed in *E. coli* under the control of a T7 promoter (Baum *et al.*, 1991). Active protein product was purified to homogeneity from resolubilized inclusion bodies, even though some soluble material was also present after induction. A similar procedure lead to the production of sufficient high-quality polio virus 3C protease (Mosimann *et al.*, 1997) and mutants of the hepatitis A 3C protease (Bergmann *et al.*, 1997) for crystallographic studies.

CORONAVIRIDAE PROTEASES

Coronaviruses are human pathogens of significance. Similar to picornaviruses, they translate the ssRNA genome into a single polyprotein precursor that requires proteolytic processing. The genome encodes both a papain-like cysteine endoprotease and a protease with homology to the 3C protease of picornaviruses (Gorbalenya *et al.*, 1989).

The 3C-like protease, 3CLP, from coronavirus strain 229E has been cloned and expressed in *E. coli* using the pQE10 His-tag vector from Qiagen (Ziebuhr *et al.*, 1995). Initial attempts to express just 3CLP domain with only a minimal number of additional flanking residues failed. The authors then used a larger fusion construction that included the N-terminal half of β-galactosidase, followed by a segment of the viral genome encoding the 3CLP domain flanked by 58 amino-proximal and 12 carboxy-proximal residues. Expression of this larger fusion protein was observed, and efficient processing in the bacteria lead to the isolation of the correctly processed 34 kDa 3CLP protease. Enzymatic activity of the purified protease was confirmed by *trans* cleavage of an *in vitro* translated polyprotein substrate.

RETROVIRIDAE PROTEASES

The retroviridae family of viruses encompasses several genera such as the type B, C and D viruses, the lentiviruses and the T-cell lymphoma viruses, all with numerous animal and human pathologies (Coffin, 1996). HIV-1 is the most studied lentivirus to date, but we shall provide a sampling of expression systems that cover proteases from the various genera. Retroviruses express their capsid proteins and enzymes as polyproteins, that later require processing by the single virus encoded protease. This processing is unique, in that it occurs once the virions are assembled, as they bud off

from an infected cell, with the exception of the type C and D animal virus, where assembly and processing occur in the cytoplasm. The retroviral proteases are classified in the aspartyl protease family but, unlike the mammalian enzymes that are composed of two domains, the viral enzymes are obligate homodimers with one catalytic Asp provided by each monomer (Miller *et al.*, 1989a).

Protease from the type C retrovirus Moloney murine leukaemia virus (Mo-MuLV) was expressed in fusion with glutathione S-transferase using the pGEX-ZT (Pharmacia LKB) vector (Menéndez-Arias *et al.*, 1993). The protein appeared in the soluble fraction of *E. coli* cells, and could be purified to homogeneity after thrombin cleavage.

The protease from the type D retrovirus Mason-Pfizer monkey virus was expressed as a 26 kDa precursor in *E. coli* (Hrusková-Heidingsfeldová *et al.*, 1995). The protease undergoes rapid self-processing following renaturation under reducing conditions from solubilized inclusion bodies.

Human T-cell leukaemia virus type 1 (HTLV-1) is an oncovirus associated with adult T-cell leukaemia. The HTLV-1 protease was expressed fused to a decahistidine-containing leader peptide using the pET19 vector (Ding *et al.*, 1998). Insoluble protein was denatured in urea prior to chromatography on a His-bind column (Novagen). The purified fusion protein was refolded by sequential dialysis, allowing autoprocessing to occur and yielding mature active enzyme.

Foamy viruses (FV) have several features not shared by other retroviruses, such as the expression of the Pro-Pol proteins by a spliced mRNA. The FV human spumaretrovirus protease was expressed as a fusion protein using a modified thioredoxin fusion vector pTRxFus (Invitrogen) that carries a His-tag in the centrally located surface loop of the *E. coli* trxA protein (Pfrepper *et al.*, 1997). The protease fusion was purified from the soluble fraction of bacterial cells, and the N-terminal fusion partner was removed by incubation with enteropeptidase.

Numerous lentiviruses besides HIV-1 have been studied at the biochemical level, HIV-2 and SIV in particular. An example of bacterial expression of HIV-2 protease involves its expression as a fusion with the human superoxide dismutase sequence (Salto *et al.*, 1994). Expression from a pTacTac vector results in rapid processing of the precursor fusion protein to yield mature enzyme that accumulates as soluble material in the cytoplasm. Expression yields were approximately 1 mg/litre of bacterial culture, and the purified protease served to characterize a series of nonpeptide irreversible inhibitors that arose from structure-based computer searches.

The SIV_{mac} protease has been produced in *E. coli* as a mature enzyme that is autoprocessed from a viral precursor form (Grant *et al.*, 1991). Mature protease was obtained from the soluble fraction and purified to homogeneity. Subsequent comparative studies of protease inhibitors served to suggest that potent molecules could show efficacy in monkey models, that could in turn be translated to human clinical trials.

Too numerous are the bacterial expression systems used to produce the HIV-1 protease. Thus, we shall mention only a representative list. The majority of the early expression vectors designed to produce active and soluble HIV-1 protease relied on the cloning of large portions of the gag-pol sequence and relied on subsequent autoprocessing by the protease (Debouck *et al.*, 1987). These early attempts lead to the confirmation of the precise protease primary structure and its cleavage sites within the viral polyprotein precursor. Later, the DNA encoding only the predicted 11 kDa protease (one monomer) was cloned, bypassing the need for autoprocessing. This

protease was expressed to a high level in *E. coli* as an active dimer (Graves *et al.*, 1988; Darke *et al.*, 1989). It has been noted that HIV-1 production under a strong promoter such as *trp* leads to inhibition of *E. coli* cell growth upon induction (Darke *et al.*, 1989), limiting the time of culture post-induction, and reducing recombinant protein yields. Other methods of expression in *E. coli* have used fusion partners such as *lac*Z (Giam and Boros, 1988) or human superoxide dismutase (Babé *et al.*, 1990). It was noted that a portion of the purified protease underwent inactivation due to autolysis after residue Tyr5 (Babé *et al.*, 1990). The 3-dimensional structure of the HIV-1 protease was solved after successful crystallization of the protease obtained by total chemical synthesis (Miller *et al.*, 1989b), mostly to circumvent the lack of sufficient quantities of highly purified material with a homogeneous and intact N-terminus.

Another approach to express the HIV-1 protease has been to construct a tethered protease dimer containing short glycine linkers (DiIanni *et al.*, 1990; Babé *et al.*, 1992; Griffiths *et al.*, 1994). All these systems produced soluble proteins that were purified to homogeneity. Expression of tethered dimers has provided the freedom to explore structure-function relationships when mutations are introduced in a single protease monomer, or when chimeras of the HIV-1 and HIV-2 proteases are expressed.

In yet another attempt to over-express the HIV-1 protease, the vector pTacTac was used to clone the viral protease downstream of the gene for the highly transcribed protein CheY (Rozzelle *et al.*, 2000). To further enhance translation efficiency, an intervening bicistronic linker was placed between the two protein coding regions. This vector effectively yields large quantities of soluble protein and has been used to express numerous HIV-1 protease mutants with compromised catalytic efficiency as well as tethered versions of heterodimers. Since autoprocessing is no longer required, inactive proteases can be easily purified.

As mentioned previously, high level expression of soluble and active HIV-1 protease in *E. coli* has been shown to be toxic to the cells. Advantage was taken of this property to design a screening assay for HIV-1 inhibitors in intact bacterial cells expressing the enzyme (Ast *et al.*, 1998). Co-expression of the protease with a series of different inhibitory peptides leads to a recovery of bacterial cell viability in a dose-dependent and sequence-dependent fashion. This was a creative way to exploit an expression system as an assay platform. Another *E. coli* expression assay and screen of HIV-1 protease variants was developed to evaluate the susceptibility to the inhibitor indinavir (Melnick *et al.*, 1998). Measurement of a downstream processing event that produces the reverse transcriptase, which is then assayed with a sensitive substrate, produced a screening assay that identified numerous HIV-1 protease mutants with varying resistance to the inhibitor. Such a system offers a way to screen numerous protease sequences and also numerous protease inhibitors without the need to purify the enzyme and set-up a separate activity assay.

The HIV-1 protease also provides an example of how a protease with exquisite substrate sequence selectivity can be used to produce other proteins in the desired form. In this case, a fusion protein was constructed where a protein of interest (interleukin-6, IL-6) was fused to thioredoxin via a sequence that contained an HIV-1 protease cleavage site and also a $(His)_6$ sequence. A cDNA for HIV-1 protease was cloned into a compatible plasmid and, when both proteins were expressed in *E. coli*, the result was soluble IL-6 with a $(His)_6$-tag at the N-terminus (Han *et al.*, 1998).

HERPESVIRIDAE PROTEASES

Numerous herpesviruses are serious human pathogens and many of their enzymes are under consideration as antiviral targets. During maturation, these viruses require the assembly of scaffold proteins inside the capsid shell, prior to packaging of viral DNA. The scaffolding proteins are encoded by two overlapping 3' co-terminal genes, and their proteolytic processing is performed by a protease encoded within the N-terminus of the larger gene. Cleavages occur at the maturation (M) site, at the C-terminus of the assembly protein, and at the release (R) site, C-terminal to the protease domain. Herpesvirus proteases use an active-site serine residue, but have no homology to mammalian serine proteases. Crystal structure studies for the cytomegalovirus protease, among others, revealed a unique dimeric structure and a novel catalytic triad of His/His/Ser, instead of the well-characterized Asp/His/Ser of known serine proteases. These enzymes have a very poor turnover rate as compared to a digestive serine protease, increasing the difficulty in measuring activity for recombinant versions of the proteins (Welch *et al.*, 1991).

The herpes simplex virus 1 (HSV-1) protease was expressed in the pET3C vector to produce soluble and active protein. The region of the UL26 gene that encodes the ICP35 assembly protein was used to clone the segment representing the first 306 amino acids (residues 1–247 correspond to the protease domain). The initial expression product was the expected 32 kDa precursor. Upon purification and concentration, the protease self-processed the C-terminal extension to yield the mature and active 27 kDa form at approximately 0.4 mg/litre of culture broth (Darke *et al.*, 1994).

Like HSV, varicella-zoster virus (VZV) belongs to the herpesvirus α–subfamily. The VZV protease was expressed as an N-terminal His-tag fusion protein using a Qiagen vector system. A modification of the catalytic domain (1–236 amino acids) was used to facilitate the crystallization of the protein product. In this case, the expressed molecule contained a deletion of the first 10 amino acids, which did not affect protease activity. This protein was purified by nickel chelation and size exclusion and successfully used to solve the 3-dimensional crystal structure (Qiu *et al.*, 1997).

Cytomegalovirus (CMV) is a member of the β-subfamily of herpesviruses, and the viral protease, also known as assemblin, is encoded by the N-terminus of the UL80 gene. This 256 amino acid region was cloned into the *E. coli* expression vector pCZR332 for expression, using a temperature shift from 32°C to 42°C for induction (Burck *et al.*, 1994). The protein obtained was 28 kDa in size; its activation required denaturation in urea and refolding under reducing conditions. Addition of 50% glycerol greatly enhanced stability and activity of the protease. The N-terminal 15 kDa protein should be the mature protease domain, but this protein form had no activity after purification. Another group took the same 256 amino acid region of CMV protease and cloned it in the T7 vector pET3C (LaFemina *et al.*, 1996), but they introduced two mutations by PCR. These mutations prevent cleavage at the two major internal autocatalytic sites used by the protease, thus yielding only the full-length 28 kDa CMV protease. The enzyme was active and soluble in the *E. coli* cytoplasm, producing 10–15 mg/litre with greater than 95% purity after two purification steps. The mature CMV protease (268 residues) has also been expressed with a C-terminal $(His)_6$ extension, permitting rapid purification after refolding from *E. coli* inclusion

bodies (Tomasselli *et al.*, 1998). A mutation was introduced in the protease sequence to eliminate auto-proteolysis. Approximately 40 mg of protease was produced per litre of cell culture after purification, and the material was crystallizable.

The Kaposi sarcoma-associated virus (KSHV, HHV-8) is a relatively new member of the herpesvirus family and classified under the γ-subfamily. The 230 amino acid protease domain was expressed in *E. coli* using the pQE30 vector (Qiagen). The sequence Met-Arg-Gly-Ser-$(His)_6$-Gly-Ser replaced the initiating Met residue. Following IPTG induction, the cells were sonicated and cell lysates were resuspended in urea. This material was bound to a nickel column and subjected to a 1 litre reverse urea gradient to permit refolding, then the protein was eluted with imidazole. The purified protein is enzymatically active against small synthetic substrates and has the expected N-terminal sequence and molecular weight (Unal *et al.*, 1997).

FLAVIVIRIDAE PROTEASES

Flaviviruses have a positive strand RNA genome that encodes a single polypeptide precursor (Rice *et al.*, 1985). Structural proteins at the N-terminus are processed by host signal peptidases, while non-structural proteins at the C-terminus are processed by the two virally encoded proteases NS2-3 and NS3. NS2-3 performs a single *cis* cleavage, identified by studies of mutant viruses, but the enzyme remains poorly characterized. NS3 protease is a chymotrypsin-like serine protease that requires another viral polypeptide as a co-factor for activity. As seen with hepatitis C (HCV) NS3, its NS4A peptide co-factor intercalates within the enzyme core to complete the structure (Love *et al.*, 1996). Another unique feature of the NS3 proteases is that they are tethered at the C-terminus to the viral helicase (Tomei *et al.*, 1993).

Of all the flaviviruses, HCV has received the most attention because of its rapid spread in industrialized nations, and the threat to the blood supply (Kuo *et al.*, 1989). Numerous pharmaceutical companies have antiviral programmes that focus on inactivating this enzyme to prevent virus proliferation. The need for reagent quantities of active NS3 protease has required the design of assay methods and expression systems that take into consideration the need for a polyprotein co-factor for activity. Initially, the NS3 domain of HVC (residues 1050–1214) was cloned in frame with the sequence for the nitrile hydratase leader (Mori *et al.*, 1996). This N-terminal extension has a stretch of histidine residues that allow protein purification by nickel-agarose column. Soluble protein was obtained and, although not mentioned, the yields were likely low. The authors detected the predicted cleavage of one peptidic substrate in the absence of the NS4A co-factor. A similar approach was later taken expressing the NS3 domain (residues 1027–1218 in this instance) following an N-terminal $(His)_6$-tag in the pTrcHisB (Invitrogen) vector (Vishnuvardham *et al.*, 1997). After a single affinity purification step, the authors obtained 3 mg of active enzyme per litre of culture, which was mostly soluble. Addition of NS4A peptide to the enzyme assays increased the k_{cat}/K_m by 20-fold. In a more recent effort to express active and stable HCV NS3 protease, the NS3 domain was expressed fused to an essential portion of the NS4A protein (residues 21–32). The co-factor was placed at the N-terminus of NS3, joined by a tetrapeptide linker. This single-chain recombinant protein was over-expressed as a soluble protein in *E. coli* (Taremi *et al.*, 1998). Following two purification steps, the enzyme showed full proteolytic activity, identical to the

authentic complexes of NS3-NS4A generated in eukaryotic cells. Another variation on the NS3-NS4A tethering theme was tried, where the NS3 domain was expressed in frame with the 54 residue NS4A (Gallinari *et al.*, 1999). Cleavage occurs between these two domains during expression in *E. coli*. A highly active, soluble, non-covalent complex with a subnanomolar dissociation constant was subsequently purified.

In the case of the dengue virus, its NS3 protease resembles the overall structural and functional features of the HCV NS3, but it requires a different viral co-factor, the NS2B protein. In order to produce protein for structural studies, the dengue NS3 protein was expressed using the pQE-30 (Qiagen) vector with an N-terminal $(His)_6$-tag (Murthy *et al.*, 1999). Following IPTG induction, the protein could be found in both the soluble and insoluble fractions. Both materials could be purified on nickel affinity columns, with the insoluble fraction requiring urea denaturation and dialysis for refolding. Addition of a peptide corresponding to the hydrophilic domain on the NS2B protein was essential for enzymatic activity, and the crystal structure confirmed the reason for this requirement.

ADENOVIRIDAE PROTEASE

The adenovirus protease has a cysteine protease catalytic triad in a completely novel protein fold, which also incorporates a peptide co-factor (Ding *et al.*, 1996). The apoenzyme is relatively inactive, but an eleven amino acid peptide that is a product of the protease activity on the viral polypeptide can enhance activity by several hundred-fold, primarily by assisting in the stabilization of the enzyme structure. Recently, the ubiquitin-like protein (Ubl)-specific protease, named Ulp1, was identified in yeast (Li and Hochstrasser, 1999). Ulp1 is the first example of an eukaryotic protein with homology to the adenovirus proteases, but it is not expected to require a peptide cofactor.

The 23 kDa protease encoded at the end of the L3 late mRNA from adenovirus 2 was expressed in *E. coli* (Anderson, 1990). The gene was cloned into the expression vector pET12b, introduced into BL21 cells and induced with IPTG. A fusion to the OmpT leader sequence did not seem to assist in obtaining active soluble protease. The protease coding sequence has also been subcloned into a modified pRIT2T expression vector under the control of the lp_L promoter (Rancourt *et al.*, 1994). Induction of selected plasmids by nalidixic acid or high cell density resulted in the expression of protein in inclusion bodies. Wild-type, as well as various protease mutants, were purified for enzymatic characterization and identification of critical catalytic residues. It is now clear that addition of the 11 amino acid co-factor is essential to impart enzymatic activity to the recombinant protease. To solve the 3-dimensional structure, adenovirus 2 protease purified from an *E. coli* expression system was co-crystallized with a synthetic peptide version of the co-factor (Ding *et al.*, 1996). The structure confirmed the existence of a new protease fold, and the presence of a catalytic triad (Cys-His-Glu) and an oxyanion hole arranged similarly to those in papain. The coordination of the co-factor peptide served to explain its critical structural role, and the lack of activity of the apoenzyme. This protease exemplifies why recombinant proteases may not be enzymatically active if compared to material produced in the natural environment; in this case, the virus was producing the required co-factor. Thus, care must be taken to understand the natural milieu of a protease and the role of

various structural domains when designing expression systems and analysing the products.

Conclusion

Bacteria, primarily *E. coli*, have been used successfully to produce a wide variety of mammalian and viral proteases. These heterologous expression systems have enabled researchers to study the structure and function of these important enzymes, and also have helped in the development of commercial products. In the future, other bacteria, such as the gram-positive bacterium *B. subtilis*, may become more utilized for protein production, as better strains and expression systems are developed. While other organisms, such as yeast, insect cells and mammalian cells, may be required for the production of certain proteases, bacteria are attractive thanks to their ease of genetic manipulation, robust growth and inexpensive production requirements.

References

ALVEY, J.C., WYCKOFF, E.E., YU, S.F., LLOYD, R. AND EHRENFELD, E. (1991). *cis* and *trans*-cleavage activities of poliovirus 2A protease expressed in *Escherichia coli*. *Journal of Virology* **65**, 6077–6083.

ANDERSON, C.W. (1990). The proteinase polypeptide of adenovirus serotype 2 virions. *Virology* **177**, 259–272.

ANGERMANN, A., BERGMANN, C. AND APPELHANS, H. (1989). Cloning and expression of human salivary-gland kallikrein in *Escherichia coli*. *Biochemical Journal* **262**, 787–793.

AOKI, T., TOMAKI, E., SATOH, M., TAKASHIRO, M., ONAGI, H., ITOH, M., TERAMOTO, T., MORIKAWA, J. AND WATABE, H. (1998). Purification of recombinant human pepsinogens and their application as immunoassay standards. *Biochemistry and Molecular Biology International* **45**, 289–301.

ARNAU, J., HJERL-HANSEN, E. AND ISRAELSEN, H. (1997). Heterologous gene expression of bovine plasmin in *Lactococcus lactis*. *Applied Microbiology and Biotechnology* **48**, 331–338.

AST, O., JENTSCH, K.D., SCHRAMM, H.J., HUNSMANN, G., LUKE, W. AND PETRY, H. (1998). A rapid and sensitive bacterial assay to determine the inhibitory effect of 'interface' peptides on HIV-1 protease co-expressed in *Escherichia coli*. *Journal of Virology Methods* **71**, 77–85.

BABÉ, L.M. AND CRAIK, C.S. (1997). Viral proteases: Evolution of diverse structural motifs to optimize function. *Cell* **91**, 427–430.

BABÉ, L.M., PICHUANTES, S., BARR, P.J., BATHURST, I.C., MASIARZ, F.R. AND CRAIK, C.S. (1990). HIV1 protease: Bacterial expression, purification and characterization. In: *Proteins and Pharmaceutical Engineering*. Eds. C.S. Craik, R. Fletterick, C.R. Matthews and J. Wells, pp 71–88. New York: Wiley-Liss Press.

BABÉ L.M., ROSE, J. AND CRAIK, C.S. (1992). Synthetic 'interface' peptides alter dimeric assembly of HIV1 and 2 proteases. *Protein Science* **1**, 1244–1253.

BABÉ, L.M., YOAST, S., DREYER, M. AND SCHMIDT, B.F. (1998). Heterologous expression of human granzyme K in *Bacillus subtilis* and characterization of its hydrolytic activity *in vitro*. *Biotechnology and Applied Biochemistry* **27**, 117–124.

BAJAJ, S.P. (1998). Coagulation factor IXa. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 157–160. New York: Academic Press.

BAUM, E.Z., BEBERNITZ, O.P., MUELLER, T. AND PLOTCH, S.J. (1991). Purification, properties, and mutagenesis of poliovirus 3C protease. *Virology* **185**, 140–150.

BERESFORD, P.J., KAM, C.-M., POWERS, J.C. AND LIEBERMAN, J. (1997). Recombinant human granzyme A binds to two putative HLA-associated proteins and cleaves one of them. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9285–9290.

BERGMANN, E.M., MOSSIMANN, S.C., CHERNAIA, M.M., MALCOLM, B.A. AND JAMES, M.N.G. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *Journal of Virology* **71**, 2436–2448.

BERKE, G. (1995). The CTL's kiss of death. *Cell* **81**, 9–12.

BILLMAN-JACOBE, H. (1996). Expression in bacteria other than *Escherichia coli*. *Current Opinion in Biotechnology* **7**, 500–504.

BILLMAN-JACOBE, H., WANG, L., KORTT, A., STEWART, D. AND RADFORD A. (1995). Expression and secretion of heterologous proteases by *Corynebacterium glutamicum*. *Applied Environmental Microbiology* **61**, 610–613

BINNIE, C., COSSAR, J.D. AND STEWART, D.I.H. (1997). Heterologous biopharmaceutical protein expression in *Streptomyces*. *Trends in Biotechnology* **15**, 315–320.

BIRCH, G.M., BLACK, T., MALCOLM, S.K., LAI, M.T., ZIMMERMAN, R.E. AND JASKUNAS, S.R. (1995). Purification of recombinant human rhinovirus 14 3C protease expressed in *Escherichia coli*. *Protein Expression and Purification* **6**, 609–618.

BLABER, M., ISACKSON, P.J. AND BRADSHAW, R.A. (1990). The characterization of recombinant mouse glandular kallikreins from *E. coli*. *Proteins: Structure, Function, and Genetics* **7**, 280–290.

BLIGHT, M.A. AND HOLLAND, I.B. (1994). Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. *Trends in Biotechnology* **12**, 450–455.

BRIGELIUS-FLOHÉ, R., STEFFENS, G., STRASSBURGER, W. AND FLOHÉ, L. (1992). High expression vectors for the production of recombinant single-chain urinary plasminogen activator from *Escherichia coli*. *Applied Microbiology and Biotechnology* **36**, 640–649.

BRÖMME, D. AND SCHMIDT, B.F. (1999). Functional expression of recombinant proteases. In: *Proteolytic Enzymes: Tools and Targets*, Springer Lab Manual. Eds. E.E. Sterchi and W. Stöcker (in press). Heidelberg, Germany: Springer-Verlag.

BRON, S., BOLHUIS, A., TJALSMA, H., HOLSAPPEL, S., VENEMA, G. AND VAN DIJL, J.M. (1998). Protein secretion and possible roles for multiple signal peptidases for precursor processing in bacilli. *Journal of Biotechnology* **64**, 3–13.

BURCK, P.J., BERG, D.H., LUK, T.P., SASSMANNSHAUSEN, L.M., WAKULCHIK, M., SMITH, D.P., HSIUNG, H.M., BECKER, G.W., GIBSON, W. AND VILLAREAL, E.C. (1994). Human cytomegalovirus maturation proteinase: expression in *Escherichia coli*, purification, and enzymatic characterization by using peptide substrate mimics of natural cleavage sites. *Journal of Virology* **68**, 2937–2946.

CAUGHEY, G.H. (1998). Chymase. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 66–71. New York: Academic Press.

CAWSTON, T.E. (1998). Interstitial collagenase. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1155–1162. New York: Academic Press.

CHAO, J. (1998a). Semenogelase or prostate specific antigen. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 102–104. New York: Academic Press.

CHAO, J. (1998b). Human tissue kallikrein. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 97–100. New York: Academic Press.

CHASSY, B.M. AND MURPHY, C.M. (1993). *Lactococcus* and *Lactobacillus*. In: *Bacillus subtilis and other gram-positive bacteria: biochemistry, physiology, and molecular genetics*. Eds. A. Sonenshein, J.A. Hoch and R. Losick, pp 65–82. Washington, D.C.: American Society for Microbiology.

CLARK I.M., MITCHELL, R.E., POWELL, L.K. AND BIGG, H.F. (1995). Recombinant porcine collagenase: purification and autolysis. *Archives of Biochemistry and Biophysics* **316**, 123–127.

COFFIN, J.M. (1996). Retroviridae: The viruses and their replication. In: *Fields Virology* Vol. 2, 3rd edition. Eds. B.M. Fields, D.M. Knipe and P.M. Howley, pp 1767–1848. Philadelphia: Lippincott-Raven Publishers.

COLLEN, D., ZAMARRON, C., LIJNEN, H.R. AND HOYLAERTS, M. (1986). Activation of plasminogen by pro-urokinase. II. Kinetics. *The Journal of Biological Chemistry* **261**, 1259–1266.

COLLIER, I.E. AND GOLDBERG, G.I. (1998). Gelatinase B. In: *Handbook of Proteolytic Enzymes*.

Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1205–1210. New York: Academic Press.

COLLINS-RACIE, L.A., MCCOLGAN, J.M., GRANT, K.L., DIBLASIO-SMITH, E.A., MCCOY, J.M. AND LAVALLIE, E.R. (1995). Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using a novel secretory fusion partner DsbA. *Bio/Technology* **13**, 982–987.

CONNER, G.E. (1998). Cathepsin D. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 828–836. New York: Academic Press.

CONNER, G.E. AND RICHO, G. (1992). Isolation and characterization of a stable activation intermediate of the lysosomal aspartyl protease cathepsin D. *Biochemistry* **31**, 1142–1147.

CONNER, G.E. AND UDEY, J.A. (1990). Expression and refolding of recombinant human fibroblast procathepsin D. *DNA and Cell Biology* **9**, 1–9.

CONRAD, B., SAVCHENKO, R.S., BREVES, R. AND HOFEMEISTER, J. (1996). A T7 promoter-specific, inducible protein expression system for *Bacillus subtilis*. *Molecular and General Genetics* **250**, 230–236.

CORDINGLEY, M.G., REGISTER, R.B., CALLAHAN, P.L., GARSKY, V.M. AND COLONNO, R.J. (1989). Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in *Escherichia coli*. *Journal of Virology* **63**, 5037–5045.

COREY, D.R., SHIAU, A.K., YANG, Q., JANOWSKI, B.A. AND CRAIK, C.S. (1993). Trypsin display on the surface of bacteriophage. *Gene* **128**, 129–134.

CORYDON, T.J., BROSS, P., HOLST, H.U., NEVE, S., KRISTIANSEN, K., GREGERSEN, N. AND BOLUND, L. (1998). A homologue of *Escherichia coli* ClpP caseinolytic protease: Recombinant expression, intracellular processing and subcellular localization. *Biochemical Journal* **331**, 309–316.

DANG, L.C., TALANIAN, R.V., BANACH, D., HACKETT, M.C., GILMORE, J.L., HAYS, S.J., MANKOVICH, J.A. AND BRADY, K.D. (1996). Preparation of an autolysis-resistant interleukin-1β converting enzyme mutant. *Biochemistry* **35**, 14910–14916.

DARKE, P.L., LEU, C.-T., DAVIS, L.J., HEIMBACH, J.C., DIEHL, R.E., HILL, W.S., DIXON, R.A.F. AND SIGAL, I.S. (1989). Human immunodeficiency virus protease, bacterial expression and characterization of the purified aspartic protease. *The Journal of Biological Chemistry* **264**, 2307–2312.

DARKE, P.L., CHEN, E., HALL, D.L., SARDANA, M.K., VELPSKI, A., LAFEMINA, R.L., SHAFER, J.A. AND KUO, L.C. (1994). Purification of active herpes simplex virus-1 protease expressed in *Escherichia coli*. *The Journal of Biological Chemistry* **269**, 18708–18711.

DATAR, R.V., CARTWRIGHT, T. AND ROSEN, C.-G. (1993). Process economics of animal and bacterial fermentations: a case study analysis of tissue plasminogen activator. *Bio/Technology* **11**, 349–357.

DAVIS, G.J., WANG, Q.M., COX, G.A., JOHNSON, R.B. (1997). Expression and purification of recombinant rhinovirus 14 3CD proteinase and its comparison to the 3C proteinase. *Archives of Biochemistry and Biophysics* **346** 125–130.

DEBOUCK, C., GORNIAK, J.G., STRICKLER, J.E., MEEK, T.D., METCALF, B.W. AND ROSENBERG, M. (1987). Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 8903–8906.

DIBELLA, E.E., MAURER, M.C. AND SCHERAGA, H.A. (1995). Expression and folding of recombinant bovine prethrombin-2 and its activation to thrombin. *The Journal of Biological Chemistry* **270**, 163–169.

DIIANNI, C.L., DAVIS, L.J., HOLLOWAY, M.K., HERBER, W.K., DARKE, P.L., KOHL, N.E. AND DIXON, R.A.F. (1990). Characterization of an active single polypeptide form of the human immunodeficiency virus type 1 protease. *The Journal of Biological Chemistry* **265**, 17348–17354.

DING, J., MCGRATH, W.J., SWEET, R.M. AND MANGEL, W.F. (1996). Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO Journal* **15**, 1778–1783.

DING, Y.S., OWEN, S.M., LAL, R.B. AND IKEDA, R.A. (1998). Efficient expression and rapid purification of human T-cell leukemia virus type 1 protease. *Journal of Virology* **72**, 3383–3386.

DOLINAR, M., MAGANJA, D.B. AND TURK, V. (1995). Expression of full-length human procathepsin L cDNA in *Escherichia coli* and refolding of the expression product. *Biological Chemistry Hoppe Seyler* **376**, 385–388.

EDGE, M., FORDER, C., HENNAM, J., LEE, I., TONGE, D., HARDERN, I., FITTON, J., ECKERSLEY, K., EAST, S., SHUFFLEBOTHAM, A., BLAKEY, D. AND SLATER, A. (1998). Engineered human carboxypeptidase B enzymes that hydrolyse hippuryl-L-glutamic acid: Reversed-polarity mutants. *Protein Engineering* **11**, 1229–1234.

ELLIS, V. AND DANO, K. (1998). u-Plasminogen activator In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 177–184. New York: Academic Press.

EMTAGE, J.S., ANGAL, S., DOEL, M.T., HARRIS, T.J.R., JENKINS, B., LILLEY, G. AND LOWE, P.A. (1983). Synthesis of calf prochymosin (prorenin) in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3671–3675.

FERNANDES-ALNEMRI, T., LITWACK, G. AND ALNEMRI, E.S. (1995a). Mch2, a new member of the apoptotic *Ced-3/Ice* cysteine protease gene family. *Cancer Research* **55**, 2737–2742.

FERNANDES-ALNEMRI, T., TAKAHASHI, A., ARMSTRONG, R., KREBS, J., FRITZ, L., TOMASELLI, K.J., WANG, L., YU, Z., CROCE, C.M., SALVESEN G., EARNSHAW, W.C., LIWACK, G. AND ALNEMRI, E.S. (1995b). Mch3, a novel human apoptotic cysteine protease highly related to CPP32. *Cancer Research* **55**, 6045–6052.

FERNANDES-ALNEMRI, T., ARMSTRONG, R.C., KREBS, J., SRINIVASULA, S.M., WANG, L., BULLRICH, F., FRITZ, L.C., TRAPANI, J.A., TOMASELLI, K.J., LITWACK, G. AND ALNEMRI, E.S. (1996). In vitro activation of CPP32 and Mch3 by Mch4, a novel human apoptotic cysteine protease containing two FADD-like domains. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 7464–7469.

FERRARI, E., JARNAGIN, A.S. AND SCHMIDT, B.F. (1993). Commercial production of extracellular enzymes. In: *Bacillus subtilis and other gram-positive bacteria: biochemistry, physiology, and molecular genetics*. Eds. A. Sonenshein, J.A. Hoch and R. Losick, pp 917–937. Washington, D.C.: American Society for Microbiology.

FOLTMAN, B. AND SZECSI, P.B. (1998). Chymosin. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 815–819. New York: Academic Press.

FOWLER, S.D., KAY, J., DUNN, B.M. AND TATNELL, P.J. (1995). Monomeric human cathepsin E. *FEBS Letters* **366**, 72–74.

FREIMARK, B.D., FEESER, W.S. AND ROSENFELD, S.A. (1994). Multiple sites of the propeptide region of human stromelysin-1 are required for maintaining a latent form of the enzyme. *The Journal of Biological Chemistry* **269**, 26982–26987.

GALLINARI, P., PAOLINI, C., BRENNAN, D., NARDI, C., STEINKUHLER, C. AND DE FRANCESCO, R. (1999). Modulation of hepatitis C virus NS3 protease and helicase activities through the interaction with NS4A. *Biochemistry* **38**, 5620–5632.

GAN, Z.-R., LEWIS, S.D., STONE, J.R. AND SHAFER, J.A. (1991). Reconstitution of catalytically competent human ζ-thrombin by combination of ζ1-thrombin residues A1–36 and B1–148 and an *Escherichia coli* expressed polypeptide corresponding to ζ2-thrombin residues B149–259. *Biochemistry* **30**, 11694–11699.

GEORGIOU, G. (1996). Expression of proteins in bacteria. In: *Protein Engineering: Principles and Practice*. Eds. J.L. Cleland and C.S. Craik, pp 101–127. New York: Wiley-Liss.

GEORGIOU, G. AND VALAX, P. (1996). Expression of correctly folded proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **7**, 190–197.

GIAM, C.-Z. AND BOROS, I. (1988). *In vivo* and *in vitro* autoprocessing of human immunodeficiency virus protease expressed in *Escherichia coli*. *The Journal of Biological Chemistry* **263**, 14617–14620.

GORBALENYA, A.E., KOONIN, E.V., DONCHENKO, A.P. AND BLINOV, V.M. (1989). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acid Research* **17**, 4847–4861.

GRÁF, L., CRAIK, C.S., PATTHY, A., ROCZNIAK, S., FLETTERICK, R.J. AND RUTTER, W.J. (1987). Selective alteration of substrate specificity by replacement of aspartic acid-89 with lysine in the binding pocket of trypsin. *Biochemistry* **26**, 2616–2623.

GRANT, S.K., DECKMAN, I.C., MINNICH, M.D., CULP, J., FRANKLIN, S., DREYER, G.B.,

TOMASZEK, T.A., JR., DEBOUCK, C. AND MEEK, T.D. (1991). Purification and biochemical characterization of recombinant simian immunodeficiency virus protease and comparison to human immunodeficiency virus type 1 protease. *Biochemistry* **30**, 8424–8434.

GRAVES, M.C., LIM, J.J., HEIMER, E.P. AND KRAMER, R.A. (1988). An 11-kDa form of human immunodeficiency virus protease expresssed in *Escherichia coli* is sufficient for enzymatic activity. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2449–2453.

GRIFFITHS, J.T., TOMCHAK, L.A., MILLS, J.S., GRAVES, M.C., COOK, N.D., DUNN, B.M. AND KAY, J. (1994). Interactions of substrates and inhibitors with a family of tethered HIV-1 and HIV-2 homo- and heterodimeric proteinases. *The Journal of Biological Chemistry* **269**, 4787–4793.

GRONSKI, T.J., MARTIN, R.L., KOBAYASHI, D.K., WALSH, B.C., HOLMAN, M.C., HUBER, M., VAN WART, H.E. AND SHAPIRO, S.D. (1997). Hydrolysis of a broad spectrum of extracellular matrix proteins by human macrophage elastase. *The Journal of Biological Chemistry* **272**, 12189–12194.

GUISE, A.D., WEST, S.M. AND CHAUDHURI, J.B. (1996). Protein folding *in vivo* and renaturation of recombinant proteins from inclusion bodies. *Molecular Biotechnology* **6**, 53–64.

GUMPERT, J. AND HOISCHEN, C. (1998). Use of cell wall-less bacteria (L-forms) for efficient expression and secretion of heterologous gene products. *Current Opinion in Biotechnology* **9**, 506–509.

HALFON, S. AND CRAIK, C.S. (1998). Trypsin. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 12–21. New York: Academic Press.

HALL, D.J. AND PALMENBERG, A.C. (1996). Mengo virus 3C proteinase: recombinant expression, intergenus substrate cleavage and localization in vivo. *Virus Genes* **13**, 99–110.

HAN, B.G., MA, X.K., MENG, L., SONG, X.G., PENG, S.Y., WANG, J.X. AND LING, S.G. (1998). Thioredoxin fusion/HIV-1 protease coexpression system for production of soluble human IL-6 in *E. coli* cytoplasm. *Biochemistry Molecular Biology International* **46**, 839–846.

HANNIG, G. AND MAKRIDES, S.C. (1998). Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology* **16**, 54–60.

HARWOOD, C. (1992). *Bacillus subtilis* and its relatives: molecular and industrial workhorses. *Trends in Biotechnology* **10**, 247–256.

HENKART, P.A. (1994). Lymphocyte-mediated cytotoxicity: two pathways and multiple effector molecules. *Immunity* **1**, 343–346.

HENNER, D.J. (1990). Inducible expression of regulatory genes in *Bacillus subtilis*. *Methods in Enzymology* **185**, 223–228.

HIGAKI, J.N., EVNIN, L.B. AND CRAIK, C.S. (1989). Introduction of a cysteine protease active site into trypsin. *Biochemistry* **28**, 9256–9263.

HILL, J., MONTGOMERY, D.S. AND KAY, J. (1993). Human cathepsin E produced in *E. coli*. *FEBS Letters* **326**, 101–104.

HO, T.F., QORONFLEH, M.W., WAHL, R.C., PULVINO, T.A., VAVRA, K.J., FALVO, J., BANKS, T.M., BRAKE, P.G. AND CICCARELLI, R.B. (1994). Gene expression, purification and characterization of recombinant human neutrophil collagenase. *Gene* **146**, 297–301.

HOCH, J.A. (1991). Genetic analysis in *Bacillus subtilis*. *Methods in Enzymology* **204**, 305–320.

HOCKNEY, R.C. (1994). Recent developments in heterologous protein production in *Escherichia coli*. *Trends in Biotechnology* **12**, 456–463.

HOLMES, W.E., PENNICA, D., BLABER, M., REY, M.W., GUENZLER, W.A., STEFFENS, G.J. AND HEYNEKER, H.L. (1985). Cloning and expression of the gene for pro-urokinase in *Escherichia coli*. *Bio/Technology* **3**, 923–929.

HOPFNER, K.-P., BRANDSTETTER, H., KARCHER, A., KOPETZKI, E., HUBER, R., ENGH, R.A. AND BODE, W. (1997). Converting blood coagulation factor IXa into factor Xa: dramatic increase in amidolytic activity identifies important active site determinants. *The EMBO Journal* **16**, 6626–6635.

HRUSKOVÁ-HEIDINGSFELDOVÁ, O., ANDREANSKY, M., FÁBRY, M., BLÁHA, I, STROP, P. AND HUNTER, E. (1995). Cloning, bacterial expression and characterization of the Mason-Pfizer monkey virus proteinase. *The Journal of Biological Chemistry* **270**, 15053–15058.

HUA, Z.C., DONG, C., ZHU, D.X. (1996). Renaturation of recombinant human pro-urokinase

expressed in *Escherichia coli*. *Biochemistry and Biophysics Research Communications* **220**, 131–136.

IMAI, T., CHO, T., TAKAMATSU, H., HORI, H., SAITO, M., MASUDA, T., HIROSE, S. AND MURAKAMI, K. (1986). Synthesis and characterization of human prorenin in *Escherichia coli*. *Journal of Biochemistry* **100**, 425–432.

ITOH, M., MASUDA, K., ITO, Y., AKIZAWA, T., YOSHIOKA, M., IMAI, K., OKADA, Y., SATO, H. AND SEIKI, M. (1996). Purification and refolding of recombinant human proMMP-7 (promatrilysin) expressed in *Escherichia coli* and its characterization. *Journal of Biochemistry* **119**, 667–673.

JANNIÈRE, L., GRUSS, A. AND EHRLICH, S.D. (1993). Plasmids. In: *Bacillus subtilis and other gram-positive bacteria: biochemistry, physiology, and molecular genetics*. Eds. A. Sonenshein, J.A. Hoch and R. Losick, pp 625–644. Washington, D.C.: American Society for Microbiology.

JENG, A.Y., WONG, M., DUELFER, T., SHAPIRO, S.D., KRAMER, R.A. AND HU, S. (1995). Mouse macrophage metalloelastase expressed in bacteria absolutely requires zinc for activity. *Journal of Biochemistry* **117**, 216–221.

KAGITANI, H., TAGAWA, M., HATANAKA, K., IKARI, T., SAITO, A., MANDO, H., OKADA, K. AND MATSUO, O. (1985). Expression in *E. coli* of finger-domain lacking tissue-type plasminogen activator with high fibrin affinity. *FEBS Letters* **189**, 145–149.

KAMENS, J., PASKIND, M., HUGUNIN, M., TALANIAN, R.V., ALLEN, H., BANACH, D., BUMP, N., HACKETT, M., JOHNSTON, C.G., LI, P., MANKOVICH, J.A., TERRANOVA, M. AND GHAYUR, T. (1995). Identification and characterization of ICH-2, a novel member of the interleukin-1β-converting enzyme family of cysteine proteases. *The Journal of Biological Chemistry* **270**, 15250–15256.

KAPRÁLEK, F., JECMEN, P., SEDLÁCEK, J., FÁBRY, M. AND ZADRAZIL, S. (1990). Fermentation conditions for high-level expression of the *tac*-promoter-controlled calf prochymosin cDNA in *Eshericia coli* HB101. *Biotechnology and Bioengineering* **37**, 71–79.

KAWAGUCHI, Y., YANAGIDA, N., UOZUMI, T. AND BEPPU, T. (1986). Improved direct expression of prochymosin cDNA through changing SD-ATG codon length. *Agricultural and Biological Chemistry* **50**, 499–500.

KAY, J. AND TATNELL, P.J. (1998). Cathepsin E. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 819–823. New York: Academic Press.

KINOSHITA, T., SATO, H., OKADA, A., OHUCHI, E., IMAI, K., OKADA, Y. AND SEIKI, M. (1998). TIMP-2 promotes activation of progelatinase A by membrane-type 1 matrix metalloproteinase immobilized on agarose beads. *The Journal of Biological Chemistry* **273**, 16098–16103.

KLESSEN, C., SCHMIDT, K.-H., GUMPERT, J., GROSSE, H.-H. AND MALKE, H. (1989). Complete secretion of activable bovine prochymosin by genetically engineered L forms of *Proteus mirabilis*. *Applied and Environmental Microbiology* **55**, 1009–1015.

KNOTT, J.A., ORR, D.C., MONTGOMERY, D.S., SULLIVAN, C.A. AND WESTON, A. (1989). The expression and purification of human rhinovirus protease 3C. *European Journal of Biochemistry* **182**, 547–555.

KOPITAR, G., DOLINAR, M., STRUKELJ, B., PUNGERCAR, J. AND TURK, V. (1996). Folding and activation of human procathepsin S from inclusion bodies produced in *Escherichia coli*. *European Journal of Biochemistry* **236**, 558–562.

KRÖGER, M. AND TSCHESCHE, H. (1997). Cloning, expression and activation of a truncated 92-kDa gelatinase minienzyme. *Gene* **196**, 175–180.

KUHELJ, R., DOLINAR, M., PUNGERCAR, J. AND TURK, V. (1995). The preparation of catalytically active human cathepsin B from its precursor expressed in *Escherichia coli* in the form of inclusion bodies. *European Journal of Biochemistry* **229**, 553–539.

KUO, G., CHOO, Q.L., ALTER, H.J., GITNICK, G.L., REDEKER, A.G., PURCELL, R.H., MIYAMURA, T., DIENSTAG, J.L., ALTER, M.J., STEVENS, C.E., TEGTMEIER, G., BONINO, F., COLOMBO, M., LEE, W.-S., KUO, C., BERGER, K., SHUSTER, J.R., OVERBY, L.R., BRADLEY, D.W. AND HOUGHTON, M. (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362–364.

LAFEMINA, R.L., BAKSHI, K., LONG, W.J., PRAMANIK, B., VELOSKI, C.A., WOLANSKI, B.S.,

MARCY, A.I. AND HAZUDA, D.J. (1996). Characterization of a soluble stable human cytomegalovirus protease and inhibition by M-site peptide mimetics. *Journal of Virology* **70**, 4819–4824.

LAM, K.H., CHOW, K.C. AND WONG, W.K. (1998). Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous proteins. *Journal of Biotechnology* **63**, 167–177.

LAVALLIE, E.R. AND MCCOY, J.M. (1995). Gene fusion expression systems in *Escherichia coli*. *Current Opinion in Biotechnology* **6**, 501–506.

LAWSON, M.A. AND SEMLER, B.L. (1990). Picornavirus protein processing: enzymes, substrates and genetic regulation. *Current Topics in Microbiology and Immunology* **161**, 49–87.

LE GRICE, S.F.J. (1990). Regulated promoter for high-level expression of heterologous genes in *Bacillus subtilis*. *Methods in Enzymology* **185**, 201–214.

LEONG, L.E.-C., WALKER, P.A. AND PORTER, A.G. (1992). Efficient expression and purification of a protease from the common cold virus, human rhinovirus type 14. *Journal of Crystal Growth* **122**, 246–252.

LI, S.-L. AND HOCHSTRASSER, M. (1999). A new protease required for cell-cycle progression in yeast. *Nature* **398**, 246–251.

LIBBY, R.T., COSMAN, D., COONEY, M.K., MERRIAM, J.E., MARCH, C.J AND HOPP, T.P. (1988). Human Rhinovirus 3C protease: cloning and expression of an active form in *Escherichia coli*. *Biochemistry* **27**, 6262–6268.

LICHTE, A., KOLKENBROCK, H. AND TSCHESCHE, H. (1996). The recombinant catalytic domain of membrane-type matrix metalloproteinase-1 (MT1-MMP) induces activation of progelatinase A and progelatinase A complexed with TIMP-2. *FEBS Letters* **397**, 277–282.

LIEBIG, H.D., ZIEGLER, E., YAN, R., HARTMUTH, K., KLUMP, H., KOWALSKI, H., BLAAS, D., SOMMERGRUBER, W., FRASEL, L., LAMPHERA, B., RHOADS, R., KUECHLER, E. AND SKERN, T. (1993). Purification of two picornaviral 2A proteinases: interaction with eIF-4γ and influence on *in vitro* translation. *Biochemistry* **32**, 7581–7588.

LIJNEN, H.R. AND COLLEN, D. (1998). t-Plasminogen activator. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 184–190. New York: Academic Press.

LIJNEN, H.R., VAN HOEF, B. AND COLLEN, D. (1986). Comparative kinetic analysis of the activation of human plasminogen and recombinant single-chain urokinase-type plasminogen activator. *Biochimica et Biophysica Acta* **884**, 402–408.

LILIE, H. SCHWARZ, E. AND RUDOLPH, R. (1998). Advances in refolding of proteins produced in *E. coli*. *Current Opinion in Biotechnology* **9**, 497–501.

LIN, X.L., WONG, R.N.S. AND TANG, J. (1989). Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen. *The Journal of Biological Chemistry* **264**, 4482–4489.

LIN, X.L., LIN, Y.-Z. AND TANG, J. (1994). Relationships of human immunodeficiency virus protease with eukaryotic aspartic proteases. *Methods in Enzymology* **241**, 195–224.

LLANO, E., PENDÁS, A.M., KNÄUPER, V., SORSA, T., SALIDO, E., MURPHY, G., SIMMER, J.P., BARTLETT, J.D. AND LOPEZ-OTÍN, C. (1997). Identification and structural and functional characterization of human enamelysin (MMP-20). *Biochemistry* **36**, 15101–15108.

LOVE, R.A., PARGE, H.E., WICKERSMAN, J.A., HOSTOMSKY, Z., HABUKA, N., MOOMAW, E.W., ADACHI, T. AND HOSTOMSKA, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **87**, 331–342.

LU, H.S., HSU, Y.-R., LU, L.I., RUFF, D., LYONS, D. AND LIN, F.-K. (1996). Isolation and characterization of human tissue kallikrein produced in *Escherichia coli*: biochemical comparison to the enzymatically inactive prokallikrein and methionyl kallikrein. *Protein Expression and Purification* **8**, 215–226.

MAKRIDES, S.C. (1996). Strategies of achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews* **60**, 512–538.

MALCOLM, B.A., CHIN, S.M., JEWELL, D.A., STRATTON-THOMAS, J.R., THUDIUM, K.B., RALSTON, R. AND ROSENBERG, S. (1992). Expression and characterization of recombinant hepatitis A virus 3CD proteinase. *Biochemistry* **31**, 3358–3363.

MALINOWSKI, J.J., GRASBERGER, B.L., TRAKSHEL, G., HUSTON, E.E., HELASZEK, C.T., SMALLWOOD, A.M., ATOR, M.A., BANKS, T.M., BRAKE, P.G., CICCARELLI, R.B., JONES, B.N., KOEHN, J.A., KRATZ, K., LUNDBERG, N., STAMS, T., RUBIN, B., ALEXANDER, R.S. AND STEVIS, P.E. (1995). Production, purification, and crystallization of human interleukin-1β converting enzyme derived from an *Escherichia coli* expression system. *Protein Science* **4**, 2149–2155.

MARCY A.I., EIBERGER, L.L., HARRISON, R., CHAN, H.K., HUTCHINSON, N.I., HAGMANN, W.K., CAMERON, P.M., BOULTON, D.A. AND HERMES, J.D. (1991). Human fibroblast stromelysin catalytic domain: expression, purification, and characterization of a C-terminally truncated form. *Biochemistry* **30**, 6476–6483.

MARSTON, F.A.O., LOWE, P.A., DOEL, M.T., SCHOEMAKER, J.M., WHITE, S. AND ANGAL, S. (1984). Purification of calf prochymosin (prorenin) synthesized in *Escherichia coli*. *Bio/Technology* **2**, 800–804.

MATTHEWS, D.A., SMITH W.W., FERRE, R.A., CONDON, B., BUDAHAZI, G., SISSON, W., VILLAFRANCA, J.E., JANSON, C.A., MCELROY, H.E., GRIBSKOV, C.L. AND WORLAND, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**, 761–771.

MATRISIAN, L.M. (1998). Stromelysin-1. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1172–1178. New York: Academic Press.

MCCAMAN, M.T., ANDREWS, W.H. AND FILES, J.G. (1985). Enzymatic properties and processing of bovine prochymosin in *Esherichia coli*. *Journal of Biotechnology* **2**, 177–190.

MCGRATH, M.E., OSAWA, A.E., BARNES, M.G., CLARK, J.M., MORTARA, K.D. AND SCHMIDT, B.F. (1997). Production of crystallizable human chymase from a *Bacillus subtilis* system. *FEBS Letters* **413**, 486–488.

MCKIE, N., DANDO, P.M., BROWN, M.A. AND BARRETT, A.J. (1995). Rat thimet oligopeptidase: large-scale expression in *Escherichia coli* and characterization of the recombinant enzyme. *Biochemical Journal* **309**, 203–207.

MEERMAN, H.J. AND GEORGIOU, G. (1994). Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. *Bio/Technology* **12**, 1107–1110.

MELNICK, L., YANG, S.S., ROSSI, R., ZEPP, C. AND HEEFNER, D. (1998). An *Escherichia coli* expression assay and screen for human immunodeficiency virus protease variants with decreased susceptibility to indinavir. *Antimicrobiological Agents Chemotherapy* **42**, 3256–3265.

MENÉNDEZ-ARIAS, L., GOTTE, D. AND OROSZLAN, S. (1993). Moloney murine leukemia virus protease: Bacterial expression and characterization of the purified enzyme. *Virology* **196**, 557–563.

MILLER, M., JASKOLSKI, M., RAO, J.K.M., LEIS, J. AND WLODAWER, A. (1989a). Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **337**, 576–579.

MILLER, M., SCHNEIDER, J., SATHYANARAYANA, B.K., TOTH, M.V., MARSHALL, G.R., CLAWSON, L., SELK, L., KENT, S.B.H. AND WLODAWER, A. (1989b). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* **246**, 1149–1152.

MITTA, M., MIYAGI, M., KATO, I. AND TSUNASAWA, S. (1998). Identification of the catalytic triad residues of porcine liver acylamino acid-releasing enzyme. *Journal of Biochemistry* **123**, 924–931.

MOLINEAUX, S.M., CASANO, F.J., ROLANDO, A.M., PETERSON, E.P., LIMJUCO, G., CHIN, J., GRIFFIN, P.R., CALAYCAY, J.R., DING, G.J.-F., YAMIN, T.-T., PALYHA, O.C., LUELL, S., FLETCHER, D., MILLER, D.K., HOWARD, A.D., THORNBERRY, N.A. AND KOSTURA, M.J. (1993). Interleukin 1β (IL-1β) processing in murine macrophages requires a structurally conserved homologue of human IL-1β converting enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 1809–1813.

MORI, A., YAMADA, K., KIMURA, J., KOIDE, T., YUASA, S., YAMADA, E. AND MIYAMURA, T. (1996). Enzymatic characterization of purified NS3 serine proteinase of hepatitis C virus expressed in *Escherichia coli*. *FEBS Letters* **378**, 37–42.

MOSIMANN, S.C., CHERNEY, M.M., SIA, S., PLOTCH, S. AND JAMES, M.N.G. (1997). Refined x-ray crystallographic structure of the poliovirus 3C gene product. *Journal of Molecular Biology* **273**, 1032–1047.

MURBY, M., UHLÉN, M. AND STÅHL, S. (1996). Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*. *Protein Expression and Purification* **7**, 129–136.

MURPHY, G. (1998). Gelatinase A. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1199–1205. New York: Academic Press.

MURTHY. H.M.K., CLUM, S. AND PADMANABHAN, R. (1999). Dengue virus NS3 serine protease. *The Journal of Biological Chemistry* **274**, 5573–5580.

NAGARAJAN, V. (1990). System for secretion of heterologous proteins in *Bacillus subtilis*. *Methods in Enzymology* **185**, 214–223.

NAGARAJAN, V. (1993). Secretion. In: *Bacillus subtilis and other gram-positive bacteria: biochemistry, physiology, and molecular genetics*. Eds. A. Sonenshein, J.A. Hoch and R. Losick, pp 713–726. Washington, D.C.: American Society for Microbiology.

NAGASE, H. AND WOESSNER, J.F. (1999). Matrix Metalloproteases. *The Journal of Biological Chemistry* **274**, 21491–21494.

NISHIMORI, K., KAWAGUCHI, Y., HIDAKA, M., UOZUMI, T. AND BEPPU, T. (1982). Expression of cloned calf prochymosin in *Escherichia coli*. *Gene* **19**, 337–344.

NISHIMORI, K., SHIMIZU, N., KAWAGUCHI, Y., HIDAKA, M., UOZUMI, T. AND BEPPU, T. (1984). Expression of cloned calf prochymosin cDNA under control of the tryptophan promoter. *Gene* **29**, 41–49.

OBUKOWICZ, M.G., GUSTAFSON, M.E., JUNGER, K.D., LEIMGRUBER, R.M., WITTWER, A.J., WUN, T.-C., WARREN, T.G., BISHOP, B.F., MATHIS, K.J., MCPHERSON, D.T., SIEGEL, N.R., JENNINGS, M.G., BRIGHTWELL, B.B., DIAZ-COLLIER, J.A., BELL, L.D., CRAIK, C.S. AND TACON, W.C. (1990). Secretion of active kringle-2-serine protease in *Escherichia coli*. *Biochemistry* **29**, 9737–9745.

O'FARRELL, T.J. AND POURMOTABBED, T. (1998). The fibronectin-like domain is required for the type V and XI collagenolytic activity of gelatinase B. *Archives of Biochemistry and Biophysics* **354**, 24–30.

O'HARE, M.C., CLARKE, N.J. AND CAWSTON, T.E. (1992). Production by *Escherichia coli* of porcine type-I collagenase as a fusion protein with β -galactosidase. *Gene* **111**, 245–248.

O'HARE, M.C., CURRY, V.A., MITCHELL, R.E. AND CAWSTON, T.E. (1995). Stabilization of purified human collagenase by site-directed mutagenesis. *Biochemical and Biophysical Research Communications* **216**, 329–337.

PARENTE, D., DE FERRA, F., GALLI, G. AND GRANDI, G. (1991). Prochymosin expression in *Bacillus subtilis*. *FEMS Microbiology Letters* **77**, 243–250.

PENDÁS, A.M., KNÄUPER, V., PUENTE, X.S., LLANO, E., MATTEI, M.-G., APTE, S., MURPHY, G. AND LÓPEZ-OTÍN, C. (1997). Identification and characterization of a novel human matrix metalloproteinase with unique structural characteristics, chromosomal location, and tissue distribution. *The Journal of Biological Chemistry* **272**, 4281–4286.

PENNICA, D., HOLMES, W.E., KOHR, W.J., HARKINS, R.N., VEHAR, G.A., WARD, C.A., BENNETT, W.F., YELVERTON, E., SEEBURG, P.H., HEYNEKER, H.L., GOEDDEL, D.V. AND COLLEN, D. (1983). Cloning and expression of human tissue-type plasminogen activator cDNA in *E. coli*. *Nature* **301**, 214–221.

PEREGO, M. (1993). Integrational vectors for genetic manipulation in *Bacillus subtilis*. In: *Bacillus subtilis and other gram-positive bacteria: biochemistry, physiology, and molecular genetics*. Eds. A. Sonenshein, J.A. Hoch and R. Losick, pp 615–624. Washington, D.C.: American Society for Microbiology.

PERO, J. AND SLOMA, A. (1993). Proteases. In: *Bacillus subtilis and other gram-positive bacteria: biochemistry, physiology, and molecular genetics*. Eds. A. Sonenshein, J.A. Hoch and R. Losick, pp 939–952. Washington, D.C.: American Society for Microbiology.

PERONA, J.J. AND CRAIK C.S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Science* **4**, 337–360.

PETANCESKA, S. AND DEVI, L. (1992). Sequence analysis, tissue distribution, and expression of rat cathepsin S. *The Journal of Biological Chemistry* **267**, 26038–26043.

PETERSEN, J.F., CHERNEY, M.M., LIEBIG, H.D., SKERN, T., KUECHLER, E. AND JAMES, M.N. (1999). The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J.* **18**, 5463–5475.

PFREPPER, K.-I., LÖCHELT, M., SCHNÖLZER, M. AND FLÜGEL, R.M. (1997). Expression and molecular characterization of an enzymatically active recombinant human spumaretrovirus protease. *Biochemical and Biophysical Research Communications* **237**, 548–553.

PILON, A., YOST, P., CHASE, T.E., LOHNAS, G., BURKETT, T., ROBERTS, S. AND BENTLEY, W.E. (1997). Ubiquitin fusion technology: bioprocessing of peptides. *Biotechnology Progress* **13**, 374–379.

PINES, O. AND INOUYE, M. (1997). Expression and secretion of proteins in *E. coli*. *Methods in Molecular Biology* **62**, 73–87.

POURMOTABBED, T., SOLOMON, T.L., HASTY, K.A. AND MAINARDI, C.L. (1994). Characteristics of 92 kDa type IV collagenase/gelatinase produced by granulocytic leukemia cells: Structure, expression of cDNA in *E. coli* and enzymic properties. *Biochimica et Biophysica Acta* **1204**, 97–107.

QIU, J., SWARTZ, J.R. AND GEORGIOU, G. (1998). Expression of active human tissue-type plasminogen activator in *Escherichia coli*. *Applied and Environmental Microbiology* **64**, 4891–4896.

QIU, X., JANSON, C.A., CULP, J.S., RICHARDSON, S.B., DEBOUCK, C., SMITH, W.W. AND ABDEL-MEGUID, S.S. (1997). Crystal structure of varicella-zoster virus protease. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 2874–2879.

RAMAGE, P., CHENEVAL, D., CHVEI, M., GRAFF, P., HEMMIG, R., HENG, R., KOCHER, H.P., MACKENZIE, A., MEMMERT, K., REVESZ, L. AND WISHART, W. (1995). Expression, refolding, and autocatalytic proteolytic processing of the interleukin-1β-converting enzyme precursor. *The Journal of Biological Chemistry* **270**, 9378–9383.

RANCOURT, C., TIHANYI, K., BOURBONNIERE, M. AND WEBER, J.M. (1994). Identification of active-site residues of the adenovirus endopeptidase. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 844–847.

RAWLINGS, N.D. (1998). Introduction: Clan CA containing papain and its relatives. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 546–555. New York: Academic Press.

RICE, C.M., LENCHES, E.M., EDDY, S.R., SHIN, S.R., SHEETS, R.L. AND STRAUSS, J.H. (1985). Nucleotide sequence of yellow fever virus: implication of flavivirus gene expression and evolution. *Science* **229**, 726–733.

ROSÉ, J.R. AND CRAIK, C.S. (1996). The art of expression: sites and strategies for heterologous expression. In: *Protein Engineering and Design*. Ed. P.R. Carey, pp 75–104. New York: Academic Press.

ROSENFELD, S.A., ROSS, O.H., CORMAN, J.I., PRATTA, M.A., BLESSINGTON, D.L., FEESER, W.S. AND FREIMARK, B.D. (1994). Production of human matrix metalloproteinase 3 (stromelysin) in *Escherichia coli*. *Gene* **139**, 281–286.

ROTONDA, J., NICHOLSON, D.W., FAZIL, K.M., GALLANT, M., GAREAU, Y., LABELLE, M., PETERSON, E.P., RASPER, D.M., RUEL, R., VAILLANCOURT, J.P., THORNBERRY, N.A. AND BECKER, J.W. (1996). The three-dimensional structure of apopain/CPP32, a key mediator of apoptosis. *Nature Structural Biology* **3**, 619–625.

ROZZELLE, J.E., DAUBER, D.S., KELLEY, R. AND CRAIK, C.S. (2000) Macromolecular inhibitors of HIV-1 protease: characterization of designed heterodimers. *The Journal of Biological Chemistry* **275**, 7080–7086.

RUECKERT, R.R. (1996). Picornaviridae: the viruses and their replication. In: *Fields Virology* Vol. 1, 3rd editon. Eds. B.N. Fields, D.M. Knipe and P.M. Howley, pp 609–654. Philadelphia: Lippincott-Raven Publishers.

SAAVEDRA-ALANIS, V.M., RYSAVY, P., ROSENBERG, L.E. AND KALOUSEK, F. (1994). Rat liver mitochondrial processing peptidase. *The Journal of Biological Chemistry* **269**, 9284–9288.

SACHDEV, D. AND CHIRGWIN, J.M. (1998). Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expression and Purification* **12**, 122–132.

SAITO, Y., ISHII, Y., SASAKI, H., HAYASHI, M., FUJIMURA, T., IMAI, Y., NAKAMURA, S., SUZUKI,

S., NOTANI, J., ASADA, T., HORIAI, H., KOBAYASHI, M. AND NIWA, M. (1994). Production and characterization of a novel tissue-type plasminogen derivative in *Escherichia coli*. *Biotechnology Progress* **10**, 472–479.

SALTO, R., BABÉ, L.M., LI, J., ROSÉ, J.R., YU, Z., BURLINGAME, A., DE VOSS, J.J., SUI, Z., ORTIZ DE MONTELLANO, P. AND CRAIK, C.S. (1994). *In vitro* characterization of nonpeptide irreversible inhibitors of HIV proteases. *The Journal of Biological Chemistry* **269**, 10691–10698.

SANDKVIST, M. AND BAGDASARIAN, M. (1996). Secretion of recombinant proteins by gram-negative bacteria. *Current Opinion in Biotechnology* **7**, 505–511.

SANTAMARÍA, I., VELASCO, G., CAZORLA, M., FUEYO, A., CAMPO, E. AND LOPEZ-OTÍN, C. (1998a). Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. *Cancer Research* **58**, 1624–1630.

SANTAMARÍA, I., VELASCO, G., PENDÁS, A.M., FUEYO, A. AND LOPEZ-OTÍN, C. (1998b). Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *The Journal of Biological Chemistry* **273**, 16816–16823.

SANTAMARÍA, I., VELASCO, G., PENDÁS, A.M., PAZ, A. AND LOPEZ-OTÍN, C. (1999). Molecular cloning and structural and functional characterization of human cathepsin F, a new cysteine proteinase of the papain family with a long propeptide domain. *The Journal of Biological Chemistry* **274**, 13800–13809.

SARMIENTOS, P., DUCHESNE, M., DENÈFLE, P., BOIZIAU, J., FROMAGE, N., DELPORTE, N., PARKER, F., LELIÈVRE, Y., MAYAUX, J.-F. AND CARTWRIGHT, T. (1989). Synthesis and purification of active human tissue plasminogen activator from *Escherichia coli*. *Bio/Technology* **7**, 495–501.

SATO, H., KINOSHITA, T., TAKINO, T., NAKAYAMA, K. AND SEIKI, M. (1996).Activation of a recombinant membrane type 1-matrix metalloproteinase (MT1-MMP) by furin and its interaction with tissue inhibitor of metalloproteinases (TIMP)-2. *FEBS Letters* **393**, 101–104.

SCARBOROUGH, P.E. AND DUNN, B.M. (1994). Redesign of the substrate specificity of human cathepsin D: the dominant role of position 287 in the S_2 subsite. *Protein Engineering* **7**, 495–502.

SCHNIERER, S., KLEINE, T., GOTE, T., HILLEMANN, A., KNÄUPER, V. AND TSCHESCHE, H. (1993). The recombinant catalytic domain of human neutrophil collagenase lacks type I collagen substrate specificity. *Biochemical and Biophysical Research Communications* **191**, 319–326.

SCHWEIZER, H.P. AND KARKHOFF-SCHWEIZER, R.R. (1997). Design of bacterial hosts for *lac*-based expression vectors. *Methods in Molecular Biology* **62**, 17–28.

SEIKI, M. (1998). Membrane-type matrix metalloproteinase. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1192–1195. New York: Academic Press.

SENIOR, R.M. AND SHAPIRO, S.D. (1998). Macrophage elastase. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1180–1183. New York: Academic Press.

SHAPIRO, S.D., KOBAYASHI, D.K. AND LEY, T.J. (1993). Cloning and characterization of a unique elastolytic metalloproteinase produced by human alveolar macrophages. *The Journal of Biological Chemistry* **268**, 23824–23829.

SIMONS, G., RUTTEN, G., HORNES, M., NIJHUIS, M. AND VAN ASSELDONK, M. (1991). Production of prochymosin in *lactococci*. *Advances in Experimental Medical Biology* **306**, 115–119.

SMITH, S.M. AND GOTTESMAN, M.M. (1989). Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *The Journal of Biological Chemistry* **264**, 20487–20495.

SOMMER, J. (1993). Synthesis of mammalian prolyl endopeptidase in *Escherichia coli* and analysis of the recombinant protein. *Biochimica et Biophysica Acta* **1173**, 289–293.

SOMMERGRUBER, W., ZORN, M., BLAAS, D., FESSL, F., VOLKMANN, P., MAURER-FOGY, I., PALLAI, P., MERLUZZI, V., MATTEO, M., SKERN, T. AND KUECHLER, E. (1989). Polypeptide 2A of human rhinovirus type 2: identification as a protease and characterization by mutational analysis. *Virology* **169**, 68–77.

SRINIVASULA, S., AHMAD, M., FERNANDES-ALNEMRI, T., LITWACK, G. AND ALNEMRI, E.S. (1996). Molecular ordering of the Fas-apoptotic pathway: The Fas/APO-1 protease Mch5 is a CrmA-inhibitable protease that activates multiple Ced-3/ICE-like cysteine proteases. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 14486–14491.

STÅHL, S., NYGREN, P-Å. AND UHLÉN, M. (1997). Strategies for gene fusions. *Methods in Molecular Biology* **62**, 37–54.

STENFLO, J. (1998). Coagulation factor X. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 163–167. New York: Academic Press.

STENNICKE, H.R., DEVERAUX, Q.L., HUMKE, E.W., REED, J.C., DIXIT, V.M. AND SALVESEN, G.S. (1999). Caspase-9 can be activated without proteolytic processing. *The Journal of Biological Chemistry* **274**, 8359–8362.

STONE, S.R. AND LEBONNIEC, B.F. (1998). Thrombin. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 168–174. New York: Academic Press.

STORER, A.C. AND MÉNARD, R. (1994). Catalytic mechanism in papain family of cysteine proteases. *Methods in Enzymology* **244**, 486–500.

STUDIER, F.W., ROSENBERG, A.H., DUNN, J.J. AND DUBENDORFF, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185**, 60–89.

SUZUKI, F., MURAKAMI, K., NAKAMURA, Y. AND INAGAMI, T. (1998a). Renin. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 851–856. New York: Academic Press.

SUZUKI, K., KAN, C.-C., HUNG, W., GEHRING, M.R., BREW, K. AND NAGASE, H. (1998b). Expression of human pro-matrix metalloproteinase 3 that lacks the N-terminal 34 residues in *Escherichia coli* and interaction with tissue inhibitor of metalloproteinase 1 (TIMP-1). *Biological Chemistry* **379**, 185–191.

TAKAO, M., MORIOKA, T., YAMAGATA, H., TSUKAGOSHI, N. AND UDAKA, U. (1989). Production of swine pepsinogen by protein-producing *Bacillus brevis* carrying swine pepsinogen cDNA. *Applied Microbiology and Biotechnology* **30**, 75–80.

TAKAYAMA, T.K., FUJIKAWA, K. AND DAVIE, E.W. (1997). Characterization of the precursor of prostate-specific antigen. *The Journal of Biological Chemistry* **272**, 21582–21588.

TALANIAN, R.V., DANG, L.C., FERENT, C.R., HACKETT, M.C., MANKOVICH, J.A., WELCH, J.P., WONG, W.W. AND BRADY, K.D. (1996). Stability and oligomeric equilibria of refolded interleukin-1β converting enzyme. *The Journal of Biological Chemistry* **271**, 21853–21858.

TANAKA, T. AND YADA, R.Y. (1996). Expression of soluble cloned porcine pepsinogen A in *Escherichia coli*. *Biochemical Journal* **315**, 443–446.

TANG, J. (1998). Pepsin A. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 805–814. New York: Academic Press.

TANG, W., SUN, Z.Y., PANNELL, R., GUREWICH, V. AND LIU, J.N. (1997). An efficient system for production of recombinant urokinase-type plasminogen activator. *Protein Expression and Purification* **11**, 279–283.

TAREMI, S.S., BEYER, B., MAHER, M., YAO, N., PROSISE, W., WEBER, P. AND MALCOLM, B.A. (1998). Construction, expression and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease. *Protein Science* **7**, 2143–2149.

TATNELL, P.J., LEES, W.E. AND KAY, J. (1997). Cloning, expression and characterisation of murine procathepsin E. *FEBS Letters* **408**, 62–66.

THORNBERRY, N.A. (1998). Caspase-1. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 732–737. New York: Academic Press.

TOMASSELLI, A.G., PADDOCK, D.J., CURRY, K.A., GARLICK, R.L., LEONE, J.W., LULL, J.M., MUTCHLER, V.T. AND BAKER, C.A. (1998). Recombinant human cytomegalovirus protease with a C-terminal (His) 6 extension: purification, autocatalytic release of the mature enzyme, and biochemical characterization. *Protein Expression and Purification* **14**, 343–352.

TOMEI, L., FAILLA, C., SANTOLINI, E., DE FRANCESCO, R. AND LA MONICA, N. (1993). NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *Journal of Virology* **67**, 4017–4026.

TSCHESCHE, H. AND PIEPER, M. (1998). Neutrophil collagenase. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1162–1167. New York: Academic Press.

TSUKAGOSHI, N., ANDO, Y., TOMITA, Y., UCHIDA, R., TAKEMURA, T., SASAKI, T., YAMAGATA, H., UDAKA, S., ICHIHARA, Y. AND TAKHASHI, K. (1988). Nucleotide sequence and expression in *Escherichia coli* of cDNA of swine pepsinogen: involvement of the amino-terminal portion of the activation peptide segment in restoration of the functional protein. *Gene* **65**, 285–292.

UDAKA, S. AND YAMAGATA, H. (1993). High-level secretion of heterologous proteins by *Bacillus brevis*. *Methods in Enzymology* **21**, 23–33.

UNAL, A., PRAY, T.R., LAGUNOFF, M., PENNINGTON, M.W., GANEM, D. AND CRAIK, C.S. (1997). The protease and the assembly protein of Kaposi sarcoma-associated herpevirus (human herpesvirus 8). *Journal of Virology* **71**, 7030–7038.

URADE, R., ODA, T., ITO, H., MORIYAMA, T., UTSUMI, S. AND KITO, M. (1997). Functions of characteristic Cys-Gly-His-Cys (CGHC) and Gln-Glu-Asp-Leu (QEDL) motifs of microsomal ER-60 protease. *Journal of Biochemistry* **122**, 834–842.

VASQUEZ, J.R., EVNIN, L.B., HIGAKI, J.N. AND CRAIK, C.S. (1989). An expression system for trypsin. *Journal of Cellular Biochemistry* **39**, 265–276.

VELASCO, G., FERRANDO, A.A., PUENTE, X.S., SÁNCHEZ, L.M. AND LOPEZ-OTÍN, C. (1994). Human cathepsin O: Molecular cloning from a breast carcinoma, production of the active enzyme in *Escherichia coli*, and expression analysis in human tissues. *The Journal of Biological Chemistry* **269**, 27136–27142.

VISHNUVARDHAN, D., KAKIUCHI, N., URVIL, P.T., SHIMOTOHNO, K., KUMAR, P.K.R. AND NISHIKAWA, S. (1997). Expression of highly active recombinant NS3 protease domain of hepatitis C virus in *E. coli*. *FEBS Letters* **402**, 209–212.

VOSS, T., MEYER, R. AND SOMMERGRUBER, W. (1995). Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for the structural intergrity. *Protein Science* **4**, 2526–2531.

WALKER, N.P.C., TALANIAN, R.V., BRADY, K.D., DANG, L.C., BUMP, N.J., FERENZ, C.R., FRANKLIN, S., GHAYUR, T., HAKETT, M.C., HAMMILL, L.D., HERZOG, L., HUGUNIN, M., HOUY, W., MANKOVICH., J.A., MCGUINESS, L., ORIEWICZ, E., PASKIND, M., PRATT, C.A., REIS, P., SUMMANI, A., TERRANOVA, M., WELCH, J.P., XIONG, L., MÖLLER, A., TRACEY, D.I., KAMEN, R. AND WONG, W.W. (1994a). Crystal structure of the cysteine protease interleukin-1β converting enzyme: a (p20/p10) 2 homodimer. *Cell* **78**, 343–352.

WALKER, P.A., LEONG, L.E.-C., NG, P.W.P., TAN, S.H., WALEER, S., MURPHY, D. AND PORTER, A.G. (1994b). Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Bio/Technology* **12**, 601–605.

WALL, J.G. AND PLÜCKTHUN, A. (1995). Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **6**, 507–516.

WANG, G., LIU, N. AND YANG, K. (1995a). High-level expression of prochymosin in *Escherichia coli*: effect of the secondary structure of the ribosome-binding site. *Protein Expression and Purification* **6**, 284–290.

WANG, J., CHAO, J. AND CHAO, L. (1991). Purification and characterization of recombinant tissue kallikrein from *Escherichia coli* and yeast. *Biochemical Journal* **276**, 63–71.

WANG, L.-F. AND DOI, R.H. (1992). Heterologous gene expression in *Bacillus subtilis*. *Biotechnology* **22**, 63–104.

WANG, L.-F., HUM, W.T., KALYAN, N.K., LEE, S.G., HUNG, P.P. AND DOI, R.H. (1989). Synthesis and refolding of human tissue-type plasminogen activator in *Bacillus subtilis*. *Gene* **84**, 127–133.

WANG, Q.M., JOHNSON, R.B., COX, G.A., VILLAREAL, E.C., CHURGAY, L.M. AND HALE, J.E. (1998). Enzymatic characterization of refolded human rhinovirus type 14 2A protease expressed in *Escherichia coli*. *Journal of Virology* **72**, 1683–1687.

WANG, Z.-M., RUBIN, H. AND SCHECHTER, N.M. (1995b). Production of active recombinant human chymase from a construct containing the enterokinase cleavage site of trypsinogen in place of the native propeptide sequence. *Biological Chemistry Hoppe-Seyler* **376**, 681–684.

WEICKERT, M.J., DOHERTY, D.H., BEST, E.A. AND OLINS, P.O. (1996). Optimization of heterologous protein production in *Escherichia coli*. *Current Opinion in Biotechnoloy* **7**, 494–499.

WELCH, A.R., WOODS, A.S., MCNALLY, L.M., COTTER, R.J. AND GIBSON, W. (1991). A herpesvirus maturation proteinase, assemblin: identification of its gene, putative active site domain, and cleavage site. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10792–10796.

WELCH, A.R., HOLMAN, C.M., BROWNER, M.F., GEHRING, M.R., KAN, C.C. AND VAN WART, H.E. (1995). Purification of human matrilysin produced in *Escherichia coli* and characterization using a new optimized fluorogenic peptide substrate. *Archives of Biochemistry and Biophysics* **324**, 59–64.

WILK, S., WILK, E. AND MAGNUSSON, R.P. (1998). Purification, characterization, and cloning of a cytosolic aspartyl aminopeptidase. *The Journal of Biological Chemistry* **273**, 15961–15970.

WILL, H., ATKINSON, S.J., BUTLER, G.S., SMITH, B. AND MURPHY, G. (1996). The soluble catalytic domain of membrane type I matrix metalloproteinase cleaves the propeptide of progelatinase A and initiates autocatalytic activation. *The Journal of Biological Chemistry* **271**, 17119–17123.

WINDSOR, L.J., BIRKEDAL-HANSEN, H., BIRKEDAL-HANSEN, B. AND ENGLER, J.A. (1991). An internal cysteine plays a role in the maintenance of the latency of human fibroblast collagenase. *Biochemistry* **30**, 641–647.

WINKLER, M.E. AND BLABER, M. (1986). Purification and characterization of recombinant single-chain urokinase produced in *Escherichia coli*. *Biochemistry* **25**, 4041–4045.

WOESSNER, J.F. (1998). Matrilysin. In: *Handbook of Proteolytic Enzymes*. Eds. A.J. Barrett, N.D. Rawlings and J.F. Woessner, pp 1183–1187. New York: Academic Press.

WONG, S.-L. (1995). Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. *Current Opinion in Biotechnology* **6**, 517–522.

WU, X.-C., LEE, W., TRAN, L. AND WONG, S.L. (1991). Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases. *Journal of Bacteriology* **173**, 4952–4959.

WYCKOFF, E.E., HERSHEY, J.W.B. AND EHRENFELD, E. (1990). Eukaryotic initiation factor 3 is required for poliovirus 2A protease-induced cleavage of the p220 component of eukaryotic initiation factor 4F. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 9529–9533.

XUE, Y., PENG, G., MA, Z., CHEN, Y. AND ZHU, D. (1997). High-level expression and purification of human pro-UK cDNA in *Escherichia coli*. *Chinese Journal of Biotechnology* **13**, 233–238.

YE, Q.-Z., JOHNSON, L.L., HUPE, D.J. AND BARAGI, V. (1992). Purification and characterization of the human stromelysin catalytic domain expressed in *Escherichia coli*. *Biochemistry* **31**, 11231–11235.

YE, Q.-Z., JOHNSON, L.L., YU, A.E. AND HUPE, D. (1995). Reconstructed 19 kDa catalytic domain of gelatinase A is an active proteinase. *Biochemistry* **34**, 4702–4708.

YEE, L. AND BLANCH, H.W. (1993). Recombinant trypsin production in high cell density fed-batch cultures in *E. coli*. *Biotechnology and Bioengineering* **41**, 781–790.

YU, P.L. (1994). Production of chymosin for the dairy industry by recombinant DNA technology. *Australas Biotechnology* **4**, 19–23.

ZHANG, Y., OLSEN, D.R., NGUYEN, K.B., OLSON, P.S., RHODES, E.T. AND MASCARENHAS, D. (1998). Expression of eukaryotic proteins in soluble form in *Escherichia coli*. *Protein Expression and Purification* **12**, 159–165.

ZIEBUHR, J., HEROLD, J. AND SIDDELL, G. (1995). Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *Journal of Virology* **69**, 4331–4338.