Recombinant Production and Characterisation of two *Fasciola hepatica* Cathepsin L Proteases

Peter R. Collins

Ph.D.

2005

Recombinant Production and Characterisation of two *Fasciola hepatica* Cathepsin L Proteases

Peter R. Collins, B.Sc.

A thesis submitted in partial fulfilment of the requirements for the degree of Ph.D.

Supervisors:

Dr. Sandra O'Neill, School of Nursing, Dublin City University, Ireland

Prof. John P. Dalton, Institute for the Biotechnology of Infectious Diseases, University of Technology, Sydney, Australia

Submitted June 2005

Declaration of Originality

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Peter N Cobbin Student ID No.: 96065125

Date: 13th JANUARY 2006

Acknowledgments

I would like to thank my supervisors, Dr. Sandra O' Neill and Prof. John Dalton for all their help and encouragement over the course of this project and, of course, for the opportunity to work on it in the first place. I would also like to especially thank Dr. Sheila Donnelly, Dr. Colin Stack, and Dr. Mary Sekiya for all their hands-on help and understanding. Thanks also to everyone I've worked with during this project at DCU, UCD and UTS – there's too many of you to list, but you all know who you are, and you've all helped enormously. Finally, thanks to IBID and IDP Education Australia for the great opportunity I was given to spend a year in UTS as part of this project.

I would like to dedicate this thesis to my parents, who have always been there when I needed support and encouragement and have made this thesis possible.

Abstract

The cysteine proteases of the liver fluke, *Fasciola hepatica*, are of great importance to the virulence of the organism. Previous work has established the involvement of these enzymes in catabolism, migration through host tissues and defence against host immune attack. They have therefore been identified as important vaccine targets.

In this study, sequence analyses compare the *F. hepatica* cathepsins L1 and L2 with each other and with other cathepsin Ls of *F. hepatica* and the related parasite, *F. gigantica*. Groupings of enzymes based on propeptide and active site cleft sequences are established, indicating possible functional diversity in both species. Comparisons with other trematode and mammalian cathepsin Ls were also performed, focusing on a propeptide motif involved in enzyme processing.

Production of the recombinant enzymes was by expression in the *Pichia pastoris* yeast expression system, providing sufficient quantities of enzyme for characterisation studies. Enzyme activity against fluorogenic substrates and native collagens was compared, demonstrating differences in substrate specificity and therefore possible function between the two enzymes. Processing and activation of the enzymes was also investigated, with inactive mutant enzymes produced for this purpose. Finally, details are given of various collaborations with other laboratories utilising the recombinant enzymes produced during this project.

Abbreviations

3D: three dimensional

AA or aa: amino acid(s)

ABD: antibody diluent

AE: acetate-EDTA buffer

AMC: 7-amino-4-methylcoumarin

APS: ammonium persulphate

BCA: bicinchoninic acid

β-Gal: β-galactosidase

BMGY: buffered glycerol complex medium

BMMY: buffered methanol complex medium

Boc-Ala-Gly-Pro-Arg-AMC or Boc-AGPR-AMC:

t-butyloxycarbonyl-L-alanyl-L-glycyl-L-prolyl-L-arginine-4-amido-7-methylcoumarin

Boc-Val-Leu-Lys-AMC or Boc-VLK-AMC:

t-butyloxycarbonyl-L-valyl-L-leucyl-L-lysine-4-amido-7-methylcoumarin

Boc-Val-Pro-Arg-AMC or Boc-VPR-AMC:

t-butyloxycarbonyl-L-valyl-L-prolyl-L-arginine-4-amido-7-methylcoumarin

BSA: Bovine serum albumin

CA074: [Propylamino-3-hydroxy-butan-1,4-dionyl]-isoleucyl-proline

CaS: calcium-sorbitol solution

CaT: calcium-tris buffer

cDNA: complementary deoxyribonucleic acid

Cel: *Caenorhabditis elegans*

CL or CatL: cathepsin L

CF: cathepsin F

CP: cysteine peptidase

Cpa: *Carica papaya*

Cs: *Clonorchis sinensis*

DAB: 3,3'-diaminobenzidine

dATP: deoxyadenosine triphosphate

DC: detergent compatible

dCTP: deoxycytidine triphosphate

DEPC: diethyl pyrocarbonate

dGTP: deoxyguanosine triphosphate

dH_2O: distilled water

DMF: *N,N'*-Dimethyl-formamide

DMSO: Dimethyl sulphoxide

DNA: deoxyribonucleic acid

dNTP: deoxynucleoside triphosphate

dsDNA: double stranded deoxyribonucleic acid

DTT: DL-Dithiothreitol

dTTP: deoxythymidine triphosphate

ES products: excretory-secretory products

E64: L-*trans*-epoxysuccinyl-leucylamido[4-guanidino]butane

EDTA: Ethylenediaminetetraacetic acid

ERFNAQ motif: Glu-Xaa-Xaa-Xaa-Arg-Xaa-Xaa-(Val or Ile)-Phe-Xaa-Xaa-Asn-Xaa-Xaa-Xaa-Ala-Xaa-Xaa-Xaa-Gln

ERFNIN motif: Glu-Xaa-Xaa-Xaa-Arg-Xaa-Xaa-(Val or Ile)-Phe-Xaa-Xaa-Asn-Xaa-Xaa-Xaa-Ile-Xaa-Xaa-Xaa-Asn

Fg or Fgi: *Fasciola gigantica*

Fh or Fhe:	*Fasciola hepatica*
GS-PAGE:	gelatine substrate polyacrylamide gel electrophoresis
GTA:	glutaraldehyde
G*X*N*X*F*X*D:	Gly-Xaa-Asn-Xaa-Phe-Xaa-Asp
GY:	glycerol
HEPES:	*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulphonic acid
His_6:	His-His-His-His-His-His tag sequence
Hsa:	*Homo sapiens*
HSA:	Human serum albumin
Human C3:	complement
IgG:	immunoglobulin G
IPTG:	Isopropyl-β-D-1-thiogalactopyranoside
K_i:	inhibition constant
LB:	Luria broth
LDL:	low density lipoprotein
MD:	minimal dextrose medium
MEGA:	Molecular Evolutionary Genetics Analysis
MHC Class II:	Major Histocompatability Complex Class II
MM:	minimal methanol medium
Mmu:	*Mus musculus*
mRNA:	messenger ribonucleic acid
Mut^+:	Methanol utilisation plus
Mut^S:	Methanol utilisation slow
NCM:	nitrocellulose membrane
NEJ:	newly excysted juvenile
nfH_2O:	nuclease-free water

NTA:	nitrilotriacetic acid
NK cells:	Natural Killer cells
OD_X:	optical density at X nm wavelength
PAGE:	polyacrylamide gel electrophoresis
PBS:	Phosphate buffered saline
PCR:	polymerase chain reaction
PEG:	polyethylene glycol
PFA:	paraformaldehyde
PIPES:	Piperazine-1,4-bis[2-ethanesulphonic acid]
$polyA^+$:	polyadenine positive
Pre:	prepeptide signal sequence
Pro:	propeptide
PVDF:	polyvinylidene difluoride
Pw:	*Paragonimus westermani*
RD:	regeneration dextrose medium
RDB:	regeneration dextrose medium with D-biotin
RDHB:	regeneration dextrose medium with L-histidine and D-biotin
RNA:	ribonucleic acid
Rno:	*Rattus norvegicus*
RT-PCR:	reverse transcriptase polymerase chain reaction
SCE:	sorbitol-citrate-EDTA buffer
sdH_2O:	sterile distilled water
SDS:	sodium dodecyl sulphate
SDS-PAGE:	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE:	sorbitol-EDTA solution
SED:	sorbitol-EDTA-DTT solution

Sj: *Schistosoma japonicum*

Sm or Sma: *Schistosoma mansoni*

SOE: splicing by overlap extension

SSC: salt-sodium citrate solution

TAE: Tris-acetate-EDTA buffer

TE: Tris-EDTA buffer

TEMED: *N,N,N',N'*-Tetramethylethylenediamine

T_H1: Helper T lymphocyte type 1

Tos-Gly-Pro-Arg-AMC or Tos-GPR-AMC:

4-toluenesulphonyl-L-glycyl-L-prolyl-L-arginine-7-amido-4-methylcoumarin

Tos-Gly-Pro-Lys-AMC or Tos-GPK-AMC:

4-toluenesulphonyl-L-glycyl-L-prolyl-L-lysine-7-amido-4-methylcoumarin

Tris: Tris(hydroxymethyl)aminomethane

TRITC: tetramethyl rhodamine isothiocyanate

UV: ultraviolet

Xaa: any amino acid

X-Gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

YNB: yeast nitrogen base

YPD: yeast extract-peptone-dextrose

Z-Arg-AMC or Z-R-AMC:

benzyloxycarbonyl-L-arginine-4-amido-7-methylcoumarin

Z-Arg-Arg-AMC or Z-RR-AMC:

benzyloxycarbonyl-L-arginyl-L-arginine-4-amido-7-methylcoumarin

Z-Leu-Arg-AMC or Z-LR-AMC:

benzyloxycarbonyl-L-leucyl-L-arginine-4-amido-7-methylcoumarin

Z-Phe-Ala-diazomethylketone or Z-FA-CHN_2:

benzyloxycarbonyl-L-phenylalanyl-L-alanine-diazomethylketone

Z-Phe-Arg-AMC or Z-FR-AMC:

benzyloxycarbonyl-L-phenylalanyl-L-arginine-4-amido-7-methylcoumarin

Z-Pro-Arg-AMC or Z-PR-AMC:

benzyloxycarbonyl-L-prolyl-L-arginine-4-amido-7-methylcoumarin

Chapter 1: Introduction – Cysteine Proteinases of Trematodes

1.1 The Trematode Parasites

1.1.1 Introduction to Trematodes and Trematode biology

The digenetic trematodes, commonly known as the flukes, are parasitic flatworms of the phylum Platyhelminthes. As the name 'digenetic' implies, the digeneans generally have complex life-cycles involving at least two hosts. However, this is not always the case and there is great variation in life-cycles from species to species. Trematodes are parasites of all classes of vertebrates, with a particular bias among many of the species towards the parasitism of marine fishes. Nearly every organ of the vertebrate body may be parasitized by the adult or juvenile stages of some trematode species. Many of the fluke species are of great economic and medical importance, with losses of both productivity and life among domestic animals and chronic illness amongst humans (Roberts and Janovy, 2000). Of the species found in the subclass Digenea, the family generally regarded as the most serious problem for human health are the blood flukes, or schistosomes. In the case of livestock, the liver flukes are of great economic importance, although they too can affect the human population.

The general life cycle of a trematode parasite involves three hosts. Eggs are released into the environment from the definitive host and hatch into the first larval stage, the miracidium, which penetrates the first intermediate host. Development takes place within this host culminating in the release of a second larval stage, the cercaria, which penetrate the second intermediate host. Here, the cercaria encysts into the infectious metacercaria stage, which lies dormant until consumed along with the second intermediate host by the definitive host. The metacercariae then excyst within the definitive host and develop to the adult stage, migrating through the host's body to

reside in different locations, depending on the species. Release of eggs continues the cycle. Variations on this basic life cycle are found among the trematode parasite species, with differing degrees of truncation such as the deletion of the second intermediate or even the definitive host (Poulin and Cribb, 2002). The following section is intended to provide some brief background on the major digenetic trematodes of economic and medical importance, in particular the liver fluke, *Fasciola hepatica*, as the organism and its enzymes will be the main focus of this thesis.

1.1.2 The liver flukes: *Fasciola* spp.

The liver flukes, *Fasciola hepatica* and *F. gigantica* are helminth parasites of the family Fasciolidae, which also contains an as yet unclassified Japanese isolate liver fluke, often referred to in the literature simply as '*Fasciola* sp.', as well as the giant intestinal fluke, *Fasciolopsis buski*. Both *F. hepatica* and *F. gigantica* are the causative agents of fasciolosis in agricultural animals, such as sheep and cattle, and have recently emerged as important pathogens of humans (O'Neill et al., 1998; Spithill et al., 1998; Dalton and Mulcahy, 2001; Rokni et al., 2002). While *F. hepatica* is found mainly in temperate regions, its larger relative *F. gigantica* is primarily found in tropical regions.

The life cycle of *F. hepatica* (see Figure 1.1) follows the same basic plan as that of other trematode parasites, although without a second intermediate host: (a) eggs are passed into the environment from the definitive host in the faeces; (b) miracidia hatch from the eggs and penetrate the intermediate snail host, such as *Lymnaea truncatula*; (c) development and multiplication of sporocyst and redia stages to cercaria takes place in the snail before the cercaria emerge and encyst on vegetation; (d) the encysted metacercaria are consumed by the definitive host (usually a ruminant) along with the vegetation and excyst, migrate and develop into adult worms, finally residing in the host's bile duct (Andrews, 1999). *F. gigantica* shares a similar life cycle, and both

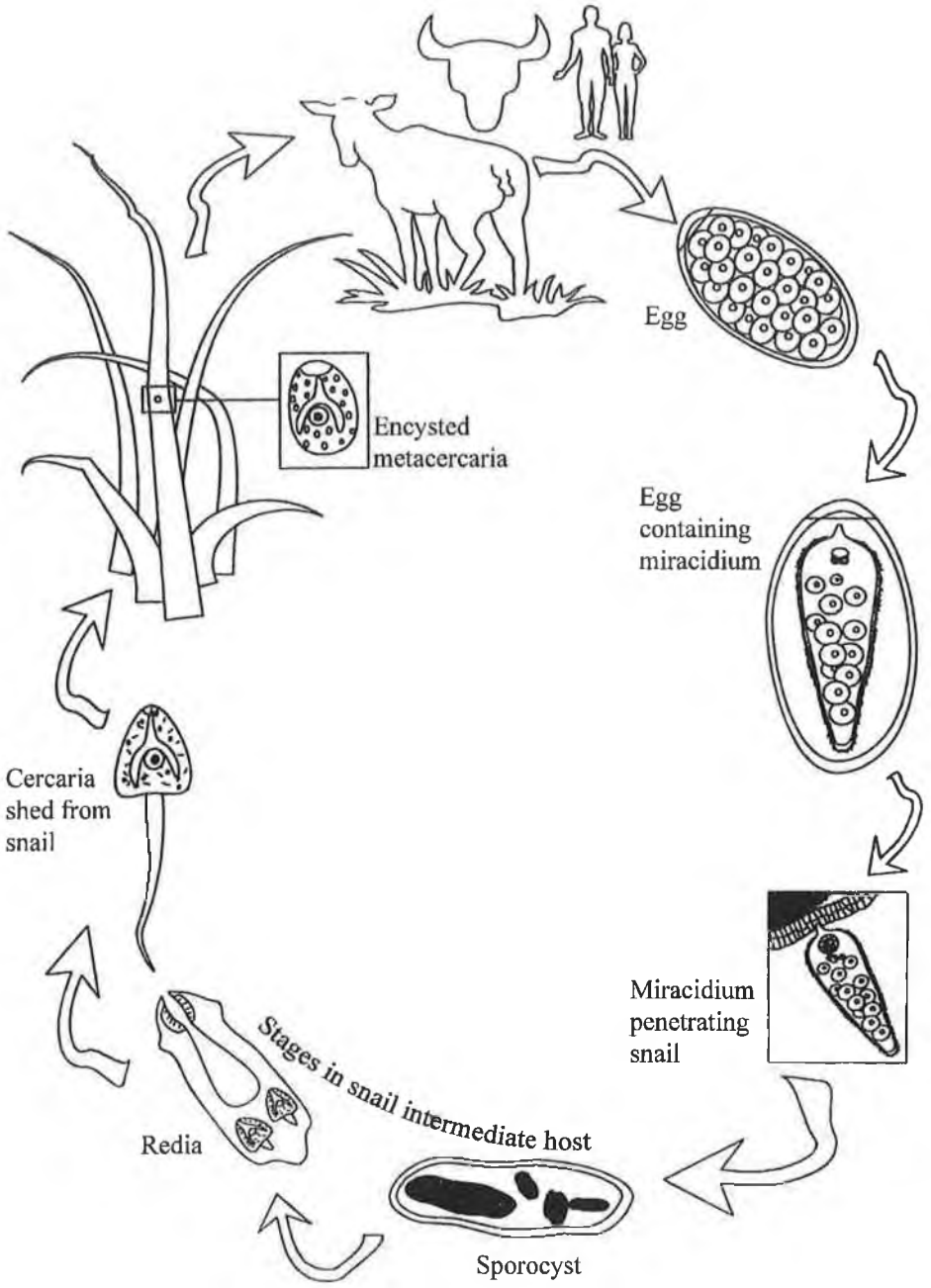


Figure 1.1: The life cycle of *Fasciola hepatica*. Eggs pass into the environment from the definitive mammalian host; miracidia hatch from the eggs and penetrate the intermediate snail host; sporocyst and redia stages develop through to cercaria in the snail before the cercaria emerge and encyst on vegetation; encysted metacercaria are consumed by the definitive host and excyst, migrate and develop into adult worms in the bile duct. Adapted from Andrews (1999).

organisms are hermaphroditic, although sexual reproduction in the definitive host is usually by cross-fertilization (Andrews, 1999). Migration of the juvenile flukes through the gut mucosa and liver parenchyma causes the symptoms of the acute stage of fasciolosis, namely extensive haemorrhaging and liver damage. Chronic symptoms occur due to damage caused to the bile ducts by the presence of the adult worms (Rokni et al., 2002).

1.1.3 The blood flukes: *Schistosoma* spp.

Helminth parasites of the family Schistosomatidae are generally referred to as the blood flukes and include organisms such as *Schistosoma mansoni*, *S. japonicum*, *S. haematobium, S. indicum*, and *S. bovis*. Schistosomiasis is regarded as the second most serious parasitic disease of humans after malaria, causing chronic morbidity and death, and affects an estimated 250 million people across around 75 countries, predominantly in tropical and subtropical regions. *S. mansoni* is widespread in Africa, the Caribbean and South America, while *S. haematobium* is found in Africa and the eastern Mediterranean and *S. japonicum* is located mainly in Asia (Roberts and Janovy, 2000; Hu et al., 2004).

The schistsomes differ from most of the platyhelminths in that they show sexual dimorphism rather then being hermaphroditic. Their life cycles are also highly truncated, with the effective removal of the definitive host and development from cercaria to egg-producing adult all occurring within the second vertebrate host (Poulin and Cribb, 2002). Briefly, the miracidia penetrate freshwater snails and develop through two generations of sporocyst to cercaria, which directly penetrate the skin of the vertebrate host once released. The cercariae lose their tails to become schistosomulae and migrate through the lungs to the hepatic portal system, to mature and pair up to produce eggs (Hu et al., 2004).

1.1.4 The Oriental lung fluke *Paragonimus westermani*

The oriental lung fluke, *Paragonimus westermani*, is a member of the Plagiorchiidae family and is the causative parasite of human lung paragonimiasis (Pezzella et al., 1981). It can also, in rare cases, cause both cerebral (Park, H. et al., 2001; Choo et al., 2003) and hepatic infections (Kim et al., 2004). It is endemic in Africa, South America and Asia, in particular in Korea and Japan (Park, H. et al., 2001).

P. westermani follows the general trematode life cycle pattern of two intermediate hosts and one definitive vertebrate host. Infection of humans is from the second intermediate decapod host, usually crustaceans such as freshwater crabs, crayfish or shrimps. The metacercariae encyst in these animals and can be passed on if they are consumed either raw or undercooked (Choo et al., 2003). The metacercariae excyst in the small intestine and migrate trough the tissues to the lungs, where the adults reside, although this migration of juveniles can be erratic leading to various ectopic infections of other organs (Pezzella et al., 1981; Kim et al., 2004; Choo et al., 2003).

1.1.5 The Chinese liver fluke: *Clonorchis sinensis*

Clonorchis sinensis, a member of the Opisthorchiidae family, is the causative agent of clonorchiasis and is found in many regions of eastern Asia, including China, Korea and Vietnam, where an estimated 7 million people are infected (Nagano et al., 2004). Like *P. westermani, C. sinensis* is a food-borne pathogen. In this case the secondary intermediate host is usually a freshwater fish, in which the cercariae encyst to metacercariae underneath the scales or within the muscle tissue (King and Scholz, 2001). Consequently, infection occurs with consumption of raw or undercooked fish. Once consumed, the excysted juveniles migrate from the intestine up through the bile duct where they mature to adults (King and Scholz, 2001). This migration therefore

shows a marked difference from the *Fasciola* spp., whose juveniles migrate through the liver tissue, although the adults reside in the bile ducts in both cases.

1.2 Cysteine Proteases of Parasitic Trematodes

1.2.1 Classification of Cysteine Proteases

The cysteine proteases of most importance secreted by trematode parasites are members of two clans – Clan CA, the papain-like cysteine proteases and Clan CD, the legumain-like cysteine proteases (Sajid and McKerrow, 2002). In the latter group, it is members of the C13 family that are of importance, including the asparaginyl endopeptidases which have been isolated from the schistosomes (Merckelbach et al., 1994; Brindley et al., 1997; Caffrey et al., 2000) and *F. hepatica* (Tkalcevic et al., 1995). These asparaginyl endopeptidases have been implicated in the breakdown of ingested host haemoglobin to usable amino acids, although this seems to be through an indirect activity, the post-translational modification of other proteases such as cathepsin L, B and D-like enzymes, rather than through direct catabolic activity (Brindley et al., 1997; Brindley and Dalton, 2004).

The larger group, Clan CA, includes the cathepsin B and cathepsin L-like family C1 and the calpain-like family C2. Calpain-like enzymes have been isolated for the two major *Schistosoma* spp. In the case of *S. mansoni*, the calpain-like enzyme has been shown to be expressed in both the sporocyst and adult stages of the parasite life cycle and is thought to be a membrane bound enzyme important in parasite metabolism (Andresen et al., 1991). A large subunit of the *S. japonicum* calpain has also been recombinantly expressed and shown to be a potential vaccine candidate, providing a partially protective response in mice (Zhang et al., 2001).

By far the largest and most important class of these enzymes for trematode parasites are the papain-like enzymes of family C1, more specifically subfamily C1A.

These are generally subdivided into cathepsin B-like and cathepsin L-like proteases, all sharing a similarity of structure and function with that of papain (Sajid and McKerrow, 2002).

Papain was the first cysteine protease to be isolated, in 1879, from the papaya fruit, *Carica papaya*, from which its name was taken. It was also the first cysteine protease to have its three dimensional structure determined (Drenth et al., 1968) and is therefore the archetype for all cysteine proteases, in particular those of the C1 family. It consists of a single chain nonglycosylated polypeptide with three disulphide bonds and which folds to form a globular protein of two domains. These two interacting domains delineate the active site cleft at the surface of the enzyme, where the substrate binds and is cleaved (Drenth et al., 1968; Ménard and Storer, 2004).

Papain has been shown to have endopeptidase, amidase and esterase activites, with a preference for residues with bulky non-polar side-chains, such as phenylalanine, in the P_2 position (Ménard and Storer, 2004). Along with the majority of the members of Clan CA, the basis of the catalytic mechanism of papain is the catalytic triad consisting of Cys^{25}, His^{159} and Asn^{175} (papain numbering). Also of importance is Gln^{19}, sometimes referred to as the 'oxyanion hole' and highly conserved across the papain-like cysteine proteases, which stabilises the oxyanion of the tetrahedral intermediate formed during peptide bond hydrolysis (Barrett, 1994; Sajid and McKerrow, 2002; Polgá, 2004).

1.2.2 General structure and function of cathepsin B and cathepsin L-like cysteine proteases (subfamily C1A)

The trematode cysteine proteases of subfamily C1A share many structural and functional characteristics with their mammalian counterparts. All are synthesised as inactive preproenzymes consisting of a prepeptide signal sequence, a propeptide and a

mature enzyme region (see Figure 1.2; also Sajid and McKerrow, 2002; Dalton et al., 2004a). In the case of the mammalian cysteine proteases, it has been shown that the prepeptide signal sequence is removed following translocation into the endoplasmic reticulum and that the *N*-terminal propeptide extension is involved in several functions (Wiederanders, 2003); these include intracellular targeting of the enzyme (Coulombe et al., 1996; Ménard et al., 1998; Rozman et al., 1999), action as an intermolecular chaperone aiding the correct folding of the mature enzyme (Baker et al., 1992; Shinde et al., 1997) and prevention of uncontrolled proteolysis by binding to the enzyme substrate cleft in a reverse, non-productive direction (Coulombe et al., 1996; Cygler and Mort, 1997). It has been demonstrated for the propeptide of human cathepsin L and other papain-like cysteine proteases that the free propeptide also acts as a specific and potent inhibitor of the active mature enzyme at neutral pH but does not bind well in alkali conditions (Coulombe et al., 1996; Baker et al., 1992; Ogino et al., 1999; Guo et al., 2000).

The cathepsin L and cathepsin B-like cysteine proteases all share a general structure with the archetype of the group, papain. The mature protein takes the form of a two-domain globular protein, with the active site cleft separating the two domains at the surface (Drenth et al., 1968; Baker, 1977; McGrath et al., 1995; Coulombe et al., 1996; Podobnik et al., 1997). The *N*-terminal domain consists mainly of bundled α-helices, while the *C*-terminal domain contains a β-barrel structure and a long helix runs through the middle of the molecule starting at the catalytic cysteine (Rawlings and Barrett, 2004). These enzymes also possess the catalytic triad of Cys^{25}, His^{159} and Asn^{175} (papain numbering) (Barrett, 1994; Rawlings and Barrett, 1994; Polgá, 2004) and most members of the family are irreversibly inhibited by E64 and its analogues (Barrett et al., 1982; Rawlings and Barrett, 2004).

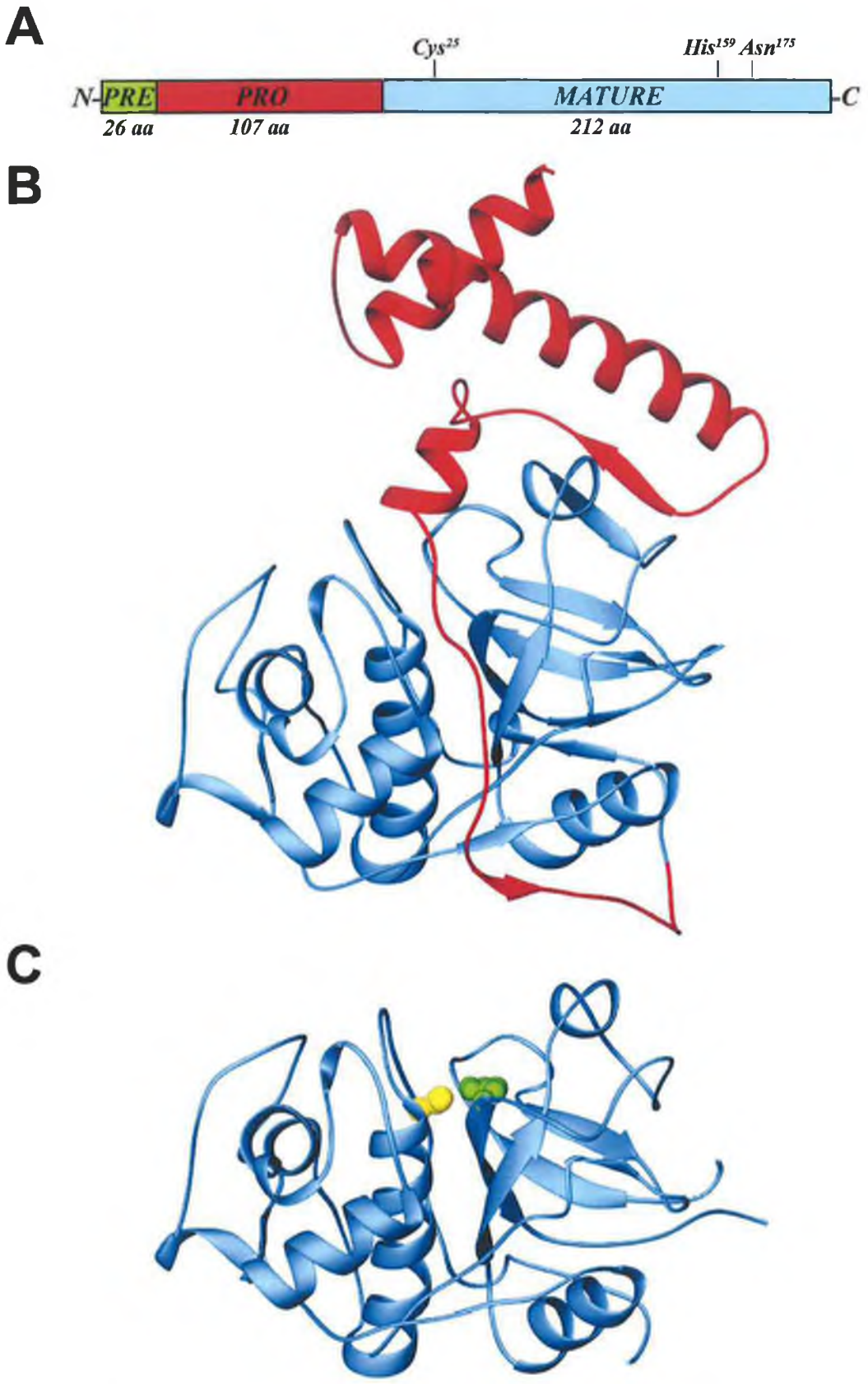


Figure 1.2: Typical structure of papain-like cysteine proteases. *A*, schematic diagram of papain showing the prepeptide signal sequence, propeptide and mature regions. The location of the residues of the active site triad are indicated (papain numbering). *B*, structure of human procathepsin L1 with the propeptide shown in *red* and the mature region shown in *blue*. *C*, structure of mature human cathepsin L with the side chains of Cys^{25} and His^{159} (human cathepsin L numbering) indicated in *yellow* and *green* respectively. Adapted from Turk and Guncar (2003).

1.2.3 Cathepsin B-like cysteine proteases

The first major category of proteases within subfamily C1A are the cathepsin B-like cysteine proteases and are present in a wide variety of species, including many trematode and other parasitic species (Karrer et al., 1993; Sajid and McKerrow, 2002).

Compared to the other major members of the papain superfamily, cathepsin B differs most in structure from the general papain-like pattern. Analyses of the crystal structure of mammalian cathepsin Bs show that it possesses an extended insertion of twenty amino acid residues, termed the occluding loop (Musil et al., 1991). This loop blocks the *C*-terminal end of the active site cleft, reducing endopeptidase activity and preventing binding of macromolecular inhibitors and contributes to the enzyme subsites, giving cathepsin Bs an extra dipeptidyl carboxypeptidase activity unique among the papain superfamily (Illy et al., 1997; Sajid and McKerrow, 2002; Mort, 2004). Two histidine residues within the loop, His^{110} and His^{111} (human cathepsin B numbering) are responsible for this exopeptidase activity, serving to anchor the carboxylate at the P'_2 position of the substrate, thereby directing the *C*-terminal dipeptide into position in the active site for cleavage (Sajid and McKerrow, 2002; Mort, 2004). The loop is also responsible for inhibition by the E-64 derived cathepsin B specific inhibitor, CA074. Studies have shown that deletion of the occluding loop prevents inhibition by CA074, abolishes the exopeptidase activity, allows bulkier inhibitors such as the cystatins to bind to the active site and increases endopeptidase activity (Illy et al., 1997; Sajid and McKerrow, 2002). The occluding loop also defines the pH dependency of inhibition of cathepsin B by its propeptide (Quarasihi et al., 1999).

In addition to the occluding loop, cathepsin B further differs from the papain model with a truncated propeptide, only two-thirds the length of that of other family C1 enzymes (Musil et al., 1991; Podobnik et al., 1997; Turk et al., 1996). However, despite this truncation, the propeptide still functions in a similar manner to that of the other

papain-like cysteine proteases, with importance in folding of the mature region, processing of the enzyme to its mature form and prevention of uncontrolled proteolysis (Mach et al., 1994a; Turk et al., 1996; Quraishi and Storer, 2001; Mort, 2004). In addition, the free propeptide has been demonstrated to form a stable complex with the mature enzyme at neutral pH, providing a mechanism for activity of the processed protease to be mediated by regional acidity and allowing it to survive extracellularly in conditions that would lead to the inactivation of other lysosomal cysteine proteases (Mach et al., 1994b; Mort 2004). Secreted cathepsin B has been implicated in tumour invasion and metastasis (Lorenzo et al., 2000).

The endopeptidase activity of cathepsin B, which may be an evolutionary holdover from its origin as a more conventional papain-like cysteine protease (Illy et al., 1997), maintains the preference for large hydrophobic residues in the P_2 position found in other members of the C1A subfamily, but uniquely can accommodate an arginine side chain in this position due to the presence of a glutamate residue within the S_2 subsite (Mort, 2004). From a practical point of view, activity against a substrate such as the fluorogenic compound Z-Arg-Arg-AMC can be used to distinguish cathepsin Bs from other members of the papain superfamily, which lack such activity.

A second cathepsin B-like enzyme, previously identified variously as cathepsin P, cathepsin Z, cathepsin Y, and cathepsin B2, is now known as cathepsin X (Ménard and Sulea, 2004). While cathepsin X most closely resembles cathepsin B, it possess several distinct and unique features: the propeptide is even further truncated then that of cathepsin B, down to only about one-third the length of other subfamily C1A propeptides; the enzymes possess a unique carboxypeptidase activity and lack any significant endopeptidase activity; finally, a unique three residue insertion, termed the mini-loop, is present within the active site and provides a signature motif for cathepsin Xs (Santamaría et al., 1998; Ménard and Sulea, 2004). Cathepsin X-like proteases have

been identified in nematode worms, such as *Caenorhabditis elegans* and *Onchocerca volvulus*, but not so far in the trematodes (Sajid and McKerrow, 2002). For example, the cathepsin X of *C. elegans*, referred to as *Ce*-CPZ-1 (reflecting the alternate nomenclature of cathepsin Xs as cathepsin Zs), is expressed in the hypodermal cells of all developmental stages, as well as specifically in the gonads and pharynx of the adult, and is believed to be involved in moulting and essential to the development of the organism (Hashmi et al., 2004).

The last member of this group is cathepsin C, also known as dipeptidyl peptidase I, which again differs greatly from the general pattern of papain-like cysteine proteases (Tort, et al., 1999; Turk et al., 2004). Cathepsin Cs have a large molecular size, possess an extra domain, form tetramers, have a requirement for halide ions for activity, and are aminopeptidases (Turk, et al., 2000; Turk et al., 2001; Turk et al., 2004).

1.2.4 Cathepsin L-like cysteine proteases

Enzymes of the second major category of papain-like cysteine proteases, the cathepsin L-like proteases, have structures more closely matching the general profile of the group. Again, they are present in a wide variety of species and many have been isolated from trematode and other parasite species (Sajid and McKerrow, 2002). Unlike the cathepsin Bs, these proteases have full length propeptides containing a prominent structural motif, referred to as the ERFNIN motif, which forms the main α-helical backbone of the propeptide (Karrer et al., 1993; Tort et al., 1999).

Cathepsin L itself was first isolated as a mammalian lysosomal cysteine protease, hence the 'L' designation, and conforms closely to the general structure of papain-like cysteine proteases (Coulombe, 1996; Kirschke, 2004a). Mammalian cathepsin Ls have been shown to have major biological functions in processes such as

lysosomal proteolysis, spermatogenesis, antigen presentation, and tumour invasion and metastasis (Kirschke, 2004a). While the enzyme normally resides in the lysosome, it is secreted in small amounts from normal cells, and is overexpressed and secreted in tumour cells, with involvement in cleavage of human C3 (Jean et al., 1997; Frade, 1999) and generation of endostatin (Felbor et al., 2000). Efforts have been made to use single chain variable fragment (ScFv) antibodies against cathepsin L to inhibit its secretion from human melanoma cells (Guillaume-Rousselet, et al., 2002).

Cathepsin K (previously identified variously and confusingly as cathepsin O, cathepsin X and cathepsin O2) closely resembles cathepsin L and has been isolated from the osteoclasts of many mammalian species (Drake et al., 1996; Shi et al., 1995; Brömme, 2004b). The enzyme is involved in bone resorption as well as arthritis, with high levels of expression found in synovial fibroblasts of rheumatoid arthritic joints (Hou et al., 2001; Brömme, 2004b). This is due to its ability to cleave collagens and aggrecans, the major building blocks of cartilage, resulting in the cartilage destruction of arthritic diseases (Kafienah et al., 1998; Hou et al., 2001; Hou et al., 2003). Cathepsin K activity is regulated and enhanced by glycosaminoglycans such as the chondroitin sulphates and heparin (Li et al., 2000; Li et al., 2002; Li et al., 2004). There is also evidence for the presence of kinase activity of cathepsin K (Godat et al., 2004). This wider substrate specificity compared to cathepsin Ls is due to differences in the S_2 pocket, particularly with the residues Tyr^{67} and Leu^{205} which replace the Leu^{67} and Ala^{205} present in mammalian cathepsin Ls (Lecaille et al., 2002; Lecaille et al., 2003).

Cathepsin S shows close similarity to cathepsins K and L, sharing similar S_2/P_2 subsite specificity (Kirschke, 2004b). Substrate preference in cathepsin S is for branched hydrophobic residues in the P_2 position, similar to cathepsin K, and crystal structures of human cathepsin S show a large S_2 pocket, consistent with its broader substrate specificity at this site compared to cathepsins B, L or K (McGrath et al., 1998;

Kirschke, 2004b). Cathepsin S is stable and active at pH 7.5, unlike cathepsin L and to a greater degree than cathepsin K (Kirschke, 2004b). It displays collagenolytic and elastinolytic activity and has been shown to play a role in antigen presentation (McGrath et al., 1998; Kirschke, 2004b). Levels of the enzyme are increased in the brains of Down syndrome and Alzheimer's disease patients, and may play a role in the pathogenicity of the latter disease (Lemere et al., 1995; Munger et al., 1995; Kirschke, 2004b).

Cathepsin F is unusual among the cathepsin L-like cysteine proteases in that it possesses a highly extended propeptide region, nearly three times the length of that of cathepsin L, which contains a cystatin-like domain towards the *N*-terminus (Wang et al., 1998; Brömme, 2004a). The protein sequence is also quite dissimilar from other cathepsin L-like proteases, with its closest neighbour being cathepsin W with which it shares approximately 42% similarity (Santamaría et al., 1999; Wang et al., 1998; Brömme, 2004a). The central α-helix of the propeptide is also different, produced by the so-called ERFNAQ motif which it also shares with cathepsin W, rather than the ERFNIN found in other members of the family (Wang et al., 1998; Brömme, 2004a). Human cathepsin F has the usual preference of hydrophobic residues in the P_2 position of its substrates and shows extremely high potency against the fluorogenic substrates Z-Phe-Arg-AMC and Z-Leu-Arg-AMC (Wang et al., 1998). It is produced by macrophages and is implicated in the mediation of MHC Class II maturation and peptide loading (Shi et al., 2000). It has also been shown to be critical for the modification and fusion of low density lipoprotein (LDL) particles during atherogenesis and may be a target for the treatment of atherosclerotic diseases (Öörni et al, 2004).

Cathepsin W, also known as lymphopain, is most closely related to cathepsin F, and the two may form a subgroup of cathepsin F-like proteases (Tort et al., 1999; Dalton and Brindley, 2004). Both share the modified ERFNAQ motif in their

propeptides, although the propeptide of cathepsin W is of a normal length without the extra cystatin-like domain of cathepsin F (Dalton and Brindley, 2004). The enzyme is produced in cytotoxic T cells and natural killer (NK) cells and has been isolated to the endoplasmic reticulum of the latter cell type (Wex et al., 2001; Dalton and Brindley, 2004; Ondr and Pham, 2004). Little further characterisation of the mammalian cathepsin Ws has been carried out.

Finally, cathepsin O is unique among the cathepsin-L like proteases in completely lacking the ERFNIN motif from its propeptide (Velasco and López-Otín, 2004). It is widely expressed, being overexpressed in human carcinomas, and is involved in normal lysosomal protein degradation and turnover (Velasco et al., 1994).

1.2.5 Trematode Cysteine Proteases

Papain-like cysteine proteases are of great importance to the metabolism and host-parasite interactions of the trematode parasites. Enzymes from both the cathepsin B-like and Cathepsin L-like groups have been isolated from these parasites (Tort et al., 1999; Sajid and McKerrow, 2002; Dalton et al., 2004a). The study of cysteine proteases of parasites is of great interest, particularly in providing the basis for various therapies to combat parasite infection in both humans and animals. The enzymes have been investigated as drug targets, with efforts made to design specific inhibitors to prevent their activity and hence disable a major part of the parasite metabolism (McKerrow et al., 1995; Li et al., 1996). There has also been studies of their use as vaccines, using either native purified or recombinant versions of the molecules directly (Dalton and Mulcahy, 2001; Marcet et al., 2002; Almeida et al., 2003; Pearce, 2003), or their genes as DNA vaccines (Kofta et al., 2000).

Further to their use in treatment of parasitic disease, the papain-like cysteine proteases have been used in the diagnosis of disease in both animals (Gorman et al.,

1997; Cornelissen et al., 2001) and humans (Chappell et al., 1989; O'Neill et al., 1998; Carnevale et al., 2001; Rokni et al., 2002; Nagano et al., 2004).

Cathepsin B-like enzymes which are expressed in trematodes include the cathepsin Bs of *S. mansoni* (Lipps et al., 1996; Dalton et al., 1996a), *S. japonicum* (Merckelbach et al., 1994), *F. hepatica* (Tkalcevic et al., 1995; Wilson et al., 1998; Law et al., 2003), and *C. sinensis* (Dalton et al., 2004a), and the cathepsin Cs of *S. japonicum*, *S. mansoni*, and *F. hepatica* (Tort et al., 1999; Turk et al., 2004). It seems, therefore, that cathepsin Bs and cathepsin B-like enzymes may be found throughout the trematode parasites.

In the case of the cathepsin L-like proteases, however, there is a strong divergence in the predominance of the two main trematode examples, cathepsin Ls and cathepsin Fs. Cathepsin Ls have been identified in *F. hepatica*, *F. gigantica* (see Section 2.3.4, Table 2.2 for references), *S. mansoni* (Dalton et al., 1996; Brindley et al., 1997; Brady et al., 2000) and *S. japonicum* (Day et al., 1995) whereas cathepsin Fs are found in *P. westermani* (Park, H. et al., 2001; Park, H. et al., 2002; Shin et al., 2005) and *C. sinensis* (Park, S. Y. et al., 2001; Lee et al., 2003; Nagano et al., 2004). These enzymes form separate clades (see Figure 1.3), trematode cathepsin L and trematode cathepsin F. Members of the cathepsin L clade have been found in *Fasciola* and *Schistosoma* spp. This lineage also contains the digestive enzymes of insects and crustaceans and has probably evolved to perform this function, later giving rise to the mammalian lysosomal cathepsins L, S and K. (Tort et al., 1999). It is also of note that the cathepsin L clade of the *Fasciola* spp. shows significant diversity, with multiple cathepsin L sequences identified in both *F. hepatica* and *F. gigantica* as belonging to a monophyletic group (Irving et al., 2003). This large gene family may enable production of greater quantities of proteolytic enzyme, thereby increasing the parasite's efficiency in penetrating and feeding on host tissues (Dalton et al., 2004b). The divergent

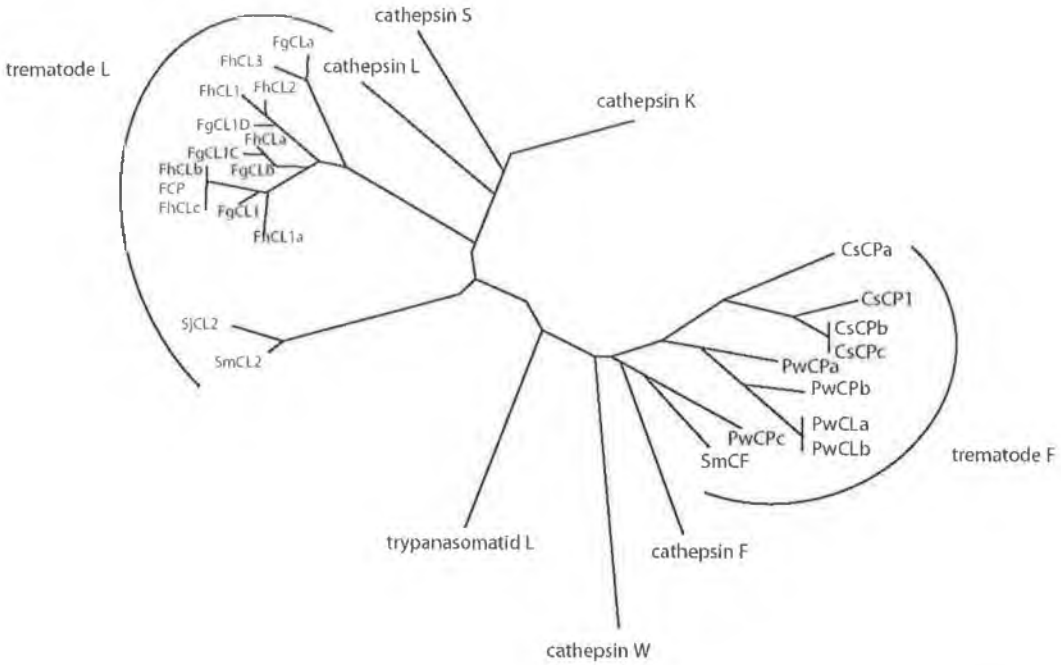


Figure 1.3: Rooted phylogentic comparison of the cathepsin L superfamily. Key: Cs, *Clonorchis sinensis*; Pw, *Paragonimus westermani*; Fg, *Fasciola gigantica*; Fh, *Fasciola hepatica*; Sj, *Schistosoma japonicum*; Sm, *Schistosoma mansoni*; CL, cathepsin L; CF, cathepsin F; CP, cysteine peptidase. The terms cathepsin, trypanosomatid and filarial refer to mammalian, *Trypanosoma cruzi* and *Onchorcerca volvulus* peptidase orthologues, respectively. The peptidase acronyms correspond to the following Protein Data Bank Accession nos. (shown in parentheses). Cathepsin F (NM_003793), cathepsin K (NM_000396), cathepsin L (M20496), cathepsin S (M90696), cathepsin W (NM_001335), CsCP1 (AF093243), CsCPa (U85984), CsCPb (U85983), CsCPc (AB020036), CwCLa (AF362769), filarial L (AF331036), PwCPa (D21124), PwCPb (AF71801), PwCLb (U70537), PwCPc (U69120), SmCF (U07345), trypanasomatid L (M84342), FgCLa (AF419329), FgCLb (AF239264), FgCL1 (AF112566), FhCLa (AF271385), FhCLb (AB009306), FhCLc (L33772), FhCL1 (Z22765), FhCL1a (U62288), FhCL1C (AF239625), FhCL1D (AF239266), FhCL2 (U62280), FhCL3 (AJ270093), FCP (S70380), SjCL2 (U38476) and SmCL2 (Z32529). Adapted from Dalton et al. (2005; unpublished).

evolution of these sequences may also have been an adaptive response to the evolution of new host species, as the approximate time of divergence coincides with the appearance of ancestor species to modern ruminants (Irving et al., 2003).

While members of the trematode cathepsin F clade have been identified in *C. sinensis*, *P. westermani* and *S. mansoni*, none have so far been identified in *Fasciola* spp. (see Figure 1.3). These enzymes appear to be much more closely related than members of the cathepsin L clade, with the possibility of a corresponding greater conservation of function (Park et al., 2002; Kang et al., 2004). These enzymes eventually gave rise to the mammalian cathepsin F clade (see Section 1.2.4), although the majority of the trematode cathepsin Fs lack the cystatin-like domain *N*-terminal extension to their propeptides, suggesting alternate functions or localizations for the trematode enzymes.

1.2.6 Processing of Papain-like cysteine proteases

Human lysosomal procathepsin L provides a good example of the activation of a papain-like cysteine protease which has been widely studied. The enzyme is stabilized at high pH by the propeptide, which protects the protein from the denaturing effects of the alkali (Ménard et al., 1998). The generation of the fully mature and active enzyme occurs with the removal of the propeptide at the lower lysosomal pH of 5.5 (Coulombe et al., 1996; Ménard et al., 1998; Mason and Massey, 1992; Ishidoh and Kominami, 1994).

This cleavage of the proregion of cysteine proteases has been shown to occur autocatalytically *in vitro* under acidic conditions, with examples found among the human cathepsins L, K and B, and other subfamilies (Ménard et al., 1998; Mach et al., 1994b; McQueney et al., 1997; Yamasaki et al., 2002). The stability of the propeptide-mature protease complex is dependent on electrostatic interactions and therefore

reduction of the environmental pH weakens the bonds between the propeptide and the catalytic site. As a consequence of this, the proenzyme may adopt a looser conformation in which the propeptide is less tightly bound into the active site cleft and, therefore, is more susceptible to proteolysis (Rozman et al., 1999; Jerala et al., 1998).

The fine details and precise mechanism of this proteolytic conversion of papain-like cysteine proteases from proenzyme to mature active enzyme is still actively debated. It has been shown from the three-dimensional structures of procathepsin L (Coulombe et al., 1996) and procathepsin B (Musil et al., 1991; Turk et al., 1996; Podobnik et al., 1997) that the *N*-terminus of the mature enzyme is quite far removed from the active site, thus making it difficult to visualize an autocatalytic event. Further to this, circular dichroism studies have revealed that activation does not involve significant conformational changes in the structure of procathepsin L (Ménard et al., 1998) or procathepsin B (Rozman et al., 1999). It is therefore assumed that the initial event in autocatalysis may involve an active proenzyme, possibly created by the reduced pH, which cleaves another proenzyme in the vicinity of the *N*-terminus and sets off a chain reaction (Ménard et al., 1998; Rozman et al., 1999).

However, more recent studies on procathepsin B and procathepsin S identified autoproteolytic intermediates of processing when cystatin was included in *in vitro* activation reactions, which supported the view that the segment of the propeptide that binds the active site cleft is susceptible to cleavage (Quraishi and Storer, 2001). It has therefore been suggested that an initial slow intramolecular cleavage event within this segment of the propeptide triggers a more rapid cascade of intermolecular cleavages at the *N*-terminus (Quraishi and Storer, 2001).

Earlier studies on the processing of yeast-expressed recombinant papain identified a conserved heptapeptide (Gly-Xaa-Asn-Xaa-Phe-Xaa-Asp) motif located between residues -42 and -36 in the propeptide (papain numbering) which may be a site

of initial cleavage in the pH-dependent autoactivation of the enzyme (Vernet et al., 1995). It was suggested that the lowering of the pH perturbed the negative charge of Asp^{-36} resulting in a conformational change that switched on the processing events by allowing proteolysis to occur at the Ala^{-37}/Asp^{-36} bond. Following this primary cleavage further removal of the remaining amino acids of the propeptide may result from the proteolytic activity of the intermediate species, resulting in fully active mature protease (Vernet et al., 1995).

1.3 *Fasciola hepatica* cathepsins L1 and L2

1.3.1 Functions and importance of *F. hepatica* cathepsins L1 and L2

Some of the most significant secretory products of *F. hepatica* are the cathepsin L-like cysteine proteinases (Dalton and Heffernan, 1989). These proteinases play pivotal roles in the pathogenicity of the parasite. The enzymes take part in acquisition of nutrients by the catabolism of host proteins to absorbable peptides (Tort et al., 1999). They facilitate the migration of the parasite through the host intestine and liver by cleavage of interstitial matrix proteins such as fibronectin, laminin, and native collagen (Berasain et al., 1999). They are implicated in evasion of attack by host immune effector cells through the cleavage of immunoglobulins in a similar manner to papain (Smith et al., 1993b; Carmona et al., 1993).

Recently, *F. hepatica* cathepsin L1 has been shown to be involved in immunomodulation by the suppression of T_H1 responses in infected laboratory animals, thereby making them susceptible to concurrent bacterial infections (Brady et al., 1999; O'Neill et al., 2000; O'Neill et al., 2001). It is therefore not surprising that these proteases have been recognized as important targets at which parasite intervention strategies, in particular vaccine development, should be directed (Smith et al., 1994; Dalton and Mulcahy, 2001). In this regard, it has been previously shown in our

laboratory that the induction of anti-cathepsin L immune responses by vaccination with purified enzyme prior to a challenge infection of *F. hepatica* larvae elicits high levels of protection in cattle against development of disease (Dalton et al., 1996b; Mulcahy et al., 1998; Dalton and Mulcahy 2001).

1.3.2 Structure of *F. hepatica* cathepsins L1 and L2

Isolation of the cDNAs encoding both *F. hepatica* cathepsins L1 and L2 revealed that they share a basic structure with their mammalian homologs; each is synthesized as an inactive preproenzyme consisting of a prepeptide signal sequence, a propeptide and a mature enzyme region (Roche et al., 1997; Dowd et al., 1997). It is assumed by analogy with the related mammalian cysteine proteases that the prepeptide signal sequence is removed following translocation into the endoplasmic reticulum and that the *N*-terminal propeptide extension is involved in multiple functions including intracellular targeting of the enzyme (Coulombe et al., 1996; Ménard et al., 1998; Rozman et al., 1999), correct folding of the mature enzyme (Baker et al., 1992; Shinde et al., 1997; Cappetta et al., 2002) and prevention of uncontrolled proteolysis (Coulombe et al., 1996; Cygler and Mort, 1997). The free propeptides of the *F. hepatica* procathepsin Ls also act as specific and potent inhibitors of the cognate mature enzyme at neutral pH but do not bind to the enzyme at pH 5.5–3.5 (Roche et al., 1999), another feature which these enzymes have in common with their mammalian counterparts (Coulombe et al., 1996; Baker et al., 1992; Ogino et al., 1999; Guo et al., 2000). Phylogenetic studies have shown that the *F. hepatica* cathepsin L1 belongs to an enzyme lineage that eventually gave rise to the mammalian cathepsin Ls, Ks, and Ss (Tort et al., 1999) and these structural and functional similarities reflect this.

1.4 Aims of this thesis: Sequence analysis, production, and characterization of *F. hepatica* CL1 and CL2

- Sequence analysis
    - Direct comparison of *F. hepatica* cathepsin L1 and L2 sequences
    - Comparison of *F. hepatica* and *F. gigantica* cathepsin L sequences currently deposited in the online databases
    - Identification of features of *Fasciola* cathepsin L clades, particularly propeptide sequences and S_2 subsite residues
    - Comparison of *F. hepatica* cathepsin L1 and L2 sequences with selected plant, trematode and mammalian equivalents
- Secretion of cathepsin L by the parasite
    - Immunoblot analysis of *F. hepatica* excretory-secretory (ES) products
    - Localisation of mature cathepsin Ls and cathepsin L propeptides within gut of *F. hepatica*
- Production of recombinant enzymes
    - Transfer of procathepsin L1 and L2 cDNAs from *Saccharomyces cerevisiae* expression vector to *Pichia pastoris* yeast expression system
    - Expression and purification of enzymes from yeast medium
    - Analysis of effects of environmental conditions on enzyme production
    - Generation and production of inactive procathepsin L1 and L2 mutants
- Characterization of recombinant enzymes
    - Study of autocatalytic processing of procathepsin L1 and L2
    - Use of inactive mutants to investigate intermolecular processing
    - Comparison of activity and substrate specificity of cathepsin L1 and L2 by use of fluorogenic substrate assays

- Comparison of activity of cathepsin L1 and L2 against native matrix proteins, using collagens as examples
- Determination of pH optima of cathepsin L1 and L2

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemical reagents

Amresco Inc., Solon, Ohio, USA:

High melting point agarose, low melting point agarose.

Bio-Rad Laboratories Ltd., Hemel Hempstead, Hertfordshire, UK:

40% acrylamide/bisacrylamide solution, *DC* Protein Assay Reagent A, *DC* Protein Assay Reagent B.

New England Biolabs (UK) Ltd., Hitchin, Hertfordshire, UK:

100 bp DNA ladder, 1 kb DNA ladder, Broad range prestained protein marker.

Pierce Biotechnology Ltd., Rockford, Illinois, USA:

BCA Protein Assay Reagent A, BCA Protein Assay Reagent B.

Promega UK Ltd., Southampton, UK:

Blue/orange 6X loading dye, dATP, dCTP, dGTP, dTTP, 100 bp DNA ladder, 1 kb DNA ladder, Broad range protein molecular weight markers, Nuclease-free water.

Qiagen Ltd., Crawley, West Sussex, UK:

Ni-NTA agarose.

Sigma-Aldrich Ireland Ltd., Dublin, Ireland:

Acetic acid, 7-Amino-4-methylcoumarin, Ammonium persulphate, Bromophenol blue sodium salt, Calcium chloride, Chloroform, Citric acid, Coomassie Brilliant Blue R-250, Diethyl pyrocarbonate, *N,N'*-Dimethyl-formamide, Dimethyl sulphoxide, DL-Dithiothreitol, Ethanol, Ethidium bromide solution, Ethylenediaminetetraacetic acid disodium salt dihydrate, Glass beads acid-washed 425-600 μm, D-(+)-Glucose monohydrate, Glycerol, Glycine, HEPES, Hydrochloric acid, Imidazole, Isopropyl-β-D-1-thiogalactopyranoside, Magnesium sulphate, Manganese(II) chloride, 2-Mercaptoethanol, Methanol, PIPES, Potassium acetate, Potassium chloride, Potassium hydroxide, Potassium phosphate dibasic, Potassium phosphate monobasic, 2-Propanol, Rubidium chloride, SIGMA*Fast* 3,3'-diaminobenzidine tablets, Sodium acetate, Sodium chloride, Sodium citrate monobasic, Sodium citrate tribasic dihydrate, Sodium dodecyl sulphate, Sodium hydroxide pellets, Sodium phosphate dibasic, Sodium phosphate monobasic monohydrate, *N,N,N',N'*-Tetramethylethylenediamine, Tris(hydroxymethyl)aminomethane, Triton X-100, Tween-20.

2.1.2 Enzymes

New England Biolabs (UK) Ltd., Hitchin, Hertfordshire, UK:

AvrII, BSA, NEBuffer 3, NEBuffer 4, PstI, SalI, SnaBI.

Promega UK Ltd., Southampton, UK:

ImProm-II reverse transcriptase, 5X ImProm-II reaction buffer, Magnesium chloride, Recombinant RNasin ribonuclease inhibitor.

Sigma-Aldrich Ireland Ltd., Dublin, Ireland:

Collagenase from *Clostridium histolyticum* (Sigma Blend Type H), KlenTaq LA DNA polymerase, 10X KlenTaq PCR buffer, 10X PCR buffer without $MgCl_2$, $MgCl_2$, REDTaq Genomic DNA polymerase.

2.1.3 Media components

BD Biosciences Ltd., Cowley, Oxford, UK:

Bacto agar, Bacteriological peptone, Difco yeast nitrogen base w/o amino acids and ammonium sulphate, Difco yeast nitrogen base with ammonium sulphate and w/o amino acids, Tryptone, Yeast extract.

Invitrogen Ltd., Paisley, UK:

Geneticin.

Sigma-Aldrich Ireland Ltd., Dublin, Ireland:

Ampicillin sodium salt, Ammonium sulphate, D-Biotin, L-Histidine, L-Glutamic acid, L-Methionine, L-Lysine, L-Leucine, L-Isoleucine, D-Sorbitol.

2.1.4 Proteins, peptides and antibodies

Bachem (UK) Ltd., St. Helens, Merseyside, UK:

Boc-Ala-Gly-Pro-Arg-AMC · HCl, Boc-Val-Leu-Lys-AMC, Boc-Val-Pro-Arg-AMC, Tos-Gly-Pro-Arg-AMC · HCl, Tos-Gly-Pro-Lys-AMC, Z-Phe-Ala-diazomethylketone (CHN_2), Z-Phe-Arg-AMC · HCl, Z-Pro-Arg-AMC · HCl.

Bio-Rad Laboratories Ltd., Hemel Hempstead, Hertfordshire, UK:

Goat anti-mouse IgG (H+L)-horseradish peroxidase

Peptides International Inc., Louisville, Kentucky, USA:

Z-Leu-Arg-AMC.

Roche Products (Ireland) Ltd., Dublin, Ireland:

Mouse monoclonal Anti-His_6 IgG

Sigma-Aldrich Ireland Ltd., Dublin, Ireland:

Anti-rabbit IgG (whole molecule)-peroxidase antibody produced in goat, Bovine serum albumin, Collagen from bovine nasal septum (Type II), Collagen from human placenta (Type VI) , Gelatin (porcine), Z-Arg-AMC · HCl, Z-Arg-Arg-AMC · HCl.

2.1.5 Vectors, strains and kits

Geneart GmbH, Regensburg, Germany:

*03-1024*pPCR-Script.

Invitrogen Ltd., Paisley, UK:

Escherichia coli Top 10F', Multi-Copy *Pichia* Expression Kit, pPIC9K, *Pichia* EasyComp Kit, *Pichia pastoris* GS115.

Promega UK Ltd., Southampton, UK:

pGEM-T Easy Vector System I, Wizard DNA Clean-up Kit, Wizard PCR Preps DNA Purification System.

Qiagen Ltd., Crawley, West Sussex, UK:

QiaQuick Gel Extraction Kit, Qiagen Midi Plasmid Prep Kit.

Vectors previously developed in the laboratory:

pAAH5.FheCL1, pAAH5.FheCL2.

2.1.6 Oligonucleotide primers

All oligonucletide primers were synthesised by Sigma-Genosys Ltd., Cambridge, UK.

Table 2.1: Sequences of oligonucleotide primers: restriction or mutation sites are underlined, His6-tag sequences are marked in bold, restriction enzyme clamps and linkers are marked in italics.

Name:	Sequence:	Features/Notes:
FheCL1F	5'-*GCGGC*TACGTA TCGAATGATGATTTGTGGCAT-3'	Binds 5'-end of FheCL1 and FheCL2 propeptide coding sequences. Includes SnaBI restriction site.
FheCL1R	5'-*GCG* CCTAGG TCA **GTGGTGGTGGTGGTGGTG** *GGGCCC* CGGAAATCGTGCCACCAT-3'	Binds 3'-end of FheCL1 and FheCL2 coding sequences. Includes AvrII restriction site, stop codon and His_6-tag sequence.
FheCL1mutF	5'-GTGGCTCCGGTTGGGCATTC-3'	Binds active site region of FheCL1. Includes T→G mismatch.
FheCL1mutR	5'-GAATGCCCAACCGGAGCCAC-3'	Binds active site region of FheCL1. Includes A→C mismatch.
M13F	5'-CGCCAGGGTTTTCCCAGTCACGAC-3'	Binds towards 5'-end of *lacZ* gene in M13 or pUC based vectors.
M13R	5'-TCACACAGGAAACAGCTATGAC-3'	Binds towards 3'-end of *lacZ* gene in M13 or pUC based vectors.
α-factor	5'-TACTATTGCCAGCATTGCTGC-3'	Binds to *P. pastoris* α-factor signal sequence.
5'AOX1	5'-GACTGGTTCCAATTGACAAGC-3'	Binds to 5'-end of *P. pastoris* alcohol oxidase 1 promoter.
3'AOX1	5'-GCAAATGGCATTCTGACATCC-3'	Binds to 3'-end of *P. pastoris* alcohol oxidase 1 promoter.
16dT	5'-TTTTTTTTTTTTTTTT-3'	Oligo-dT to bind poly(A) tails of mRNA in RT-PCR.

2.2 Solutions and Media

2.2.1 Solutions

50X TAE:

242 g tris

57.1 ml acetic acid

100 ml 0.5 M EDTA

Solution adjusted to pH 8.0 with HCl and brought to 1 L with dH_2O. Stored at room temperature.

TE:

10 mM tris/HCl, pH 8.0

1 mM EDTA

10X Phosphate buffered saline (PBS):

400 g NaCl

10 g KCl

58 g Na_2HPO_4

10 g KH_2PO_4

Solution adjusted to pH 7.3 with HCl or NaOH and brought to 1 L with dH_2O. Stored at room temperature.

Lysis buffer:

50 mM glucose

10 mM EDTA

25 mM tris

5 M potassium acetate solution, pH 4.8:

60 ml 5 M potassium acetate

11.5 ml glacial acetic acid

28.5 ml dH_2O

TFB1:

30 mM potassium acetate

10 mM $CaCl_2$

50 mM $MnCl_2$

100 mM RbCl

15% glycerol

Solution adjusted to pH 5.8 with 1M acetic acid and filter sterilised with a 0.2 μm filter.

Stored at room temperature.

TFB2:

10 mM PIPES, pH 6.5

75 mM $CaCl_2$

10 mM RbCl

15% glycerol

Solution adjusted to pH 6.5 with 1 M KOH and filter sterilised with a 0.2 μm filter.

Stored at room temperature.

SE:

1 M sorbitol

25 mM EDTA, pH 8.0

Provided in Multi-Copy *Pichia* Expression Kit (Invitrogen)

SED:

19 ml of SE

1 ml of 1 M DTT

SCE:

1 M sorbitol

10 mM sodium citrate buffer, pH 5.8

1 mM EDTA

Provided in Multi-Copy *Pichia* Expression Kit (Invitrogen)

CaS:

1 M sorbitol

10 mM tris/HCl, pH 7.5

10 mM $CaCl_2$

Provided in Multi-Copy *Pichia* Expression Kit (Invitrogen)

40% PEG:

40% PEG 3350 (reagent grade) in water

Provided in Multi-Copy *Pichia* Expression Kit (Invitrogen)

CaT:

20 mM tris/HCl, pH 7.5

20 mM $CaCl_2$

Provided in Multi-Copy *Pichia* Expression Kit (Invitrogen)

PEG/CaT:

Mix 40% PEG with CaT in a 1:1 ratio

Solution I:

Sorbitol solution containing ethylene glycol and DMSO for the preparation of competent *P. pastoris* cells

Provided in *Pichia* EasyComp Kit (Invitrogen)

Solution II:

PEG solution for the transformation of competent *P. pastoris* cells

Provided in *Pichia* EasyComp Kit (Invitrogen)

Solution III:

Salt solution for washing and plating transformed *P. pastoris* cells

Provided in *Pichia* EasyComp Kit (Invitrogen)

Denaturing solution:

0.5 M NaOH

1.5 M NaCl

Stored at room temperature.

Neutralising solution:

1.5 M NaCl

0.5 M tris/HCl, pH 7.4

Stored at room temperature.

20X SSC (500 ml):

87.65 g NaCl

50.25 g trisodium citrate

500 ml dH_2O

Stored at room temperature.

10X electrophoresis buffer:

250 mM tris

1.92 M glycine

1% SDS

Solution adjusted to pH 8.3 with HCl. Stored at room temperature.

5X reducing loading buffer:

0.625 M tris/HCl, pH 6.8

50% glycerol

10% SDS

0.1% bromophenol blue

5% 2-mercaptoethanol

4X non-reducing loading buffer with SDS:

0.5 M tris/HCl, pH 6.8

40% glycerol

0.08% bromophenol blue

8% SDS

4X non-reducing loading buffer without SDS:

0.5 M tris/HCl, pH 6.8

40% glycerol

0.08% bromophenol blue

Towbin transfer buffer:

6.06 g tris

28.8 g glycine

600 ml dH_20

200 ml methanol

Solution brought to 1 L with dH_20. pH should be between 8.1 and 8.5. Stored at room temperature.

Column buffer:

50 mM sodium phosphate buffer, pH 8.0

300 mM NaCl

10 mM imidazole

Solution adjusted to pH 8.0 with HCl. Stored at 4°C.

Wash buffer:

50 mM sodium phosphate buffer, pH 8.0

300 mM NaCl

20 mM imidazole

Solution adjusted to pH 8.0 with HCl. Stored at 4°C.

Elution buffer, pH 7:

50 mM sodium phosphate buffer, pH 7.0

300 mM NaCl

250 mM imidazole

Solution adjusted to pH 7.0 with HCl. Stored at 4°C.

Elution buffer, pH 6:

50 mM sodium phosphate buffer, pH 6.0

300 mM NaCl

250 mM imidazole

Solution adjusted to pH 6.0 with HCl. Stored at 4°C.

Breaking buffer:

50 mM sodium phosphate, pH 7.4

1 mM EDTA

5% glycerol

Stored at room temperature.

AE buffer:

50 mM sodium acetate, pH 5.3

1 mM EDTA

Stored at room temperature.

2.2.2 Media

Luria broth (LB):

10 g tryptone

5 g yeast extract

10 g NaCl

Dissolved in 1 L dH_20 and autoclaved for 20 min at 121°C. Stored at 4°C.

LB agar (20 plates):

5 g tryptone

2.5 g yeast extract

5 g NaCl

7.5 g agar

Dissolved in 500 ml dH_20 and autoclaved for 20 min at 121°C. Cooled to ~55°C before addition of any antibiotics and poured. Stored at 4°C.

10X YNB:

134 g yeast nitrogen base with ammonium sulphate without amino acids

OR:

34 g yeast nitrogen base without ammonium sulphate and amino acids

100 g ammonium sulphate

Dissolved in 1 L dH_20, with heating if necessary. Solution sterile filtered with a 0.2 μm filter. Stored at 4°C.

500X B:

20 mg D-biotin

100 ml dH_2O

Solution sterile filtered with a 0.2 μm filter. Stored at 4°C.

100X H:

200 g L-histidine

100 ml dH_2O

Solution heated to no more than 50°C until dissolved, then sterile filtered with a 0.2 μm filter. Stored at 4°C.

10X D:

200 g D-glucose

Dissolved in 1 L dH_20 and autoclaved for 20 min at 121°C or sterile filtered with a 0.2 μm filter. Stored at 4°C.

10X M:

50 ml methanol

950 ml dH_2O

Solution sterile filtered with a 0.2 μm filter. Stored at 4°C.

10X GY:

100 ml glycerol

900 ml dH_2O

Solution autoclaved for 20 min at 121°C or sterile filtered with a 0.2 μm filter. Stored at 4°C.

100X AA:

500 mg L-glutamic acid

500 mg L-methionine

500 mg L-lysine

500 mg L-leucine

500 mg L-isoleucine

100 ml dH_2O

Solution sterile filtered with a 0.2 μm filter. Stored at 4°C.

Yeast extract peptone dextrose medium (YPD):

10 g yeast extract

20 g peptone

Dissolved in 900 ml dH_2O, autoclaved for 20 min at 121°C and cooled. The following was then added:

100 ml 10X D

YPD agar (20 plates):

5 g yeast extract

10 g peptone

10 g agar

Dissolved in 450 ml dH_2O, autoclaved for 20 min at 121°C and cooled to ~55°C. The following was then added:

50 ml 10X D

Any antibiotics required were added and the plates poured.

SOS medium:

1 M sorbitol

0.3X YPD

10 mM $CaCl_2$

Provided in Multi-Copy *Pichia* Expression Kit (Invitrogen)

Regeneration dextrose (RD) top agar:

186 g sorbitol

10 g agar

Dissolved in 700 ml dH_2O and autoclaved for 20 min at 121°C. Placed in a water bath at 60°C

The following mixture was prepared and prewarmed to 45°C before addition to the main mix.

100 ml 10X D

100 ml 10X YNB

2 ml 500X B

10 ml 100X AA

88 ml sdH_2O

RDB agar (20 plates):

93 g sorbitol

10 g agar

Dissolved in 350 ml dH_2O and autoclaved for 20 min at 121°C. Placed in a water bath at 60°C

The following mixture was prepared and prewarmed to 45°C before addition to the main mix.

50 ml 10X D

50 ml 10X YNB

1 ml 500X B

5 ml 100X AA

44 ml sdH_2O

Plates poured.

RDHB agar (20 plates):

93 g sorbitol

10 g agar

Dissolved in 690 ml dH_2O and autoclaved for 20 min at 121°C. Placed in a water bath at 60°C

The following mixture was prepared and prewarmed to 45°C before addition to the main mix.

50 ml 10X D

50 ml 10X YNB

1 ml 500X B

5 ml 100X AA

5 ml 100X H

44 ml sdH_2O

Plates poured.

Minimal dextrose (MD) agar (20 plates):

7.5 g agar

Dissolved in 400 ml dH_2O, autoclaved for 20 min at 121°C and cooled to 60°C. The following was then added:

50 ml 10X YNB

1 ml 500X B

50 ml 10X D

Plates poured.

Minimal methanol (MM) agar (20 plates):

7.5 g agar

Dissolved in 400 ml dH_2O, autoclaved for 20 min at 121°C and cooled to 60°C. The following was then added:

50 ml 10X YNB

1 ml 500X B

50 ml 10X M

Plates poured.

YNB/2% methanol agar (20 plates):

7.5 g agar

Dissolved in 250 ml dH_2O, autoclaved for 20 min at 121°C and cooled to 60°C. The following was then added:

50 ml 10X YNB

200 ml 10X M

Plates poured.

Buffered glycerol-complex medium (BMGY):

10 g yeast extract

20 g peptone

Dissolved in 700 ml dH_2O, autoclaved for 20 min at 121°C and cooled. The following was then added:

100 ml 1 M potassium phosphate buffer, pH 6.0

100 ml 10X YNB

2 ml 500X B

100 ml 10X GY

Buffered methanol-complex medium (BMMY):

10 g yeast extract

20 g peptone

Dissolved in 700 ml dH_2O, autoclaved for 20 min at 121°C and cooled. The following was then added:

100 ml 1 M potassium phosphate buffer, pH 6.0 (or other required pH)

100 ml 10X YNB

2 ml 500X B

100 ml 10X M

2.3 Methods

2.3.1 *In-vitro* cultivation of parasites and preparation of excretory-secretory (ES) products

Adult *F. hepatica* were obtained from infected cattle at a local abattoir and cultures *in vitro* in RPMI 1640 containing 30 mM HEPES, 1 % glucose and 25 mg/ml gentamycin as described by Dalton and Heffernan (1989). The medium was collected after 6 h, cleared by centrifugation at 14,000 x *g* for 30 min at 4°C and stored at -20°C. This preparation was termed excretory-secretory (ES) products.

2.3.2 Preparation of anti-propeptide and anti-mature cathepsin L antisera

Native mature cathepsin L1 (nFheCL1) and L2 (nFheCL2) was previously purified in the laboratory from excretory-secretory (ES) products of adult *F. hepatica* by gel permeation and ion exchange chromatography and antiserum prepared in rabbits (Smith et al., 1993a; Dowd et al., 1994). Recombinant cathepsin L1 propeptide was also previously generated and purified in the laboratory (Roche et al., 1999). Antiserum to this propeptide was prepared by immunizing New Zealand White rabbits five times with 20 μg of protein formulated in Freund's Complete and Incomplete Adjuvant.

2.3.3 Immunofluorescence and immunoelectron microscopy

Immunofluorescence and immunoelectron microscopy experiments were performed in Prof. Aaron Maule's laboratory at Queen's University Belfast, UK, using anti-propeptide and anti-mature cathepsin L1 antisera from our laboratory (Collins et al., 2004). Adult *F. hepatica* were recovered from infected cattle at a local abattoir, washed, and transported to the laboratory in mammal saline (0.9% NaCl) at 37 °C. Parasites were rinsed in mammal saline and allowed to regurgitate their gut contents before being flat-fixed in 4% paraformaldehyde (PFA) in PBS (pH 7.2) for 4 h. They

were washed in antibody diluent (ABD: PBS with, 0.1% bovine serum albumin, 0.3% Triton X-100, 0.1% sodium azide) for 24 h before being incubated for 48 h at 4 °C in antiserum prepared against purified mature cathepsin L1 (diluted 1:3000) and subsequently washed in ABD (24 h, 4 °C). Swine anti-rabbit tetramethyl rhodamine isothiocyanate (TRITC; 1:100; Dako Ltd.) was used to visualize bound primary antibody before the worms were washed in ABD (24 h, 4 °C) and mounted on glass microscope slides in PBS/glycerol (1:9) containing 2.5% 2,4-diazabicylco 2.2.2 octane. Specimens were viewed using a Leica TCS-NT confocal scanning laser microscope. For electron microscopy worms were washed in mammal saline and fixed for 1 h in 2% double-distilled glutaraldehyde (GTA) (Agar Scientific) in 0.1 M sodium cacodylate buffer (pH 7.2) containing 3% sucrose at 4 °C. Following thorough washing in buffer, specimens were dehydrated through graded ethanol to propylene oxide, infiltrated and embedded in Agar 100 resin (Agar Scientific). Ultrathin sections (80–90 nm) were cut on a Reichert Ultracut E ultramicrotome, collected on bare 200-mesh nickel grids and dried at room temperature. For immunogold labeling sections were etched with 10% hydrogen peroxide for 5 min and rinsed thoroughly with 20 mM Tris-HCl buffer (pH 8.2) containing 0.1% bovine serum albumin and Tween 20 (1:40 dilution). Grids were incubated in normal goat serum (1:20 dilution) for 30 min and then transferred to primary antibody diluted to 1:20,000 with 0.1% bovine serum albumin/Tris-HCl buffer for 12–18 h. Grids were then washed in bovine serum albumin/Tris-HCl and transferred to a 20 μl drop of 10 nm gold-conjugated goat anti-rabbit IgG (Bio Cell International) for 2 h at room temperature. Following another buffer wash, grids were lightly fixed with 2% double-distilled GTA for 3 min, and finally washed with buffer and rinsed with distilled water. Grids were double stained with uranyl acetate (5 min) and lead citrate (3 min) and examined in a FEI (Philips) CM100 transmission electron microscope, operating at 100 keV. Controls consisted of (i) incubation of whole-mounts/sections

with secondary antibody in the absence of primary antibody and (ii) incubation with preimmune serum followed by the secondary antiserum.

2.3.4 Sequence analysis

The *F. hepatica* cathepsin L1 protein sequence was aligned with several related cathepsin sequences using ClustalX 1.81. Protein sequences used included *Carica papaya*, *Fasciola hepatica* cathepsin L1, *F. hepatica* cathepsin L2, *F. gigantica* cathepsin L1, *Schistosoma mansoni* cathepsin L2, *Caenorhabditis elegans* CPL-1, mouse cathepsin L, rat cathepsin L, and human cathepsin L.2 Sequences were numbered according to the papain numbering used by Vernet et al. (1995) where the propeptide residues are recorded as a negative beginning from the cleavage site between propeptide and mature enzyme. Sequences were also compared with ClustalW (http://npsa-pbil.ibcp.fr) to determine similarity and identity to FheCL1 and FheCL2.

Multiple cathepsin L protein sequences from *F. hepatica* and the closely related organism *F. gigantica* were aligned using ClustalX 1.81. Phylogenetic trees were generated from the alignment by the boot-strapped (1000-trial) neighbour-joining method using MEGA (Kumar et al., 2001). Accession numbers for all analysed sequences are given in Table 2.2.

Table 2.2: Accession numbers of sequences analysed. For *F. hepatica* and *F. gigantica* sequences, the alternate name is from Irving et al, 2003.

Name:	Species:	GenBank Accession no.:	Alternate names:	Reference:
FheCL1	*Fasciola hepatica*	U62288	FhCatL1	Roche et al. (1997)
FheCL2	*Fasciola hepatica*	U62289	FhCatL2	Dowd et al. (1997)
FheCL3	*Fasciola hepatica*	AJ279091	FhCatL8	Harmsen et al. (2004)
FheCL4	*Fasciola hepatica*	L33772	FhCatL4	Wijffels et al. (1994)
FheCL5	*Fasciola hepatica*	AF271385	FhCatL5	Smooker et al. (2000)
FheCL6	*Fasciola hepatica*	AB009306	FhCatL6	Yamasaki and Aoki (1993)
FheCL7	*Fasciola hepatica*	Z22765	FhCatL7	Heussler and Dobbelaere (1994)
FheCL8	*Fasciola hepatica*	L33771	FhCatL3	Wijffels et al. (1994)

FheCL9	*Fasciola hepatica*	AJ279092	FhCatL9	Cornelissen et al. (2001)
FheCL10	*Fasciola hepatica*	AJ279093	FhCatL10	Harmsen et al. (2004)
FheCL11	*Fasciola hepatica*	AY029229	FhCatL11	Carnevale et al. (2001)
FheCL12	*Fasciola hepatica*	S43991	---	Wijffels et al. (1994)
FheCL13	*Fasciola hepatica*	AY277628	---	Kofta et al. (2000)
FheCL14	*Fasciola hepatica*	Z22763	---	Heussler and Dobbelaere (1994)
FheCL15	*Fasciola hepatica*	Z22764	---	Heussler and Dobbelaere (1994)
FheCL16	*Fasciola hepatica*	Z22766	---	Heussler and Dobbelaere (1994)
FheCL17	*Fasciola hepatica*	Z22767	---	Heussler and Dobbelaere (1994)
FheCL18	*Fasciola hepatica*	Z22769	---	Heussler and Dobbelaere (1994)
FheCL19	*Fasciola hepatica*	AY519971	---	---
FheCL20	*Fasciola hepatica*	AF490984	---	---
FheCL21	*Fasciola hepatica*	AY519972	---	---
FgiCL1A	*Fasciola gigantica*	AF112566	FgCatL-A	Grams et al. (2001)
FgiCL1B	*Fasciola gigantica*	AF239264	FgCatL-B	Grams et al. (2001)
FgiCL1C	*Fasciola gigantica*	AF239265	FgCatL-C	Grams et al. (2001)
FgiCL1D	*Fasciola gigantica*	AF239266	FgCatL-D	Grams et al. (2001)
FgiCL1E	*Fasciola gigantica*	AF239267	FgCatL-E	Grams et al. (2001)
FgiCL1F	*Fasciola gigantica*	AF239268	FgCatL-F	Grams et al. (2001)
FgiCL2	*Fasciola gigantica*	AF510856	FgCatL2	Irving et al. (2003)
FgiCL3	*Fasciola gigantica*	AF419329	---	---
FgiCL4	*Fasciola gigantica*	AB010923	---	Yamasaki et al. (2002)
FgiCL5	*Fasciola gigantica*	AB010924	---	Yamasaki et al. (2002)
SmaCL2	*Schistosoma mansoni*	Z32529	SmCL2	Michel et al. (1995)
CelCPL1	*Caenorhabditis elegans*	NP_507199	CPL-1	Hashmi et al. (2002)
MmuCL	*Mus musculus*	P06797	---	Joseph et al. (1988)
RnoCL	*Rattus norvegicus*	KHRTL	---	Ishidoh et al. (1987)
HsaCL	*Homo sapiens*	M20496	---	Joseph et al. (1988)
CpaPap	*Carica papaya*	P00784	Papain	Cohen et al. (1986)

2.3.5 Polymerase chain reaction (PCR)

PCR was used for the amplification of DNA sequences using specific oligonucleotide primers. Cloning and mutation of gene sequences, addition of restriction sites and His_6-tags and analysis of transformants were performed by PCR. In general, 25 or 50 µl reaction mixes were prepared in either 200 µl thin-walled PCR tubes or 500 µl tubes, depending on the model of thermal cycler used. Two different *Taq* polymerase enzyme preparations were used: (a) for Sigma REDTaq, reaction mixes

were prepared with final concentrations of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix, 0.8 μM of each primer, 0.02-0.05 U/μl of REDTaq DNA polymerase; (b) for the high fidelity polymerase KlenTaq LA, reaction mixes were prepared with final concentrations of 1X KlenTaq PCR buffer, 0.2 mM dNTP mix, 0.8 μM of each primer, 0.1 U/μl of KlenTaq LA DNA polymerase. DNA template quantities varied depending on the nature of the reaction. All reactions were brought to a total or 25 or 50 μl with nfH_2O. The standard PCR programme used was a hot start of 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 1 min and finished with a further 7 min at 72°C.

Details of any variations on the general method are given below where appropriate.

2.3.6 Agarose gel electrophoesis

Separation and visualisation of DNA and RNA was carried out by agarose gel electrophoresis (Sambrook et al., 1989). For solution and media compositions, see section 2.2. Gels between 50 and 120 ml were prepared, depending on number and volume of samples. This was done by adding 1% agarose to the correct volume of 1X TAE and heating in a microwave on low power, mixing occasionally, until the agarose was fully dissolved. The agarose solution was allowed to cool until hand hot. For every 20 ml of agarose solution prepared, 1 μl of 10 mg/ml ethidium bromide was added and mixed in well. The solution was poured into an appropriately sized gel rig and combs added to form wells. The solution was allowed to set at room temperature for standard gels and at 4°C for low melting point agarose gels. Once set, gels were transferred to a gel tank the tank filled with 1X TAE until the gel was covered. 100 bp and 1 kb ladders were prepared by mixing 5 μl of DNA with 5 μl of 1X TAE and 2 μl of 6X blue/orange loading dye. Samples were similarly prepared. Wells were loaded

and electrophoresis performed at 120 to 150 V for standard gels, 100 V for low melting point agarose gels, for 30-60 min. Gels were recorded under a UV lamp and camera system.

2.3.7 Preparation of competent *E. coli* cells

The rubidium chloride protocol used was taken from the Promega Protocols and Application Guide (3rd edition) and is adapted from Hanahan (1985). See section 2.2 for solution and media compositions. LB medium (2.5 ml) was inoculated with a single colony of *E. coli* Top 10F' cells from an LB agar plate and grown overnight at 37°C with shaking at 225 rpm. The entire overnight culture was then used to inoculate 250 ml of LB medium containing 20 mM $MgSO_4$. This culture was then grown in a 1L flask until the OD_{600} reached between 0.4 and 0.6. The cells were pelletted by centrifugation at 4,500 x *g* for 5 min at 4°C. For this step, the culture was split into two 250 ml centrifuge bottles and centrifuged in a Sorvall GSA rotor.

The cell pellet was gently resuspended in 100 ml of ice-cold TFB1, 50 ml per centrifuge bottle. The resuspended cells were combined and incubated on ice for 5 min at 4°C. The cells were again pelletted by centrifugation in a Sorvall GSA rotor at 4,500 x *g* for 5 min at 4°C. The pellet was gently resuspended on ice in 10 ml ice-cold TFB2 and then incubated on ice for 60 min. Finally, 200 µl aliquots were placed in 1.5 ml tubes and stored at -80°C until needed.

2.3.8 Transformation of competent *E. coli* cells

See section 2.2 for media compositions. Competent *E. coli* Top 10F' cells were thawed slowly on ice and divided into 100 µl aliquots in 1.5 ml tubes. DNA was added to a maximum of 20 µl or 0.5 µg, and mixed. Positive controls were prepared with either pGEM-T Easy or pPIC9K plasmid DNA as appropriate. Competent cells alone

were used as negative controls. Tubes were incubated on ice for 30 min, then heat shocked in a heating block at 42°C for 2 min. Cells were then cooled on ice for 1 min. Cells were added to 1 ml LB medium without antibiotics in a 20 ml tube and grown for 1 h at 37°C with shaking at 225 rpm. 100 µl of cells were plated in duplicate on LB agar plates containing 100 µg/ml ampicillin and incubated overnight at 37°C.

2.3.9 Small scale isolation of plasmid DNA from *E. coli* cultures by alkaline lysis

Plasmid DNA isolation was carried out using the alkaline lysis method (Sambrook et al., 1989). See section 2.2 for solution and media compositions. Between 2 and 5 ml LB medium containing 100 µg/ml ampicillin in a 50 ml conical tube was inoculated with one colony from an agar plate and grown overnight at 37°C with shaking at 225 rpm. The culture was transferred in 1 ml aliquots into 1.5 ml tubes and centrifuged in a benchtop microcentrifuge at 12,000 x *g* for 1 min. The supernatant was decanted and the cell pellet resuspended with 100 µl of lysis buffer, with vigorous vortexing. Tubes were left at room temperature for 5 to 10 min and then 200 µl of freshly prepared 0.2 N NaOH/1% SDS was added at room temperature. Following an incubation on ice for 10 min, 200 µl of 5 M potassium acetate, pH 4.8, was added and the tubes gently inverted to mix. Tubes were incubated on ice for 10 min, followed by centrifugation at 12,000 x *g* for 5 min. The supernatant was transferred to a new 1.5 ml tube and the pellet discarded.

Phenol-chloroform extraction was performed by adding 400 µl phenol:chloroform (1:1) and vortexing the tube for 1 min. Tubes were then centrifuged at 12,000 x *g* for 5 min. The upper aqueous layer was transferred to a new 1.5 ml tube and 200 µl of chloroform added. Tubes were vortexed for 30 s and centrifuged at 12,000 x *g* for 2 min. The upper aqueous layer was transferred to a new 1.5 ml tube.

The purified DNA was concentrated by ethanol precipitation, resuspended in 40 µl of nfH_2O and stored at -20°C until required.

2.3.10 Concentration of DNA solutions by ethanol precipitation

At least 2.5 volumes of ice-cold 95% ethanol was added to each tube containing DNA to be concentrated and the tubes frozen at -20°C for at least 1 h. Tubes were centrifuged at 12,000 x *g* for 10 min. The supernatant was discarded and the pellets washed with 100 µl of 70% ethanol. Tubes were centrifuged at 12,000 x *g* for 5 min and the supernatants again discarded. The pellets were allowed to air dry before being resuspended with the required volume of nfH_2O (usually 5 to 50 µl).

2.3.11 Determination of DNA and RNA concentrations

A spectrophotometric method was used to determine the concentrations of DNA and RNA preparations (Sambrook et al., 1989). A 50 µg/ml solution of dsDNA has an $OD_{260} = 1$, to within a small degree of error. As the ratio of absorbance to DNA concentration is linear up to an $OD_{260} = 2$, the concentration of any sample may be calculated directly from the absorbance. A similar principal may be applied to RNA, where a 40 µg/ml solution has an $OD_{260} = 1$.

A diluted sample of DNA or RNA was prepared in nfH_2O or DEPC-treated water. The OD_{260} of the diluted sample was determined and the concentration calculated based on the following equations:

$$\mu g/ml\ DNA = \frac{50\mu g/ml}{1} \times (OD_{260}\ measured) \times (dilution\ factor)$$

$$\mu g/ml\ RNA = \frac{40\mu g/ml}{1} \times (OD_{260}\ measured) \times (dilution\ factor)$$

Purity of the sample was also determined by reading the OD_{280} and calculating the ratio of OD_{260}:OD_{280} (values of between 1.7 and 2.0 are expected for pure nucleic acid).

2.3.12 Preparation of *E. coli* and *P. pastoris* glycerol stocks

See section 2.2 for media compositions. For all *E. coli* Top 10F' stocks, 10 ml LB medium containing 100 μg/ml ampicillin in a 50 ml conical tube was inoculated with a single colony from an LB agar plate, also containing 100 μg/ml ampicillin. This culture was grown overnight at 37°C with shaking at 225 rpm. 200 μl of sterile filtered glycerol was added to 800 μl aliquots of culture in 1.5 ml tubes. Tubes were then vortexed briefly and stored at -80°C until required.

For all *P. pastoris* stocks, 20ml YPD medium in a 100 ml baffled flask was inoculated with a single colony from a YPD agar plate. This culture was grown for 48 h at 30°C with shaking at 250 rpm. 200 μl of sterile filtered glycerol was added to 800 μl aliquots of culture in 1.5 ml tubes. Tubes were then vortexed briefly and stored at -80°C until required.

2.3.13 Construction of expression vector encoding cDNA for wildtype procathepsin L1 (FheproCL1)

The full length *F. hepatica* preprocathepsin L1 cDNA was previously cloned into the *S. cerevisiae* expression vector pAAH5 in our laboratory (Roche et al., 1997). The procathepsin L1 encoding cDNA was amplified from the pAAH5.FheCL1 vector by PCR with Sigma REDTaq. Primers were used to incorporate a SnaBI restriction site at the 5'-end of the gene and an AvrII restriction site and His_6-tag sequence at the 3'-end. These primers were labelled FheCL1F and FheCL1R for the forward and reverse primers respectively (see Table 2.1). The PCR reaction mix was prepared in 0.5 ml tubes, with 5 reaction tubes prepared. A reaction using primers specific for the pAAH5

plasmid was used as a positive control, and one replacing the plasmid with nfH_2O was used as a negative control. The PCR was run at 95°C for 5 min, followed by 40 cycles at 95°C for 1.5 min, 50°C for 1.5 min and 72°C for 1.5 min and finished with a further 7 min at 72°C.

The PCR products were separated on a 1% low melting point agarose gel and bands of approximately 980 kb were cut out and purified using the Promega Wizard PCR Preps DNA purification system. Both pPIC9K vector DNA and the FheproCL1 PCR product were double digested with the SnaBI and AvrII restriction enzymes using appropriate buffering conditions as per the manufacturers instructions. Restriction digestion was carried out in 0.5 ml tubes in a water bath at 37°C for 3 h. The mixes were then cleaned with a Promega Wizard DNA clean-up kit, with final elution in 40 µl nfH_2O. The DNA concentration of both the vector and insert digest preparations was determined.

The vector and insert were ligated in two reaction with T4 DNA ligase. The first of these contained a molar ratio of vector:insert of 1:3, while the second was at 1:1. A re-ligation control mix containing double digested vector but no insert was also prepared. The ligation reactions were incubated overnight at 4°C. The ligation reactions were then transformed into competent *E. coli* Top 10F' cells and grown on LB agar plates containing 100 µg/ml ampicillin overnight at 37°C. Several transformant colonies were selected from each ligation and replica plated on LB agar plates containing 100 µg/ml ampicillin and grown at 37°C overnight. Each selected colony was also inoculated into 2ml LB medium containing 100 µg/ml ampicillin and grown overnight at 37°C with shaking at 225 rpm. Plasmid DNA was purified from these cultures by the alkaline lysis method.

The selected transformants were screened for correct insertion of the FheproCL1 sequence by several methods: (a) purified plasmid DNA, along with control samples of

pPIC9K vector, was single digested with PstI restriction enzyme, using appropriate buffering conditions as per the manufacturer's instructions; restriction digestion was carried out in 0.5 ml tubes in a water bath at 37°C for 1.5 h; these samples were then separated on a 1% agarose gel along with a sample of the FheproCL1 PCR product to allow comparison; (b) PCR reactions with the pPIC9K α-factor forward and 3'AOX1 reverse primers (see Table 2.1) using Sigma REDTaq were performed; reaction mixes were again separated on 1% agarose gels to confirm appropriately sized products; (c) samples of purified plasmids which indicated correct insertion of the FheproCL1 DNA were sent for sequencing by Oswel DNA Sequencing, University of Southampton, UK, to confirm the insert was in the correct orientation and open reading frame and had no significant errors. Glycerol stocks of each transformant corresponding to plasmid DNA sent for sequencing were prepared. One plasmid was selected for transformation into *P. pastoris*.

2.3.14 Transformation of *P. pastoris* GS115 with pPIC9K.FheproCL1 construct

Transformation of *P. pastoris* with the pPIC9K.FheproCL1 expression vector construct was performed by the spheroplasting method (Higgins and Cregg, 1998). The transformation method was based on that provided in the Multi-Copy *Pichia* Expression Kit manual, Version F (Invitrogen). See section 2.2 for solution and media compositions.

pPIC9K.FheproCL1 and pPIC9K control DNA samples were linearised by restriction digested with SalI using appropriate buffering conditions as per the manufacturer's instructions, with tubes incubated at 37°C in a water bath for 2.5 h. Linearised DNA was cleaned using the Promega Wizard DNA clean up kit and the concentration determined. Yeast was prepared for spheroplasting by two growth steps. Firstly, a starter culture of 10 ml YPD in a 100 ml baffled flask was inoculated with a

single colony of *P. pastoris* GS115 from a YPD agar plate and grown overnight at 30°C with shaking at 250 rpm. Three 500 ml baffled flasks containing 200 ml of YPD were inoculated with 5, 10 and 20 µl respectively of cells from the starter culture. Secondly, these cultures were grown overnight as before and the culture with an OD_{600} of between 0.2 and 0.3 the following morning was harvested by centrifugation at 1,500 x *g* for 10 min at room temperature. The supernatant was decanted and discarded.

The cells were washed by gentle resuspension with 20 ml of sdH_2O, transfer to a new sterile 50 ml conical tube and centrifugation at 1,500 x *g* for 5 min at room temperature. The supernatant was discarded and the cells washed again, replacing the sdH_2O with 20 ml fresh SED. A third wash was performed with 20 ml of 1 M sorbitol and the cells gently resuspended by swirling with 20 ml of SCE buffer. The suspension was split between two sterile 50 ml conical tubes. One tube was used to determine optimal incubation time with zymolase. For this, a spectrophotometer was blanked at OD_{800} with 800 µl 5% SDS and 200 µl SCE. 200 µl of cells were added to a tube containing 800 ml 5% SDS as a zero time point. 7.5 µl of zymolase slurry was added to the remaining cells. The cells were incubated at 30°C and 200 µl samples taken at time points up to 50 min. These samples were also added to tubes containing 800 µl 5% SDS. The OD_{800} was read for all tubes and % spheroplasting was determined for each time point using the equation:

$$\%\ spheroplasting = 100 - \left[\left(\frac{OD_{800}\ at\ time\ t}{OD_{800}\ at\ time\ 0}\right) \times 100\right]$$

The time point which gave approximately 70% spheroplasting was chosen as optimal. 7.5 µl of zymolase slurry was added to the second tube of cells, which were incubated at 30°C for the optimal time as determined. Spheroplasts were harvested by centrifugation

at 750 x *g* for 10 min at room temperature and the supernatant discarded. Spheroplasts were washed very gently with 10 ml 1 M sorbitol and collected by centrifugation at 750 x *g* for 10 min at room temperature. The wash was repeated with 10 ml of CaS solution and the spheroplasts gently resuspended with 0.6 ml of CaS solution. Spheroplasts were transferred in 100 µl aliquots in 20 ml tubes and used for transformation immediately.

Linearised pPIC9K.FheproCL1 and pPIC9K control plasmids were added to separate spheroplast tubes and incubated at room temperature for 10 min. Then, 1 ml of PEG/CaT solution was added to each tube and mixed gently. Tubes were incubated for a further 10 min at room temperature. The suspension was centrifuged at 750 x *g* for 10 min at room temperature and the supernatant carefully removed. The transformed cells were resuspended in 150 µl of SOS medium and incubated at room temperature for 20 min. Finally, 850 µl of 1 M sorbitol was added to each tube.

Transformed spheroplasts were plated by mixing 100 µl aliquots of the spheroplast suspension with 10 ml of molten RD top agar which was poured onto RDB agar plates. Once the top agarose was hardened, plates were inverted and grown at 30°C for 7 days before further screening was carried out. For cell viability controls, 100 µl of untransformed spheroplasts was mixed with 900 µl of 1 M sorbitol. 100 µl aliquots were plated in RD top agar on RDB and RDHB agar plates and grown as before.

2.3.15 Screening of *P. pastoris* GS115 pPIC9K.FheproCL1 transformants

After initial screening on histidine-deficient (RDB) agar plates, transformants were screened for multiple inserts and for Mut phenotype based on the protocols provided in the Multi-Copy *Pichia* Expression Kit manual, Ver. F (Invitrogen). See section 2.2 for media compositions.

For each RDB agar plate of transformants, the layer of top agar containing the transformants was removed with a sterile spreader and transferred to a sterile 50 ml conical tube containing 15 ml sdH_2O. This tube was vortexed for 2 min and the agar allowed to settle. The cell density of the supernatant was determined by the use of a haemocytometer and the supernatant diluted to a cell density of 5×10^5 cells/ml with sdH_2O. Spread plates were prepared with 200 µl aliquots of cells on four YPD agar plates containing 0.25 mg/ml of geneticin and four containing 0.5 mg/ml of geneticin. Plates were inverted and grown at 30°C for 7 days.

Individual colonies from the YPD agar plates containing 0.5 mg/ml of geneticin were selected as they were likely to have greater numbers of inserts present. These colonies were replica plated onto pairs of MM and MD agar plates, along with colonies of *P. pastoris* GS115 HSA as a control for the Mut^S phenotype and *P. pastoris* GS115 β-Gal as a control for the Mut^+ phenotype. The grids printed in the Multi-Copy *Pichia* Expression Kit manual, Ver. F (Invitrogen), page 43 were used as a guide. Plates were inverted and grown at 30°C for 4 days. Suitable colonies were then selected for induction to confirm expression of recombinant protein.

2.3.16 Construction of expression vector encoding cDNA for inactive mutant procathepsin L1 ([Gly^{26}]FheproCL1) and transformation into *P. pastoris* GS115

Mutants were generated using the pPIC9K.FheproCL1 construct as a template by a PCR-based site-directed mutagenesis method known as gene splicing by overlap extension (SOEing) (Ho et al., 1989). The mutants were generated in our laboratory by Dr. Colin Stack (Collins et al., 2004).

The construction of the inactive FheproCL1 mutant involved changing the active site cysteine (Cys^{26}) residue to a glycine residue in a two-step PCR process. The primers used were FheCL1F, FheCL1R, FheCL1mutF and FheCL1mutR (see Table

2.1). Each reaction used one flanking primer that hybridized at one end of the target sequence (primer FheCL1F or FheCL1R) and one overlapping internal primer that hybridizes at the site of the mutation and contains the mismatched base (primer FheCL1mutR and FheCL1mutF). In the first round of amplification, two sections of the cDNA were amplified using primers FheCL1F+FheCL1mutR and FheCL1mutF+FheCL1R. These two PCR products, with an overlap of 21 bp at one end of each fragment, were then combined in a second PCR to amplify the entire [Gly^{26}]FheproCL1 cDNA. Primers for this reaction were the two outside primers used in each of the first round reactions (primers FheCL1F and FheCL1R). All PCRs used high-fidelity *Taq* polymerase and run for 25 cycles at 94°C for 30 s, 55°C for 1 min and 72°C for 2 min.

The PCR product was then inserted into the AvrII/SnaBI site of expression vector pPIC9K and the plasmid insert was sequenced to verify the presence of the correct gene sequence and mutation. Two selected plasmids were linearised by digestion with SalI and introduced to *P. pastoris* cells by spheroplasting (Higgins and Cregg, 1998). Transformants were selected for their ability to grow on histidine-deficient agar plates and on agar plates containing minimal media and methanol. Insertion of the gene into *P. pastoris* was confirmed by PCR using primers specific to the yeast genome (Linder et al., 1996).

During the screening process, another procathepsin L mutant was identified that exhibited a slower migration on the SDS-PAGE. The cDNA encoding this proenzyme was sequenced and shown to have in addition to the Cys^{26} to Gly^{26} mutation a proline substituted for a leucine residue at position –12 in the propeptide. This clone was termed [$Pro^{-12}Gly^{26}$]FheproCL1 and used for further studies

2.3.17 Construction of expression vector encoding cDNA for wildtype procathepsin L2 (FheproCL2)

The full length *F. hepatica* preprocathepsin L2 cDNA was previously cloned into the *S. cerevisiae* expression vector pAAH5 in the laboratory (Dowd et al., 1997). The procathepsin L2 encoding cDNA was amplified from the pAAH5.FheCL2 vector by PCR with Sigma REDTaq. As with the amplification of FheproCL1, primers were used to incorporate a SnaBI restriction site at the 5'-end of the gene and an AvrII restriction site and His_6-tag sequence at the 3'-end. Primers FheCL1F and FheCL1R were again used for this purpose. Five identical PCR reactions were prepared as for the amplification of FheproCL1 from pAAH5.FheCL1. A reaction replacing the pAAH5 plasmid with pPIC9K.FheproCL1 was used as a positive control, and one replacing the plasmid with nfH_2O was used as a negative control. The standard PCR programme was used (see Section 2.3.5).

The PCR products were separated on a 1% low melting point agarose gel and bands of approximately 980 kb were cut out and purified using the Promega Wizard PCR Preps DNA purification system. Both pPIC9K vector DNA and the FheproCL2 PCR product were double digested with the SnaBI and AvrII restriction enzymes, cleaned and the DNA concentrations determined as described for FheproCL1.

The vector and insert were ligated in two reactions with T4 DNA ligase, with molar ratios of vector:insert of 1:3 and 1:2 respectively. A re-ligation control using double digested vector but no insert was also prepared. The ligation reactions were incubated overnight at 4°C and transformed into competent *E. coli* Top 10F' cells. Several transformants were selected and grown as for FheproCL1 and plasmid DNA was purified from these cultures by the alkaline lysis method.

The selected transformants were screened for correct insertion of the FheproCL2 sequence by both single digestion with PstI and double digestion with SnaBI and AvrII

as for FheproCL1 (see Section 2.3.13). Samples of pPIC9K vector were also digested as controls. All samples were separated on a 1% agarose gel and compared with the pPIC9K vector banding pattern. Samples of purified plasmids which indicated correct insertion of the FheproCL2 DNA were sent for sequencing by Oswel DNA sequencing, University of Southampton, UK, to confirm the insert was in the correct orientation and open reading frame and had no significant errors. Glycerol stocks of each transformant corresponding to plasmid DNA sent for sequencing were prepared. Two plasmids were selected for transformation into *P. pastoris*.

2.3.18 Construction of expression vector encoding cDNA for inactive mutant procathepsin L2 ([Gly^{26}]FheproCL2)

The inactive mutant form of FheproCL2, in which the active site cysteine (Cys^{26}) residue was replaced with a glycine residue, was synthesised by Geneart GmbH, Regensburg, Germany. The synthetic gene (labelled *03-1024*) was assembled from synthetic oligonucleotides based on the original full length FheCL2 cDNA sequence (GenBank accession number U62289) with a single base change corresponding to the desired mutation. The fragment was cloned into the pPCR-Script Amp plasmid (Stratagene, CA, USA) and amplified by transformation into a bacteria. The purified *03-1024*pPCR-Script plasmid DNA construct was verified by sequencing and 10 µg lyophilised for shipping. On receipt, the lyophilised plasmid was resuspended with 100 µl of TE, pH 8.0 to give a 100 ng/µl stock and stored at -20°C.

The inactive procathepsin L2 encoding cDNA was amplified from the *03-1024*pPCR-Script vector by PCR with KlenTaq LA. As with the amplification of FheproCL1, primers were used to incorporate a SnaBI restriction site at the 5'-end of the gene and an AvrII restriction site and His_6-tag sequence at the 3'-end. Primers

FheCL1F and FheCL1R were again used for this purpose. The standard PCR programme was used (see Section 2.3.5).

The PCR products were separated on a 1% low melting point agarose gel and bands of approximately 980 kb were cut out and purified using the QiaQuick Gel Extraction Kit. To allow for ligation into the pGEM-T Easy vector, a poly-A overhang was added to the PCR products by preparing a PCR reaction with Sigma REDTaq and the entire DNA preparation, but not including primers and only incubating at 72°C for 20 min in place of the normal PCR programme. The mix was then ligated directly into the pGEM-T Easy vector with T4 DNA ligase, incubated overnight at 4°C. The ligation reaction was then transformed into *E. coli* Top 10F' and grown on LB agar plates containing 100 μg/ml of ampicillin, 0.1 mM IPTG and 40 μg/ml of X-Gal. Transformants containing the insert were selected by colony pick PCR using the M13F and M13R primers (see Table 2.1) and Sigma REDTaq, and replacing the DNA template with individual colonies. A transformant including an insert of the correct size was selected and inoculated into two 12.5 ml aliquots of LB medium containing 100 μg/ml of ampicillin in 50 ml conical tubes and grown overnight at 37°C with shaking at 225 rpm. Plasmid DNA was purified from these cultures using the Qiagen Midi Plasmid Prep Kit.

Both pPIC9K vector DNA and the pGEM-T.[Gly^{26}]FheproCL2 construct were double digested with the SnaBI and AvrII restriction enzymes, cleaned and the DNA concentrations determined as described for FheproCL1 (see Section 2.2.13). The vector and insert were ligated with T4 DNA ligase, with a molar ratios of vector:insert of 1:3. The ligation reactions were incubated overnight at 4°C and transformed into competent *E. coli* Top 10F' cells. Several transformants were selected and screened for insertion of the [Gly^{26}]FheproCL2 sequence by colony pick PCR using the 3'AOX1 and 5'AOX1 primers (see Table 2.1). Several transformants which contained inserts of the correct

size were grown as before and plasmid DNA was purified using the Qiagen Midi Plasmid Prep Kit. Samples of purified plasmids which indicated correct insertion of the [Gly^{26}]FheproCL2 DNA were sent for sequencing by Australian Genome Research Facility (AGRF) Ltd., Brisbane, Australia, to confirm the insert was in the correct orientation and open reading frame and had no significant errors. Glycerol stocks of each transformant corresponding to plasmid DNA sent for sequencing were prepared. Two plasmids were selected for transformation into *P. pastoris*.

2.3.19 Transformation of *P. pastoris* GS115 with pPIC9K.FheproCL2 and pPIC9K.[Gly^{26}]FheproCL2 constructs

P. pastoris GS115 competent cells were prepared and transformed using the *Pichia* EasyComp Kit (Invitrogen). Methods used were based on the protocols provided in the *Pichia* EasyComp Kit manual, Ver. E (Invitrogen). See section 2.2 for media compositions. All provided solutions were equilibrated to room temperature before use.

pPIC9K expression vector constructs as well as pPIC9K and pPIC9K.FheproCL1 control DNA samples were linearised by restriction digested with SalI using appropriate buffering conditions as per the manufacturer's instructions, with tubes incubated at 37°C in a water bath for 2.5 h. Linearised DNA was cleaned using the Promega Wizard DNA clean up kit and the concentration determined. The DNA was concentrated by ethanol precipitation and resuspended in 5 µl of nfH_2O.

A starter culture was prepared by inoculating 10 ml of YPD medium in a 100 ml baffled flask with a single colony of *P. pastoris* GS115 from a YPD agar plate. The culture was grown overnight at 30°C with shaking at 250 rpm. Cells were then diluted to an OD_{600} of between 0.1 and 0.2 with fresh YPD medium and 10 ml transferred grown in a 100 ml baffled flask at 30°C with shaking at 250 rpm until an OD_{600} of between 0.6 and 1.0 had been reached. Cells were pelletted by centrifugation at 500 x *g*

for 5 min at room temperature and the supernatant discarded. Cells were resuspended with 10 ml of Solution I (provided) and the centrifugation repeated. Cells were resuspended in 1 ml of Solution I and sterile 1.5 ml screw-cap tubes prepared with 200 ml aliquots of the now-competent cells.

Linearised pPIC9K expression vector constructs and pPIC9K control plasmids were added to separate 50 µl aliquots of competent *P. pastoris* GS115 cells and gently mixed with 1 ml of Solution II (provided). Tubes were incubated at 30°C in a water bath for 1 h, with mixing every 15 min. The cells were heat shocked in a heating block at 42°C for 10 min. Cells were pelletted by centrifugation at 3,000 x *g* for 5 min at room temperature and the supernatant discarded. The pellets were washed by resuspension in 1 ml of Solution III (provided) followed by centrifugation at 3,000 x *g* for 5 min at room temperature and the supernatant discarded. The cells were finally resuspended in 100 µl of Solution III. Each suspension was then plated onto an RDB agar plate using a sterile spreader and grown at 30°C for between 3 and 6 days.

2.3.20 Screening of *P. pastoris* GS115 pPIC9K.FheproCL2 and pPI9K.[Gly^{26}]FheproCL2 transformants

After initial screening on histidine-deficient (RDB) agar plates, transformants were screened for expression by the colony blot rapid screening method (Higgins and Cregg, 1998). This involves growth and colony lysis on nitrocellulose membrane, followed by antibody probing of bound proteins. See section 2.2 for media compositions.

Individual colonies from the RDB agar plates were replica plated onto duplicate YPD agar plates, along with a colony per plate of *P. pastoris* GS115 transformed with pPIC9K.FheproCL1 as an expression control. The grids printed in the Multi-Copy *Pichia* Expression Kit manual, Ver. F (Invitrogen), page 43 were used as a guide.

These plates were inverted and grown at 30°C for 2 days. Nitrocellulose membrane was cut to size for a Petri dish and colonies transferred from one set of YPD agar plates by gently laying the membrane onto the plate and rubbing down with a sterile spreader. Each membrane was then lifted and placed colony side up on a YNB/2% methanol agar plate. Plates were inverted and incubated at 30°C for 2 days. The remaining set of duplicate YNB agar plates were stored at 4°C.

Five Petri dishes were prepared containing 3mm filter paper soaked, but not saturated, in colony lysis solutions (see below). Each nitrocellulose membrane was the transferred colony side up the Petri dishes as follows, ensuring no bubbles were formed: (a) 10% SDS for 10 min, (b) Denaturing solution for 5 min, (c) Neutralising solution for 5 min, repeated with a second fresh dish for 5 min and (d) 2X SSC for 15 min. All steps were incubated at room temperature with gentle agitation. The membranes were washed twice for 10 min each with 15 ml of PBS/0.5% Tween-20. They were then blocked by incubation for 1 h at room temperature in 15 ml of 5% milk in PBS/0.5% Tween-20, followed by two washes of 5 min each with 15 ml PBS/0.5% Tween-20. The membranes were probed with 15 ml of a 1/1000 dilution of mouse monoclonal anti-His_6 IgG in PBS/0.5% Tween-20 for 1 h at room temperature, followed by two washes of 5 min each with 15 ml PBS/0.5% Tween-20. A 1/1000 dilution of goat anti-mouse IgG-peroxidase conjugate in 15 ml PBS/0.5% Tween-20 was added and the membranes incubated at room temperature for a further 1 h. The membranes were washed three times for 5 min each with 15 ml PBS/0.5% Tween-20 and bound antibody visualised with 3,3'-diaminobenzidine (DAB) in urea buffer. The membranes were rinsed with water and allowed to dry.

The most strongly stained colonies were selected for induction to confirm expression of recombinant protein.

2.3.21 Determination of protein concentration

Determination of protein concentrations was carried out by two similar colourimetric methods, both using a clear 96-well micro-titre plate. Standards for both assays were prepared in 0.2 mg/ml increments between 0 and 2 mg/ml with bovine serum albumin (BSA) in an appropriate buffer. Buffers were chosen to match the samples being tested to take into account any variations in the colour reactions due to the buffer rather than the protein present.

In the bicinchoninic acid (BCA) protein assay, BSA standards were pipetted in 10 µl aliquots in duplicate or triplicate into a clear 96-well micro-titre plate. A 1/10 dilution of each sample was prepared if concentration of protein was expected to be higher than 2 mg/ml. 10 µl aliquots of samples in duplicate or triplicate where quantities of sample allowed were pipetted onto the plate. BCA working reagent was prepared by mixing BCA Reagent A: Reagent B in the ratio 50:1. BCA working reagent was then added to both standard and sample wells at 200 µl/well. Plates were incubated at 37°C for 30 min. Absorbances at 562 nm were read in a micro-titre plate reader. Duplicate and triplicate values were averaged. Standard curves were generated and protein concentrations determined using Microsoft Excel.

For the *DC* (detergent compatible) assay, BSA standards were pipetted in 5 µl aliquots in duplicate or triplicate into a clear 96-well micro-titre plate. A 1/10 dilution of each sample was prepared if concentration of protein was expected to be higher than 2 mg/ml. 5 µl aliquots of samples in duplicate or triplicate where quantities of sample allowed were pipetted onto the plate. *DC* Reagent A was added to both standard and sample wells at 25 µl/well, followed by *DC* Reagent B at 200 µl/well. Plates were left on the bench at room temperature for 15 min. Absorbances at 562 nm were read in a micro-titre plate reader. Duplicate and triplicate values were averaged. Standard curves were generated and protein concentrations determined using Microsoft Excel.

2.3.22 Expression of recombinant proteins by methanol induction

The basic methanol induction protocol was based on the Mut^S intracellular or secreted protocol provided in the Multi-Copy *Pichia* Expression Kit manual, Ver. F (Invitrogen). See section 2.2 for media compositions.

In the basic protocol, a 1 L baffled flask containing 100 ml of BMGY broth, buffered to pH 6.0, was inoculated with a single colony of *P. pastoris* transformant or from a glycerol stock. The culture was grown overnight at 30°C with shaking at 250 rpm until the OD_{600} reached between 2 and 6. The cells were harvested by centrifugation at 2000 x *g* for 5 min and the supernatant decanted and discarded. The pellet was resuspended with 20 ml of BMMY broth, buffered to pH 6.0, in a 100 ml baffled flask to induce expression. The methanol concentration was brought to 1% by the addition of 100 µl sterile filtered methanol. The induced cultures were grown at 30°C with shaking at 250 rpm for 6 days. Samples of 1 ml of culture were taken at daily intervals, centrifuged at 12,000 x *g* and the pellet and supernatant stored separately at -80°C. Sterile filtered methanol was added daily to a final concentration of 1%. The culture was harvested by centrifugation at 5000 x *g* for 10 min and recombinant protein purified by Ni-NTA affinity chromatography. Daily samples were analysed by 12% SDS-PAGE and fluorogenic substrate assay (see Section 2.3.31)

The basic methanol induction protocol was modified as necessary, with scale-up of cultures performed maintaining the proportion of total initial growth culture to total induction culture. The initial growth cultures were scaled up to a maximum of 35% of flask volume and the induction cultures to a maximum of 20% of flask volume to maintain sufficient aeration. Any cultures with greater volumes were split between multiple flasks. Environmental conditions and induction times were changed as necessary. After pilot expressions, the number of samples taken was reduced, with only analysis by 12% SDS-PAGE being performed.

2.3.23 Effect of environmental conditions on recombinant protein expression

In order to assess the effect on production of various environmental conditions, the basic methanol induction protocol was modified accordingly. In each case, a single initial growth culture was prepared and split evenly before centrifugation to ensure variations in conditions were checked against identical cultures. For variations in medium pH, cells were resuspended in BMMY broth, buffered to pH 4.5, 6.0, 7.0 and 8.0 in various experiments. Inhibition was investigated by daily addition of the cathepsin L inhibitor Z-Phe-Ala-diazomethylketone (Z-FA-CHN_2) to the induction culture to a final concentration of 25 μM. For temperature variations, the initial growth culture was grown at 30°C as normal, but the induction cultures were grown at 20°C, 25°C and 30°C with shaking at 250 rpm. Samples were analysed by 12% SDS-PAGE.

2.3.24 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed based on the method of Laemmli (1970) and carried out using either the Atto or Bio-Rad mini-gel systems. See section 2.2 for solution compositions. Plates were first arranged in the relevant casting rig for the system being used and checked for leaks. Either 12% or in some cases 10% resolving gels were prepared based on the recipes given in Table 2.3, with the addition of 10% ammonium persulphate and TEMED being left until just before pouring. The resolving gel was poured, allowing room for the stacking gel, and topped with isopropanol to ensure a flat surface and prevent air from interfering with the polymerisation.

5% stacking gels were prepared based on the recipe in Table 2.3, again leaving addition of 10% ammonium persulphate and TEMED until just before pouring. Once the resolving gels were fully polymerised, isopropanol was decanted from the gels and the stacking gel poured. Appropriate combs were then used to form loading wells and the stacking gel allowed to polymerise. The combs were removed from the gels and the

gels transferred from the casting rig to the electrophoresis chamber. The chamber was then filled with 1X electrophoresis buffer.

Table 2.3: Components of SDS-PAGE resolving and stacking gels, with volumes suitable for the preparation of up to four complete mini-gels.

	10% resolving	12% resolving	5% stacking
40% acrylamide/bisacrylamide	7.5 ml	9 ml	1.25 ml
1M tris/HCl pH 8.8	11.2 ml	11.2 ml	---
1M tris/HCl pH 6.8	---	---	1.25 ml
dH_2O	10.8 ml	9.3 ml	7.25 ml
10% SDS	300 µl	300 µl	100 µl
10% ammonium persulphate	150 µl	150 µl	100 µl
TEMED	50 µl	50 µl	50 µl

Samples were prepared by mixing 20 µl of sample with 5 µl 5X SDS-PAGE reducing buffer and boiled for 2 min. Markers were prepared by boiling for 2 min and 10 µl aliquots were loaded onto the gel. Appropriate volumes of prepared sample, usually 15 µl, were then loaded onto the gel and electrophoresis performed at between 100 V and 150 V for 60-90 min. Gels were carefully removed and stained with 0.1% Coomassie Brilliant Blue R-250 in 40% methanol/10% acetic acid for at least 2 h, more usually overnight. Gels were destained with 40% methanol/10% acetic acid for 1 h, followed by slow destaining in water. Stained gels were photographed and digitally scanned on a flatbed scanner between two clear acetates to record results.

2.3.25 Native and Semi-native polyacrylamide gel electrophoresis

See section 2.2 for solution compositions. For native PAGE, 10% resolving gels were prepared as for SDS-PAGE, replacing 10% SDS with dH_2O. No stacking gels were used. The electrophoresis chamber was filled with 1X electrophoresis buffer without SDS. Samples were prepared by mixing 15 µl of sample with 5 µl 4X SDS-PAGE non-reducing buffer without SDS, with no boiling step. No markers were used,

as they would not run correctly in a native gel. Appropriate volumes of prepared sample, usually 15 µl, were then loaded onto the gel and electrophoresis performed at between 100 and 150 V for 60-90 min. Gels were carefully removed and stained and destained as for SDS-PAGE. Stained gels were photographed and scanned on a flatbed scanner between two clear acetates to record results.

For semi-native PAGE, all gels were prepared as for standard SDS-PAGE, and 1X electrophoresis buffer was used as normal. Samples were prepared by mixing 15 µl of sample with 5 µl 4X SDS-PAGE non-reducing buffer with SDS, with no boiling step. No markers were used, as these would not match the semi-native preparation of the samples. Appropriate volumes of prepared sample, usually 15 µl, were then loaded onto the gel and electrophoresis performed at between 100 and 150 V for 60-90 min. Gels were carefully removed and stained and destained as for SDS-PAGE. Stained gels were photographed and digitally scanned on a flatbed scanner between two clear acetates to record results.

2.3.26 Gelatin substrate polyacrylamide gel electrophoresis (GS-PAGE)

GS-PAGE was carried out using either the Atto or Bio-Rad mini-gel systems. See section 2.2 for solution compositions. Plates were first arranged in the relevant casting rig for the system being used and checked for leaks. Either 10% or in some cases 12% resolving gels were prepared based on the recipes given in Table 2.4. The 3% gelatin was melted in a microwave on low power or water bath at 37°C before addition to the mixture, which was then cooled briefly on ice. As for SDS-PAGE gels, the addition of 10% APS and TEMED was left until just before pouring. The resolving gel was poured, allowing room for the stacking gel, and topped with isopropanol to ensure a flat surface and prevent air from interfering with the polymerisation.

Table 2.4: Components of GS-PAGE resolving and stacking gels, with volumes suitable for the preparation of up to four complete mini-gels.

	10% resolving	12% resolving	5% stacking
40% acrylamide/bisacrylamide	7.5 ml	9 ml	1.25 ml
1M tris/HCl pH 8.8	11.2 ml	11.2 ml	---
1M tris/HCl pH 6.8	---	---	1.25 ml
dH_2O	9.8 ml	8.3 ml	7.25 ml
3% gelatin	1 ml	1 ml	---
10% SDS	300 µl	300 µl	100 µl
10% APS	150 µl	150 µl	100 µl
TEMED	50 µl	50 µl	50 µl

5% stacking gels were prepared as for SDS-PAGE, based on the recipe in Table 2.4. The combs were removed from the gels and the gels transferred from the casting rig to the electrophoresis chamber. The chamber was then filled with 1X electrophoresis buffer. Samples were prepared by mixing 15 µl of sample with 5 µl 4X SDS-PAGE non-reducing buffer with SDS, with no boiling step. Appropriate volumes of prepared sample, usually between 5 and 15 µl, were then loaded onto the gel and electrophoresis performed at between 100 and 150 V for 60-90 min.

Gels were carefully removed and washed with 0.1M sodium phosphate pH 6.0 with 2.5% Triton X-100 for 30 min to remove SDS. This wash was repeated twice more with fresh buffer. Gels were then rinsed with 0.1M sodium phosphate pH 6.0 to remove excess Triton X-100. Gels were incubated overnight at 37°C in 0.1M sodium phosphate pH 6.0 and stained and destained as for SDS-PAGE. Stained gels were photographed and digitally scanned on a flatbed scanner between two clear acetates to record results.

For experiments in which activity at different pH was required, gels were loaded with identical samples and cut into strips with one sample per strip after electrophoresis had been performed. Each strip was treated individually as above, with the exception

that 0.1M sodium phosphate pH 6.0 was replaced with either a 0.1M sodium phosphate or a 0.1M sodium acetate buffer at an appropriate pH.

2.3.27 Purification of His_6-tagged recombinant proteins by Ni-NTA affinity chromatography

All recombinant proteins were purified from yeast media by affinity chromatography using Ni-NTA agarose. All procedures were performed at 4°C. See section 2.2 for solution compositions. Columns were prepared with between 1 ml and 15 ml Ni-NTA agarose resin, depending on the volume of the culture and the predicted yield. The resin bed was allowed to settle and the ethanol from storage of the resin allowed to run through. The column was equilibrated by passing through 10 column volumes of column buffer, pH 8.0.

Culture was prepared by centrifugation at 5000 x *g* for 10 min in 250 ml centrifuge bottles in a Sorvall GSA rotor. This was followed by filtration of the supernatant through a 0.45 μn filter, using a syringe filter system for smaller volumes and a vacuum filtration system for larger volumes. Filtered supernatant was then combined with column buffer, pH 8.0 in the ratio 1:4 (supernatant:buffer). This was passed over the column and the run-through collected. The column was washed with 15 column volumes of wash buffer, pH 8.0. Bound protein was eluted using between 1 and 5 column volumes of elution buffer, pH 7.0, followed by a second elution with elution buffer, pH 6.0.

Purified recombinant proteases were dialysed against either PBS, pH 7.3 or 20mM HEPES, pH 7.5 and stored at -80°C. Samples (250 μl) were taken at each stage of the process and analysed by SDS-PAGE to confirm removal of impurities. Concentration of the purified recombinant protein was determined by either the BCA or *DC* assay.

2.3.28 Immunoblotting (Western blots)

Immunoblots were prepared by transferring proteins to nitrocellulose membrane. See section 2.2 for solution compositions. Proteins to be probed were first resolved by 12% SDS-PAGE. The gel was carefully removed and rinsed in Towbin transfer buffer for 5-10 min. In many cases, a duplicate gel was prepared and stained for comparison with the final blot. Whattman paper was cut into 8 pieces, each slightly larger than the gel and soaked in Towbin transfer buffer for 5 min. Nitrocellulose membrane (NCM) was cut to size and soaked in Towbin transfer buffer for 5 min. The layers were arranged on a semi-dry transfer cell as follows, ensuring no bubbles were formed: 4 sheets of Whattman paper, NCM, polyacrylamide gel, 4 sheets of Whattman paper. Proteins were transferred at 15 V for 20-30 min.

The NCM was then blocked for between 45 min and 1 h at room temperature with 15 ml of either 5% milk or 0.5% BSA in PBS/0.5% Tween-20. Either 1% milk/PBS/0.5% Tween-20 or 0.5% BSA/PBS/0.5% Tween-20 was prepared as a wash buffer. The NCM was then probed with 15 ml of 1° antibody diluted in wash buffer for between 45 min and 1 h at room temperature. The NCM was washed three times for 5 min each, followed by incubation in 15 ml of 2° antibody-peroxidase conjugate diluted in wash buffer for between 45 min and 1 h at room temperature. The NCM was again washed three times for 5 min each and bound antibody was visualised with DAB in urea buffer as described by the manufacturer's instructions. Blots were rinsed with water and allowed to dry. Blots were digitally scanned using a flatbed scanner to record results.

2.3.29 Preparation of recombinant protein for *N*-terminal sequencing

See section 2.2 for solution compositions. Recombinant proteins were resolved by 12% SDS-PAGE. The gel was carefully removed and rinsed in Towbin transfer

buffer for 10-15 min. Whattman paper was cut into 8 pieces, each slightly larger than the gel and soaked in Towbin transfer buffer for 5 min. Polyvinylidene difluoride (PVDF) membrane was cut to size and soaked briefly in methanol, followed by Towbin transfer buffer for 5 min. The layers were arranged on a semi-dry transfer cell as follows, ensuring no bubbles were formed: 4 sheets of Whattman paper, PVDF membrane, polyacrylamide gel, 4 sheets of Whattman paper. Proteins were transferred at 15 V for 20 min. The membrane was then washed with dH_2O and stained with 0.025% Coomassie Brilliant Blue R-250 in 40% methanol (Higgins and Cregg, 1998). Protein bands of interest were subjected to *N*-terminal sequencing at Genoshpere Biotechnologies, Paris, France.

2.3.30 Comparison of rFheproCL1 with *F. hepatica* ES products

Aliquots of *F. hepatica* ES products were compared with samples taken from *P. pastoris* medium after induction of rFheproCL1 expression for 72 h at pH 8.0. Samples were analysed by 12% SDS-PAGE and by immunoblotting. Proteins were transferred to NCM and probed with either a 1/800 dilution of rabbit anti-FheCL1 serum, a 1/1200 dilution of rabbit anti-FheCL1 propeptide serum, or a 1/1,000 dilution of control preimmunised rabbit serum. The secondary antibody used was a 1/10,000 dilution of goat anti-rabbit IgG-peroxidase conjugate.

2.3.31 Fluorometric substrate assay for recombinant cathepsin L activity

Cathepsin L activity was determined by fluorometric substrate assay, based on that described by Dowd et al. (1994). Samples to be tested were assayed in a total volume of 1 ml 0.1 M sodium phosphate buffer, pH 6.0 containing 2.5 mM EDTA, 2 mM DTT and 10 μM Z-Phe-Arg-AMC (Z-FR-AMC) in 1.5 ml tubes. The reaction was incubated at 37°C for between 10 and 30 min and stopped by the addition of 200 μl of

10% acetic acid. Fluorescence was recorded at an excitation wavelength of 370 nm and an emission wavelength of 440 nm. The activity of the samples were calculated from a standard curve of AMC ranging from 0 to 10 μM in concentration. AMC standards were prepared from a 500μM stock in 100% DMF by dilution with 0.1M sodium phosphate buffer, pH 6.0 containing 2.5 mM EDTA, 2 mM DTT. Standards were incubated and 10% acetic acid added as for the samples. Standard curves were generated and activity determined using Microsoft Excel. Activity was calculated and presented as nmol AMC min^{-1} ml^{-1}.

For many experiments, a variation on this assay was performed using opaque white or black 96-well microtitre plates. The assay was carried out as for the larger scale version, but with a total volume of 200 μl substrate/buffer and addition of 50 μl of 10% acetic acid to stop the reaction.

2.3.32 *In vitro* autoactivation of recombinant cathepsin Ls

Processing of rFheproCL1 was carried out by incubating 5 μg of the purified recombinant enzyme in 0.1 M sodium citrate buffer, pH 5.0 containing 1 mM DTT and 2 mM EDTA at 37°C. Samples were taken at various time points up to 2 h and the proteolytic cleavage of the propeptide visualised by 12% SDS-PAGE. Samples were also assayed for enzyme activity with the standard fluorogenic substrate assay described above. Cleavage products in both purified recombinant and 2 h activated rFheproCL1 were subjected to *N*-terminal sequencing by Genosphere Biotechnologies, Paris, France.

Processing of rFheproCL2 was carried out in a similar manner, replacing the buffer with 0.1 M sodium acetate, pH 4.5. Samples were visualised on 12% SDS-PAGE and assayed for activity with the standard fluorogenic substrate assay.

For both enzymes, a control experiment was carried out using either 0.1 M sodium acetate buffer, pH 4.5 or PBS, pH 7.3 and taking samples at time points up to 24

h. For later experiments, 2h and 24 h activated recombinants were prepared by incubation at a final concentration of 100 μg/ml in 0.1 M sodium acetate buffer, pH 4.5 containing 1 mM DTT and 2 mM EDTA at 37°C for the relevant time.

2.3.33 *In vitro* processing of inactive mutants by rFheproCL1 and rFheproCL2

Exogenous processing of r[Gly^{26}]FheproCL1 and r[$Pro^{-12}Gly^{26}$]FheproCL1 by wildtype rFheproCL1 was carried out by mixing 7.5 μg of purified mutant enzyme with 0.5 μg of 2h activated wildtype enzyme. The mix was incubated in 0.1 M sodium acetate buffer, pH 5.0 containing 1 mM DTT and 2 mM EDTA at 37°C and samples taken at time points up to 3 h. Samples were analysed by 12% SDS-PAGE and cleavage products subjected to *N*-terminal sequencing by Genosphere Biotechnologies, Paris, France.

To confirm that the mutant enzymes were expressed as correctly folded proteins, a similar experiment was performed in parallel with mutant enzyme preparations that had been unfolded by heating at 95°C for 3 min before addition of wildtype enzyme.

Cross-processing was investigated by incubating all three mutants (r[Gly^{26}]FheproCL1, r[$Pro^{-12}Gly^{26}$]FheproCL1 and r[Gly^{26}]FheproCL2) with wildtype 24 h activated rFheproCL1 and rFheproCL2 enzymes in similar conditions. After incubation for 3 h at 37°C, samples were visualised by 12% SDS-PAGE.

2.3.34 Inhibition of rFheproCL1 activity by recombinant propeptide and Z-Phe-Ala-diazomethylketone

Recombinant cathepsin L1 propeptide was previously generated and purified in the laboratory as described by Roche et al. 1999. A fluorogenic substrate assay was carried out in a white 96-well microtitre plate. Purified rFheproCL1 was assayed in triplicate in the standard buffer with the addition of between 0 and 5 μM of recombinant

cathepsin L1 propeptide or between 0 and 5 μM of the cathepsin L inhibitor Z-FA-CHN_2. Plates were incubated for 30 min at 37°C and reactions stopped by the addition of 50 μl/well of 10% acetic acid. Fluorescence was recorded at an excitation wavelength of 370 nm and an emission wavelength of 440 nm. Average fluorescence values were compared directly using Microsoft Excel.

Effect of buffer pH on inhibition of protease activity was investigated by preincubating 2 h activated rFheproCL1 with 0.1M sodium citrate buffers, pH between 3.0 and 7.0, containing 2.5 mM EDTA, 2 mM DTT in addition to either 0, 1 or 0.1 μM propeptide or 0, 1.5 or 0.15 μM Z-FA-CHN_2 for 30 min at room temperature. These samples were then assayed in triplicate in 200 μl total volume of buffer containing 10 μM Z-FR-AMC. Plates were incubated for 15 min at 37°C and reactions stopped by the addition of 50 μl/well of 10% acetic acid. Fluorescence was recorded at an excitation wavelength of 370 nm and an emission wavelength of 440 nm. Percentage inhibition values were calculated from the fluorescence values directly using Microsoft Excel.

2.3.35 Activity profiles for recombinant proteins at varying pH values

Fluorogenic substrate assays were performed with buffers at a range of pHs replacing the 0.1 M sodium phosphate, pH 6.0 in the standard assay. Assays were carried out in a black 96-well microtitre plate. The three sets of buffers used were as follows: 0.1 M sodium acetate, pH 3.0, 3.5, 4.0, 4.5, 5.0, 5.5 and 6.0, 0.1 M sodium phosphate, pH 5.5, 6.0, 6.5, 7.0, 7.5 and 8.0 and 0.1 M tris/HCl, pH 7.5, 8.0, 8.5, 9.0, 9.5, 10.0 and 10.5. All buffers included 50 mM NaCl to give an even ionic strength. 24 h activated rFheproCL1 or rFheproCL2 was assayed with each buffer in triplicate. AMC standards were prepared as for the standard assay. Fluorescence was recorded at an excitation wavelength of 370 nm and an emission wavelength of 440 nm after 15 min of incubation at 37°C. Standard curves were generated and activity determined using

Microsoft Excel. Activity was calculated and presented as nmol AMC min^{-1} ml^{-1} and graphed as activity vs. buffer pH.

Samples of 24 h activated recombinants were also analysed by 10% GS-PAGE, with overnight incubations at 37°C in 0.1 M sodium acetate buffers, pH 4.0, 5.0, 6.0, 7.0 and 7.75, containing 50 mM NaCl and 1 mM DTT.

2.3.36 Analysis of mRNA produced by *P. pastoris* transformants

For each recombinant enzyme and mutant, a 6 day 2.5X scaled up methanol induction was performed, with 2 ml samples take at time points 0 and 96 h. Samples were centrifuged at 12,000 x *g* for 2 min and the supernatants transferred to new 10 ml conical tubes. Total RNA was isolated from the cell based on the protocol provided in the Multi-Copy *Pichia* Expression Kit manual, Ver. F (Invitrogen) and Schmitt et al. 1990. See section 2.2 for solution compositions. All solutions were prepared with DEPC-treated water. Cell pellets were resuspended in 400 µl of AE buffer and transferred to 1.5 ml tubes. 40 µl of 10% SDS was added and the tubes vortexed for 20 s, followed by the addition of 600 µl of phenol and a further 20 s of vortexing. The tubes were incubated at 65°C in a water bath for 4 min, followed by incubation in a dry ice/ethanol bath for 1 min. The tubes were vortexed at 12,000 x *g* for 2 min at 4°C and the upper aqueous phase transferred to new 1.5 ml tubes. 600 µl of phenol:chloroform (1:1) was added and the tubes vortexed for 20 s. After centrifugation at 12,000 x *g* for 2 min at 4°C, the upper phase was transferred to new 1.5 ml tubes and 40 µl of 3 M sodium acetate, pH 5.3, added. 750 µl of 100% ethanol at -20°C was added and the tubes centrifuged at 12,000 x *g* for 15 min at 4°C. The supernatant was removed and the pellets washed with 500 µl of 80% ethanol, followed by centrifugation at 12,000 x *g* for 10 min at 4°C. The pellets were allowed to air-dry and were resuspended with 40 µl each of DEPC-treated water. Concentration of RNA in each sample was determined.

First strand cDNA was synthesised by reverse transcriptase PCR (RT-PCR) on 1 μg of total RNA from each recombinant. Total RNA was combined with 20 pmol 16dT primer (see Table 2.1) in a 200 μl thin walled PCR tube and incubated at 70°C for 5 min, followed by 4°C for 5 min. To this was added 15 μl of a RT-PCR mix to give final concentrations of 1X ImProm-II reaction buffer, 6.4 mM $MgCl_2$, 0.5 mM dNTP mix, 1 U/μl RNasin and 0.05 U/μl of ImProm-II reverse transcriptase. The PCR programme was run at 25°C for 5 min, 42°C for 60 min and 70°C for 15 min. A standard PCR was performed on the synthesised cDNA using the primers FheCL1F and FheCL1R (see Table 2.1) and Sigma REDTaq. The standard PCR programme was used. Samples were analysed on a 1% agarose gel.

2.3.37 Analysis of intracellular expression of recombinant proteins

For each recombinant enzyme and mutant, a 6 day 2.5X scaled up methanol induction was performed, with 1 ml samples take at time points 0 and 72 h. Samples were centrifuged at 12,000 x *g* for 2 min and the supernatants transferred to new 1.5 ml tubes. The cells were lysed based on the protocol provided in the Multi-Copy *Pichia* Expression Kit manual, Ver. F (Invitrogen). See section 2.2 for solution compositions. The cells were resuspended in 100 μl of breaking buffer and an equal volume of acid-washed 0.5 mm diameter glass beads (estimated visually by displacement) was added. The mix was vortexed for 30 s, followed by incubation on ice for 30 s. This was repeated for a total of 8 cycles. The mix was centrifuged at 12,000 x *g* for 10 min at 4°C and the cell lysate transferred to new 1.5 ml tubes.

Cell lysates were analysed by 12% SDS-PAGE and immunoblotting with a 1/500 dilution of mouse monoclonal anti-His_6 IgG. The secondary antibody used a 1/1000 dilution of a goat anti-mouse IgG-peroxidase conjugate. Corresponding *P. pastoris* medium samples were analysed in parallel for comparison.

2.3.38 Substrate comparison

Fluorogenic substrate assays were performed with a range of fluorogenic substrates replacing Z-FR-AMC in the standard assay. Assays were carried out in a black 96-well microtitre plate. The substrates used were Boc-VLK-AMC, Z-FR-AMC, Z-LR-AMC, Z-R-AMC, Z-RR-AMC, Z-PR-AMC, Boc-VPR-AMC, Tos-GPK-AMC, Tos-GPR-AMC and Boc-AGPR-AMC. 24 h activated rFheproCL1 or rFheproCL2 was assayed with each substrate in triplicate. AMC standards were prepared as for the standard assay. Fluorescence was recorded at an excitation wavelength of 370 nm and an emission wavelength of 440 nm after 10, 20 and 30 min of incubation at 37°C. Standard curves were generated and activity determined using Microsoft Excel. A graph of relative fluorescence vs. substrate was generated for each enzyme.

2.3.39 Collagen digestion

Collagen Types II (bovine nasal septum) and VI (basement membrane) suspensions were prepared in PBS, pH 7.3 to a final concentration of 1 mg/ml. Each collagen (15 μg) was incubated with 0.5 μg of 24 h activated rFheproCL1 or rFheproCL2 in PBS, pH 7.3, overnight at 37°C. Negative controls of each collagen in PBS alone and positive controls with 0.5 μg collagenase blend H were similarly prepared. Samples were analysed by 12% SDS-PAGE and on semi-native gels.

Chapter 3: Sequence analysis and cloning of *F. hepatica* procathepsins L1 and L2

3.1 Sequence analysis of pre-procathepsins L1 and L2

The main focus of this thesis is the characterisation and analysis of the major *Fasciola hepatica* cysteine proteases, cathepsins L1 and L2. Before cloning of the genes encoding these proteins was carried out, the amino acid sequences of pre-procathepsins L1 and L2 (accession numbers U62288 and U62289 respectively) were aligned using ClustalX 1.81 (see Figure 3.1). The sequences were found to have 78.22% identity and 92.63% similarity to each other (see Table 3.3), indicating a very close relationship between the two enzymes and a commonality in function within the parasite.

The sequences were also aligned with other *F. hepatica* and *F. gigantica* sequences. Details of the 21 *F. hepatica* and 10 *F. gigantica* sequences used are given in Table 2.2. A bootstrap neighbour-joining phylogenetic tree was generated from the alignment using MEGA2 (Kumar et al., 2001). This tree indicated that the sequences fell into 4 distinct clades; importantly, three of these contained sequences of cDNAs isolated from adult parasites, whereas one contained sequences derived from metacercarial or newly excysted juvenile stage cDNAs (see Figure 3.2). Therefore, this analysis has identified stage-specific cathepsin L proteases. For the purposes of this thesis, these groups were labelled as 'Adult Group 1', 'Adult Group 2', 'Adult Group 3' and 'Metacercarial Group'. *F. hepatica* pre-procathepsin L1 was located in the Adult Group 1, while pre-procathepsin L2 was located in the Adult Group 3.

The position of three of the sequences in the tree (FgiCL5, FheCL14 and FheCL17) were not reliable. The FgiCL5 sequence is complete and may be an intermediate between Adult Group 1 and Adult Group 2 as it falls between these groups.

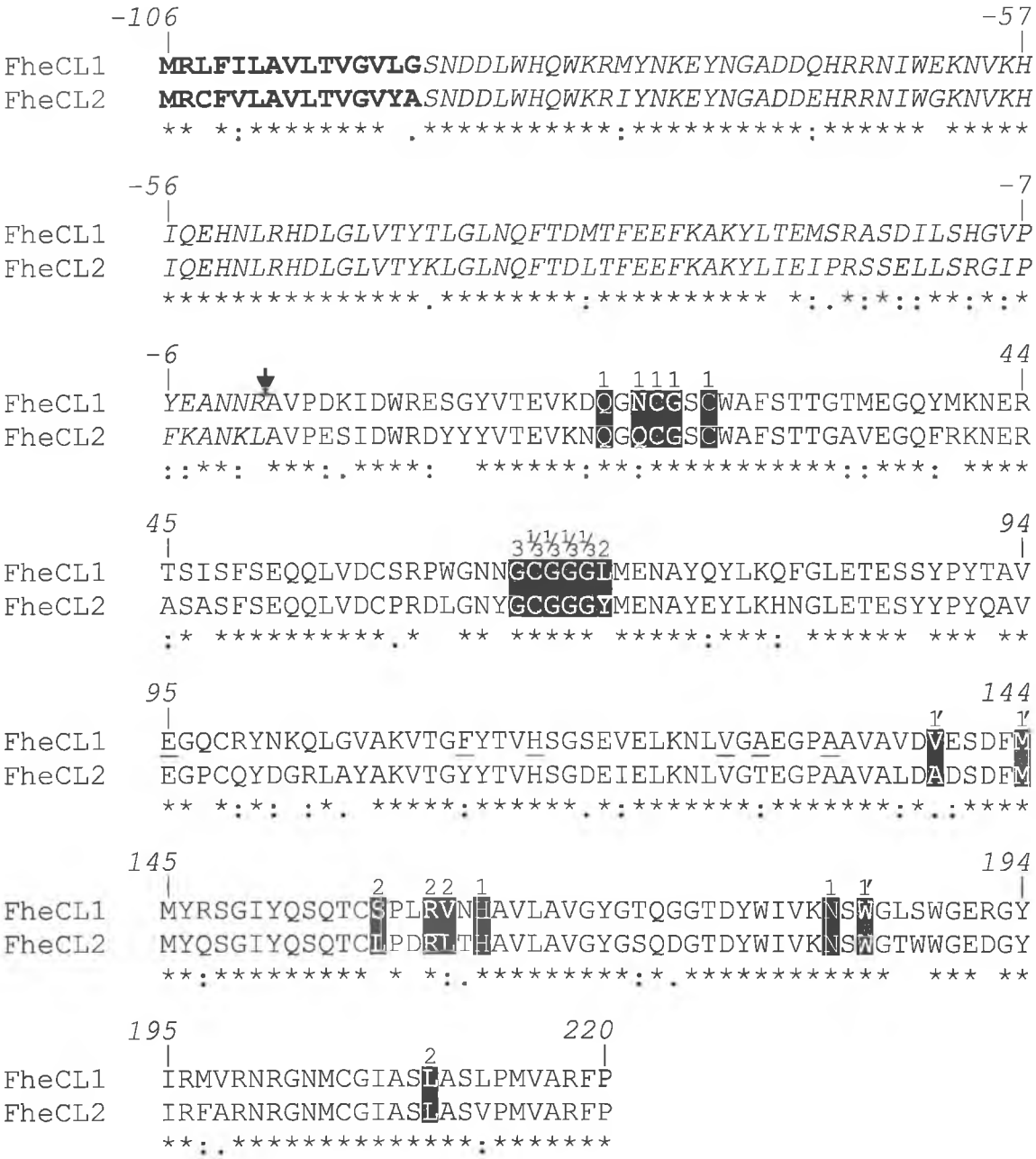


Figure 3.1: Alignment of *F. hepatica* pre-procathepsins L1 and L2 protein sequences. Alignment was performed using ClustalX 1.81. The signal sequence is indicated in bold and the propeptide is in italics. The cleavage point between the propeptide and the mature region is indicated by an arrow. The active site triad, situated in the S_1 subsite, is marked in black and residues implicated in subsite interactions, as per Turk et al. (1998), are in grey, with subsite designations listed above. FheCL1 residues which differ from the original U62288 sequence deposited in the public database are underlined.

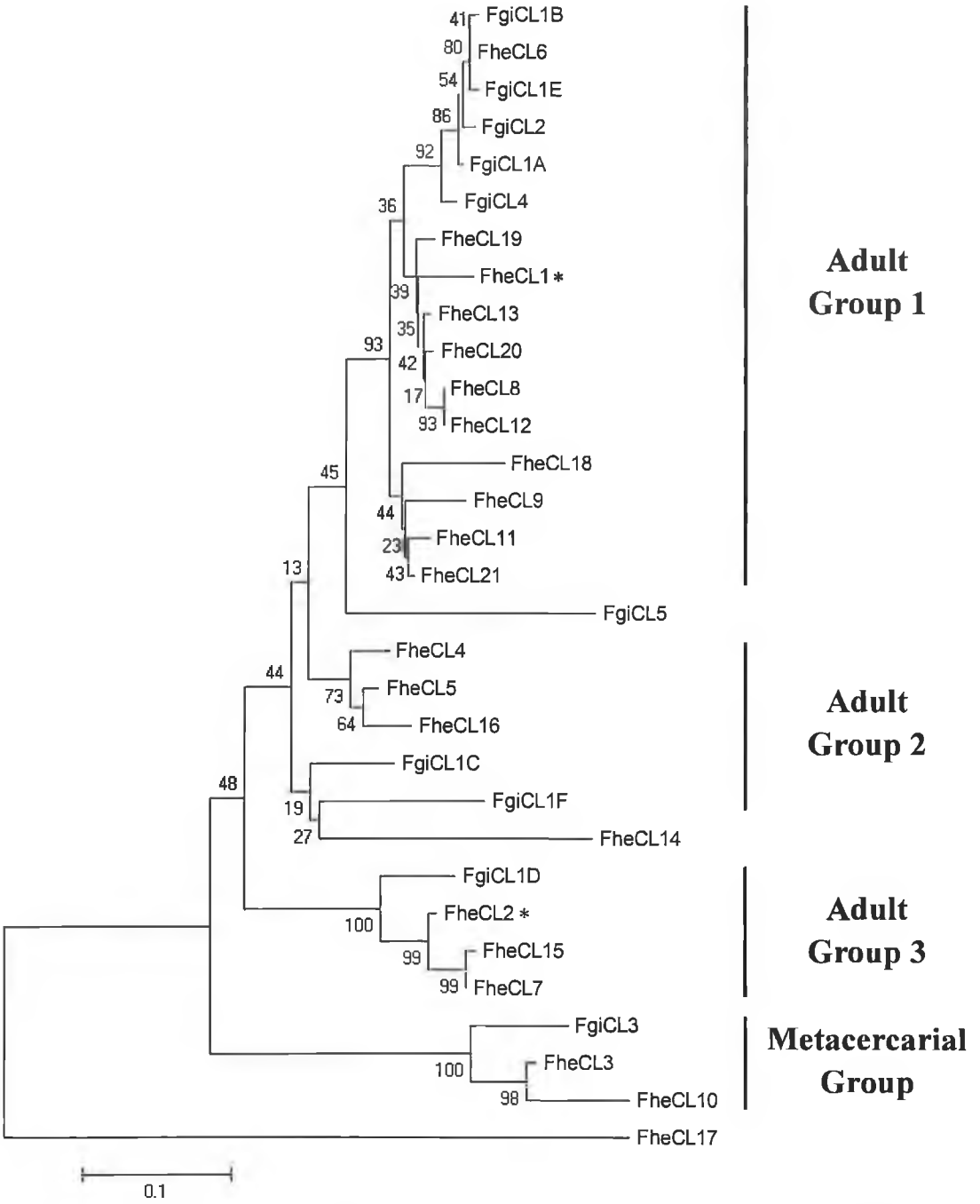


Figure 3.2: Bootstrap neighbour-joining phylogenetic tree of *Fasciola* cathepsin L sequences. The tree was generated and visualised using MEGA2. Numbers indicate bootstrap support at a given node. The major clades are indicated. Three sequences (FgiCL5, FheCL14 and FheCL17) do not fall easily into specific groups as they are either missing large sections of data or contain mixed features of different groups. The *F. hepatica* cathepsins L1 and L2 are indicated by an asterisk. Accession numbers for all sequences are given in Table 2.2.

Both FheCL14 and FheCL17 are incomplete sequences, with no data available up to position 20 of the mature protein sequence or for residues after position 185, and likely account for the position ambiguity of these two sequences.

This tree was compared with a similar one generated by Irving et al. (2003), which used fewer sequences, and was divided into similar clades. The addition of extra sequences to the tree, some of which did not include the full propeptide sequence or were otherwise incomplete, reduced the bootstrap support for many of its branches. However, the general structure of the groups was the same as that noted by Irving et al. (2003).

Analysis of the alignments showed that the region of most variation in the enzymes was located adjacent to the junction between the propeptide and the mature enzyme (residues -21 to 5; see Figure 3.1). This region is highly conserved within each group, with few variations; these are listed in Table 3.1. However, differences were found between each group; for instance, the sequence from positions -5 to -1 is Glu-Ala-Asn-Asn-Arg for Adult Group 1, Lys-Ala-Asn-Lys-Arg for Adult Group 2, Lys-Ala-Asn-Lys-Leu for Adult Group 3, and Glu-Ala-Glu-Gly-Asn for the Metacercarial Group. Any variations within each group were mainly limited to differences between *F. hepatica* and *F. gigantica* (see Table 3.1).

Table 3.1: Features of cathepsin L clades. The propeptide sequence for each group leading up to the junction with the mature region is given in full along with the start of the mature region itself. The cleavage point is indicated with a downwards arrow. Variations are shown below the main sequence, with those found only in *F. gigantica* indicated in bold italics. Residues which are completely conserved across all groups are underlined.

Group:	Pro/mature junction (-21 to 5):
Adult 1	TEMPRASDILSHGIPYEANNR↓AVPDK S V T
Adult 2	TEMPRASELLSHGIPYKANKR↓AVPDR ***HR DI*** ***E***
Adult 3	IEIPRSSELLSRGIPYKANKL↓AVPES F ***P***
Metacercarial	MEMSPVSESLSDGISYEAEGN↓DVPAS ***I*** ***E*** V K

A divergence between the three Adult Groups and the Metacercarial Group was also located at position 1 of the mature region (see Table 3.1). While all three Adult Groups contained an alanine residues at this location, the Metacercarial Group sequences were unique in possessing an aspartate residue here. Notwithstanding that only three metacercarial sequences are available, this divergence was common to both *F. hepatica* and *F. gigantica* sequences.

We also examined residues within the active site of these enzymes that are involved in substrate specificity. While peptide bonds are cleaved by Cys^{26} (with the involvement of His^{163} and Asn^{183}) which is situated in the S_1 subsite, enzyme specificity is determined by residues in the S_2 subsite. Experiments published by Dowd et al. (1997) and Smooker et al. (2000) indicated that three residues were important in the determination of substrate specificity of cathepsins L. These residues are located at positions 70, 162 and 164 (*Fasciola* numbering) of the mature protein and are particularly involved in determining the preference for residues at the S_2 positions of a substrate. When the *F. hepatica* and *F. gigantica* sequences were compared at these three residues, no variations were found within each group (see Table 3.2). However, this set of three residues did differ between groups, with a pattern of Leu^{70}-Asn^{162}-Ala^{164} for Adult Group 1, Leu^{70}-Asn^{162}-Gly^{164} for Adult Group 2, Tyr^{70}-Thr^{162}-Ala^{164} for Adult Group 3 and Trp^{70}-Thr^{162}-Ala^{164} for the Metacercarial Group. It was also noted that two of these residues, at positions 70 and 162, were conserved between Adult Groups 1 and 2, while the residues at positions 162 and 164 were conserved between Adult Group 3 and the Metacercarial Group. These observations suggest that differences in the substrate specificity exists between the cathepsin Ls present in each group.

No variations in the residues at positions 70, 162 or 164 were found within groups between the *F. hepatica* and *F. gigantica* sequences. Further analysis showed

that enzymes within each group also tended to have the same residues within their subsites, based on the residues implicated in subsite interactions by Turk et al. (1998; see Table 3.2). Variations towards the end of the area of the propeptide which dips into the active site cleft (residues -21 to -13) may correspond with changes in the subsites.

Table 3.2: Subsite variations between cathepsin L clades. Variations in amino acid residues are indicated for each position of the S_1, S_2, S_3 and $S_{1'}$ pockets. Residues which are totally conserved between all groups, such as those of the active site triad, are not included. Variations found in only a minority of sequences within a group are given in brackets, with those found only in *F. gigantica* in bold italics.

Group:	S_1:	S_2:							S_3:		$S_{1'}$:				
	22	**70**	**157**	**160**	**161**	**162**	**164**	**210**	**64**	**67**	**139**	**140**	**144**	**187**	**188**
Adult 1	N (Q)	L	S/L	R/A (S)	V/L	N	A	L	N (Y/***M***)	S (G/M)	V (A)	E	M (***T***)	L (S/T)	S (Y)
Adult 2	G/Q (***D***)	L	S	R (***F***)	L	N	G	L (***M***)	Y (***F/N***)	N	V	E (***D***)	M (***V***)	T (***P***)	W
Adult 3	Q	Y	L	R	L	T	A	L	Y (***H***)	G	A (***V***)	D	M	T (***S***)	W (***S***)
Meta.	Q	W	S/T	S/R	V	T	A	V	H	G (***S***)	A	Q	Y	K	W

These comparisons all confirmed the grouping pattern indicated by the phylogenetic tree. FheCL1 was found to fall into the Adult 1 group, while FheCL2 fell into the Adult 3 group.

The pre-procathepsin L1 and L2 amino acid sequences were also aligned with several related cathepsin Ls from other species. Details of the sequences used are given in Section 2.3.4). Comparison of the sequences based on amino acid identity and similarity (see Table 3.3) indicated the closest relationship was to the related parasitic trematode, *Schistosoma mansoni*. The next highest similarity was to the non-parasitic worm *Caenorhabditis elegans*, followed by the mammalian cathepsin Ls of mouse, rat and humans, and lastly the plant cysteine protease, papain. This indicates that the *F.*

hepatica cathepsins L1 and L2 are true cathepsin Ls and represent a monophyletic group of enzymes in the parasite.

Table 3.3: Similarity and identity of related cathepsin L sequences to *F. hepatica* CL1 and CL2. Amino acid sequences were compared using ClustalW (http://npsa-pbil.ibcp.fr). Similarity was calculated by adding the values for strongly and weakly similar residues to the identity. Accession numbers for all sequences are given in Table 2.2.

Sequence:	Species:	Length:	Identity to FheCL1:	Similarity to FheCL1:	Similarity to FheCL2:	Identity to FheCL2:
FheCL1	*Fasciola hepatica*	326	100%	100%	92.63%	78.22%
FheCL2	*Fasciola hepatica*	326	78.22%	92.63%	100%	100%
SmaCL2	*Schistosoma mansoni*	317	44.65%	76.15%	42.51%	73.09%
CelCPL1	*Caenorhabditis elegans*	337	40.47%	73.02%	40.76%	73.61%
MmuCL	*Mus musculus*	334	41.47%	71.18%	41.06%	68.91%
RnoCL	*Rattus norvegicus*	334	41.18%	70%	40.76%	68.91%
HsaCL	*Homo sapiens*	333	42.86%	71.14%	39.65%	69.68%
CpaPap	*Carica papaya*	345	32.77%	61.9%	32.77%	60.78%

The nucleotide sequences of the genes encoding *F. hepatica* pre-procathepsins L1 and L2 were also compared to investigate if the same primers could be used to amplify both sequences. The 3'-ends of both sequences were found to match exactly, while the sequence encoding the start of the propeptide had two mismatches between FheCL1 and FheCL2. However, these differences were silent and did not reflect any difference in amino acid sequence, therefore the same primers were used for amplification of both cDNAs.

3.2 Construction of pPIC9K.FheproCL1 expression vector

Procathepsin L1 encoding cDNA was successfully isolated from the pAAH5.FheCL1 vector (Roche et al., 1997) containing the full pre-procathepsin

sequence by PCR. With the restriction sites, His_6-tag, spacer and restriction enzyme clamp sequences added by the PCR primers, the total length of the product was expected to be 980 bp. This was confirmed when run on a 1% agarose gel, with the band clearly appearing just below the 1 kb band of the DNA ladder (see Figure 3.3, lane 1). After purification of the PCR product from the gel, it was double digested with SnaBI and AvrII and ligated with similarly treated pPIC9K vector. The ligation mix was then used to transform *E. coli* Top 10F' cells.

Purified plasmid DNA from *E. coli* transformants was analysed by (a) PCR with the α-factor and 3'AOX primers (see Figure 3.3, lanes 2 and 3), as well as by (b) restriction digestion with PstI (see Figure 3.3, lanes 4 and 5). For the PCR with the α-factor and 3'AOX vector primers, a product of just over 1.1 kb was generated for positive colonies but not for negatives or the pPIC9K vector control. From analysis of the vector and insert sequences, PstI was expected to cut the vector in four locations but not to cut the insert. When separated on an agarose gel, the uppermost of the four bands (containing the insert) was retarded in the gel for positive transformants compared to pPIC9K control DNA.

Plasmid DNA from two colonies which were found to be positive for the FheproCL1 insert were sequenced and one transformant selected for transformation into *P. pastoris*. The sequence confirmed that the FheproCL1 cDNA was in frame with the yeast α-factor secretion signal sequence.

However, when the sequence was analysed, several differences were found compared to the original sequence deposited by our laboratory in the public database (GenBank sequence, U62288; Roche et al., 1997). While a number of these changes were silent with respect to the database sequence, some were reflected in the amino acid sequence of the protein: $Gly^{95} \rightarrow Glu^{95}$, $Tyr^{112} \rightarrow Phe^{112}$, $Gln^{116} \rightarrow His^{116}$, $Ile^{127} \rightarrow Val^{127}$,

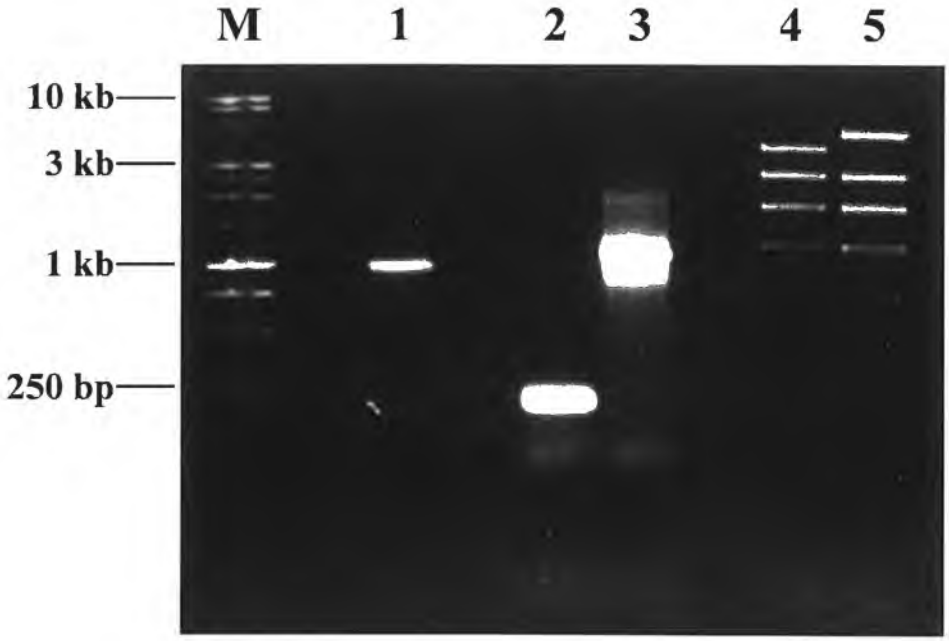


Figure 3.3: 1% agarose gel showing construction of pPIC9K.FheproCL1 expression vector. *Lane M*, 1 kb DNA ladder; *lane 1*, FheproCL1 PCR product insert before ligation into pPIC9K; *lanes 2 and 3*, pPIC9K and pPIC9K.FheproCL1 vectors amplified by PCR with α-factor and 3'AOX primers; *lanes 4 and 5*, pPIC9K and pPIC9K.FheproCL1 vectors restriction digested with PstI.

$Ser^{129} \rightarrow Ala^{129}$, $Ser^{133} \rightarrow Ala^{133}$. These changes in the sequence relative to the database are underlined in Figure 3.1.

On comparing the new sequence with others from *F. hepatica* and *F. gigantica* by alignment using ClustalX 1.81, it was found that the changes matched more closely with the related enzymes than the original sequence. It was therefore assumed that errors had been made in the original sequencing which have now been corrected by the better sequencing technologies available.

3.3 Transformation and screening of *P. pastoris* GS115 with pPIC9K.FheproCL1 construct

P. pastoris GS115 spheroplasts were prepared and transformed with the pPIC9k.FheproCL1 expression vector construct restriction-digested with SalI. Transformants were initially screened by growth on histidine-deficient plates; therefore only those showing the His^+ phenotype, and having the construct integrated into their chromosomal DNA, showed growth. His^+ transformants were grown on YPD plates containing varying concentrations of geneticin to isolate those clones with higher copy numbers of the construct integrated. Colonies were selected from the plates with the highest concentration of geneticin, and therefore the highest copy number, as these would potentially give higher levels of recombinant protein expression when induced with methanol.

One hundred of these high copy His^+ colonies were replica plated onto minimal dextrose and minimal methanol plates to determine their Mut phenotype. Colonies which grew well on both media were determined to be Mut^+, while those which grew poorly on minimal methanol medium were determined to be Mut^S phenotype. Greater numbers of Mut^S phenotype colonies than Mut^+ phenotype colonies were found. This was unexpected as linearization of the plasmid with SalI before transformation favours

integration within the *His4* gene producing Mut^+ phenotype. It was decided that three His^+ Mut^S colonies (labelled A14, A22 and A38) would be selected for further expression studies.

3.4 Construction of pPIC9K.FheproCL2 expression vector

As with FheproCL1, the procathepsin L2 encoding cDNA was successfully cloned from the pAAH5.FheCL2 vector (Dowd et al., 1997) containing the full pre-procathepsin sequence by PCR. When run on an agarose gel, the expected 980 bp band was visualised (see Figure 3.4A). The purified, SnaBI/AvrII digested insert was ligated into pPIC9K vector and transformed into *E. coli* Top 10F'.

Purified plasmid DNA from *E. coli* transformants was screened by restriction digestion with the SnaBI/AvrII combination (see Figure 3.4B) and with PstI (see Figure 3.4C). As with pPIC9K.FheproCL1, for the SnaBI/AvrII combination a 980 bp fragment was generated for positive colonies but not for negatives. Similarly, PstI was expected to cut the vector in four locations but not to cut the insert, with the uppermost of the four bands (containing the insert) running higher in an agarose gel for positive transformants compared to pPIC9K control DNA. Following this analysis, plasmid DNA from two colonies (labelled A3 and B1) which were found to be positive for an insert in both digests (see Figure 3.4) were selected for transformation into *P. pastoris* and sent for sequencing. The sequences were confirmed to match the database sequence and to be in frame with the yeast α-factor secretion signal sequence.

3.5 Transformation and screening of *P. pastoris* GS115 with pPIC9K.FheproCL2 construct

The two pPIC9K.FheproCL2 constructs were linearised with SalI and transformed into competent *P. pastoris* GS115 cells. His^+ phenotype was confirmed by

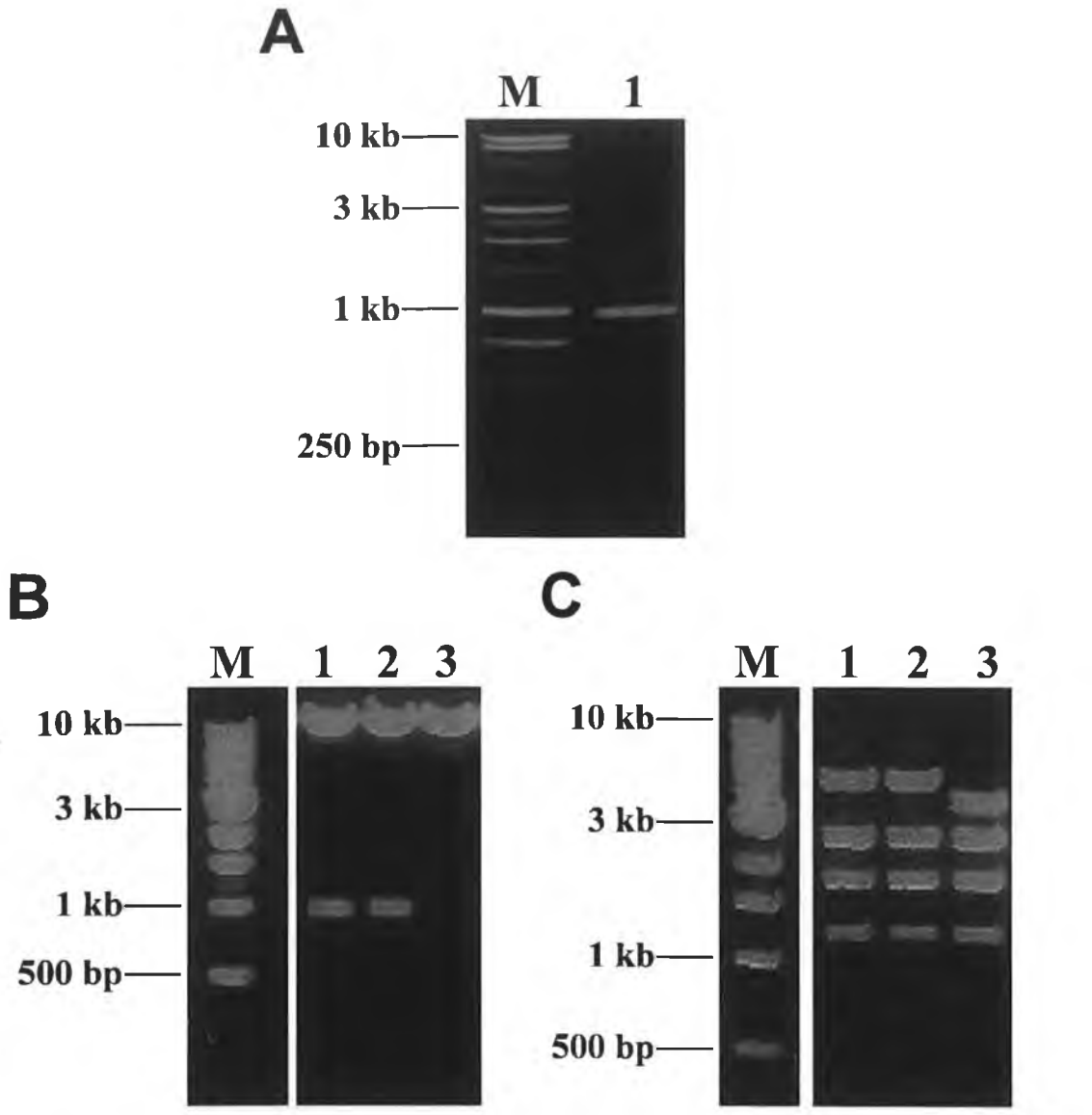


Figure 3.4: 1% agarose gels showing construction of pPIC9K.FheproCL2 expression vector. *A*, procathepsin L2 PCR product before ligation into pPIC9K. *Lane M*, 1 kb DNA ladder; *lane 1*, FheproCL2 PCR product. *B*, purified plasmid from selected *E. coli* transformants restriction digested with SnaBI and AvrII. *Lane M*, 1 kb DNA ladder; *lanes 1 and 2*, transformants A3 and B1; *lane 3*, pPIC9K vector control. *C*, purified plasmid from selected *E. coli* transformants restriction digested with PstI. *Lane M*, 1 kb DNA ladder; *lanes 1 and 2*, transformants A3 and B1; *lane 3*, pPIC9K vector control.

growth on histidine deficient medium and fifty colonies from each transformation were screened directly for expression by the colony blot rapid screening method. Two colonies (labelled A3.1 and B1.28) which were found to be strongly labelled with anti-His_6 antibody and therefore showed potentially good expression levels were selected for further expression studies (see Figure 3.5).

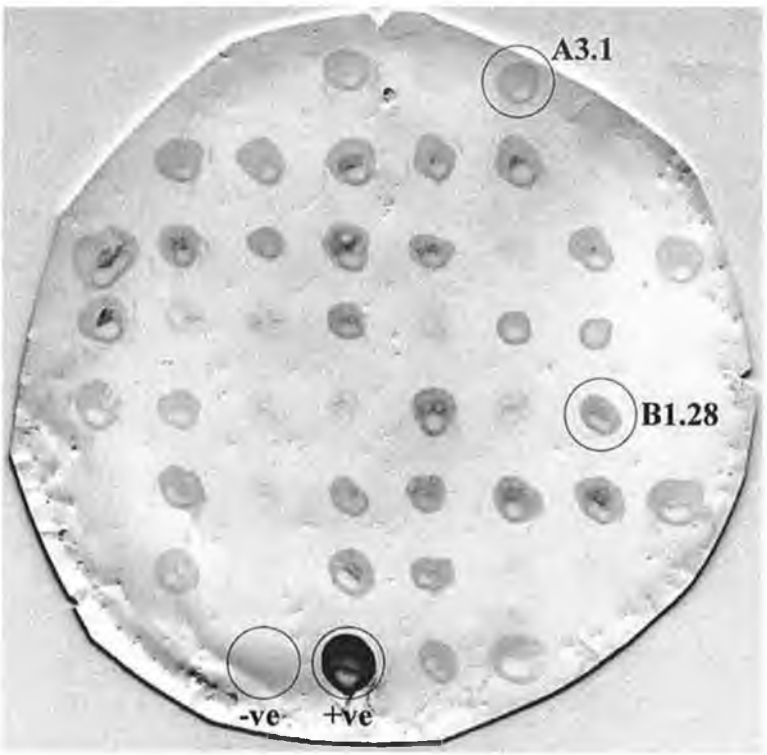


Figure 3.5: Colony blot screening of *P. pastoris* pPIC9K.FheproCL2 transformants for expression. Colonies were transferred onto nitrocellulose membrane and grown on YNB/2% methanol plates. Bound protein after cell lysis was probed with mouse monoclonal anti-His_6 antibody. Positive colonies show as a dark patch. The positive (transformed with pPIC9K.FheproCL1) and negative (transformed with pPIC9K) control colonies are marked with a ring, as are the two colonies chosen for further expression studies.

Chapter 4: Expression and characterisation of wildtype procathepsin L1

4.1 Expression of procathepsin L1

Pilot expression experiments were carried out with the three His^+ Mut^S colonies of *P. pastoris*, labelled A14, A22 and A38, which were transformed with the pPIC9K.FheproCL1 expression vector construct. These were induced for 6 days by 1% methanol in BMMY medium buffered to pH 6.0. Samples of the culture were taken daily and the supernatants were analysed by 12% SDS-PAGE. Cultures of *P. pastoris* transformed with pPIC9K alone and with the human serum albumin construct (provided by Invitrogen in the Multi-Copy *Pichia* Expression Kit) were induced as negative and positive controls respectively. All cultures showed high levels of secreted recombinant protein in two bands at 30 kDa and 37 kDa (see Figure 4.1). These bands both ran higher than native mature cathepsin L1 or L2 (see Figure 4.3) but the 37 kDa band was close to the size expected for unprocessed procathepsin L1. As expression levels were approximately equal for each clone, one (A14) was selected and grown on YPD agar plates for stocks. To check for consistency of expression, three colonies from these plates, labelled A14A, A14B and A14C, were again induced for 6 days with 1% methanol in BMMY medium buffered to pH 6.0. Samples taken daily were analysed by 12% SDS-PAGE and showed consistent levels of protein expression. Clone A14A was selected arbitrarily as the final expression clone and glycerol stocks prepared for future inoculations.

The effect of various environmental conditions on rFheproCL1 production was examined. Firstly, three identical cultures were induced with 1% methanol in BMMY medium buffered to pH 6.0, 7.0 and 8.0 respectively. Samples of culture supernatant taken after 48 and 72 h of induction were analysed by 12% SDS-PAGE (see Figure 4.2). While all three cultures produced roughly equal amounts of protein at all three pHs, the

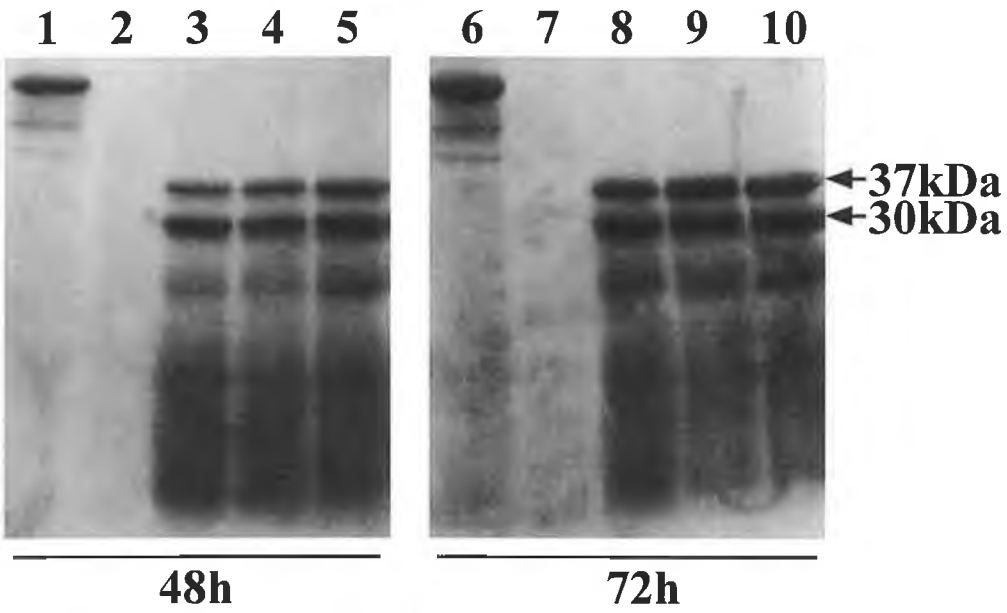


Figure 4.1: SDS-PAGE analysis of *F. hepatica* procathepsin L1 expression from *P. pastoris*. Cultures of *P. pastoris* transformed with a cDNA encoding procathepsin L1 were induced with 1% methanol in BMMY medium buffered to pH 6.0 at 30°C for 6 days. *P. pastoris* transformed with human serum albumin (HSA) and pPIC9K alone were induced as positive and negative controls respectively. Aliquots (10 µl) removed at 48 and 72 h after induction were analysed by 12% SDS-PAGE. The 37 and 30 kDa components are indicated. *Lanes 1 and 6*, HSA secretion control; *lanes 2 and 7*, pPIC9K background control; *lanes 3 and 8*, clone A14; *lanes 4 and 9*, clone A22; *lanes 5 and 10*, clone A38.

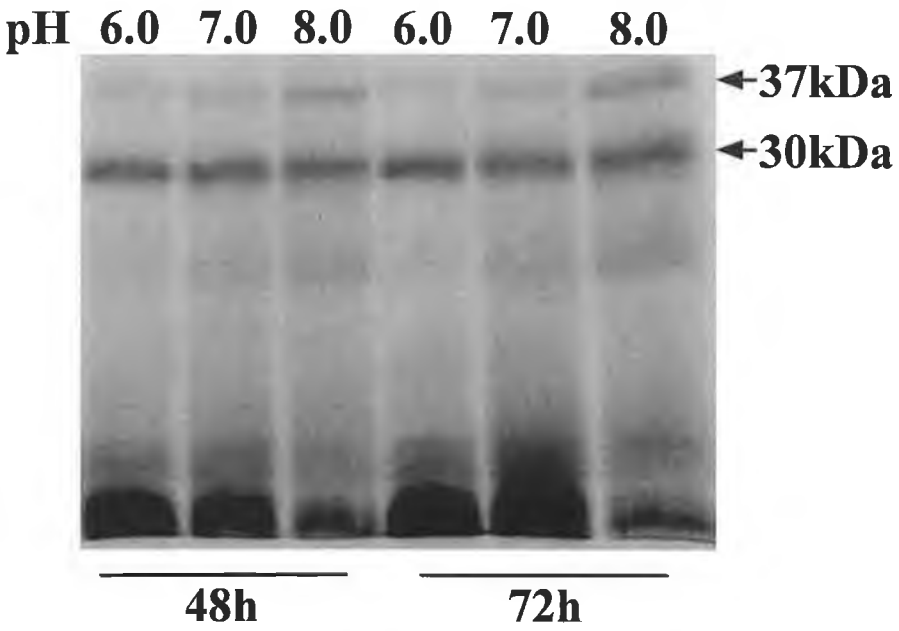


Figure 4.2: SDS-PAGE analysis of *F. hepatica* procathepsin L1 expression from *P. pastoris*. Cultures of *P. pastoris* transformed with a cDNA encoding procathepsin L1 were induced with 1% methanol. Fermentations were carried out at 30°C in media buffered to pH 6.0, 7.0, and 8.0. Aliquots (10 µl) removed at 48 and 72 h after induction were analyzed by 12% SDS-PAGE. The 37 and 30 kDa components are indicated. Adapted from Collins et al. (2004).

proportion of recombinant protein produced in the 37 kDa form increased with increasing pH.

4.2 Immunoblot analysis of expression of procathepsin L1

Samples of the supernatant taken after 72 h of induction from the pH 8.0 buffered culture were analysed by 12% SDS-PAGE and by immunoblotting alongside ES products prepared from adult *F. hepatica* (see Figure 4.3). The blots were probed with antisera prepared against native mature cathepsin L1 and against recombinant propeptide of cathepsin L1 (see Section 2.3.2). Both the 37 and 30 kDa bands were found to be immunoreactive with sera prepared against both native mature enzyme and recombinant propeptide. This suggested that the identity of the 37 kDa band as unprocessed procathepsin L1 and indicated that the 30 kDa represented a semi-processed form of the enzyme rather than fully mature cathepsin L1.

4.3 Effect of the cysteine protease inhibitor Z-Phe-Ala-diazomethylketone on procathepsin L1 production

To determine if the processing of the 37 kDa band to the 30 kDa band was occurring in culture due to the activity of CL1 or yeast proteases, cultures were induced with 1% methanol in BMMY medium buffered to pH 8.0 in the presence or absence of 25 μM of the potent cysteine protease inhibitor, Z-FA-CHN_2. Samples were again analysed by 12% SDS-PAGE after 48 and 72 h of induction (see Figure 4.4A). As with increasing pH, a greater proportion of the 37 kDa band was produced in the presence of the inhibitor. Cathepsin L activity of the culture supernatants taken at various time points up to 72 h after induction was analysed with the fluorogenic substrate Z-FR-AMC (see Figure 4.4B). This indicated that cathepsin L activity was greatly reduced in

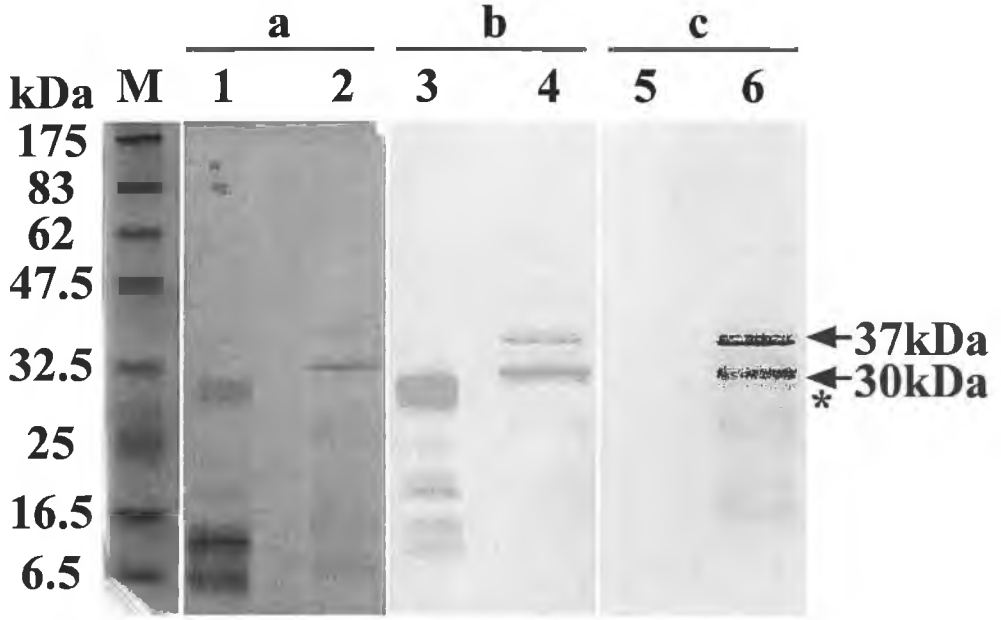


Figure 4.3: SDS-PAGE and immunoblot analysis of *F. hepatica* procathepsin L1 expression from *P. pastoris*. Comparison of *P. pastoris*-expressed procathepsin L with native *F. hepatica*-secreted mature cathepsin L1. Aliquots of adult *F. hepatica* ES products (*lanes 1, 3, and 5*) were compared with samples taken from *P. pastoris* medium after induction for 72 h at pH 8.0 (*lanes 2, 4, and 6*). *Lane M*, molecular size markers; *panel a*, Coomassie Blue-stained 12% SDS-PAGE gel; *panel b*, samples probed with antiserum prepared in rabbits against mature portion of cathepsin L1; *panel c*, samples probed with antiserum prepared in rabbits against recombinant propeptide portion of procathepsin L1. Note, the anti-propeptide serum reacts with the 37 and 30 kDa *P. pastoris*-expressed components (indicated with arrows) but not with parasite-produced mature cathepsin L1 or cathepsin L2 (indicated by an asterisk). Control preimmunized rabbit serum was not reactive with any proteins (not shown). Adapted from Collins et al. (2004).

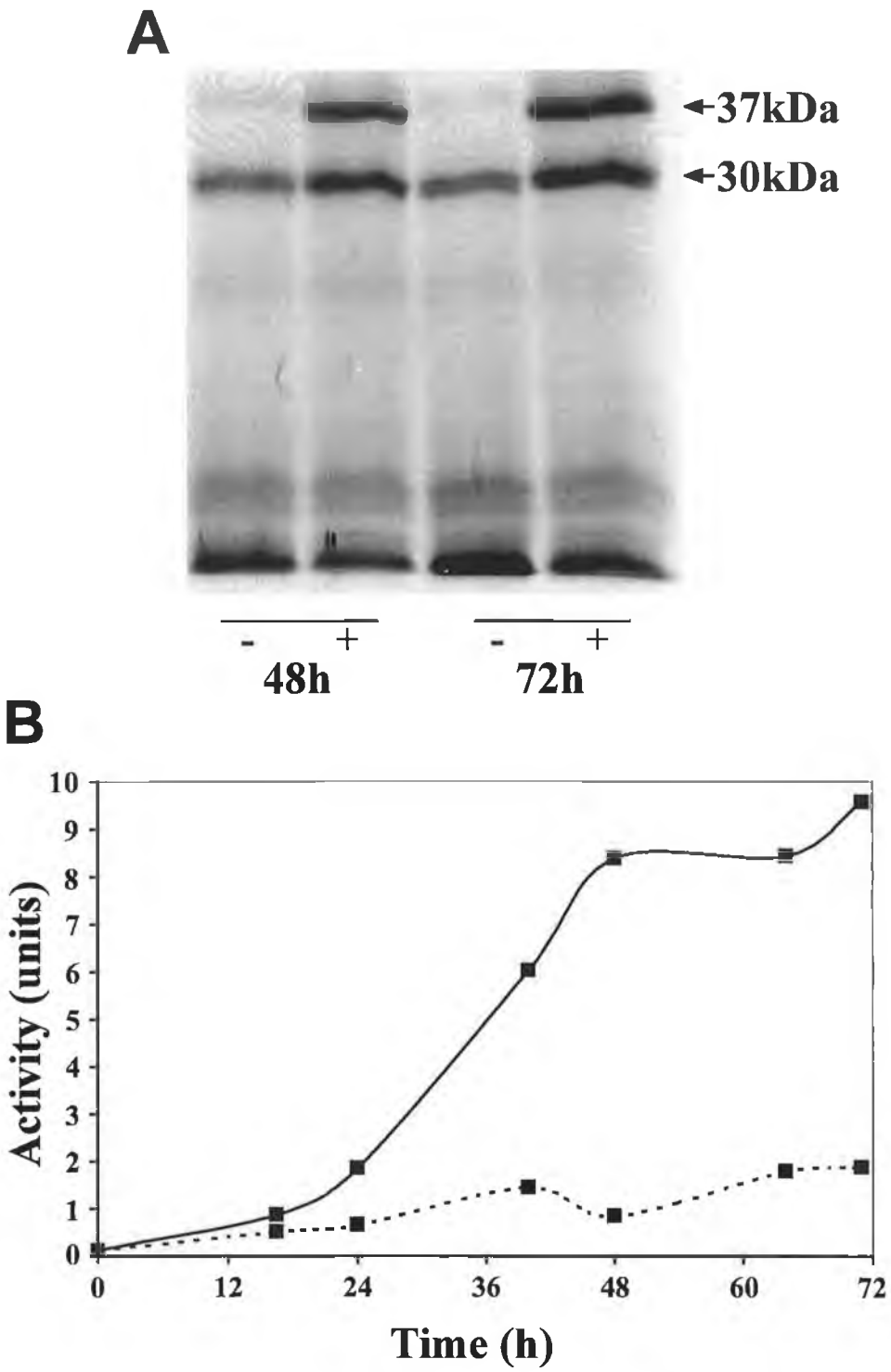


Figure 4.4: Production of procathepsin L1 by transformed *P. pastoris* in the presence and absence of Z-Phe-Ala-diazomethylketone. *A*, transformed *P. pastoris* were cultured at 30°C and pH 8.0 in the presence (+) or absence (-) of the cysteine protease inhibitor Z-FA-CHN$_2$ (25 µM). Aliquots of culture supernatant, removed after 48 and 72 h, were analyzed by SDS-PAGE. The 37 and 30 kDa bands are indicated. *B*, recombinant yeast cultured in the presence (*dotted line*) and absence (*solid line*) of 25 µM Z-FA-CHN$_2$. Cathepsin L activity in the culture supernatant was measured with the fluorogenic substrate Z-FR-AMC. Activity units are presented as nmol of AMC released min^{-1} ml^{-1}. Adapted from Collins et al. (2004).

the culture containing the inhibitor, while protein production remained largely unaffected.

4.4 Effect of pH and temperature on procathepsin L1 production

Production of protein at low pH was investigated by the induction of cultures by 1% methanol in BMMY medium buffered at pH 4.5, 6.0 and 8.0. Samples were compared by 12% SDS-PAGE (see Figure 4.5A). Reduction in pH favoured an almost complete elimination of the 37 kDa band, with an increased production of a lower band, possibly representing the mature enzyme, with decreasing medium pH.

Effect of culture temperature was investigated by induction of cultures by 1% methanol in BMMY medium buffered to pH 8.0 in the presence and absence of Z-FA-CHN_2, with growth at 20 and 25°C instead of the normal 30°C. Samples were analysed by 12% SDS-PAGE (see Figure 4.5B). Lowering of induction temperature reduced the production of breakdown products, particularly in the presence of inhibitor. Reduction in temperature also affected the appearance and clarity of the culture supernatant. While cultures induced at 30°C were a yellow-brown colour and quite turbid, cultures induced at 25 or 20°C were clearer and more yellow in colour.

4.5 *N*-terminal sequencing of procathepsin L1 components

The 37 and 30 kDa components were purified from yeast culture supernatant by Ni-NTA agarose affinity chromatography, transferred to PVDF membrane and subjected to *N*-terminal sequencing (see Section 2.3.29). For the 37 kDa band, a sequence of Glu-Ala-Glu-Ala-Tyr was determined, corresponding to 5 amino acids positioned one residue from the end of the yeast α-factor secretory signal sequence. This indicated that the 37 kDa band represented recombinant procathepsin L1 plus the last 6 amino acids (Glu-Ala-Glu-Ala-Tyr-Val) of the yeast α-factor secretory signal

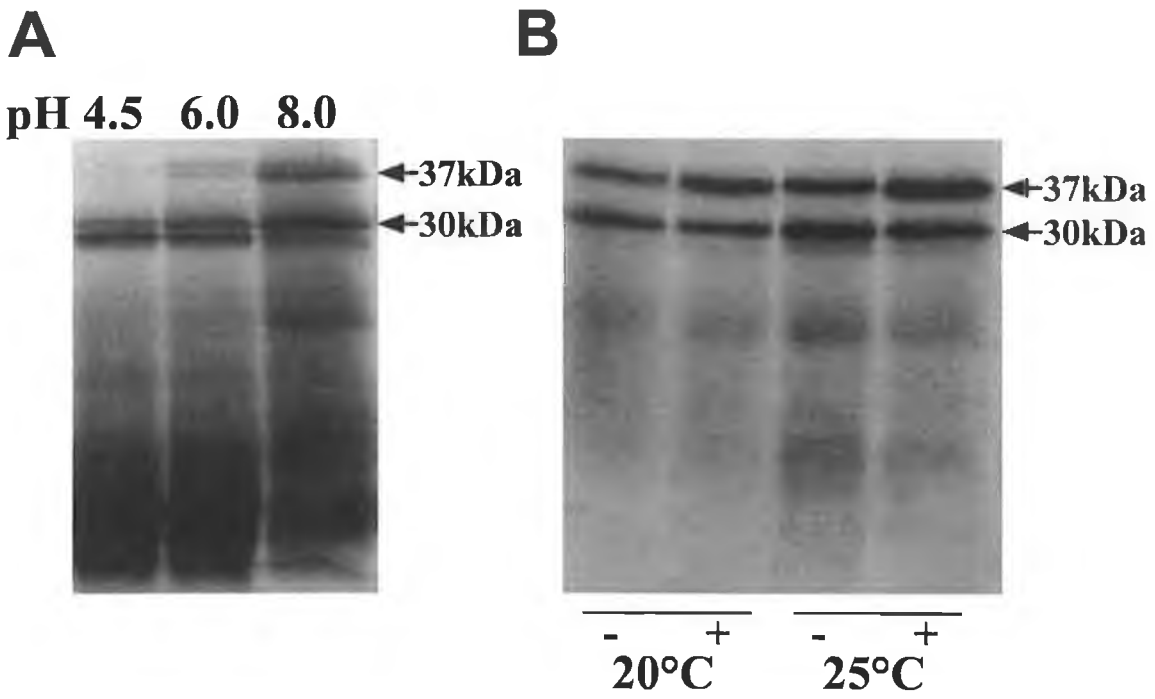


Figure 4.5: Effect of medium pH and temperature on procathepsin L1 production. *A*, transformed *P. pastoris* were cultured at 30°C in media buffered to pH 4.5, 6.0, and 8.0. Aliquots of culture supernatant, removed after 72 h, were analysed by 12% SDS-PAGE. The 37 and 30 kDa bands are indicated. *B*, transformed *P. pastoris* were cultured in media buffered to pH 8.0 at 20 and 25°C in the presence (+) or absence (-) of the cysteine protease inhibitor Z-FA-CHN_2 (25 μM). Aliquots of culture supernatant, removed after 24 h, were analysed by 12% SDS-PAGE. The 37 and 30 kDa bands are again indicated.

sequence before the Ser^{-91} start of the propeptide (see Figure 4.6). For the 30 kDa band, a sequence of Leu-Asn-Gln-Phe-Thr was determined, indicating that this band represented an intermediate processed form of cathepsin L1 beginning at residue Leu^{-38} (see Figure 4.8; Collins et al., 2004). Also, based on these two *N*-terminal sequences, the predicted average molecular weights for the 37 kDa and 30 kDa forms were 36.74 and 29.62 kDa respectively, matching closely to the original estimates.

4.6 *In vitro* autoactivation of recombinant procathepsin L1

Processing of the rFheproCL1 from the 37 kDa band to the 30 kDa band was further investigated by incubating purified rFheproCL1 in 0.1 M sodium citrate buffer, pH 5.0, containing 1 mM DTT and 2 mM EDTA at 37°C. Samples were taken at various time points up to 2 h and analysed by 12% SDS-PAGE (see Figure 4.6). Over the course of the incubation, the 37 kDa procathepsin L was processed to the 30 kDa intermediate form and then through various intermediates to a single protein migrating at 24.5 kDa. The samples were also assayed for cathepsin L activity with the fluorogenic substrate Z-FR-AMC and showed a gradual 5-fold increase in the activity of rFheproCL1 (Collins et al., 2004). The 24.5 kDa product was subjected to *N*-terminal sequencing. This revealed a mixture of two sequences, Asn-Arg-Ala-Val-Pro and Arg-Ala-Val-Pro-Asp, corresponding to the sequence just 2 and 1 amino acids from the native mature protein processing site respectively (see Figure 4.6).

A control experiment compared processing of purified rFheproCL1 when incubated in PBS, pH 7.3 with the previous low pH conditions. The incubation was carried out for 24 h, with samples taken at various time points and analysed by 12% SDS-PAGE (see Figure 4.7). While the samples at low pH processed as before, the samples in PBS remained unprocessed, with only some breakdown of the lower

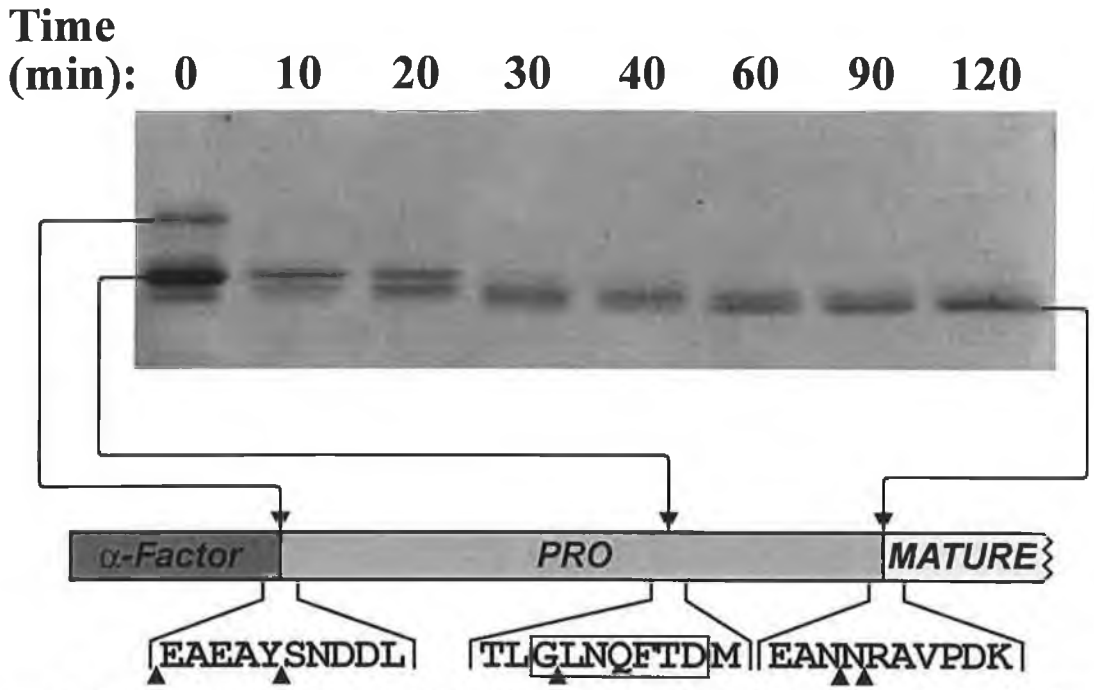


Figure 4.6: Autocatalytic processing of procathepsin L1 at pH 5.0. Purified procathepsin L1 (5.0 µg) was incubated at 37°C for 2 h in 0.1 M sodium citrate buffer, pH 5.0, containing 2 mM DTT and 2.5 mM EDTA. Samples removed at various time points were analyzed by SDS-PAGE. The predominant proteins of 37, 30, and 24.5 kDa were subjected to *N*-terminal sequencing. The schematic indicates the amino acid sequences surrounding the intermolecular processing sites (*arrowed*) that generated these components and their location within the propeptide of procathepsin L1. The box indicates the conserved motif G*X*N*X*F*X*D. Adapted from Collins et al. (2004).

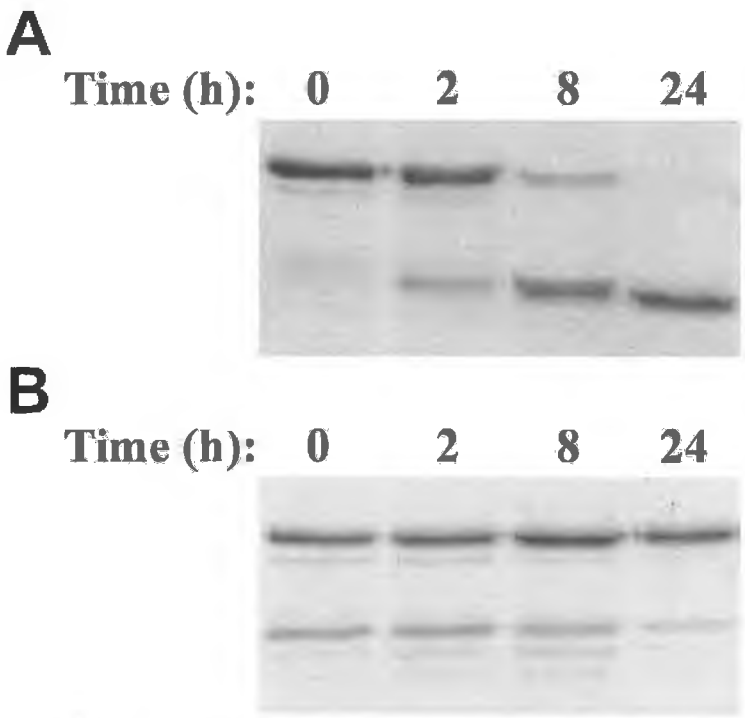


Figure 4.7: Control incubation of procathepsin L1 in PBS. Purified procathepsin L1 was incubated at 37°C for 24 h in either 0.1 M sodium acetate buffer, pH 4.5 (*A*) or PBS, pH 7.3 (*B*), both containing 2 mM DTT and 2.5 mM EDTA. Samples removed at various time points were analysed by SDS-PAGE. Processing of the 30 kDa band did not occur at the higher pH, although the lower intermediates were broken down.

intermediates and mature protein. This also demonstrates the stability of the zymogen form of the enzyme. Extension of the incubation time to 24 h also allowed complete processing of the multiform rFheproCL1 to the mature form in the low pH buffer.

4.7 Alignment of cathepsin L propeptides: identification of G*X*N*X*F*X*D motif

The cleavage point between Gly^{-39} and Leu^{-38} which gave rise to the 30 kDa intermediate form corresponded closely to the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} (papain numbering) motif which was previously found to be pivotal in the intermolecular processing of papain expressed in the yeast *S. cerevisiae* (Vernet et al., 1995). Therefore, an amino acid sequence alignment was performed with ClustalX 1.81 of the region around this motif in selected plant, helminth parasite and mammal cathepsin L sequences (see Figure 4.8, also Table 2.2 for sequence details). The G*X*N*X*F*X*D motif was found to be completely conserved between papain and all the helminth parasite sequences, with the *F. hepatica* CL1 sequence of Gly-Leu-Asn-Gln-Phe-Thr-Asp conserved completely with the *F. hepatica* CL2, and *F. gigantica* CL1 sequences. Furthermore, on examining the alignment prepared for the generation of the phylogenetic tree of *Fasciola spp.* Cathepsin Ls (see Figure 3.2), this sequence was also completely conserved in all but 2 of the Adult 1 group *F. hepatica* and *F. gigantica* sequences, with FheCL6 and FgiCL4 possessing a leucine in place of the phenylalanine. It was also noted that the Gly^{-42}-Leu^{-41} (papain numbering) bond was preserved in all helminth parasite sequences checked, as well as in papain and *C. elegans* CPL1. This bond was not conserved in the mammalian sequences checked, with Glu^{-42}-Met^{-41} or Ala^{-42}-Met^{-41} (papain numbering) replacing it.

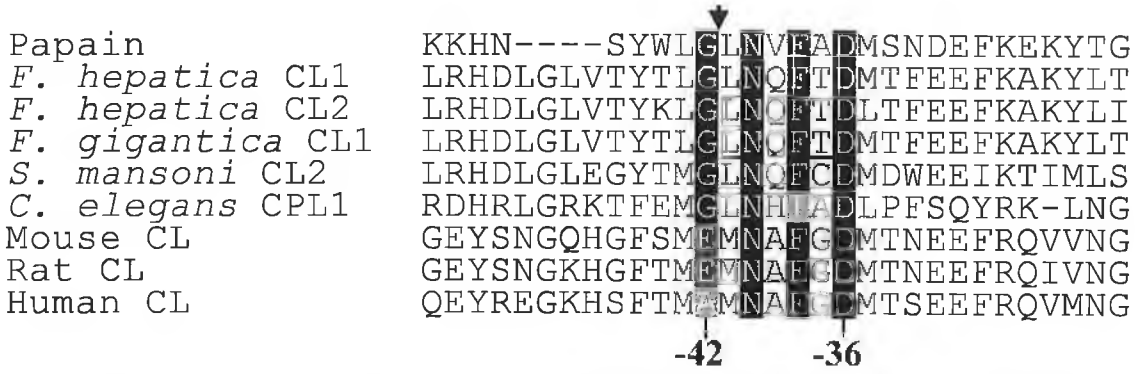


Figure 4.8: Comparison of the G*X*N*X*F*X*D motif between papain and cathepsin L proteases of helminths and mammals. The G*X*N*X*F*X*D motif sequence of papain is identical to that of *F. hepatica* cathepsin L1 and cathepsin L2 and to the cathepsin Ls of related the trematodes *F. gigantica* and *S. mansoni* (identical residues highlighted in black). The arrow indicates the position of the intermolecular cleavage site, between the Gly^{-42}-Leu^{-41} bond (papain numbering), that generates the intermediate 30 kDa component of the *F. hepatica* procathepsin L when it is expressed in *P. pastoris*. The Gly^{-42} is substituted for Glu in the rat and mouse cathepsin L sequences, and for Ala in the human cathepsin L sequence (highlighted in gray, papain numbering). Accession numbers for all sequences are given in Table 2.2. Adapted from Collins et al. (2004).

4.8 Inhibition of rFheproCL1 activity by recombinant propeptide and Z-Phe-Ala-diazomethylketone

Purified rFheproCL1 was mixed with a range of concentrations between 0 and 5 μM of recombinant cathepsin L1 propeptide or Z-FA-CHN_2 and assayed for activity against the fluorogenic substrate Z-FR-AMC. Inhibition titre curves of fluorescence versus inhibitor concentration were prepared and compared (see Figure 4.9A). The recombinant propeptide was found to be a potent inhibitor of rFheproCL1, with greater inhibition at a lower concentration than that achieved with equal concentrations of Z-FA-CHN_2.

Effect of pH on inhibitor binding was investigated by preincubating 2 h activated rFheproCL1 with either 0, 1 or 0.1 μM recombinant propeptide or 0, 1.5 or 0.15 μM Z-FA-CHN_2 in 0.1M sodium citrate buffers, pH between 3.0 and 7.0, containing 2.5 mM EDTA, 2 mM DTT for 30 min at room temperature. These samples were then assayed as before and percentage inhibition values calculated (see Figure 4.9B). At the higher concentrations, inhibition of greater than 95% was achieved at all pH values for both inhibitors. At the lower concentrations, however, inhibition was significantly lower. This was particularly noticeable for the recombinant propeptide, which showed a large drop in inhibition for pH 6.0 and 8.0 buffers.

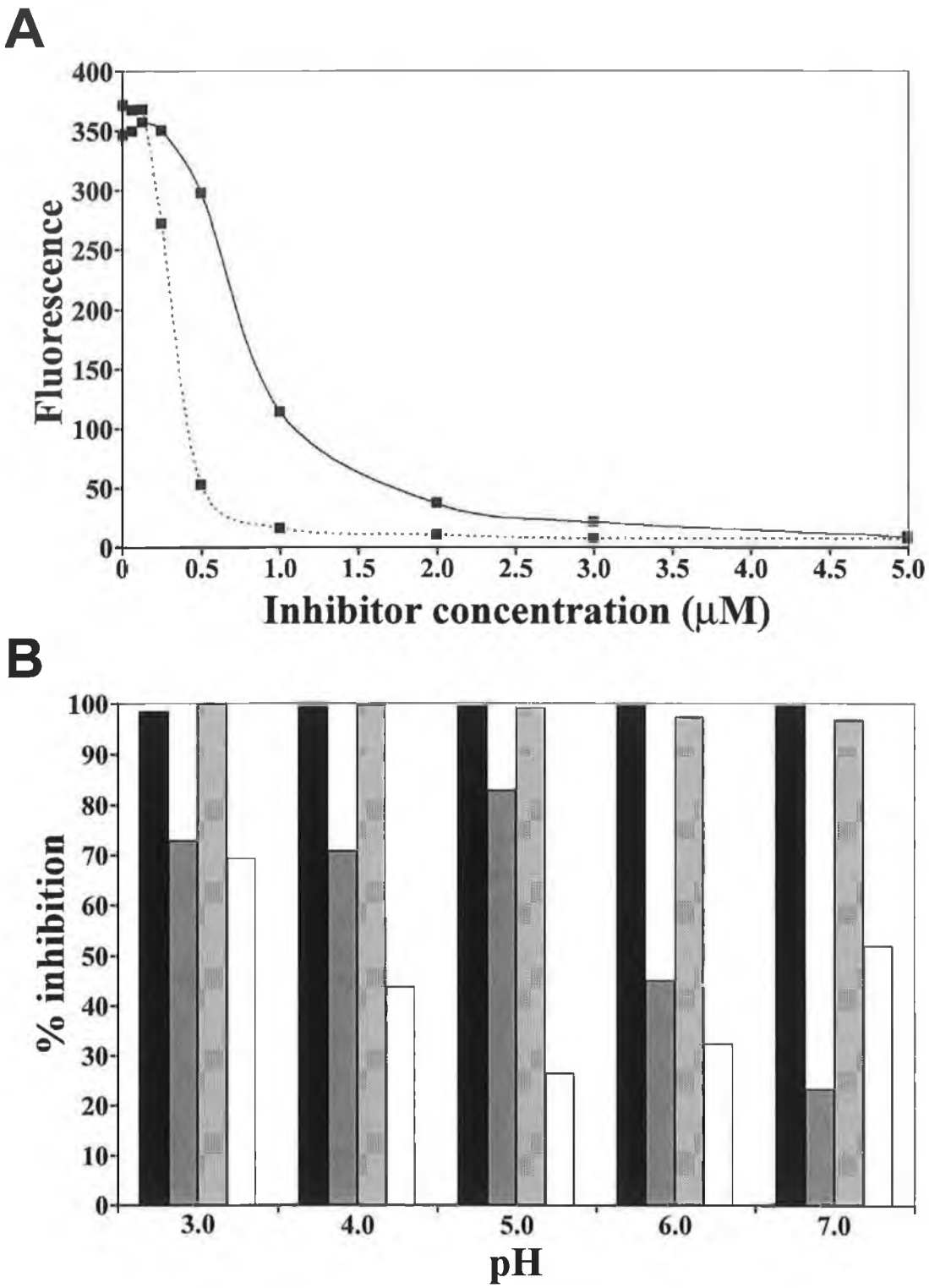


Figure 4.9: Inhibition of recombinant procathepsin L1 activity. *A*, purified rFheproCL1 was mixed with between 0 and 5 μM of recombinant cathepsin L1 propeptide (*dotted line*) or Z-FA-CHN_2 (*solid line*) and assayed with the fluorogenic substrate Z-FR-AMC. *B*, 2 h activated rFheproCL1 was preincubated with two concentrations of each of the inhibitors in 0.1 M sodium citrate buffers with pH between 3.0 and 7.0. Results are presented as percentage inhibition for each inhibitor concentration compared with uninhibited enzyme; *black bars* represent 1 μM propeptide, *grey* represents 0.1 μM propeptide, *diagonal lines* represent 1.5 μM Z-FA-CHN_2 and *white* represents 0.15 μM Z-FA-CHN_2.

Chapter 5: Expression and characterisation of wildtype procathepsin L2

5.1 Expression of wildtype procathepsin L2

Pilot expression studies were carried out with the two *P. pastoris* colonies, labelled A3.1 and B1.28, which were transformed with the pPIC9K.FheproCL2 expression vector construct and previously selected for expression by the colony blot method (see Figure 3.5). These were induced for 3 days by 1% methanol in BMMY medium buffered to pH 6.0, with samples taken daily and analysed by 12% SDS-PAGE. A culture of *P. pastoris* transformed with the pPIC9K.FheproCL2 construct was also induced and sampled as a positive expression control. Unlike the production of rFheproCL1, rFheproCL2 was produced almost entirely in the 37 kDa procathepsin form, with only small amounts produced as the intermediate 30 kDa or mature cathepsin forms (see Figure 5.1A).

Cathepsin L activity of the supernatant samples was determined with the fluorogenic substrate Z-FR-AMC, using the rFheproCL1 supernatants as a control (see Figure 5.1B). Although activity was found for the rFheproCL2 supernatants, it was between 10- and 14-fold lower than for the equivalent rFheproCL1 cultures after 48 h of methanol induction. However, this was consistent with both the much lower levels of intermediate and fully processed protein found in the rFheproCL2 supernatants, as well as the lower activity levels for native FheCL2 versus native FheCL1 (Dowd et al., 1994).

As protein levels produced were visibly higher for the A3.1 *P. pastoris* clone, this was selected as the final expression clone. This was grown on YPD agar plates and glycerol stocks prepared for future inoculations.

The effect of environmental pH on rFheproCL2 production was examined by inducing three identical cultures with 1% methanol in BMMY medium buffered to pH 4.5, 6.0

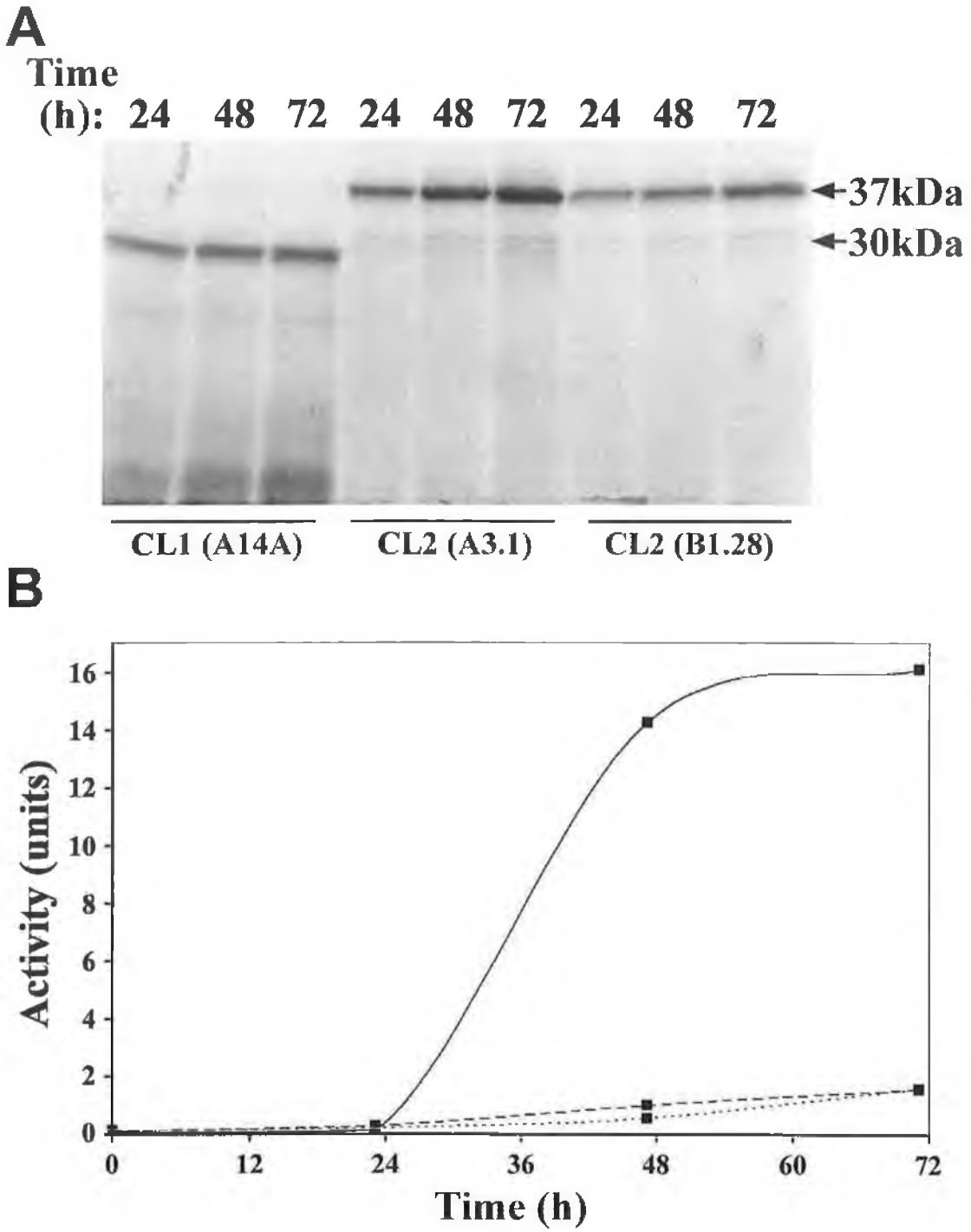


Figure 5.1: SDS-PAGE analysis of *F. hepatica* procathepsin L2 expression from *P. pastoris*. *A*, Cultures of two *P. pastoris* clones (A3.1 and B1.28) transformed with a cDNA encoding procathepsin L2 were induced with 1% methanol in BMMY medium buffered to pH 6.0 at 30°C. *P. pastoris* transformed with procathepsin L1 was induced as a positive control. Aliquots (10 µl) removed at 24, 48 and 72 h after induction were analysed by 12% SDS-PAGE. The 37 and 30 kDa components are indicated. *B*, cathepsin L activity in each culture supernatant was measured with the fluorogenic substrate Z-FR-AMC. *Solid line*, FheproCL1 clone A14A control; *dashed line*, FheproCL2 clone A3.1; *dotted line*, FheproCL2 clone B1.28. Activity units are presented as nmol of AMC released min^{-1} ml^{-1}.

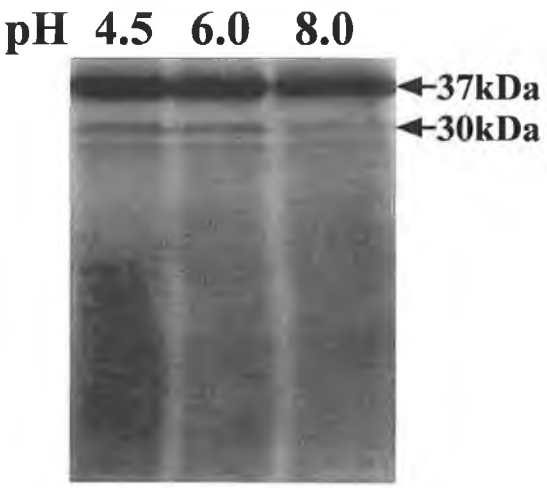


Figure 5.2: Effect of medium pH on procathepsin L2 production. Transformed *P. pastoris* were cultured at 30°C in media buffered to pH 4.5, 6.0, and 8.0. Aliquots of culture supernatant, removed after 72 h, were analysed by 12% SDS-PAGE. The 37 and 30 kDa bands are indicated.

and 8.0 respectively. Samples of culture supernatant taken after 48 and 72 h of induction were analysed by 12% SDS-PAGE (see Figure 5.2). Production levels of protein remained largely unaffected, with roughly equal levels for each medium pH. As for production of rFheproCL1 (see Figure 4.6A), the low level of processing which occurred was reduced at increased pH and increased at lower pH. Unlike for production of rFheproCL1, however, the majority of the protein remained in the 37 kDa procathepsin form in all cultures. The rFheproCL2 protein was purified from yeast culture supernatant by Ni-NTA agarose affinity chromatography for further studies.

5.2 *In vitro* autoactivation of recombinant procathepsin L2

As only very limited amounts of processed protein with low activity levels were found in rFheproCL2 supernatants, *in vitro* autoactivation of rFheproCL2 was investigated using a similar protocol to that used previously. Purified rFheproCL2 was incubated in 0.1 M sodium acetate buffer, pH 4.5, containing 1 mM DTT and 2 mM EDTA at 37°C. Samples were taken at various time points up to 2 h and analysed by 12% SDS-PAGE (see Figure 5.3A). As for rFheCL1 (see Figure 4.8), the 37 kDa procathepsin L2 was processed through intermediates to the fully mature form migrating at approximately 24-25 kDa. A similar incubation was performed with samples taken at time points up to 3 h and analysed for cathepsin L activity with the fluorogenic substrate Z-FR-AMC (see Figure 5.3B). This showed an approximately 5-fold increase in activity, similar to that found for activated rFheproCL1.

A control experiment compared processing of purified rFheproCL2 when incubated in PBS, pH 7.3 with processing when incubated in 0.1 M sodium acetate, pH 4.5. The incubation was carried out for 24 h, with samples taken at various time points and analysed by 12% SDS-PAGE (see Figure 5.4). As for rFheproCL1, the low pH samples processed as expected, while the samples in PBS remained unprocessed, with

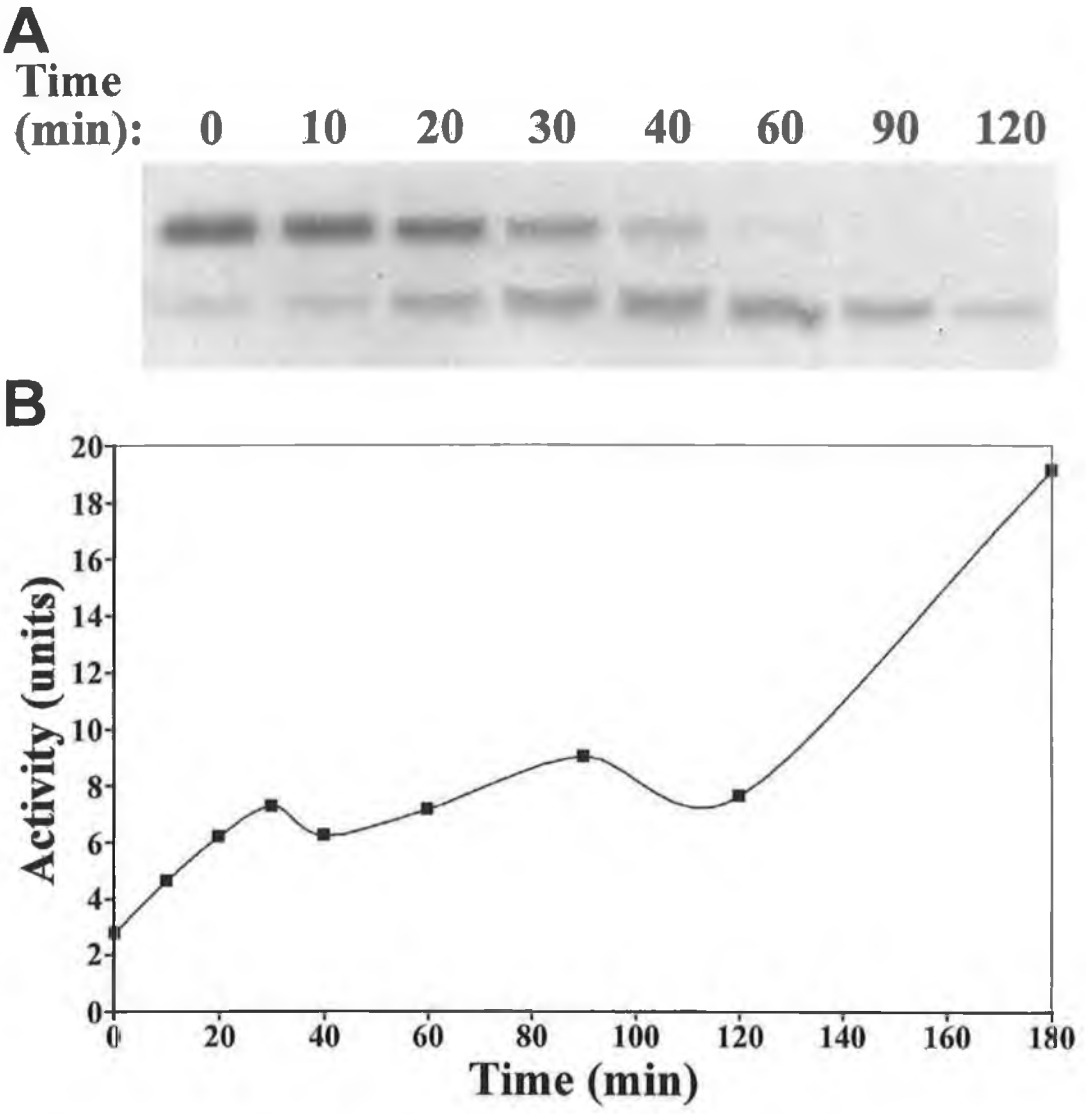


Figure 5.3: Autocatalytic processing of procathepsin L2 at pH 4.5. *A*, Purified procathepsin L2 (5.0 μg) was incubated at 37°C for 2 h in 0.1 M sodium acetate buffer, pH 4.5, containing 2 mM DTT and 2.5 mM EDTA. Samples removed at various time points were analyzed by SDS-PAGE. A pattern of predominant proteins of 37, 30, and 24-25 kDa was found similar to that for the autocatalytic processing of rFheproCL1. *B*, cathepsin L activity of samples taken at time points up to 3 h was measured with the fluorogenic substrate Z-FR-AMC. Activity units are presented as nmol of AMC released min^{-1} ml^{-1}.

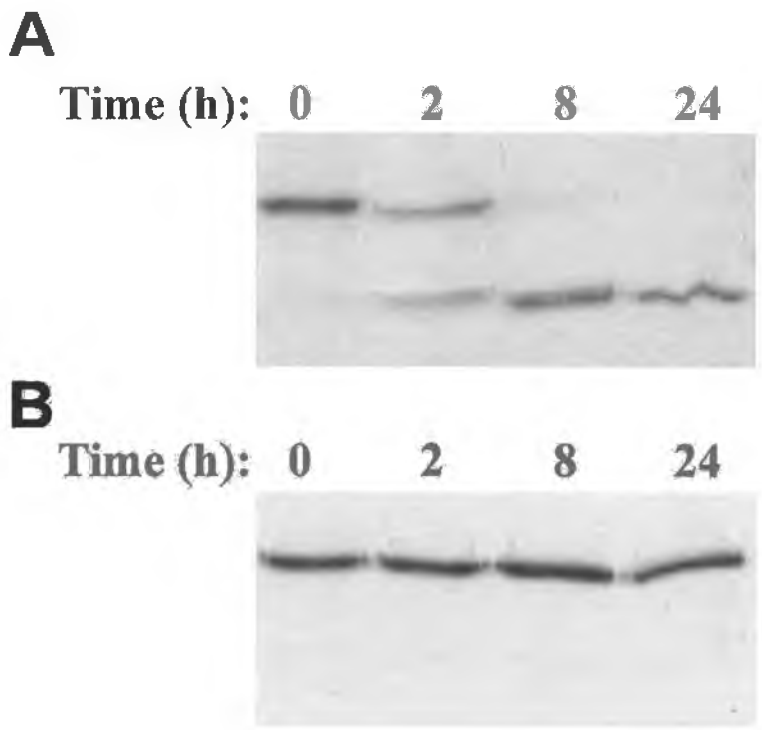


Figure 5.4: Control incubation of procathepsin L2 in PBS. Purified procathepsin L2 was incubated at 37°C for 24 h in either 0.1 M sodium acetate buffer, pH 4.5 (*A*) or PBS, pH 7.3 (*B*), both containing 2 mM DTT and 2.5 mM EDTA. Samples removed at various time points were analysed by SDS-PAGE. Processing of the 30 kDa band did not occur at the higher pH.

the small amount of lower intermediates and mature protein being broken down completely by 24 h. This again demonstrated the stability of the zymogen form of the enzyme. Extension of the incubation time to 24 h also allowed complete processing of rFheproCL2 to the mature active form in the lower pH buffer.

Chapter 6: Preparation, expression and analysis of inactive mutants

6.1 Expression of inactive procathepsin L1 mutants

The inactive mutant procathepsin L1 construct (pPIC9K.[Gly^{26}]FheproCL1) was prepared from the pPIC9K.FheproCL1 template in our laboratory by Dr. Colin Stack (Collins et al., 2004) by the method of gene SOEing (Ho et al., 1989). This mutant was produced by the substitution of the active site cysteine at position 26, with a glycine. As this cysteine residue is crucial to the cleavage of substrate peptide bonds, its replacement with glycine renders the protein inactive.

After transformation of the pPIC9K.[Gly^{26}]FheproCL1 construct into *P. pastoris* GS115 and screening of transformants, a number of colonies were analysed by SDS-PAGE for production of [Gly^{26}]FheproCL1. Two colonies, labelled 2.1 and 5.1 were selected for pilot expression studies. When supernatants were analysed by 12% SDS-PAGE, it was noted that the protein produced by one clone, labelled 5.1, exhibited slower migration than either wildtype rFheproCL1 or the other mutant, produced by clone 2.1 (see Figure 6.1). The cDNA of both clones was sequenced and the 5.1 clone was shown to have an additional mutation to $Cys^{26} \rightarrow Gly^{26}$ located within the propeptide and that was not found in the protein produced by the 2.1 clone. This mutation substituted a proline in place of a leucine at position -12 in the propeptide, therefore this clone was labelled [$Pro^{-12}Gly^{26}$]FheproCL1.

Effect of pH on production was examined by inducing cultures of both mutants along with wild-type CL1 as a control by 1% methanol in BMMY medium buffered to pH 6.0 and 8.0. Samples of supernatant were taken at various time points and analysed by 12% SDS-PAGE (see Figure 6.1). Both mutants were unaffected by medium pH, with r[Gly^{26}]FheproCL1 appearing as a single strong band co-migrating with the 37 kDa procathepsin L component of wildtype rFheproCL1 and the r[$Pro^{-12}Gly^{26}$]FheproCL1

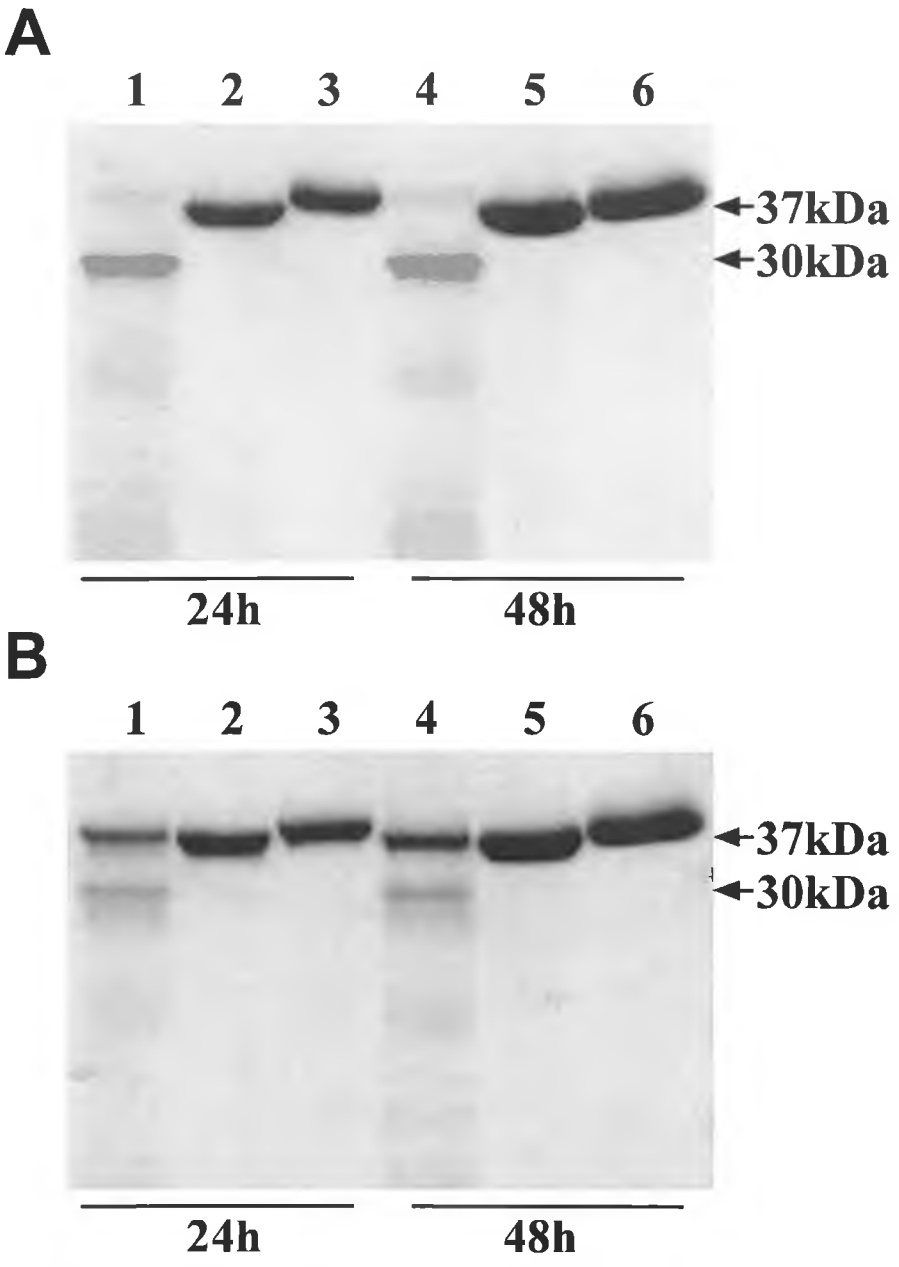


Figure 6.1: Expression of procathepsin L1 inactive mutants by *P. pastoris*. *P. pastoris* cells transformed with either wild-type procathepsin L1 or the procathepsin L1 clones that had the active site Cys^{26} replaced by Gly were cultured at 25°C in media buffered to pH 6.0 (*A*) and 8.0 (*B*). After 24 h and 48 h of induction with 1% (v/v) methanol, samples of culture supernatant were analysed by 12% SDS-PAGE. The 30-kDa intermediate form of procathepsin L was observed only in the wild-type culture. Mutant clone 5.1, which was later found to contain an additional mutation from Leu^{-12} to Pro ran slightly higher than the wild-type or single mutant procathepsin L1. *Lanes 1 and 4*, rFheproCL1 wild-type; *lanes 2 and 5*, r[Gly^{26}]FheproCL1 inactive mutant; *lanes 3 and 6*, r[$Pro^{-12}Gly^{26}$]FheproCL1 double mutant.

appearing as a single strong band migrating slightly higher in the gel. No other bands or breakdown products were formed and the production of both mutants was almost 4-fold greater than that of the combined 37 and 30 kDa components of rFheproCL1.

The two mutant proteins were purified from yeast culture supernatant by Ni-NTA agarose affinity chromatography and subjected to *N*-terminal sequencing. In both cases, a sequence of Glu-Ala-Glu-Ala-Tyr was determined, corresponding to 5 amino acids positioned one residue from the end of the yeast α-factor secretory signal sequence and matching results for the 37 kDa component of the wild-type. This indicated that the both bands represented recombinant inactive mutant procathepsin L1 plus the last 6 amino acids (Glu-Ala-Glu-Ala-Tyr-Val) of the yeast α-factor secretory signal sequence before the Ser^{-91} start of the propeptide (see Figure 6.5). As expected, both mutant proteins lacked activity against the fluorogenic substrate Z-FR-AMC.

6.2 *In vitro* processing of inactive procathepsin L1 mutants

In order to investigate processing of procathepsin L1 to its mature form, both mutant proteins were incubated at pH 4.5 alongside wildtype rFheproCL1 as a control in a similar manner to previous incubations (see Section 4.6) As in these previous incubations, purified wildtype rFheproCL1 was autocatalytically processed through various intermediates to a mature form. However, neither of the inactive mutant enzymes were capable of processing to lower molecular size forms (see Figure 6.2). This was to be expected due to lack of proteolytic activity present in the mutant enzymes.

Further investigations were performed with the incubation of exogenously added wildtype rFheproCL1 which had been activated for 2 h mixed with each of the two mutant enzymes. Addition of activated wild-type rFheproCL1 to purified r[Gly^{26}]FheproCL1 mutant at pH 5.0 resulted in the progressive appearance of a minor

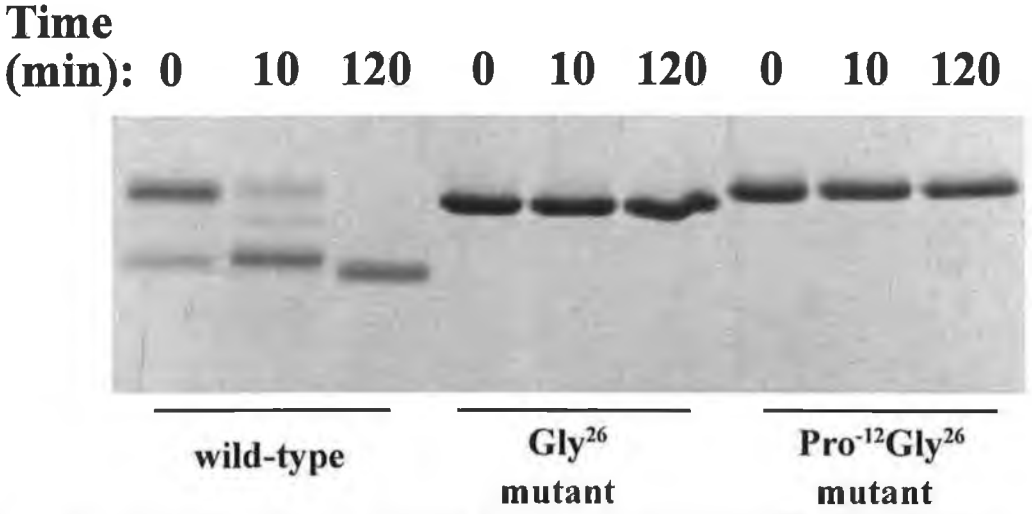


Figure 6.2: Incubation of procathepsin L1 inactive mutants at pH 4.5. Purified rFheproCL1, r[Gly^{26}]FheproCL1 and [Pro^{-12}Gly^{26}]FheproCL1 were incubated at 37°C for 2 h in 0.1 M sodium acetate buffer, pH 4.5, containing 2 mM DTT and 2.5 mM EDTA. Samples removed at various time points were analysed by SDS-PAGE. While autoactivation of rFheproCL1 occurred as before, the inactive mutants remained unaffected by the incubation.

band at ~35 kDa over an incubation period of 4 h and a second major band at ~25 kDa that co-migrated with fully active cathepsin L1, with the intensity of this band increasing over the incubation time (see Figure 6.3A). In parallel experiments with purified r[$Pro^{-12}Gly^{26}$]FheproCL1, no significant processing of the double mutant occurred, although the major band progressively shifted to a lower position in the gel over time to co-migrate with the 37 kDa procathepsin form of rFheproCL1 (see Figure 6.3B).

Effect of pH on exogenous processing of the mutants was investigated by repeating the above incubations for 2 h in 0.1 M sodium acetate buffers at pH 3.0, 4.0 and 5.0. Similar results were found at all pH values, with the double mutant again showing only a small shift in position with no significant intermolecular processing (see Figure 6.3C). Experiments in which activated rFheproCL1 was added to heat-denatured r[Gly^{26}]FheproCL1 showed that the mutant was sensitive to complete degradation by active cathepsin L1, confirming that the expressed mutant was correctly folded (see Figure 6.4).

6.3 *N*-terminal sequencing of exogenously processed forms of inactive mutant procathepin L1

Reaction mixes containing exogenously added activated wildtype rFheproCL1 and each of the inactive mutants were incubated as before (see Section 6.2), analysed by SDS-PAGE and transferred to PVDF membrane. The membrane was then sent for *N*-terminal sequencing of the exogenously processed forms of each mutant enzyme. Attempts were made to obtain the *N*-terminal amino acid sequence for the ~35 kDa intermediate form of [Gly^{26}]FheproCL1, but were unsuccessful. For the fully processed ~25 kDa form, a sequence of His-Gly-Val-Pro-Tyr (with less than 10 % Gly-Val-Pro-

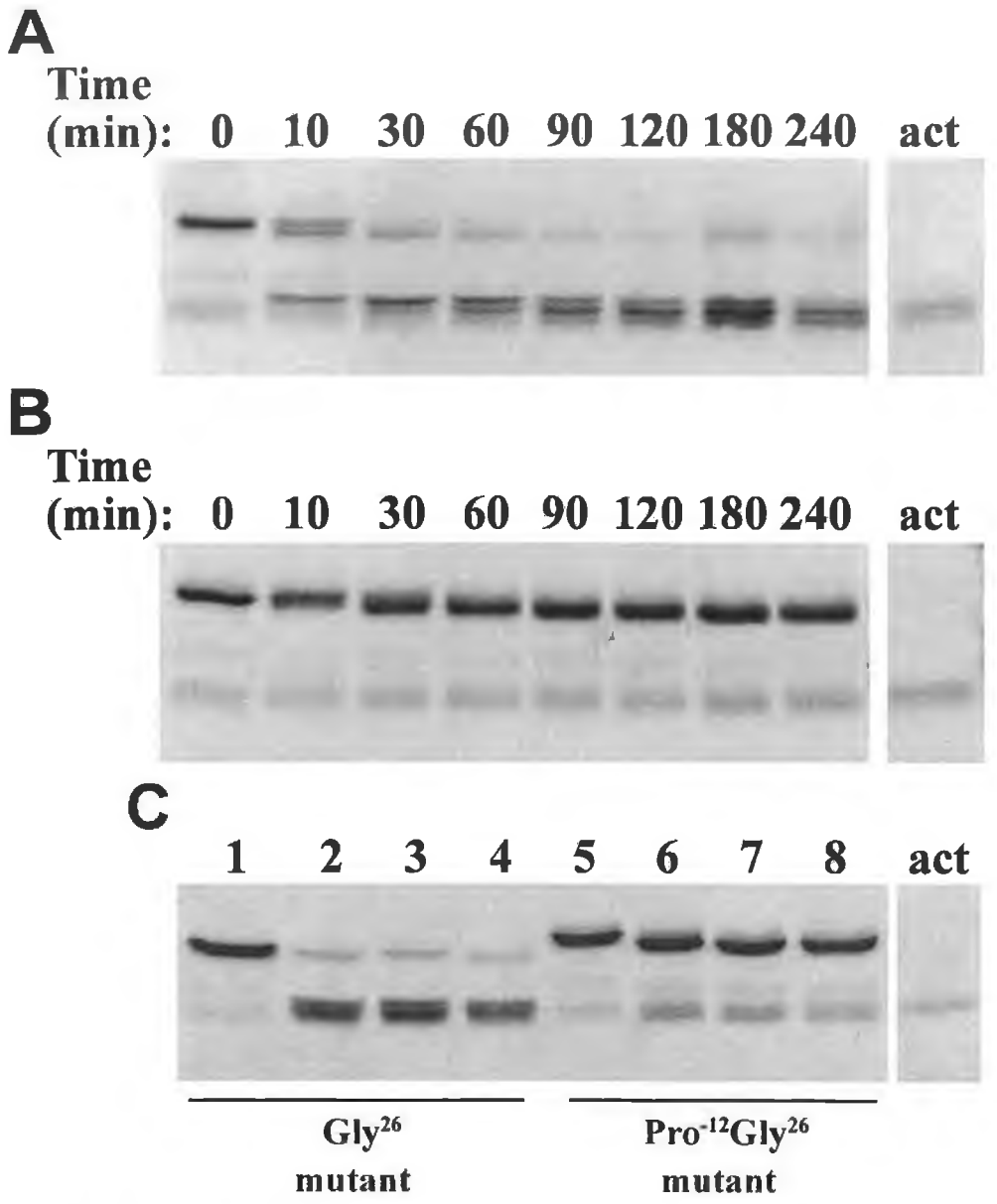


Figure 6.3: Incubation of procathepsin L1 inactive mutants with exogenously added active cathepsin L1. Purified r[Gly^{26}]FheproCL1 (5 µg) (*A*) or r[Pro^{-12}Gly^{26}]FheproCL1 (5 µg) (*B*) were incubated for up to 4 h in 0.1 M sodium acetate buffer, pH 5.0, containing 2 mM DTT and 2.5 mM EDTA at 37°C in the presence of 0.5 µg of 2 h activated rFheproCL1 (*act*). Note that the double mutant remains unprocessed. *C*, parallel experiments were carried out as above but replacing the buffer with 0.1 M sodium acetate, pH 3.0, 4.0 and 5.0, and an incubation time of 2 h. Processing of the r[Gly^{26}]FheproCL1 mutant occurred at all pH values, while the double mutant remained unprocessed. *Lanes 1 and 5*, no incubation; *lanes 2 and 6*, pH 3.0; *lanes 3 and 7*, pH 4.0; *lanes 4 and 8*, pH 5.0.

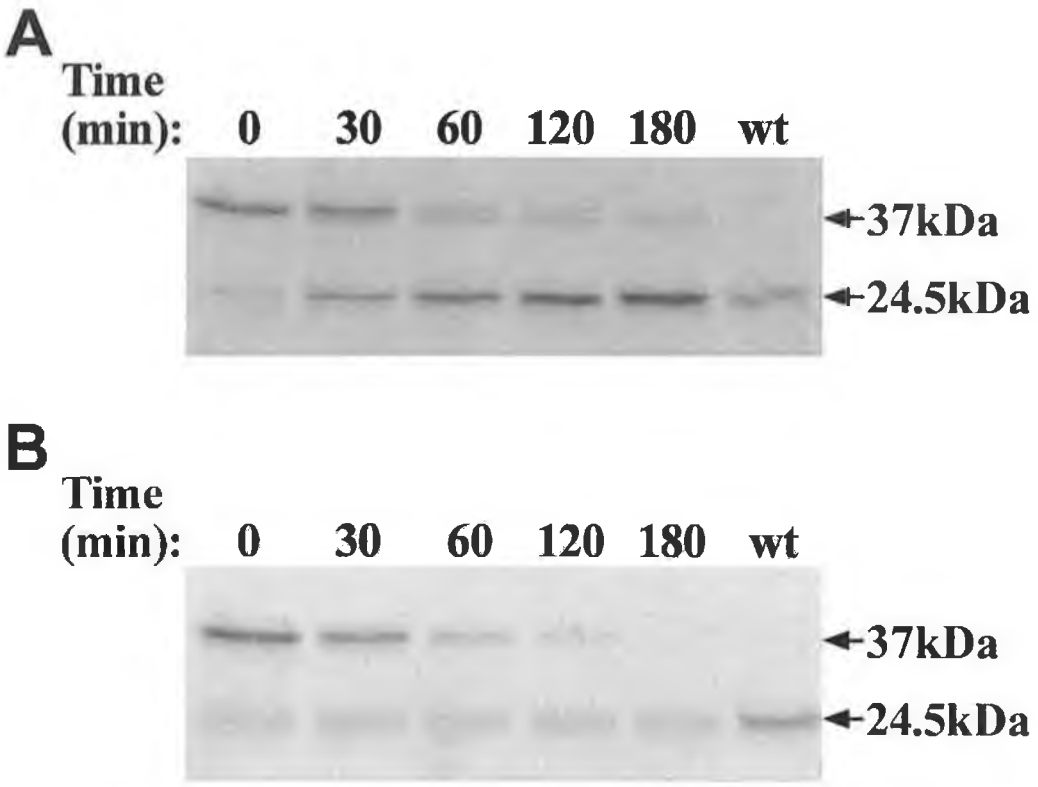


Figure 6.4: Intermolecular processing of r[Gly^{26}]FheproCL1 inactive mutant by exogenously added active cathepsin L1. *A*, purified r[Gly^{26}]FheproCL1 (5 µg) was incubated for up to 3 h in 0.1 M sodium citrate buffer, pH 5.0, containing 2 mM DTT and 2.5 mM EDTA at 37°C in the presence of 0.5 µg of 2 h activated rFheproCL1. The position of the mature wild-type cathepsin L1 at 24.5 kDa (2.5 µg loading) is indicated in the lane labelled wt with the arrow. *B*, parallel experiments were carried out as above but in this case purified r[Gly^{26}]FheproCL1 was unfolded by heating at 95°C for 3 min prior to the addition of activated wild-type CL1. Adapted from Collins et al. (2004).

Tyr-Glu), indicating a cleavage site located 10 amino acids prior to the *N*-terminus of mature cathepsin L1 (see Figure 6.5).

N-terminal analysis of the double mutant r[Pro^{-12}Gly^{26}]FheproCL1 reactions revealed several sequences, with high levels of Val-Xaa-Asp, Asn-Asp and Leu-Xaa found. These sequences match positions found at the end of the yeast α-factor secretion signal sequence and the early residues of the propeptide (see Figure 6.5). These data showed that while the exogenously added wildtype cathepsin L1 was capable of degrading the r[Pro^{-12}Gly^{26}]FheproCL1 sequentially from the *N*-terminus, it was incapable of processing the protein into a mature form.

6.4 Construction of pPIC9K. [Gly^{26}]FheproCL2 inactive mutant expression vector

The cDNA encoding the inactive mutant form of FheproCL2, in which the active site cysteine (Cys^{26}) residue was replaced with a glycine residue, was produced as a synthetic gene in the pPCR-Script plasmid by Geneart GmbH, Regensburg, Germany. The gene was successfully amplified from the *03-1024*pPCR-Script plasmid by PCR as for both the wild-type cathepsins L1 and L2. When run on a 1% agarose gel, the expected 980 bp band was visualised (see Figure 6.6A). The purified product was first ligated into the pGEM-T Easy vector and transformed into *E. coli* Top 10F'. Transformants were screened by the colony pick PCR method and one transformant containing the correct sized insert was selected for digestion and insertion into the pPIC9K expression vector (see Figure 6.6B). The purified, SnaBI/AvrII digested insert was ligated into pPIC9K vector and retransformed into *E. coli* Top 10F'.

Purified plasmid DNA from *E. coli* transformants was screened by colony pick PCR with the standard cathepsin L primers (see Figure 6.6C). Plasmid DNA from two colonies (labelled C5 and C9) which were found to be positive for an insert were

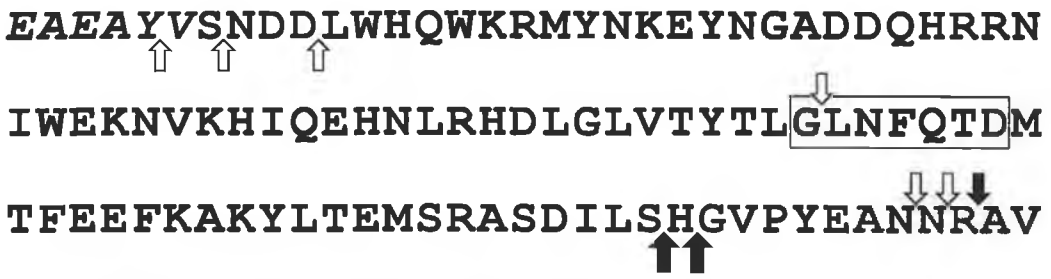


Figure 6.5: Amino acid sequence of the propeptide of *F. hepatica* cathepsin L1 with cleavage sites that lead to activation indicated. The downward solid arrow indicates the cleavage site observed when native mature cathepsin L1 is secreted in vitro by adult *F. hepatica*. The first downward open arrow indicates the site of cleavage within the GXNXFXD motif (*boxed*) that gives rise to the 30 kDa intermediate component when procathepsin L1 is expressed in *P. pastoris*, and the second and third downward open arrows show the sites of cleavage when this is further activated in vitro at pH 5.0 (see Figure 4.8). The upward solid arrows show the sites of cleavage of r[Gly^{26}]FheproCL1 mutant by exogenously added mature cathepsin L1. The upward open arrows show the probable sites of cleavage of r[Pro^{-12}Gly^{26}]FheproCL1 double mutant by exogenously added mature cathepsin L1. The final residues of the yeast α-factor secretion signal sequence are marked in italics. Adapted from Collins et al. (2004).

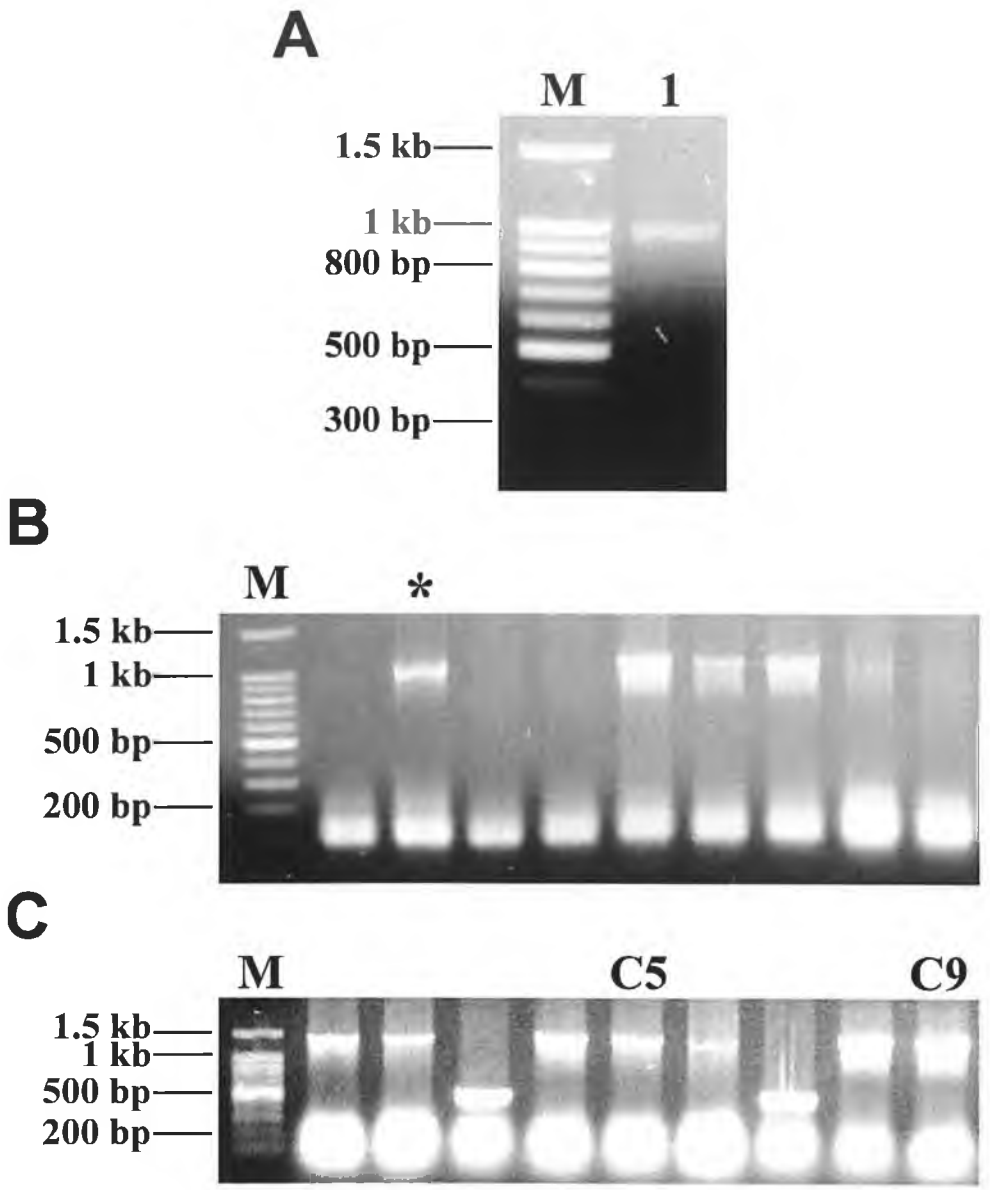


Figure 6.6: 1% agarose gels showing construction of pPIC9K.[Gly^{26}]FheproCL2 expression vector. *A*, inactive mutant procathepsin L2 PCR product before ligation into pGEM-T. *Lane M*, 100 bp DNA ladder; *lane 1*, [Gly^{26}]FheproCL2 PCR product. *B*, colony pick PCR of selected *E. coli* pGEM-T. [Gly^{26}]FheproCL2 transformants with 5'AOX and 3'AOX primers. *Lane M*, 100 bp DNA ladder; *asterisk*, transformant selected for digestion and ligation into pPIC9K. *C*, colony pick PCR of selected *E. coli* pPIC9K. [Gly^{26}]FheproCL2 transformants with 5'AOX and 3'AOX primers. *Lane M*, 100 bp DNA ladder; clones C5 and C9, used for transformation of *P. pastoris*, are indicated.

selected for transformation into *P. pastoris* and sent for sequencing. The sequences were confirmed to match the database sequence (with the exception of the active site $Cys^{26} \rightarrow Gly^{26}$ mutation) and to be in frame with the yeast α-factor secretion signal sequence.

6.5 Transformation and screening of *P. pastoris* GS115 with pPIC9K.[Gly^{26}]FheproCL2 construct

The two pPIC9K.[Gly^{26}]FheproCL2 constructs were linearised with SalI and transformed into competent *P. pastoris* GS115 cells. His^+ phenotype was confirmed by growth on histidine deficient medium and between 16 and 20 colonies from each transformation were screened directly for expression by the colony blot rapid screening method. All colonies showed low expression of a His_6-tagged protein. One colony, labelled C5.8, which showed slightly stronger labelling with anti-His_6 antibody was selected for pilot expression studies (see Figure 6.7). No colonies showed antibody labelling as strongly as the rFheproCL1 control.

6.6 Expression of inactive mutant procathepsin L2

The *P. pastoris* GS115 clone C5.8 which was transformed with the pPIC9K.[Gly^{26}]FheproCL2 expression vector construct was induced for 6 days by 1% methanol in BMMY medium buffered to pH 6.0, with samples taken daily and analysed by 12% SDS-PAGE. Cultures of the wild-type procathepsins L1 and L2 as well as the double mutant inactive procathepsin L1 were also induced as controls and to gauge comparative expression levels (see Figure 6.8). Although the inactive procathepsin L2 was produced with a predominant 37 kDa band as for rFheproCL2 and the procathepsin L1 mutants, it was produced in several fold lower quantities. The protein was purified from culture supernatant by Ni-NTA agarose affinity chromatography. Purified protein

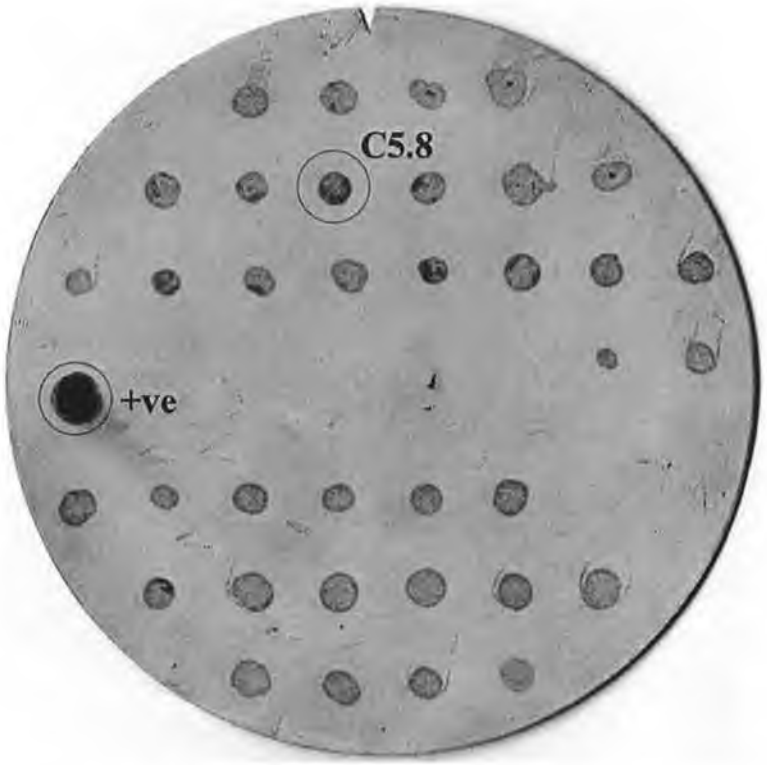


Figure 6.7: Colony blot screening of *P. pastoris* pPIC9K.[Gly^{26}]FheproCL2 transformants for expression. Colonies were transferred onto nitrocellulose membrane and grown on YNB/2% methanol plates. Bound protein after cell lysis was probed with mouse monoclonal anti-His_6 antibody. Positive colonies show as a dark patch. The positive control colony, transformed with pPIC9K.FheproCL1, and the colony chosen for pilot expression studies (C5.8) are marked with a ring.

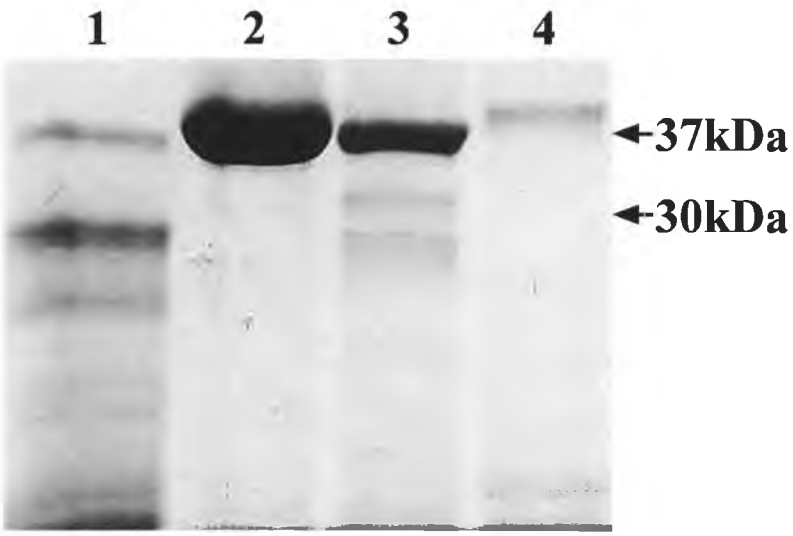


Figure 6.8: Expression of procathepsin L2 inactive mutant by *P. pastoris*, compared with previous recombinants. *P. pastoris* cells transformed with either the procathepsin L2 clone that had the active site Cys^{26} replaced by Gly or the previous recombinants were cultured at 30°C in media buffered to pH 6.0. After 6 days of induction with 1% (v/v) methanol, samples of culture supernatant were analysed by 12% SDS-PAGE. *Lane 1*, rFheproCL1 wild-type; *lane 2*, r[$Pro^{-12}Gly^{26}$]FheproCL1 inactive mutant; *lane 3*, rFheproCL2 wild-type; *lane 4*, r[Gly^{26}]FheproCL2 inactive mutant.

was transferred to PVDF membrane and sent for *N*-terminal sequence analysis. A sequence of Glu-Ala-Glu-Ala-Tyr-Val was determined, matching with the last 6 amino acid residues of the yeast α-factor secretory signal sequence before the Ser^{-91} start of the propeptide. As expected, the mutant protein lacked activity against the fluorogenic substrate Z-FR-AMC.

6.7 Determination of the 3D structure of *F. hepatica* procathepsins L1 and L2

In collaboration with Dr. Colin Stack in our laboratory, 30 mg of purified inactive mutant r[Gly^{26}]FheproCL1 was produced and sent to Dr. Linda Brinen of the *Fasciola* Structural Biology Laboratory at UCSF, San Francisco, USA for determination of the 3D structure of the protein by X-Ray crystallography. The structure is still being completed, but a preliminary model has been produced (see Figure 6.10) whose general structure compares well with the crystal structure of human cathepsin L (Coulombe et al., 1996).

Due to the low levels of production of r[Gly^{26}]FheproCL2 from transformed *P. pastoris* cultures, a large scale culture, induction and purification of the inactive mutant protein was performed. This culture was started with 7.7 L of BMGY medium for the initial growth stage and resuspended in 1.55 L BMMY buffered to pH 6.0 containing 1% methanol for induction of protein expression (see Section 2.3.22). The r[Gly^{26}]FheproCL2 protein was then purified in a large scale process by Ni-NTA affinity chromatography and concentrated approximately 20-fold with a centrifuge concentrator. However, when the concentrated protein preparation was analysed by SDS-PAGE it was found that much degradation had occurred over the course of the process (see Figure 6.9), likely due to the longer times involved for the processing of such a large volume of culture supernatant. Also, only a relatively small quantity of protein was produced for the culture volume used; only 8 mg of protein was obtained

from the 1.55 L induction culture, compared to 110 mg of rFheproCL1 produced from a previous 400 ml induction culture. Further optimisation of the induction and purification process of this protein, including re-transformation of *P. pastoris* in order to identify higher copy number integrants with greater levels of protein production, would be required before more experimentation could be performed and the 3D structure of procathepsin L2 determined. This work is ongoing in our laboratory.

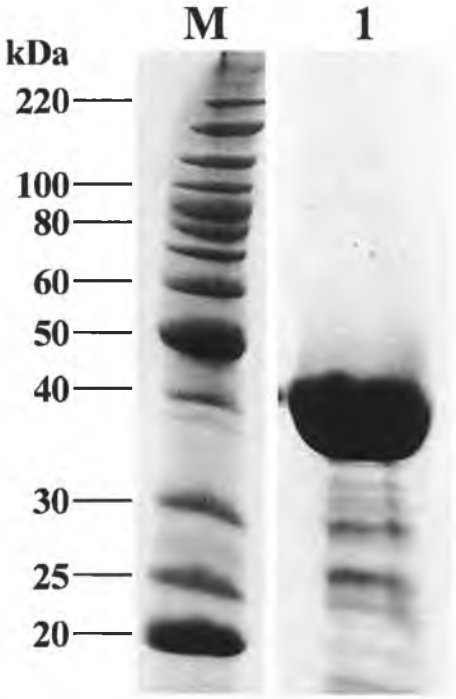


Figure 6.9: SDS-PAGE analysis of purified inactive mutant procathepsin L2. *P. pastoris* cells transformed with the procathepsin L2 clone that had the active site Cys^{26} replaced by Gly were cultured at 30°C in media buffered to pH 6.0 in a large scale culture. After 3 days of induction with 1% (v/v) methanol, the protein was purified from the culture medium by Ni-NTA affinity chromatography and 20X concentrated with a centrifuge concentrator. The concentrated r[Gly^{26}]FheproCL2 inactive mutant was analysed by 12% SDS-PAGE and found to have been partially degraded during the preparation process. *Lane M*, molecular size markers; *lane 1*, concentrated purified r[Gly^{26}]FheproCL2.

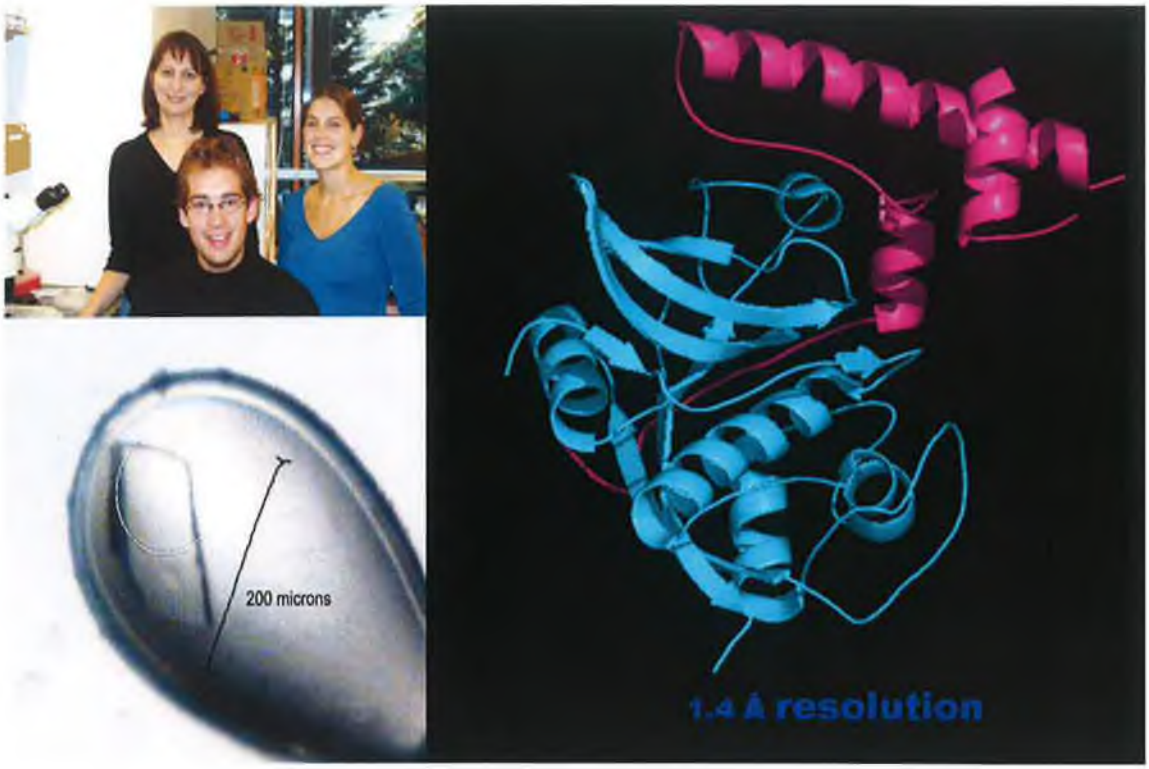


Figure 6.10: Collaboration with the *Fasciola* Structural Biology Laboratory, UCSF on the determination of the 3D structure of *F. hepatica* procathepsin L1. Purified r[Gly^{26}]FheproCL1 prepared in our laboratory was successfully crystallised by Dr. Linda Brinen and her team (pictured top left) in the *Fasciola* Structural Biology Laboratory, UCSF, San Francisco, USA. An example of the crystals formed is presented in the bottom left. While the determination of the 3D structure of procathepsin L1 is still being elucidated by X-ray crystallography, a preliminary model (right) has been prepared whose general structure matches well with that of human procathepsin L.

Chapter 7: Comparison of recombinant cathepsin L1 and L2

7.1 Secretion of cathepsin L proteases by *F. hepatica* parasites

Adult *F. hepatica* maintained *in vitro* secrete two major proteases previously characterised as 27.5 kDa cathepsin L1 and 29 kDa cathepsin L2 in the laboratory (Smith et al., 1993a; Dowd et al., 1994). Immunoblotting of adult *F. hepatica* ES products demonstrated that both enzymes are reactive with sera prepared against purified native mature cathepsin L1. Neither reacts with sera prepared specifically against a recombinant cathepsin L1 propeptide (see Figure 7.1, also Figure 4.2B; Collins et al., 2004).

Immunofluorescence and immunoelectron microscopy experiments were carried out by Dr. Aaron Maule's laboratory at Queen's University Belfast, UK, using anti-propeptide and anti-mature cathepsin L1 antisera from our laboratory (Collins et al., 2004). Confocal laser immunocytochemistry using the anti-mature cathepsin L1 antiserum localised the protease to the gastrodermal epithelial cells throughout the gut of adult *F. hepatica* (see Figure 7.2B). Labelling was found to be more dense toward the apical end of these gastrodermal cells, suggesting that the enzyme is stored in secretory vesicles within the cells (see Figure 7.2C). This was confirmed by electron microscopy (see Figure 7.2D-F), with the densely-stained secretory vesicles showing strong immunoreactivity with both anti-mature and anti-propeptide sera. The reactivity of the vesicles with anti-mature and anti-propeptide antibodies indicates that the enzyme is stored in these vesicles in its inactive procathepsin form.

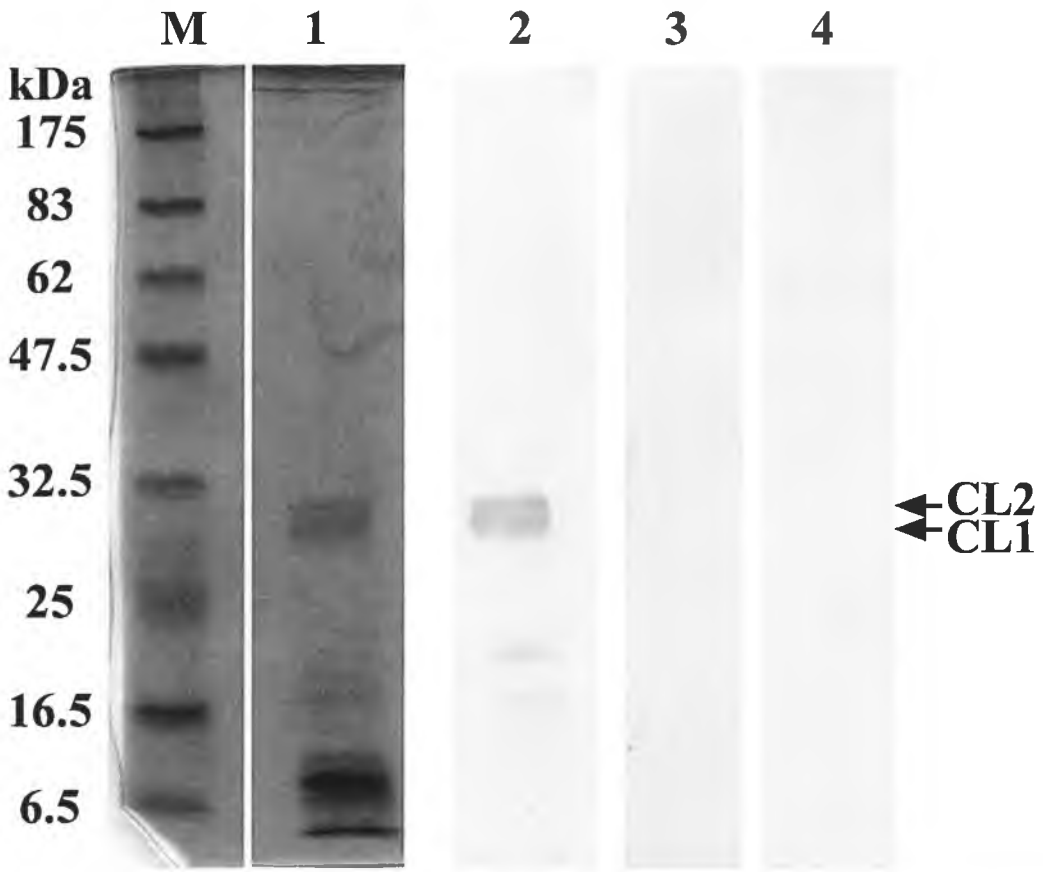


Figure 7.1: SDS-PAGE and immunoblot analysis of *F. hepatica* ES products. ES products obtained from medium in which *F. hepatica* were cultured were analysed by SDS-PAGE and immunoblotting. *Lane M*, molecular size markers; *lane 1*, Coomassie Blue-stained 12% SDS-PAGE gel of ES products; *lane 2*, ES products probed with antiserum prepared in rabbits against native mature cathepsin L1; *lane 3*, ES products probed with antiserum prepared in rabbits against recombinant propeptide of procathepsin L1; *lane 4*, ES products probed with normal rabbit serum. Arrows indicate position of major secreted and fully processed cathepsin L1 and cathepsin L2. Adapted from Collins et al. (2004).

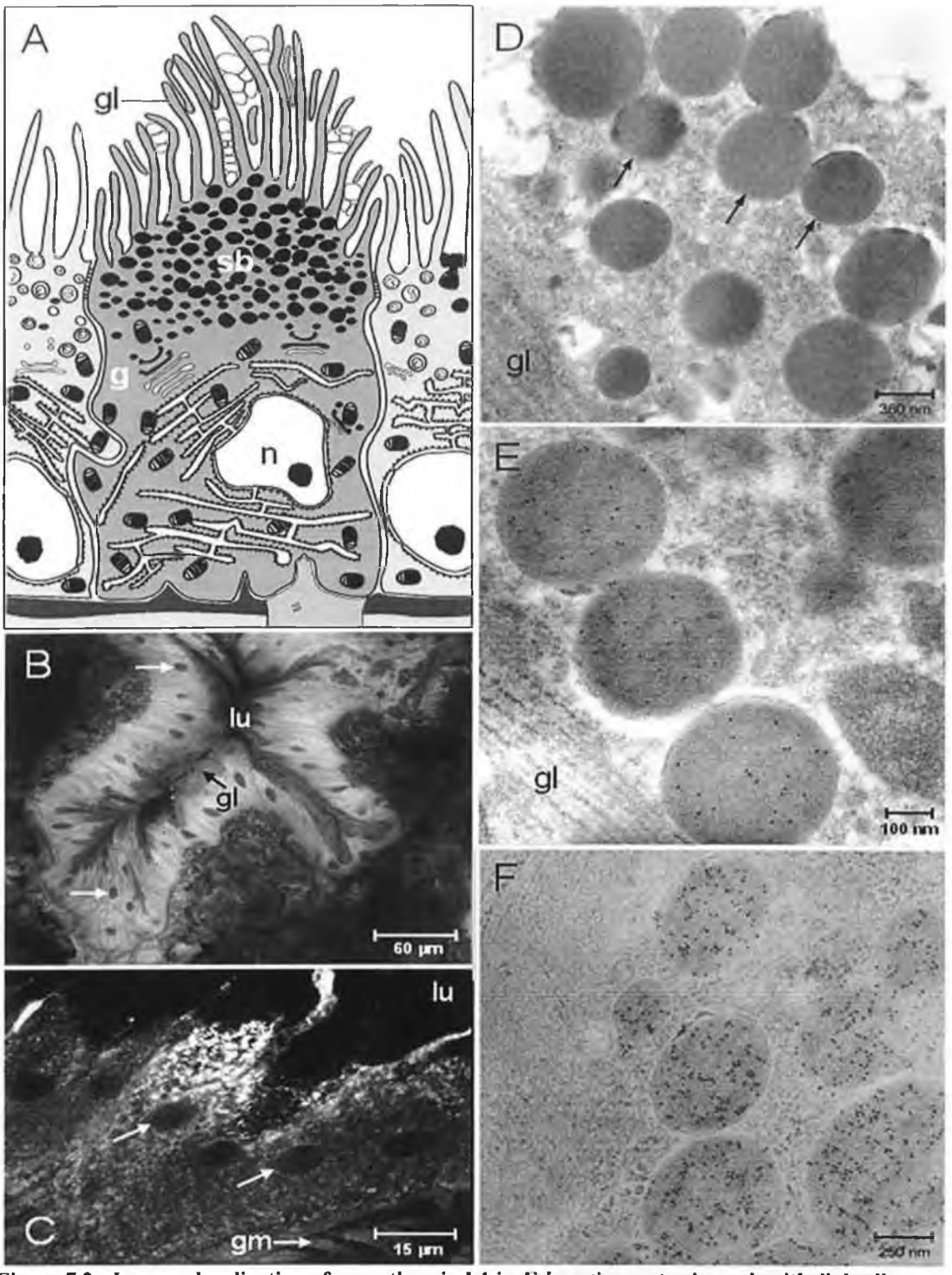


Figure 7.2: Immunolocalization of procathepsin L1 in *F. hepatica* gastrodermal epithelial cells. *A*, schematic representation of gastrodermal epithelial cells in *F. hepatica* (after Smyth and Halton, 1983). *g*, Golgi; *gl*, gut lamellae; *n*, nuclei; *sb*, secretory bodies. *B*, confocal scanning laser micrograph showing immunoreactivity for mature cathepsin L1 within the gastrodermis of adult *F. hepatica*; the nuclei (*arrows*) of the gastrodermal cells are clearly visible; *gl*, gut lamellae; *lu*, gut lumen. *C*, confocal scanning laser micrograph showing immunoreactivity for mature cathepsin L1 within the gastrodermal cells of adult *F. hepatica*; the nuclei (*arrows*) of the gastrodermal cells are clearly visible. Immunoreactivity appears as a punctuate pattern at the apical end of the cells where the secretory vesicles are located; *gm*, gut muscle; *lu*, gut lumen. *D*, electron micrograph showing non-reactivity of secretory bodies (*arrows*) within the epithelial cells of the gastrodermis with control preimmunized rabbit serum; *gl*, gut lamellae. *E*, electron micrograph showing the localization *F. hepatica* procathepsin L1 to secretory bodies of the gastrodermal epithelial cells with antiserum prepared against the recombinant propeptide of procathepsin L1; *gl*, gut lamellae. *F*, electron micrograph showing the localization of *F. hepatica* cathepsin L1 to secretory bodies of the gastrodermal epithelial cells with antiserum prepared against purified mature portion of cathepsin L1. Note that the labeling is confined to the contents of the secretory bodies. Adapted from Collins et al. (2004).

7.2 Analysis of mRNA production and intercellular expression by *P. pastoris* transformants

Total RNA was extracted from *P. pastoris* cells transformed with each of the recombinant proteins (with the exception of r[Gly^{26}]FheproCL1) after they had been induced for 4 days by 1% methanol in BMMY medium buffered to pH 6.0. The RNA was quantified and a RT-PCR reaction prepared with 1 μg template to synthesise first strand cDNA. This cDNA should correspond to the total $polyA^+$ mRNA produced by the cells. PCR was used to probe the cDNA for the presence of the cathepsin L gene transcripts and the reaction products visualised on a 1% agarose gel (see Figure 7.3A). All transformants showed a clear band at ~980 bp with roughly the same intensity.

Samples of the same cultures were taken after 3 days of induction and cell lysates were prepared. These lysates were analysed directly by 12% SDS-PAGE and by immunoblotting (see Figure 7.3B). In the latter case, the lysates were probed with mouse monoclonal anti-His_6 antibody to locate His_6-tagged recombinants. SDS-PAGE analysis showed many bands in all supernatants; however, bands representing the recombinant proteins could not be discerned from background proteins generated by *P. pastoris*. In the immunoblot analysis, both the 37 and 30 kDa bands reacted with the anti-His_6 antibody in the rFheproCL1 lysate (see Figure 7.3B, Lane 7); bands of approximately 37 kDa were reactive for each of the other lysates (see Figure 7.3B, Lanes 8 to 10). No anti-His_6 reactive bands were present in the non-induced control (see Figure 7.3B, Lane 6). These reactive bands present in induced *P. pastoris* cell lysates correspond to those present in culture supernatants.

7.3 Gelatin-substrate activity of recombinant cathepsins L1 and L2

The effect of pH on the activity of the two active recombinants, rFheproCL1 and rFheproCL2, was first investigated by gelatin gel zymography. Samples of the fully

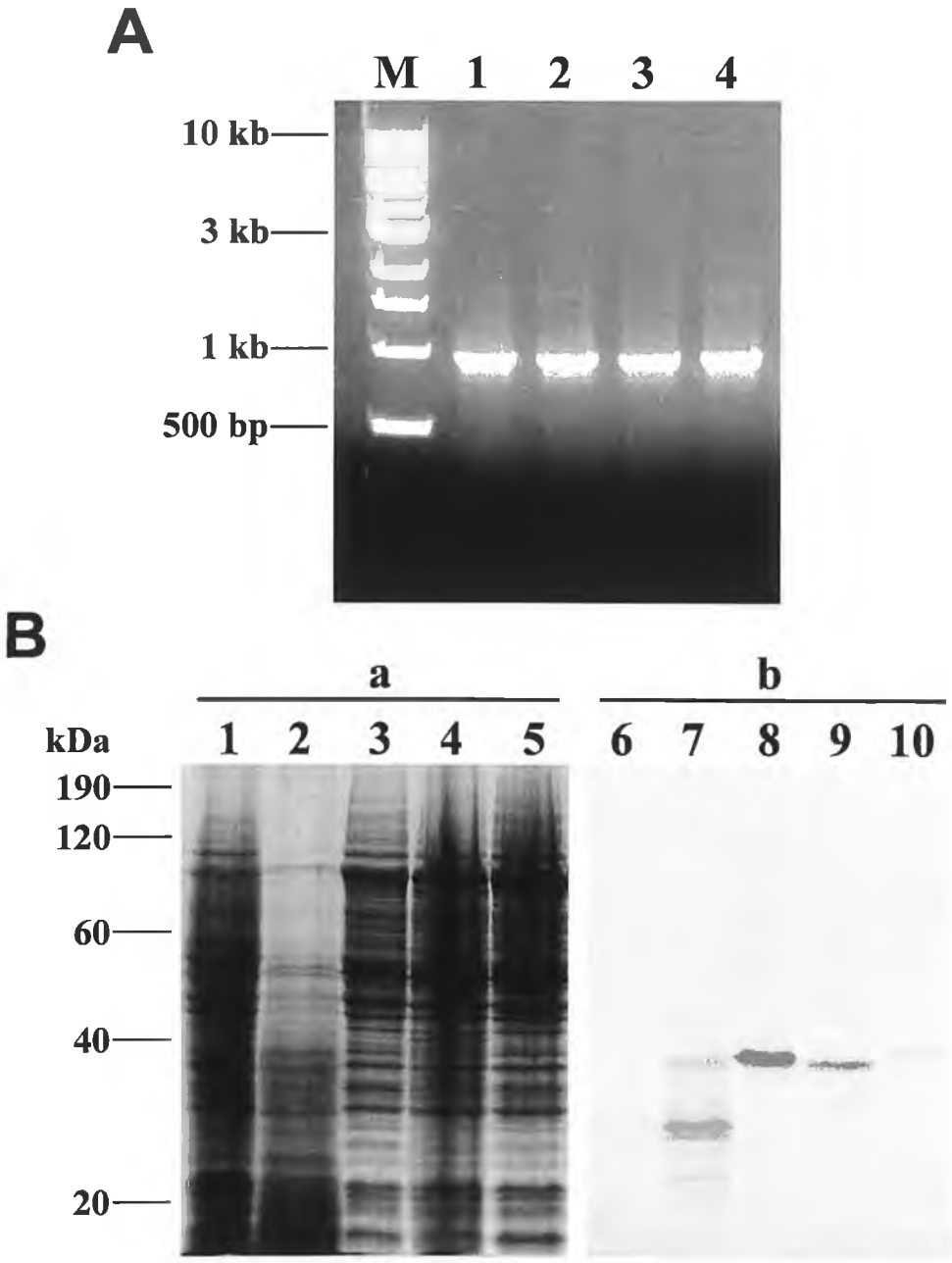


Figure 7.3: Analysis of mRNA production and intercellular expression by *P. pastoris*. *A*, total RNA was extracted from *P. pastoris* cells transformed with the various recombinants and induced for 4 days by 1% methanol in BMMY medium buffered to pH 6.0. First strand cDNA was synthesised for each sample by RT-PCR using 1 µg template. The cDNA was analysed by PCR with the FheCL1F and FheCL1R primers (see Table 2.1) and visualised on a 1% agarose gel. *Lane M*, 1 kb DNA ladder; *lane 1*, rFheproCL1; *lane 2*, r[Pro^{-12}Gly^{26}]FheproCL1; *lane 3*, rFheproCL2; *lane 4*, r[Gly^{26}]FheproCL2. *B*, similar cultures were induced for 3 days and cell lysates analysed. *Panel a*, 12% SDS-PAGE gel of cell lysates; *panel b*, immunoblot of cell lysates probed with mouse monoclonal anti-His_6 antibody; *lane M*, molecular size markers; *lanes 1 and 6*, non-induced control; *lanes 2 and 7*, rFheproCL1; *lanes 3 and 8*, r[Pro^{-12}Gly^{26}]FheproCL1; *lanes 4 and 9*, rFheproCL2; *lanes 5 and 10*, r[Gly^{26}]FheproCL2.

mature enzymes, prepared by activation for 24 h, were analysed by 10% GS-PAGE with an overnight incubation in 0.1 M sodium acetate buffers, pH 4.0, 5.0, 6.0, 7.0 and 7.75 in the presence of 1 mM DTT (see Figure 7.4). Of immediate note was the difference in separation between the two proteins. Under the non-reducing conditions of the gelatin substrate gel, the migration of mature cathepsin L1 was very much slower than that of mature cathepsin L2, with the former running about 1/3 of the way into the gel and the latter running almost all the way to the bottom. It was also noted that mature cathepsin L1 migrated as two distinct bands, normally unresolved on 12% SDS-PAGE gels. Mature cathepsin L1 showed strongest gelatinolytic activity at pH 4.0, with a decrease in activity with increasing pH. Mature cathepsin L2 showed strongest activity at pH 5.0, with good activity at all other pH values, including at the higher pHs of 7.0 and 7.75 where CL1 activity had greatly reduced.

7.4 Determination of pH optima of recombinant cathepsins L1 and L2

Next, pH activity profiles were generated by assaying the mature enzymes for activity against the fluorogenic substrate Z-FR-AMC in buffers at a range of pH values (see Figure 7.5). For both enzymes, a major peak of activity was found at pH 7.0, with a steady decrease in activity at higher pH values. In the case of mature cathepsin L1, a possible second peak was observed around pH 4.5. This peak may also be present to a lesser extent for mature cathepsin L2, although there was a clearer steady drop in activity below pH 7.0.

7.5 Substrate preference of recombinant wild-type cathepsins L1 and L2

Substrate preferences of the two mature recombinant enzymes were investigated by assaying them for activity against a variety of fluorogenic substrates, with particular emphasis on the P_2 position amino acid residues (see Figure 7.6). Both enzymes

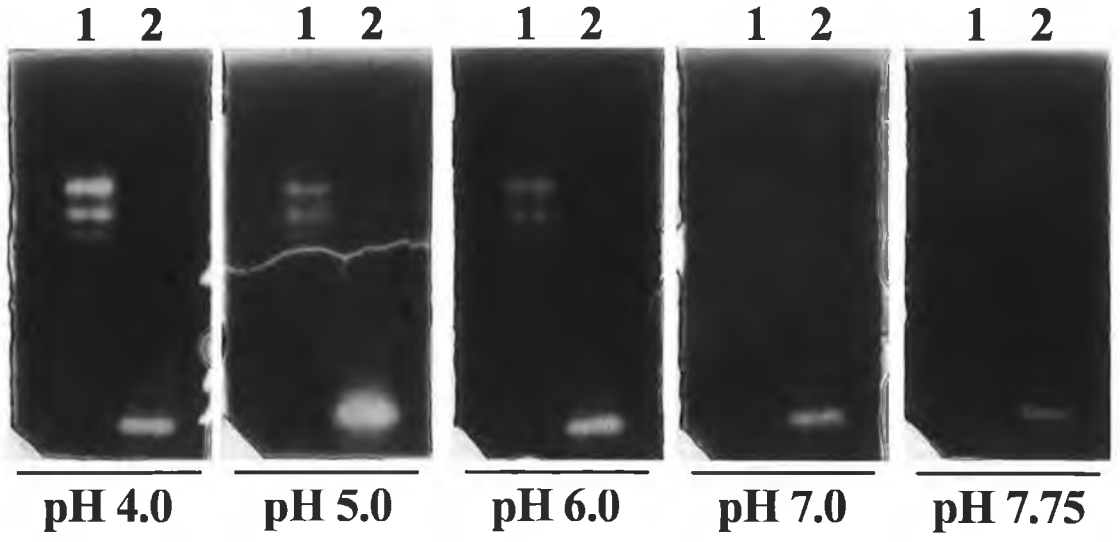


Figure 7.4: Gelatine gel zymography of recombinant wild-type cathepsins L1 and L2. Samples of 24 h activated recombinants were analysed by 10% GS-PAGE with overnight incubations at 37°C in 0.1 M sodium acetate buffers, pH 4.0, 5.0, 6.0, 7.0 and 7.75, containing 50 mM NaCl and 1 mM DTT. *Lanes 1*, mature recombinant cathepsin L1; *lanes 2*, mature recombinant cathepsin L2.

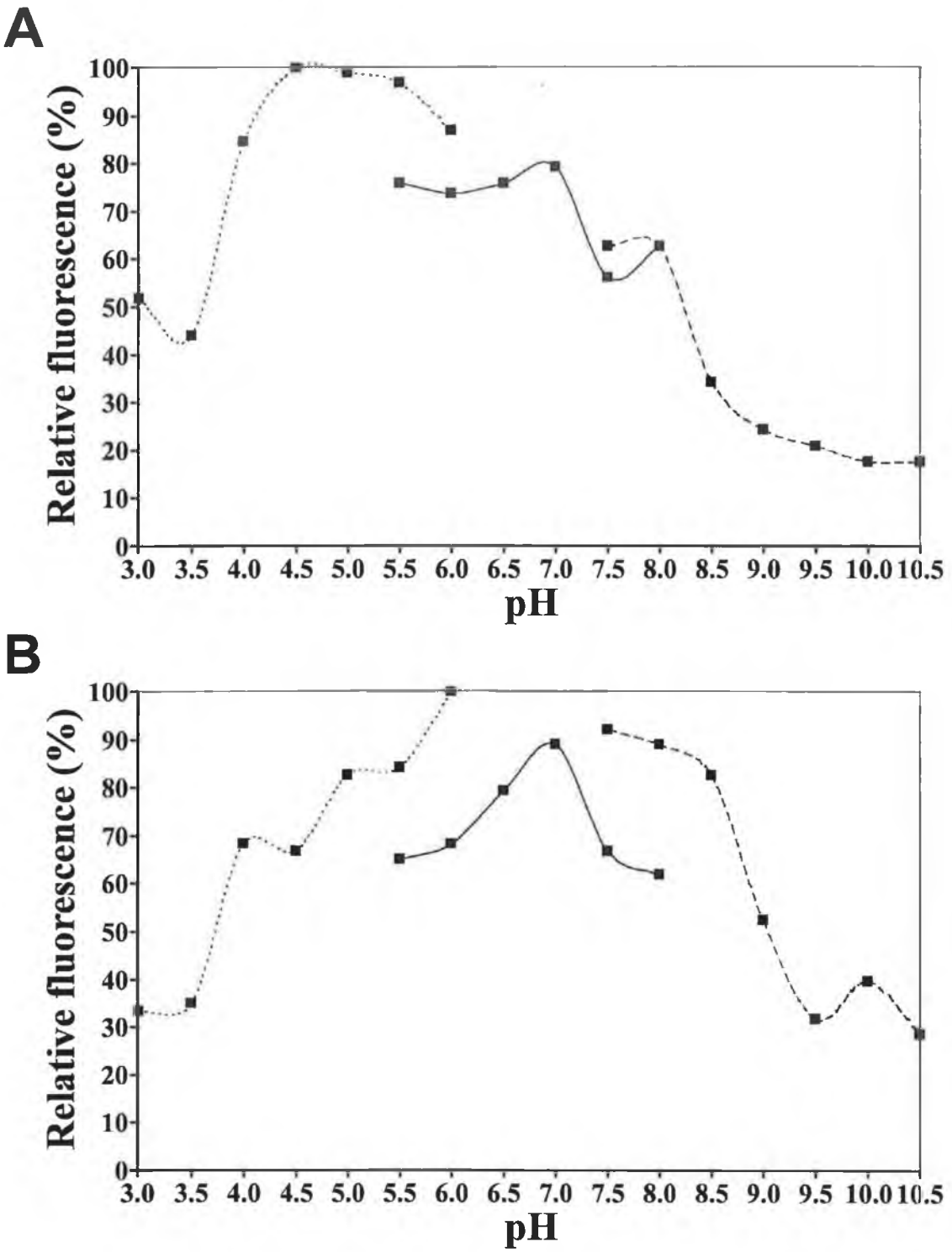


Figure 7.5: Activity profiles for recombinant proteins at varying pH values. 24 h activated recombinant cathepsin L1 (*A*) and cathepsin L2 (*B*) were assayed for activity against the fluorogenic substrate Z-FR-AMC in a range of buffers. *Dotted line*, 0.1 M sodium acetate buffers, pH 3.0 to 6.0; *solid line*, 0.1 M sodium phosphate buffers, pH 5.5 to 8.0; *dashed line*, 0.1 M tris/HCl buffers, pH 7.5 to 10.5.

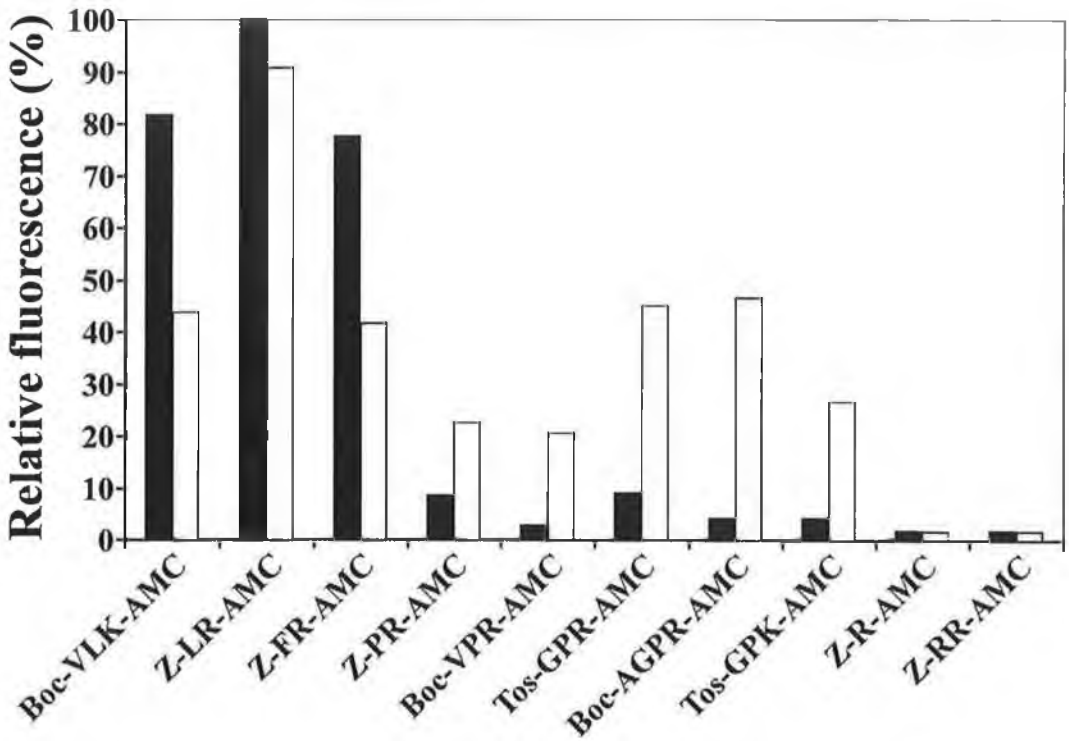


Figure 7.6: Substrate preference of recombinant wild-type cathepsins L1 and L2. Mature recombinant cathepsins L1 (*black bars*) and L2 (*white bars*) were assayed for activity against a variety of fluorogenic substrates: Boc-VLK-AMC, Z-LR-AMC, Z-FR-AMC, Z-PR-AMC, Boc-VPR-AMC, Tos-GPR-AMC, Boc-AGPR-AMC, Tos-GPK-AMC, Z-R-AMC and Z-RR-AMC. Cathepsin L2 is noted to have a much greater preference for proline in the P2 position than cathepsin L1.

showed strongest activity for substrates with leucine in the P_2 position, particularly when coupled with arginine in the P_1 position, as in the case of Z-LR-AMC. The next highest activity was against Z-FR-AMC, an activity characteristic of cathepsin L proteases. For peptides that included a proline in the P_2 position, a clear difference was seen between the two enzymes. While mature cathepsin L1 showed little or no activity against these substrates, mature cathepsin L2 showed good activity, ranging from between 20% and nearly 50% of the activity found against the most preferred substrate, Z-LR-AMC. Neither enzyme was active against the substrates Z-R-AMC or Z-RR-AMC as expected, again a characteristic feature of cathepsin L activity.

7.6 Collagen digestion by recombinant wild-type cathepsins L1 and L2

Activity of recombinant cathepsins L1 and L2 against interstitial materials was investigated by analysing their digestion of collagen as a representative structural protein. Collagen Types II and VI were incubated overnight in PBS, pH 7.3, with fully mature recombinant cathepsin L1 or L2 (previously activated for 24 h). Samples of the collagen Type II mixes were analysed by 12% SDS-PAGE (see Figure 7.7) and displayed a marked difference in collagenolytic activity between the two enzymes. Mature cathepsin L1 showed little activity against collagen Type II, while mature cathepsin L2 completely digested the collagen over the course of 24 h. In the case of collagen Type VI, samples were analysed by both 12% SDS-PAGE and on semi-native gels (see Figure 7.8). Both the enzymes showed similar patterns of degradation of this collagen, although mature cathepsin L2 progressed much further in this process over the course of 24 h.

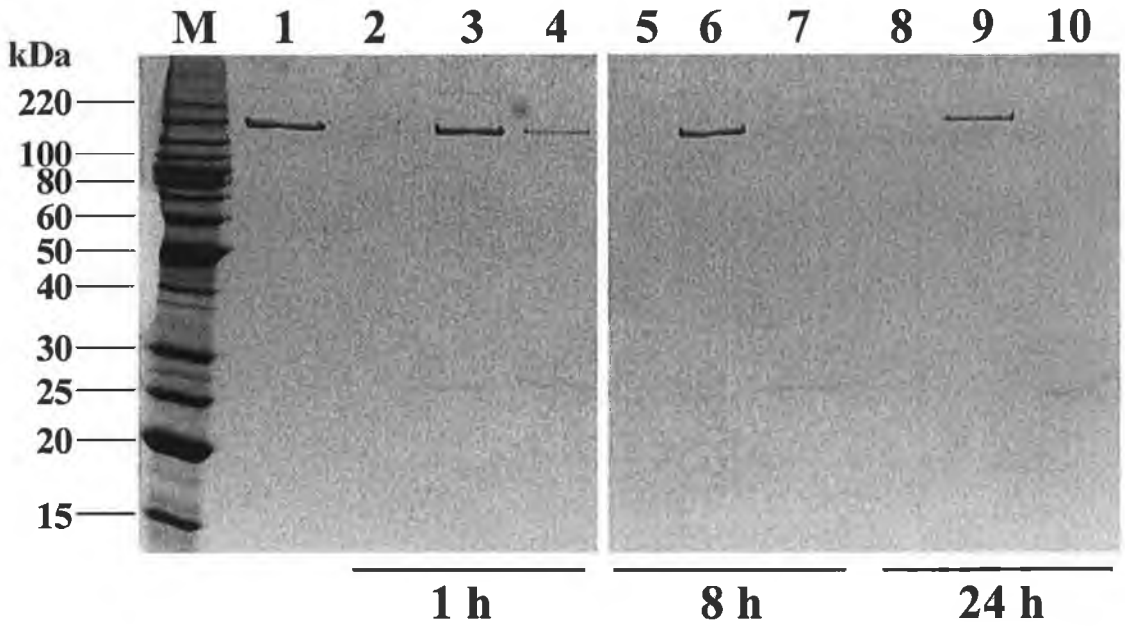


Figure 7.7: Digestion of collagen Type II by mature recombinant cathepsins L1 and L2. Samples of collagen Type II (bovine nasal septum; 15 μg) were incubated with 0.5 μg mature recombinant enzyme or collagenase in PBS, pH 7.3, overnight at 37°C. Samples taken at various time points were analysed by 12% SDS-PAGE. *Lane M*, molecular size markers; *lane 1*, collagen Type II in PBS as negative control; *lanes 2, 5 and 8*, collagen Type II digested with collagenase blend; *lanes 3, 6 and 9*, collagen Type II digested with mature cathepsin L1; *lanes 4, 7 and 10*, collagen Type II digested with mature cathepsin L2.

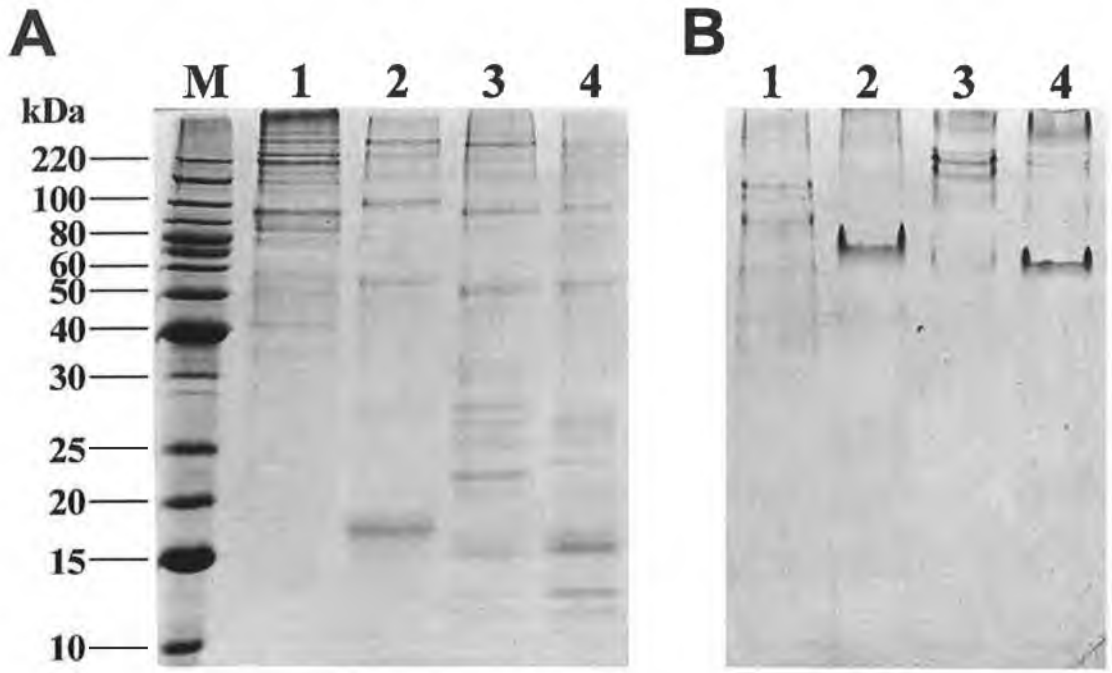


Figure 7.8: Digestion of collagen Type VI by mature recombinant cathepsins L1 and L2. Samples of collagen Type VI (basement membrane; 15 µg) were incubated with 0.5 µg mature recombinant enzyme or collagenase in PBS, pH 7.3, overnight at 37°C. Samples were analysed by 12% SDS-PAGE (*A*) and on semi-native gels (*B*). *Lane M*, molecular size markers; *lane 1*, collagen Type VI in PBS as negative control; *lane 2*, collagen Type VI digested with collagenase blend; *lane 3*, collagen Type VI digested with mature cathepsin L1; *lane 4*, collagen Type VI digested with mature cathepsin L2.

7.7 Exogenous processing of inactive mutant cathepsin L1s by activated wild-type cathepsins L1 and L2

To further investigate the blocking of exogenous processing by the $Leu^{-12}\rightarrow Pro^{-12}$ in the double mutant procathepsin L1, both mutant proteins were "cross-processed" by addition of 24 h activated rFheproCL2 in parallel with incubation with activated rFheproCL1 as in previous experiments (see Figure 7.9). In this case it was found that rFheproCL2, which has a high substrate preference for proline in the P_2 position not present in cathepsin L1, was capable of processing the double mutant to a ~25 kDa form co-migrating with that of the single mutant [Gly^{26}]FheproCL1. It was also noted that an equal concentration of activated rFheproCL2 was capable of more fully processing [Gly26]FheproCL1 over the course of the 3 h incubation than activated rFheproCL1.

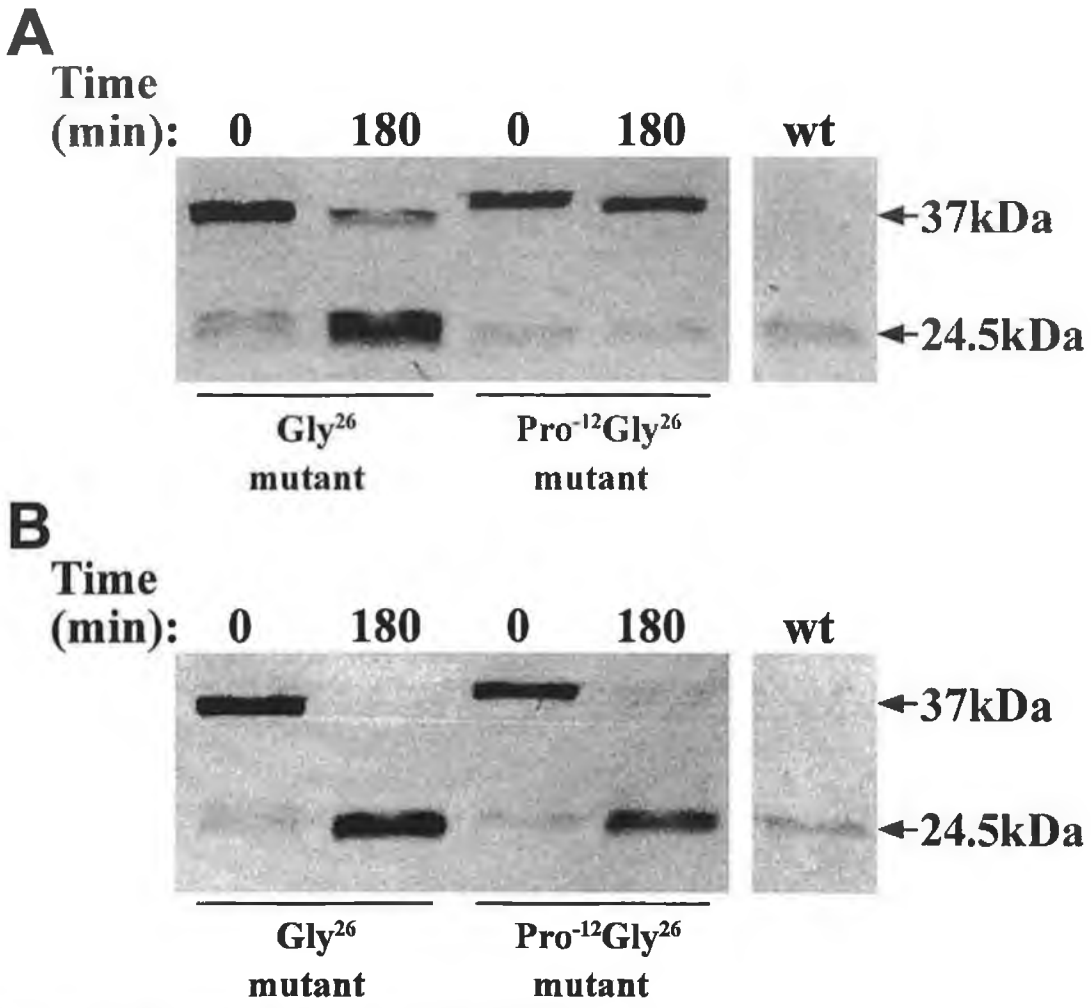


Figure 7.9: Exogenous processing of procathepsin L1 inactive mutants by active cathepsins L1 and L2. Purified r[Gly^{26}]FheproCL1 and r[$Pro^{-12}Gly^{26}$]FheproCL1 were incubated for 3 h in 0.1 M sodium acetate buffer, pH 4.5, containing 2 mM DTT and 2.5 mM EDTA at 37°C in the presence of either 24 h activated rFheproCL1 (*A*) or rFheproCL2 (*B*). Mature CL2, which has a high substrate preference for proline in the P_2 position, is capable of processing the r[$Pro^{-12}Gly^{26}$]FheproCL1 double mutant which remains unprocessed by mature cathepsin L1.

Chapter 8: Discussion

8.1 *F. hepatica* procathepsin L1 and L2 and related cysteine proteases

Comparison of the two major *F. hepatica* cysteine proteases, cathepsins L1 and L2, at the amino acid sequence level indicates a very close relationship between the two. The similarity of the amino acid sequences is very high at 92.63%, with the identity also high at 78.22% (see Section 3.1; also Table 3.3). Clearly, a significant commonality in function of the two enzymes within the parasite is likely. This high level of sequence identity is also responsible for cross reactivity of the enzymes with antiserum prepared against each other (see Figures 4.3 and 7.1).

A phylogenetic analysis of cathepsin L sequences from both *F. hepatica* and *F. gigantica* indicated that the sequences fall into 4 distinct clades. Of particular importance was that while three of these clades contained only sequences of cDNAs obtained from adult parasites, the fourth only contained sequences derived from metacercarial or newly excysted juvenile stage flukes (see Figure 3.2). It is therefore clear that there are stage-specific cathepsin Ls produced in both *F. hepatica* and *F. gigantica*. These data are supported by a previous phylogenetic analysis performed by Irving et al. (2003) in which the *Fasciola* sequences could be similarly divided into 4 clades. However, the addition of further cathepsin L cDNA sequences to the analysis retains this division of three adult and one metacercarial group. Only 2 largely incomplete sequences and one complete sequence out of 31 total cDNAs did not fit easily into the groupings; in the former case, the data is insufficient to allow a reliable alignment; in the latter, the sequence may represent an intermediate between two of the adult groups.

Analysis of amino acid sequence alignments of the cathepsin L sequences within each group showed some characteristic features. Firstly, while the majority of the

propeptide sequence matched closely for all enzymes, the *C*-terminal portion from residue -21 up to and including the 5th amino acid of the mature region showed distinct variation between groups (see Table 3.1). Most notably here, the Metacercarial Group proteins were unique in possessing an aspartate residue at position 1 of the mature region, contrasting with the alanine found in all three adult groups. This feature was also common to both *F. hepatica* and *F. gigantica* and may represent a characteristic signature of such early stage cathepsin Ls. Secondly, residues which have been implicated in subsite interactions by Turk et al. (1998) also display the grouping patterns between the different clades (see Figure 3.2). Three of these residues, at positions 70, 162, and 164, form part of the S_2 pocket and are key to the determination of substrate specificity in cathepsin L (Smooker et al., 2000). These three residues are quite similar in Adult Groups 1 and 2, both of which contain Leu^{70} and Asn^{162} while Adult Group 1 contains Ala^{164} in contrast to Adult Group 2's Gly^{164}. Adult Group 3 and the Metacercarial group share Thr^{162} and Ala^{164}, while Adult Group 3 contains Tyr^{70} in place of the Trp^{70} present in the Metacercarial Group. As can be seen here, the least variation is found between groups at position 163 while the greatest is at position 69. This later position is of significance as variation here directly specificity of the S_2 pocket; Smooker et al. (2000) demonstrated that mutation at this location from leucine to tyrosine in *F. hepatica* cathepsin L5 enhanced its activity against the substrate Tos-GPR-AMC 10-fold. This increased preference for proline in the P_2 position has been shown previously (Dowd et al., 1994) and confirmed in this project (see Section 8.7) to be found in *F. hepatica* cathepsin L2 (an Adult Group 3 protease) but not in cathepsin L1 (an Adult Group 1 protease). As cathepsin L5 belongs to Adult Group 2, it is possible that the Adult Group 1 and Group 2 proteases share substrate preferences to a certain degree, and mutation of $Leu^{70} \rightarrow Tyr^{70}$ of FheCL1 may also introduce the ability to cleave proteins with proline in the P_2 position as has been shown for FheCL5. It is

also likely that all members of the Adult Group 3 may share this broader substrate preference of FheCL2. The presence of Trp^{70} in the Metacercarial Group may allow an even broader range of residues to bind in the S_2 pocket, allowing cleavage of a wider variety of proteins by the NEJ. Irving et al. (2003) demonstrate that it is the divergence between the Metacercarial Group and the 3 Adult Groups which would have occurred earliest, followed by the divergence of Adult Groups 1 and 2 from Group 3, and finally the divergence of Adult Group 1 from Group 2. This data would support the idea that the *Fasciola* cathepsin Ls started with a broad specificity and have steadily evolved tighter specificities over time; the groups would also reflect an increasing specificity of function as the parasite evolved alongside new host species.

Alignment of *F. hepatica* pre-procathepsin L1 and L2 sequences with selected helminth and mammalian cathepsin Ls, along with papain, demonstrate that the proteins have closest links with the parasitic helminth, *S. mansoni*, followed by the free living nematode, *C. elegans*. They also match quite closely with the mammalian cathepsin Ls. These observations follow the pattern expected as the helminth cathepsin L lineages eventually gave rise to the mammalian cathepsins L, S and K (Dalton et al., 2004a). The *Fasciola* enzymes are clearly true cathepsin Ls and represent a monophyletic group of enzymes in the parasite.

8.2 Expression of wildtype procathepsins L1 and L2

In order to study the functional similarities and differences between *F. hepatica* cathepsins L1 and L2 the recombinant proenzymes were expressed in the methylotrophic yeast, *Pichia pastoris*. Both enzymes had been previously expressed in our laboratory from full length pre-proenzyme sequences cloned into *Saccharomyces cerevisiae* (Roche et al., 1997; Dowd et al., 1997). However, it was decided to transfer these cDNAs into the *P. pastoris* system due to the benefits potentially achievable over

the previous system; *P. pastoris* is capable of producing very high quantities of recombinant protein in a secreted form; it can be grown to very high levels of biomass, up to 50% of the total culture volume, without losses in recombinant production; finally, the organism secretes very low quantities of background native yeast proteins, reducing the need for multiple purification steps to obtain pure recombinant protein (Cereghino and Cregg, 2000; Brady et al., 2001; Aoki et al., 2003). With the presence of the yeast α-factor signal sequence within the chosen pPIC9K vector, only the cDNAs representing the proenzymes of FheCL1 and FheCL2 were amplified from the original *S. cerevisiae* vectors and transferred to the new system, rather than the full pre-proenzyme sequences. The incorporation of a *C*-terminal His_6-tag sequence to each cDNA also allowed for a simple one-step purification of the secreted recombinant proteins from the yeast medium by Ni-NTA affinity chromatography.

After the construction of the pPIC9K.FheproCL1 expression vector and its transformation into *E. coli* cells, samples of plasmid DNA from two clones confirmed to contain the FheproCL1 cDNA insert were sequenced and both found to be correctly in frame with the α-factor signal sequence. However, when the sequence was compared by alignment with the original pre-procathepsin L1 sequence deposited by our laboratory in the public database (GenBank sequence, U62288; Roche et al., 1997), several differences were found between the two. While a number of these changes were silent with respect to the database sequence, six were reflected in the amino acid sequence (see Section 3.2; also Figure 3.1). On replacement of the database sequence with the new sequence in the sequence analysis of *F. hepatica* and *F. gigantica* cathepsin Ls, the residue changes matched more closely with the related enzymes than the original sequence. As both clones also showed the same changes relative to the database in several parallel sequencing runs, it was assumed that the new sequence reflects a more accurate picture of the enzyme; errors may have been made in the

original sequencing procedure which have now been corrected with the better sequencing technologies available at this time.

Once three *P. pastoris* colonies transformed with the pPIC9K.FheproCL1 were selected, pilot expression studies were carried out, with yeast cells induced for 6 days by 1% methanol in BMMY medium buffered to pH 6.0, followed by SDS-PAGE analysis of the culture supernatants. All three cultures produced similar levels of recombinant proteins which migrated as two major bands of 37 and 30 kDa (see Figure 4.1). A final expression clone was selected from these three and the induction repeated, this time with the BMMY medium buffered at pH 6.0, 7.0, or 8.0. Again, two major proteins were secreted into the medium at 37 and 30 kDa (see Figure 4.2). Both components migrate higher than either the mature cathepsin L1 or cathepsin L2 secreted by the adult *F. hepatica* parasites when the supernatants were compared with adult parasite excretory-secretory (ES) products on 12% SDS-PAGE (see Figure 4.3). However, immunoblotting experiments in which both the recombinant supernatant and adult ES products were probed with polyclonal antisera prepared against the propeptide portion and the mature portion of procathepsin L1 showed that both react with the 37 and 30 kDa components expressed in *P. pastoris*, confirming that both components are indeed two forms of procathepsin L. By contrast to this, the fully processed native enzymes secreted by the adult parasite only showed reactivity to the antiserum prepared against the mature portion of procathepsin L1, with no reactivity with the anti-propeptide antiserum (see Figure 4.3).

After purification of the two rFheproCL1 components from yeast medium, *N*-terminal sequencing analysis of each band revealed that the 37 kDa protein represented the full procathepsin L1 (plus the last 6 amino acids of the yeast α-factor secretory signal sequence) while the 30 kDa component represented an intermediate form beginning at residue Leu^{-38} of the propeptide. These results also confirmed that the

molecular weights of the two components as previously estimated from the SDS-PAGE gel matched closely to the predicted average molecular weights of 36.74 and 29.62 kDa respectively. It could be assumed, therefore, that no post-translational modifications of the enzyme had occurred which may have caused the proteins to migrate more slowly within the gel; this was as expected due to the lack of glycosylation sites identified in the FheproCL1 sequence.

The partial processing of the rFheproCL1 to the intermediate form represented by the 30 kDa band was also clearly effected by the pH of the BMMY medium in which recombinant protein expression was induced; while the amount of recombinant protein produced in medium buffered to pH 6.0, 7.0, and 8.0 increased over the course of the induction to roughly the same levels, the relative ratio of the 37 and 30 kDa forms showed dependence on medium pH (see Figure 4.2). As medium pH increased, the proportion of unprocessed rFheproCL1 represented by the 37 kDa band also increased. A pH profile of activity for the recombinant cathepsin L1 (see Section 8.7) showed that it was similar to that of the native mature enzyme (Dowd et al., 1994; Dowd et al., 2000); the enzyme is active over a pH range of 4.0 to 8.5, with a major peak of activity between pH 6.5 and 7.0. Therefore, the increase in the pH of the yeast medium beyond the pH optimum for FheCL1 activity (*i.e.* to pH 8.0) reduced the level of processing of the rFheproCL1 to the intermediate 30 kDa form, indicating some level of autocatalytic involvement in this processing.

In order to further address this issue, yeast cells transformed with pPIC9K.FheproCL1 were induced at pH 8.0 in the presence and absence of the potent cathepsin L-specific inhibitor, Z-Phe-Ala-diazomethylketone (see Figure 4.4). The inhibitor was added twice daily at a concentration of 25 μM, a concentration which is 250-fold greater than its K_i for cathepsin L (Dowd et al., 1994; Higgins and Cregg, 1998). The proportion of the full length rFheproCL1 produced represented by the 37

kDa band was considerably greater than that in uninhibited cultures, although the partially processed 30 kDa intermediate form was still evident. These data support the idea that rFheproCL1 secreted from *P. pastoris* autoprocesses to an intermediate active form and that this autoprocessing is prevented by the inhibition of cathepsin L activity by the potent inhibitor Z-FA-CHN_2.

The cathepsin L proteases of *F. hepatica* are active within the gut of the organism, where the environmental pH is estimated to be quite low, around pH 5.5 or slightly lower (Halton, 1997). As FheCL1 maintains activity at this lower pH range and the previous inductions showed greater levels of processing from the full 37 kDa form of rFheproCL1 to the 30 kDa intermediate form at pH 6.0 than at pH 7.0 or 8.0, a further induction of yeast cells at the lower pH of 4.5 as well as at pH 6.0 and 8.0 was carried out (see Figure 4.5A). In this case, reduction of medium pH to 4.5 caused almost complete elimination of the full 37 kDa component, indicating almost complete processing of the recombinant procathepsin L1 to the 30 kDa intermediate form at low pH. It was also noted that a third, lower band was visible at pH 6.0 and pH 4.5. This band was more intense in the pH 4.5 sample and may represent either a further intermediate or the fully processed mature enzyme. These data confirmed the overall pattern of increasing levels of autoprocessing of rFheproCL1 with decreasing pH.

Reductions in the temperature at which the yeast cultures were induced to 25°C and 20°C instead of the normal 30°C showed no effect on processing of the recombinant enzyme, although the clarity of the cultures was increased and the levels of background proteins and degradative products present decreased with decreasing temperature, especially in the presence of the Z-FA-CHN_2 inhibitor (see Figure 4.5B). However, lower quantities of recombinant protein were produced at the lower temperatures.

In order to further characterise and compare the recombinant *F. hepatica* cathepsin L1 to the native enzyme, activity of the enzyme against the fluorogenic substrate Z-FR-AMC in the presence of varying levels of either the cysteine protease inhibitor Z-FA-CHN_2 or a recombinant form of the FheCL1 propeptide was assayed. Greater levels of inhibition occurred with the recombinant propeptide at lower concentrations than with Z-FA-CHN_2 and at varying pH values (see Figure 4.9), confirming the high specificity of the propeptide for cathepsin L1.

The cDNA encoding procathepsin L2 was transferred to the *P. pastoris* expression system in a similar manner to procathepsin L1 and confirmed to match the expected sequence (see Figure 3.4). Expression was induced in two pPIC9K.FheproCL2 transformed yeast clones in parallel with the pPIC9K.FheproCL1 clone by 1% methanol in medium buffered to pH 6.0. Unlike rFheproCL1, rFheproCL2 was produced almost entirely in the full length procathepsin L2 form migrating at 37 kDa in SDS-PAGE (see Figure 5.1A). Very little of the protein (only faint minor bands) was processed to lower forms such as the 30 kDa intermediate of rFheproCL1. Assays of supernatant samples against the fluorogenic substrate Z-FR-AMC showed much lower activity for rFheproCL2 supernatants compared with those containing rFheproCL2 (see Figure 5.1B). However, as the enzyme was almost entirely in the zymogen form and the activity of native cathepsin L2 against Z-FR-AMC is lower than that of cathepsin L1 (Dowd et al., 1994), these results were consistent. Yeast cultures induced in medium buffered at pH 4.5, 6.0 and 8.0 produced roughly equal amounts of rFheproCL2, although with a slight increase in the level of processing observed with decreasing pH (see Figure 5.2). While this observation matches the pattern seen for the production of recombinant procathepsin L1 at these pH values, it should be noted that the level of processing is still far lower for rFheproCL2 and the vast majority of the protein remains in the 37 kDa procathepsin form at all three pH values. This difference

suggests that the recombinant wildtype procathepsin L2 is more stable in yeast medium. It is tempting to speculate that, as cathepsin L1 is a member of the Adult Group 1 *Fasciola* cathepsin Ls while cathepsin L2 is a member of the Adult Group 3 (see Section 8.1), differences in the propeptide sequences between these two groups may account for different stabilities of the zymogen forms and/or differential processing of the two recombinant enzymes in yeast medium. However, it is also possible that these observed differences are due to the effect of components of the yeast proteome during trafficking of the recombinants for secretion from the yeast cell.

8.3 Processing of wildtype procathepsins L1 and L2

To further examine this intermolecular processing at pH values lower than 6.0, the 37 kDa and 30 kDa rFheproCL1 components were purified from a pH 8.0 yeast culture supernatant by Ni-NTA affinity chromatography and dialysed against PBS, pH 7.3. Samples of this purified protein preparation were then reduced to pH 5.0 by addition of 0.1 M sodium citrate buffer containing 1 mM DTT and 1.25 mM EDTA and incubated at 37°C for 2 h. Samples taken at various time points during the course of the incubation were analysed by SDS-PAGE. Over the course of the incubation, the 37 kDa procathepsin L1 was processed to the 30 kDa intermediate form and then through various further intermediates to eventually give rise to a single peptide migrating at 24.5 kDa (see Figure 4.6). This corresponded with a 5-fold increase in the enzyme's activity against the fluorogenic substrate Z-FR-AMC.

N-terminal sequencing analysis of the 24.5 kDa band revealed sequences of NRAVP and RAVPD, corresponding to the sequence starting just 2 and 1 amino acid residues from the start, respectively, and indicated that the 24.5 kDa band represented the fully mature cathepsin L1 enzyme (see Figure 4.6). These results were compared with a control experiment in which the pH was not reduced, with the 0.1 M sodium

citrate buffer replaced by PBS, pH 7.3. Here, the 37 kDa procathepsin L1 was not processed at all, even to the 30 kDa intermediate form (see Figure 4.7). There was some breakdown of the 30 kDa intermediate form present in the preparation but no processing to the fully mature form. These data indicate not only the stability of the zymogen form of the enzyme at neutral pH but also the pH dependence of the autocatalytic processing of the proenzyme.

Similar experiments performed using purified recombinant procathepsin L2 showed a pattern of autocatalytic processing to the mature active form of the enzyme closely matching that of procathepsin L1. Samples of purified protein were reduced to pH 4.5 by addition of 0.1 M sodium acetate buffer containing 1 mM DTT and 1.25 mM EDTA and incubated at 37°C for 2 h. Samples taken at various time points during the course of the incubation were analysed by SDS-PAGE. Over the course of the incubation, the 37 kDa procathepsin L2 was processed through various intermediates to eventually give rise to a single peptide migrating at 24.5 kDa (see Figure 5.3A), representing the fully mature active enzyme. This again corresponded with a 5-fold increase in the enzyme's activity against the fluorogenic substrate Z-FR-AMC (see Figure 5.3B). A control experiment in which the pH was not reduced, with the 0.1 M sodium acetate buffer replaced by PBS, pH 7.3 showed no processing of the zymogen to the mature form under these conditions (see Figure 5.4); this again matches the observations for the processing of procathepsin L1.

The cleavage site between Gly^{-39} and Leu^{-38} that generated the 30 kDa intermediate form of rFheCL1 was of particular interest as it occurred within the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} (papain numbering) motif which was previously found to play a pivotal role in the intermolecular processing of papain expressed in *S. cerevisiae*. Vernet et al. (1995) showed by random mutagenesis studies and by sequence comparisons of various papain-like cysteine proteases that the Gly^{-42} and Asp^{-36} (papain

numbering) residues within this G*X*N*X*F*X*D motif were the most constrained and proposed that two cleavage sites between Gly^{-42}-Leu^{-41} and between Ala^{-37}-Asp^{-36} were of importance in proenzyme processing. The group considered the latter cleavage site to be more significant, as it was predicted to perturb the negative charge of the Asp^{-36} residue (papain numbering), which is absolutely conserved in all non-cathepsin B-like cysteine proteases of helminth parasites, plants and mammals, including cathepsins L, S and K (Vernet et al., 1995; Tort et al., 1999). Studies of the crystal structure of human cathepsin L shows that this residue participates in an important salt bridge between the propeptide and the mature enzyme (Coulombe et al., 1996). In the case of both *F. hepatica* cathepsins L1 and L2, the G*X*N*X*F*X*D motif corresponds to the sequence Gly^{-39}-Leu-Asn-Gln-Phe-Thr-Asp^{-33} (FheCL1 numbering), a sequence which is completely conserved in all *F. hepatica* and *F. gigantica* cathepsin L sequences with the exception of two members of the Adult Group 1 (see Section 8.1; also Figure 4.8); only FheCL6 and FgiCL4 (see Table 2.2 for sequence details) differ, with Phe^{-35} replaced with Leu^{-35}.

The Leu^{-41} (papain numbering) residue is also conserved in another helminth parasite cathepsin L sequence, that of *S. mansoni* CL2, as well as in that of the free-living helminth, *C. elegans* (see Figure 4.8). Thus, the Gly^{-42}-Leu^{-41} (papain numbering) bond is preserved across these sequences and this implicates its importance in propeptide function in these helminths. However, these two residues are not conserved in mammalian sequences such as the cathepsin Ls of humans, mice and rats, where the glycine residue is replaced by either glutamate or alanine and the leucine residue is replaced by methionine (see Figure 4.8). This could suggest that these differences explain why an intermediate 30 kDa form of human cathepsin L was not observed in studies of its intermolecular processing (Ménard et al., 1998). However, this is unlikely as McQueney et al. (1997) showed that a cleavage between Ala^{-42} and Met^{-41} (papain numbering) was involved in the intermolecular processing of an inactive

mutant form of human procathepsin K by activated human cathepsin K. Therefore (as has been pointed out by Coulombe et al., (1996), while there may be a direct involvement of the final aspartate residue in the G*X*N*X*F*X*D motif in the pH-dependant processing of proforms of papain-like cysteine proteases, cleavage at other peptide bonds within this motif, such as is the case here in rFheproCL1, could influence the local charge state triggering processing through to the mature enzyme form.

8.4 Expression of inactive mutant procathepsin L1s

In order to further study the intermolecular processing of the *F. hepatica* cathepsin Ls, an inactive mutant form of FheproCL1 was prepared and expressed in the *P. pastoris* expression system. The activity of the enzyme was removed by replacement of the active site cysteine at position 26 with a glycine. Pilot expression studies of two transformed yeast colonies in medium buffered at pH 6.0 and 8.0 showed the recombinant protein secreted migrates as a single strong band in SDS-PAGE for both transformants and was unaffected by the pH of the medium. The inactive recombinant proteins were also produced in almost 4-fold greater quantities than the wildtype recombinant procathepsin L1 in parallel inductions (see Figure 6.1). However, it was noted that while the protein produced by one clone co-migrated with the 37 kDa component of rFheproCL1, protein produced by the second clone migrated slightly higher in the gel (see Figure 6.1). Results of *N*-terminal sequencing analysis of the two proteins showed that both represented the complete procathepsin L1 sequence (with the additional 6 amino acids of the yeast α-factor secretion signal sequence as found for the 37 kDa component of rFheproCL1). However, sequencing of the cDNA of each clone indicated that the second, more slowly migrating mutant contained a mutation within the propeptide, resulting in the replacement of a leucine at position -12 with a proline. It is presumably this mutation which leads to the protein's slower migration, possibly by

slightly altering the 3D structure of the propeptide. The double mutant r[Pro^{-12}Gly^{26}]FheproCL1 was included in further experiments to ascertain if this mutation within the propeptide would have any effect on exogenous processing of the protein. As expected, neither inactive recombinant displayed activity when assayed against the fluorogenic substrate Z-FR-AMC.

Further insight into the differences and similarities of the *F. hepatica* cathepsin Ls could be obtained by the elucidation of the 3D structure of one or more of the enzymes. As the recombinant inactive procathepsin L1 mutant, r[Gly^{26}]FheproCL1, retains the full length propeptide, a large scale induction and purification of the protein was prepared in collaboration with Dr. Colin Stack in our laboratory. The purified protein was sent to Dr. Linda Brinen of the *Fasciola* Structural Biology Laboratory at UCSF, San Francisco, USA, for the determination of the 3D structure by X-Ray crystallography. A preliminary model (see Figure 6.10), demonstrates that the general structure compares well with that of the other cathepsin Ls and cathepsin L-like cysteine proteases; once completed, more detailed analysis of the protein structure will be possible.

A third inactive mutant, this time replacing the Cys^{26} of procathepsin L2 with Gly^{26}, was prepared and expressed in the *P. pastoris* system. Although the protein is produced as a single band at around 37 kDa as expected for the full length procathepsin L2 form, it is produced in much lower quantities when compared with the previously prepared recombinant proteins in parallel expression cultures (see Figure 6.8). This is possibly due to integration of fewer copies of the cDNA into the *P. pastoris* genome, although analysis of the mRNA produced by the yeast during induction does not support this (see Section 8.7). The results of *N*-terminal sequencing analysis of the expressed protein again confirmed that the band represented the full procathepsin L2 sequence, with the now expected addition of the last 6 amino acid residues of the yeast α-factor

secretion signal sequence. Lack of activity of the protein was confirmed in assay against the fluorogenic substrate Z-FR-AMC. Attempts were made to produce the protein at a large enough scale and high purity for the determination of the crystal structure but were unsuccessful; the larger quantities of culture needed to produce this protein compared with the previous recombinants greatly increased the amount of downstream processing required, including the addition of a concentration step; the extra time and handling involved may have caused the partial degradation of the protein, reducing its suitability for X-Ray crystallography. However, further work is ongoing in the laboratory to obtain clones which express r[Gly^{26}]FheproCL2 in higher quantities more comparable with the other recombinants.

8.5 Exogenous processing of inactive mutant procathepsin L1s

When incubated in parallel with rFheproCL1 at pH 5.0, neither the inactive mutant r[Gly^{26}]FheproCL1 nor the double mutant r[$Pro^{-12}Gly^{26}$]FheproCL1 processed to lower molecular size forms (see Figure 6.2). Experiments were performed in which each of the purified inactive recombinant proteins were mixed with activated mature rFheCL1. In the case of r[Gly^{26}]FheproCL1, addition of activated rFheCL1 at pH 5.0 resulted in the progressive appearance of a minor band at around 35 kDa along with a major band at around 25 kDa that co-migrates with fully active cathepsin L1 (see Figure 6.3A) over a 4 h incubation period. In contrast to this, similarly prepared double mutant r[$Pro^{-12}Gly^{26}$]FheproCL1 showed no substantial processing, only slightly shifting over the course of the incubation to more closely match the migration of the 37 kDa component of wildtype procathepsin L1 (see Figure 6.3B). These patterns were unaffected by lowering the pH of the incubation mix to 4.0 or 3.0 (see Figure 6.3C). Furthermore, experiments in which activated wildtype cathepsin L1 was added to heat-denatured mutant procathepsin L1 showed that the latter was sensitive to complete

degradation by active cathepsin L1 and confirms that the expressed mutant is correctly folded (see Figure 6.4).

N-terminal sequencing of the components produced shed further light on the exogenous processing of the inactive mutants. While no result was obtainable for the ~35 kDa component arising from r[Gly^{26}]FheproCL1, the lower ~25 kDa component had a sequence of HGVPY (with less than 10% GVPYE), placing the cleavage site 10 amino acid residues prior to the *N*-terminus of mature cathepsin L1. In the case of the double mutant r[Pro^{-12}Gly^{26}]FheproCL1, a mixture of several sequences was found corresponding to the early residues of the propeptide (see Figure 6.5). This data indicates that while the exogenously added wildtype cathepsin L1 was incapable of processing the double mutant, it was capable of sequentially degrading the mutant from the *N*-terminus, removing the remains of the yeast α-factor secretion signal sequence as well as the first few residues of the propeptide. A low preference for proline in the P_2 position may be responsible for this lack of cleavage (see Section 8.7).

An overview of the cleavage sites observed in all the processing experiments involving both autoactivation of wildtype procathepsin L1 and the intermolecular processing of the inactive mutants by exogenously added active cathepsin L1, along with those previously determined for the native mature cathepsin L1 secreted by *F. hepatica* parasites *in vitro* is shown in Figure 6.5. It is clear from these data that autoactivation and intermolecular activation occur by cleavages at different sites and thus processing of procathepsin L1 to an active enzyme may be achieved by more than one pathway. It should be noted that the location of the Leu→Pro mutation at position -12 in the double mutant is in close proximity to the cleavage point (Ser^{-11}-His^{-10}) which occurs when the single mutant r[Gly26]FheproCL1 is exogenously processed by active wildtype cathepsin L1. Further to this, the portion of the propeptide immediately prior to and including Leu^{-12} corresponds in alignments to the location in human procathepsin

L of a short two-stranded antiparallel β-sheet formed with the mature region (between Lys^{-14}-Phe^{-12} and Phe^{112}-Asp^{114}, human cathepsin L numbering), stabilising the propeptide as it leaves the active site cleft (Coulombe et al., 1996). This region of the propeptide may therefore be a target for intermolecular processing.

Also of particular note is the absence of a cleavage site within the G*X*N*X*F*X*D motif that would generate the 30 kDa intermediate component by exogenous activation of r[Gly^{26}]FheproCL1. As has been demonstrated already (see Section 8.3), this cleavage site is dominant in yeast-expressed rFheproCL1 and the strict conservation of the motif suggests an important role in the processing of the procathepsin Ls of *F. hepatica*. Vernet et al. (1995) demonstrated the importance of this motif in papain processing by using various mutations within the G*X*N*X*F*X*D sequence; certain residue replacements prevented correct folding of the enzyme, preventing activity, while others prevented full processing of the enzyme, leaving inactive or partially active intermediates. The remaining portion of the propeptide towards its *C*-terminal which is further processed from the 30 kDa form to the mature form in rFheproCL1 is less structured and more easily accessible to proteases, as shown by the crystal structures of human procathepsin L (Coulombe et al., 1996) and cathepsin B (Turk et al., 1996; Podobnik et al., 1997). This can also be seen in the collective data which shows that papain (Vernet et al., 1995), cathepsin L (Coulombe et al., 1996), cathepsin B (Turk et al., 1996; Podobnik et al., 1997), cathepsin S (Quraishi and Storer, 1998), and cathepsin K (McQueney et al., 1997) can all be activated to mature enzymes either by *cis*- or *trans*-cleavage at various bonds within this region of the propeptide. Also, the cathepsin B of the related helminth parasite *S. mansoni* can be similarly activated (Lipps et al., 1996; Sajid and McKerrow, 2002). However, the sites of cleavage of the wild-type and mutant procathepsin L1 differed suggesting that prior cleavage within the G*X*N*X*F*X*D motif of the wild-type enzyme may have altered the accessibility of the

C-terminus to protease action. Also clear is that additional amino acid residues are left by both processing pathways at the *N*-terminus of the mature enzyme, unlike in the mature cathepsin L1 present in adult *F. hepatica* ES products which is fully processed to the characteristic *N*-terminal residue. It is therefore possible that an additional protease, such as an exopeptidase, may be required to remove these additional amino acids.

8.6 Secretion of cathepsin L proteases by *F. hepatica* parasites

The two major cysteine proteases of *F. hepatica*, cathepsins L1 and L2, are secreted when the adult parasite is maintained *in vitro* and were previously characterised in our laboratory (Smith et al., 1993a; Dowd et al., 1994). *N*-terminal sequencing of both these enzymes indicated that they were secreted as fully mature cathepsin L enzymes. Immunoblotting experiments in which adult *F. hepatica* ES products were probed with antisera prepared against purified mature cathepsin L1 and recombinant CL1 propeptide demonstrated that while both react strongly with the anti-mature antiserum, neither enzyme reacts with the anti-propeptide antiserum (see Figure 7.1; see also Figure 4.3). Comparison of the amino acid sequences of the two enzymes show a high level of identity of 78.22% (see Section 8.1), explaining their cross-reactivity with antiserum prepared against either enzyme seen here and in other studies (Roche et al., 1997; Dowd et al., 1994). Together, these data confirm that the enzymes are secreted in a fully processed form.

F. hepatica cathepsins L1 and L2 belong to a phylogenetic lineage which includes the mammalian cathepsin Ls and have substrate specificities that are typical to this group; both have a preference for positively charged residue side chains, such as arginine, in the P_1 position and bulkier hydrophobic side chains, such as phenylalanine and leucine, in the P_2 position, as evidenced by their characteristic breakdown of the

fluorogenic substrate Z-FR-AMC (Coulombe et al., 1996; Dowd et al., 1994). However, they differ from their mammalian counterparts in their ability to exhibit activity over a wide pH range, between pH 4.0 and 8.5, and remain stable at neutral pH. While human cathepsin L loses all activity after just 20 min of incubation at pH 7 and 37°C, the *F. hepatica* cathepsin Ls retain most of their activity after 24 h in the same conditions (Dowd et al., 2000). It has been suggested that this enhanced stability is important in the digestion of host proteins within the parasite gut and in the migration of the parasite through host tissues such as the intestine and liver (Dalton et al., 2003).

With the obvious importance of the cathepsin L proteases to *F. hepatica*, it is of great interest to examine how and where the enzymes are produced and secreted by the parasite before their appearance in the ES products. Previous immunolocalisation and *in situ* hybridisation studies have shown that the *F. hepatica* cathepsin L proteases are synthesised within the gastrodermis that lines the parasite gut (Smith et al., 1993a; Dalton et al., 2003). The gastrodermis is composed of epithelial cells that are columnar in shape and contain large numbers of dense vesicles, referred to as secretory vesicles, situated at the apical end (see Figure 7.2A). Despite being secretory in function, these vesicles have long been described as similar to lysosomes and have been suspected for almost thirty years to contain proteases of some form (Halton, 1967; Robinson and Threadgold, 1975; Halton, 1997). As part of this study, immunofluorescence and immunelectron microscopy experiments were carried out by Dr. Aaron Maule's laboratory at Queen's University, Belfast, UK, using the anti-propeptide and anti-mature cathepsin L1 antisera from our laboratory (Collins et al., 2004). Confocal laser immunocytochemistry using the native mature cathepsin L antiserum demonstrated that the proteases are located to the gastrodermal epithelial cells through the gut of adult *F. hepatica* (see Figure 7.2B). This labelling appeared punctate and was more dense towards the apical end of the gastrodermal cells, suggesting that the enzymes are

associated with the secretory vesicles (see Figure 7.2C). Further experiments using immunoelectron microscopy revealed that the cathepsin L proteases are indeed stored within the secretory vesicles (see Figure 7.2D-F). Also, the secretory vesicles were strongly immunoreactive with both the antiserum prepared against the recombinant cathepsin L1 propeptide (see Figure 7.2E) and the anti-mature cathepsin L antiserum (see Figure 7.2F). This dual labelling demonstrates that the enzymes are present in the secretory vesicles in the inactive procathepsin L zymogen form.

The above observations provide important insights into the regulation of cysteine protease activity within the gastrodermis of *F. hepatica*. Firstly, these observations are in agreement with those of Halton and co-workers (Halton, 1967; Halton, 1997), as it appears that the function of the secretory vesicles is to package proteolytic enzymes for delivery into the gut lumen. Secondly, it makes good sense for the enzymes to be accumulated in the secretory vesicles in the inactive precursor form rather than as active mature enzyme given the high abundance of the vesicles in the epithelial cells and the accompanying possibility of protease leakage into the cytoplasm. Finally, in combination with the data on the *in vitro* processing of these two enzymes discussed above (see Sections 8.2 and 8.3), these observations suggest that the processing and activation of the procathepsins L1 and L2 takes place following the secretion of the zymogen into the gut lumen.

Electron microscopy has shown that the contents of the secretory vesicles are extruded into the gut lumen where they mix with the ingested blood and tissue meal between the long extruding lamellae (Robinson and Threadgold, 1975; Halton, 1997). Briefly, the epithelial cells lining the gastrodermis go through a cyclical pattern of changes between a secretory phase, where the Golgi-apparatus is active and there are large numbers of secretory vesicles, and an absorptive phase, characterised by more numerous surface lamellae and the release of the contents of the secretory vesicles into

the gut (Robinson and Threadgold, 1975; Halton, 1997; see also Figure 8.1). Across the surface of the gastrodermis, both processes occur simultaneously, within adjacent cells at different points in the cycle at any one time (Halton, 1997). As the pH of the gut is estimated to be around 5.5 or slightly lower (Halton, 1997) and the *in vitro* activations of both FheproCL1 and FheproCL2 to the mature form occur best around pH 4.0-5.5 (see Sections 8.2 and 8.3), it seems likely that this is the location of enzyme activation. The enzymes could immediately begin functioning in protein catabolism on activation as it would be in direct contact with the bloodmeal within the gut.

F. hepatica, like other helminth parasites, has a blind-ending gut and removal of undigested products and accumulated waste, such as haematin, occurs by the simple voiding of the gut contents approximately every 3 h (Halton, 1997). Active mature cathepsins L1 and L2 would therefore be regularly liberated from the gut of the parasite into the surrounding intestinal or hepatic tissues allowing them to perform their other external functions, such as tissue degradation facilitating the movement of the parasite through host tissues (Halton, 1997; Fairweather et al., 1999). The enhanced stability of these enzymes would allow for high concentrations over long periods needed to digest native matrix proteins such as collagens (see Figures 7.7 and 7.8).

The pattern of storage, secretion, and activation of the *F. hepatica* procathepsins contrasts quite dramatically with the pattern found with the mammalian procathepsin Ls, which are trafficked through the Golgi apparatus to acidic lysosomes in which the processing to the mature active enzyme form occurs (Coulombe et al., 1996; Ménard et al., 1998). However, there are some examples which show similarity to the situation in *F. hepatica*; low levels of procathepsin L are known to be secreted from normal cells, playing a role in extracellular processing (Ishidoh and Kominami, 1998); high levels of secretion have been correlated with the metastatic activity of transformed cells (Lorenzo

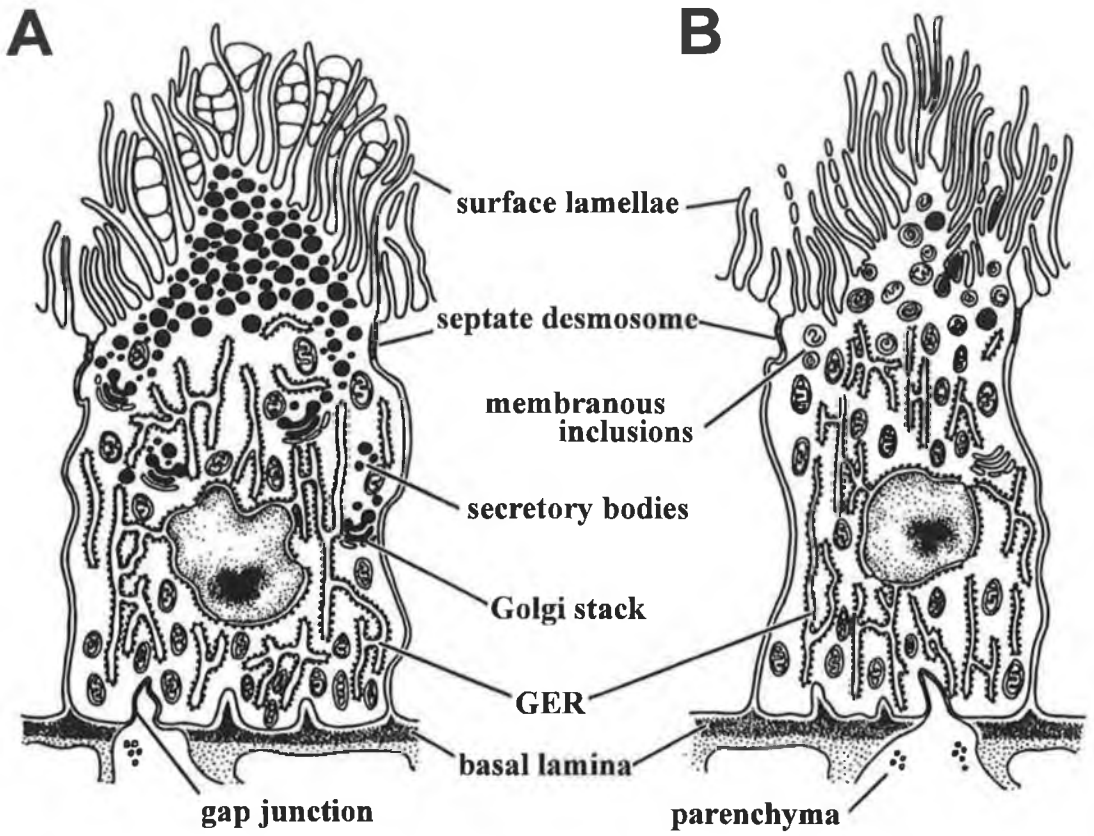


Figure 8.1: Schematic diagram of the gut cell in *Fasciola hepatica.* *A*, epithelial cell of the gastrodermis in the secretory phase, with active Golgi apparatus and numerous secretory vesicles towards the apical end containing the cathepsin Ls. *B*, epithelial cell in the absorptive phase, with an increase in the number of surface lamellae and the release of the contents of the secretory vesicles into the gut lumen. Adapted from Halton (1997).

et al., 2000; Guillaume-Rousselet et al., 2002). Similarly, the cathepsin-L like cysteine protease, cathepsin K, is secreted in humans by osteoclasts and has a function in bone resorption (Shi et al., 1995; Drake et al., 1996; McQueney et al., 1997). In this case, osteoclasts adhere to the bone surface and form an extracellular compartment, known as an absorption pit, between the cell and bone into which cathepsin K is secreted. The pH within this pit is slightly acidic and is considered to be important for both the dissolution of the bone and the activation of procathepsin K to the active mature form (Shi et al., 1995; Drake et al., 1996). Further comparisons can be made with mammalian cathepsins B, K, and L, which have been shown to be secreted into the extracellular lumen of thyroid follicles and play a role in the proteolytic processing of covalently cross-linked soluble thyroglobulin to liberate the thyroid hormone thyroxine (T_4) (Tepel et al., 2000; Friedrichs et al., 2003).

8.7 Comparison of recombinant wildtype cathepsins L1 and L2

Comparison of the levels of expression of each of the recombinants produced in parallel cultures (see Figure 6.8) indicates that, with the exception of the inactive mutant procathepsin L2, roughly equal quantities of protein are produced for each of the wildtype recombinants, rFheproCL1 and rFheproCL2, while larger amounts are produced for the inactive mutant procathepsin L1s. Analysis of the mRNA produced during parallel induction experiments, however, shows no discernible difference in levels of specific recombinant mRNA as a proportion of the total poly A^+ mRNA produced (see Figure 7.3A); it is possible, therefore, that sequence differences between the recombinant transcripts may affect levels of translation and/or secretion of the proteins. Analysis of cell lysates from the same parallel cultures indicate that some protein is retained within the cells, although the level is small and is only detectable by immunoblot with a mouse monoclonal anti-His_6 antibody (see Figure 7.3B). The bands

visualised for each recombinant on immunoblot match those which occur for the secreted proteins present in yeast medium supernatant. This observation indicates that some processing of the two wildtype procathepsin Ls may occur during trafficking within the cells before secretion into the medium.

The effect of pH on the activity of active mature rFheCL1 and rFheCL2 was investigated by two methods; determination of the pH optima of the enzymes in assay against the characteristic fluorogenic substrate Z-FR-AMC and migration and activity in gelatine substrate gels (GS-PAGE). The pH profile of activity for the recombinant mature cathepsin L1 showed that it was similar to that of the native mature enzyme (Dowd et al., 1994; Dowd et al., 2000); active rFheCL1 was active over the pH range of 3.5 to 8.5 with highest levels of activity at pH 6.5-7.0 (see Figure 7.5). Interestingly, a second peak of activity was found at a pH of 4.5, in the region of pH values at which the zymogen form has been shown to autoactivate. It is possible, therefore, that a small amounts of processed active CL1 are responsible for accelerated processing of procathepsin Ls. However, this second peak was not found to occur for active rFheCL2, and may only have a small effect. Overall, rFheCL2 displayed a clearer pattern of activity, with a clear peak of activity at pH 7.0. Gelatinolytic activity was visible across a wide pH range for both enzymes, with highest activity occurring in gels incubated at pH 4.0 and 5.0 for rFheCL1 and rFheCL2 respectively. It should be noted also that rFheCL2 retained strong gelatinolytic activity at the higher pH values of 7.0 and 7.75, where little activity was detected for rFheCL1 at pH 7.0 and none was found at pH 7.75 (see Figure 7.4). Also observed was a major difference in the migration of the two proteins; while mature CL2 migrated as a single band almost all the way to the bottom of the gel, mature CL1 migrated much more slowly, only about 1/3 of the way into the gel, in two bands. Clearly, as the GS-PAGE gels are effectively run under native conditions and separation is based on charge rather than size, there is a difference

in the overall charge of the two mature proteins. Also, while both enzymes were activated in parallel over the course of 24 h to form only one band visible on SDS-PAGE each, CL2 may be more efficiently producing its mature form while CL1 is retaining one or two amino acids of the propeptide at the *N*-terminal, leaving a mixture of two mature forms of almost identical size (making them visually inseparable on SDS-PAGE) with slightly differing charges (making them separately visible on GS-PAGE). The differences between the amino acid sequences of each enzyme immediately prior to the *N*-terminal of the mature peptide are possibly responsible for this differential cleavage.

Substrate preference of the two recombinant enzymes was investigated by comparison of their activity in assays against various fluorogenic peptide substrates (see Figure 7.6). Strong activity of both enzymes against Z-FR-AMC and no activity against Z-R-AMC and Z-RR-AMC provided the characteristic cathepsin L activity expected. In agreement with previous analyses (Dowd et al., 1994), the activity of cathepsin L2 against Z-FR-AMC was approximately half that of CL1. Unlike human cathepsin L (Lecaille et al., 2002), high levels of activity were found for both *F. hepatica* cathepsin Ls against substrates with leucine in the P_2 position, with the higher activity against Z-LR-AMC for both enzymes than for any other substrate tested. However, this additional preference for leucine in the P_2 position has also been shown in the cathepsin Ls of another helminth parasite, *S. mansoni* (Brady et al., 2000). Of particular note was the difference in preference between the two enzymes against substrates with proline in the P_2 position; while CL1 showed very low levels of activity against these substrates, CL2 showed between 2- and 10-fold more activity against them, particularly when coupled with glycine in the P3 position, such as in Tos-GPR-AMC and Boc-AGPR-AMC. This data is in agreement with previous results obtained in our laboratory (Dowd et al., 1994).

This greater preference for proline in the P_2 position is also found in human cathepsin K. Lecaille et al. (2002) have shown that replacement in cathepsin K of Tyr^{67} (human cathepsin K numbering) with Leu^{67} as found in human cathepsin L alongside a mutation of $Leu^{205} \rightarrow Ala^{205}$ altered its pattern of substrate preferences to more closely match those of cathepsin L; most notably, the mutations removed cathepsin K's preference for proline in the P_2 position and greatly enhanced its preference for phenylalanine here. The $Tyr^{67} \rightarrow Leu^{67}$ (human cathepsin K numbering) corresponds to the presence of Tyr^{70} in FheCL2 and Leu^{70} in FheCL1 which causes a similar alteration in substrate specificity (see Section 8.1). However, FheCL2 still retains a high preference for both leucine in the P2 position (found in human cathepsin K but to a lesser extent in human cathepsin L) and phenylalanine (found in human cathepsin L but not human cathepsin K) here. These data suggest that FheCL2 may possess a cathepsin K-like activity in addition to its cathepsin L activity, rather than in place of it. Cathepsin K is known to have a collagenolytic activity (Kafienah et al., 1998; Hou et al., 2001; Hou et al., 2003). Collagens contain numerous proline residues, so the preference of *F. hepatica* cathepsin L1 for proline in the P_2 position may indicate a function in collagen degradation similar to that of human cathepsin K. Collagenolytic activity of both recombinant *F. hepatica* cathepsin Ls was assayed against collagens Type II (from bovine nasal septum) and VI (from basement membranes) (see Figures 7.7 and 7.8). In both cases, active mature cathepsin L2 demonstrated greater collagenolytic activity over the course of 24 h compared to cathepsin L1; cathepsin L2 breakdown of both collagen types was comparable to that of a commercial collagenase blend. The greater ability of *F. hepatica* cathepsin L2 to cleave collagen suggests it may have a more prominent function in aiding the migration of the parasite through the host tissues of the intestine and the liver compared to cathepsin L1.

As has been discussed above (see Section 8.5), a mutation of Leu^{-12}→Pro^{-12} within the propeptide of one of the *F. hepatica* inactive procathepsin L1 mutants prevented its intermolecular processing to lower molecular size forms by exogenously added active recombinant cathepsin L1, with cleavage only of the first few amino acid residues achieved. The substrate preference assay results above demonstrate that while both FheCL1 and FheCL2 can cleave substrates with leucine in the P2 position, only FheCL2 possesses any substantial ability to cleave substrates with proline here. Intermolecular processing experiments involving the double mutant r[$Pro^{-12}Gly^{26}$]FheproCL1 were repeated to include incubation with exogenously added active mature rFheCL2 in parallel with active mature rFheCL1 (see Figure 7.9). Unlike cathepsin L1, cathepsin L2 was capable of processing the proline containing mutant cathepsin L1 to the mature form over the course of a 3 h incubation. Cathepsin L2 was also capable of more fully processing the single mutant r[Gly^{26}]FheproCL1 over the course of these parallel experiments than cathepsin L1. Both these results further demonstrate in practical terms the broader substrate specificity of cathepsin L2, as well as the possible key role of position -12 for the intermolecular processing of cathepsin L1.

The large family of cathepsin Ls found in the *Fasciola* spp. is in contrast to the small numbers of cathepsin Fs so far discovered for the trematodes *P. westermani* and *C. sinensis* (see Figure 1.3; also Dalton et al., 2005). The cathepsin Fs of both *P. westermani* and *C. sinensis* have similar functions, namely the facilitation of migration through the host and acquisition of nutrients by the hydrolysis of haemoglobin and collagens (Park, H. et al., 2001; Park, S. Y. et al., 2001; Park, H. et al., 2002; Shin et al., 2005). However, the *P. westermani* and *C. sinensis* cathepsin Fs are much more closely related to each other than to members of the *Fasciola* cathepsin L clade, indicating a corresponding greater conservation of these functions between the enzymes (Park et al.,

2002; Kang et al., 2004). While the cathepsin Fs of these species have not greatly diversified over time, the cathepsin Ls of the *Fasciola* spp. have evolved into distinct groups with increasing levels of specificity and therefore of functional capability. In the case of the *Schistosoma* spp., both cathepsin Ls and a cathepsin F have been identified (Day et al., 1995; Dalton et al., 1996; Brindley et al., 1997; Brady et al., 2000), although again there is no large diverse group of cathepsin L-like proteases as found for the *Fasciola* spp. Therefore, the cathepsin L-like cysteine proteases of the *Fasciola* spp. have evolved separately from those of the other major trematode parasite species.

The differing substrate specificities found here for *F. hepatica* cathepsins L1 and L2 demonstrate some of this evolved diversity of function. Cathepsin L2, belonging to the evolutionarily older Adult Group 3 enzyme (see Section 8.1; also Irving et al., 2003), possesses a wider substrate specificity with a greater ability to accept proline residues in the S_2 subsite. This allows it to more efficiently cleave collagens and indicates a greater role for cathepsin L2 in facilitating the migration of the parasite through host tissues than for cathepsin L1. Cathepsin L1, as an evolutionarily younger Adult Group 1 enzyme (see Section 8.1; also Irving et al., 2003), is more specific in its substrate preferences, indicating a development for more specialised functions within the parasite. Cathepsin L1 may therefore have greater involvement in processes such as cleavage of immunoglobulins and immunomodulation by suppression of the host T_H1 response, as has been seen in experiments with bacterial co-infections in laboratory animals (Brady et al., 1999; O'Neill et al., 2000; O'Neill et al., 2001). This specialisation and diversity of function within the *Fasciola* cathepsin L family has not been observed in the other trematode parasite species.

8.8 Summary and conclusions

Given the importance of cathepsin L proteases in the virulence of *F. hepatica* and other helminth pathogens it is important to understand the mechanisms of their synthesis, processing and activation.

A large number of nucleotide and protein sequences have been deposited in the online databases representing cathepsin Ls of both *F. hepatica* and *F. gigantica*. In this project, these enzymes have been shown to form four major clades in phylogenetic analyses; 3 Adult Groups and 1 Metacercarial Group. There are distinct features of the propeptides of each group and the residues responsible for S_2 subsite specificity (Turk et al., 1998; Smooker et al., 2000) also fall into the four clades, indicating diversity of function among the *Fasciola* cathepsin Ls. A pattern of evolving specialisation of these enzymes can be seen.

Although recombinant versions of the two *F. hepatica* cathepsin Ls studied here have been produced previously in *Saccharomyces cerevisiae* (Dowd et al., 1997; Roche et al., 1997), it was decided to transfer the genes to vectors suitable for the *Pichia pastoris* yeast expression system. This system can provide larger quantities of secreted recombinant from high concentrations of biomass and can allow easier, one-step purification due to its low levels of background native protein secretion (Cereghino and Cregg, 2000; Brady et al., 2001; Aoki et al., 2003). The larger quantities of protein have been used for various characterisation studies and for collaborative work with other laboratories.

As has been noted already, the propeptides of several members of the papain superfamily have been shown to contain autocatalytic cleavage sites (Ishidoh and Kominami, 1994; Ménard et al., 1998; Quraishi and Storer, 2001), most notably the well conserved G*X*N*X*F*X*D motif (Vernet et al., 1995). In this study, the transport and

processing of *Fasciola hepatica* cathepsin L1 and L2 both *in vivo* and *in vitro* was investigated using the recombinant proenzymes.

Intermolecular processing was studied by generating inactive procathepsin L mutants, where the catalytic cysteine residue at position 26 was substituted with a glycine, which was incapable of autoactivation. The results show that *P. pastoris*-expressed cathepsins L1 and L2 can autoactivate in low pH conditions to an active intermediate form by an initial cleavage within the conserved heptapeptide (Gly-Xaa-Asn-Xaa-Phe-Xaa-Asp), as previously described for papain (Vernet et al., 1995). Further activation of the enzyme occurs by cleavage in the vicinity of the *N*-terminus. In contrast, intermolecular processing of the mutant inactive procathepsin Ls to a mature product by exogenous mature cathepsin L did not involve the formation of the above intermediate but was initiated by a direct cleavage close to the *N*-terminus.

Further to this, a second cathepsin L1 inactive mutant, in which Leu^{-12} of the propeptide was replaced with Pro^{-12} in addition to the $Cys^{26} \rightarrow Gly^{26}$ mutation was generated serendipitously and shed further light on the processing and substrate specificity of the two enzymes. Exogenous processing studies show that this mutation prevents proper processing of the mutant by cathepsin L1, but not by cathepsin L2, indicating variations in specificity, and therefore function, between the two enzymes.

It has been previously shown in our laboratory that the cathepsins L1 and L2 are secreted by *F. hepatica* and therefore function extracellularly, (Smith et al., 1993a; Dalton et al., 2003) in a similar manner to the mammalian cathepsin Ls involved in the processing of the hormone thyroglobulin (Tepel et al., 2000; Friedrichs et al., 2003). In this project, antibodies against both the mature and propeptide regions of the *F. hepatica* cathepsin Ls were employed in confocal laser and electron immunocytochemistry studies to show that the enzymes are stored in secretory vesicles of the gastrodermal epithelial cells in their inactive proenzyme forms. It is suggested

that the inactive zymogens are released from the secretory vesicles into the gut, where they are processed to the mature active form by autocatalytic means at the low pH of the gut environment.

Differences in substrate specificity were further examined by investigating the activity of the recombinant cathepsins L1 and L2 against various fluorogenic substrates. These showed a marked difference, with cathepsin L2 displaying the ability to cleave a larger range of substrates, particularly those containing proline in the P_2 position. This ability is a feature of human cathepsin K, which is known to have collagenolytic activity. The recombinant enzymes were compared in their ability to cleave native structural proteins, with collagens used as examples of such proteins. While cathepsin L1 does show some collagenolytic activity, cathepsin L2 is far more efficient, more comparable to a commercial collagenase. It is therefore suggested that cathepsin L2 possesses a cathepsin K-like activity in addition to, but not in place of, its cathepsin L activity and may play an important role in migration of the parasite through host tissues.

The data obtained in this study provides a mechanism by which *F. hepatica* synthesizes and secretes two fully activated proteases that are essential to its existence as a parasite, and demonstrates differences in function between divergent groups of cathepsin Ls which have evolved in *F. hepatica*.

8.9 Related collaborative work

The production of large quantities of purified recombinant *F. hepatica* cathepsins L1 and L2 has allowed several collaborative studies both within and outside of our laboratory. As well as the immunolocalisation and *in situ* hybridisation studies performed as part of this project, work has been performed in collaboration with Dr. Aaron Maule and his laboratory in Queen's University, Belfast, UK, to investigate the role of *F. hepatica* cathepsin Ls in the processing of neuropeptides within the nervous

system of the parasite. Work is ongoing to elucidate the 3D structures of both cathepsin L1 and L2 by X-Ray crystallography in collaboration with Dr. Linda Brinen of the *Fasciola* Structural Biology Laboratory at UCSF, San Francisco, USA. Recombinant FheproCL1 has been used in the analysis of a novel multi-domain member of the cystatin superfamily recently identified in the laboratory of Dr. Nathalie Moiré of the Université François Rabelais, Tours, France (Khaznadji et al., 2005, in press). The recombinant enzymes are also being used in ongoing experimental vaccine trials with Dr. Grace Mulcahy of the Department of Veterinary Microbiology and Parasitology at University College Dublin and Dr. Sandra O'Neill of the School of Nursing, Dublin City University. Mutation studies on the *F. hepatica* cathepsin Ls in regard to their substrate specificity are also being conducted by Prof. John P. Dalton and Dr. Colin Stack at the Institute for the Biotechnology of Infectious Diseases, University of Technology, Sydney, Australia.

References

Almeida, M. S., Torloni, H., Lee-Ho, P., Vilar, M. M., Thaumaturgo, N., Simpson, A. J. G. and Tendler, M. (2003). Vaccination against *Fasciola hepatica* infection using a *Schistosoma mansoni* defined recombinant antigen, Sm14. *Parasite Immunol.*, **25**, 135-137.

Andresen, K., Tom, T. D. and Strand, M. (1991). Characterization of cDNA clones encoding a novel calsium-activated neutral proteinase from *Schistosoma mansoni*. *J. Biol. Chem.*, **266**, 15085-15090.

Andrews, S. J. (1999). The life cycle of *Fasciola hepatica*. In: *Fasciolosis* (Dalton, J. P., ed.). CABI Publishers, Wallingford, UK, 1-21.

Aoki, H., Ahsan, M. N. and Watabe, S. (2003). Heterologous expression in *Pichia pastoris* and single-step purification of a cysteine proteinase from northern shrimp. *Protein Expr. Purif.*, **31**, 213-221.

Baker, D., Silen, J. L. and Agard, D. A. (1992). Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Protein Struct. Funct. Genet.*, **12**, 339-344.

Baker, E. N. (1977). Structure of actinidin: details of the polypeptide chain conformation and active site from an electron density map at 2.8 Å resolution. *J. Mol. Biol.*, **115**, 263-277.

Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M. and Hanada, K. (1982). L-trans-epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochem. J.*, **201**, 189-198.

Barrett, A. J. (1994). Classification of peptidases. *Meth. Enzymol.*, **244**, 1-15.

Berasain, P., Goñi, F., McGonigle, S., Dowd, A. J., Dalton, J. P., Frangione, B. and Carmona, C. (1997). Proteinases secreted by *Fasciola hepatica* degrade extracellular matrix and basement membrane components. *J. Parasitol.*, **83**, 1-5.

Brady, C. P., Brindley, P. J., Dowd, A. J. and Dalton, J. P. (2000). *Schistosoma mansoni*: differential expression of cathepsins L1 and L2 suggests discrete biological functions for each enzyme. *Exp. Parasitol.*, **94**, 75-83.

Brady, C. P., Shimp, R. L., Miles, A. P., Whitmore, M. and Stowers, A. W. (2001). High-level production and purification of P30P2MSP119, an important vaccine antigen for malaria, expresses in the methylotropic yeast *Pichia pastoris*. *Protein Expr. Purif.*, **23**, 468-475.

Brady, M. T., O'Neill, S. M., Dalton, J. P. and Mills, K. H. G. (1999). *Fasciola hepatica* suppresses a protective T_H1 response against *Bordetella pertussis*. *Infect. Immun.*, **67**, 5372-5378.

Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R., Wong, J. Y. M., Smythe, M. L. and McManus, D. P. (1997). Proteolytic degradation of host haemoglobin by schistosomes. *Mol. Biochem. Parasitol.*, **89**, 1-9.

Brindley, P. J. and Dalton, J. P. (2004). Schistosome legumain. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1305-1310.

Brömme, D. (2004a). Cathepsin F. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1087-1088.

Brömme, D. (2004b). Cathepsin K. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1092-1097.

Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Salter, J. P., Sajid, M., Lucas, K. D., Franklin, C., Bogyo, M. and McKerrow, J. H. (2000). Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Lett.*, **466**, 244-248.

Cappetta, M., Roth, I., Díaz, A., Tort, J. and Roche, L. (2002). Role of the prosegment of *Fasciola hepatica* cathepsin L1 in folding of the catalytic domain. *Biol. Chem.*, **383**, 1215-1221.

Carmona, C., Dowd, A. J., Smith, A. M. and Dalton, J. P. (1993). Cathepsin L proteinase secreted by *Fasciola hepatica in vitro* prevents antibody-mediated eosinophil attachment to newly excysted juveniles. *Mol. Biochem. Parasitol.*, **62**, 9-18.

Carnevale, S., Rodríguez, M. I., Guarnera, E. A., Carmona, C., Tanos, T. and Angel, S. O. (2001). Immunodiagnosis of fasciolosis using recombinant procathepsin L cysteine proteinase. *Diagn. Microbiol. Infect. Dis.*, **41**, 43-49.

Cereghino, J. L. and Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.*, **24**, 45-66.

Chappell, C. L., Hackel, J. and Davis, A. H. (1989). Cloned *Schistosoma mansoni* proteinase (hemoglobinase) as a putative serodiagnostic reagent. *J. Clin. Microbiol.*, **27**, 196-198.

Choo, J.-D., Suh, B.-S., Lee, H.-S., Lee, J.-S., Song, C.-J., Shin, D.-W. and Lee, Y.-H. (2003). Chronic cerebral paragonimiasis combined with aneurismal subarachnoid hemorrhage. *Am. J. Trop. Med. Hyg.*, **69**, 466-469.

Cohen, L. W., Coghlan, V. M. and Dihel, L. C. (1986). Cloning and sequencing of papain-encoding cDNA. *Gene*, **48**, 219-227.

Collins, P. R., Stack, C. M., O'Neill, S. M., Doyle, S., Ryan, T., Brennan, G. P., Mousley, A., Stewart, M., Maule, A. G., Dalton, J. P. and Donnelly, S. (2004). Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J. Biol. Chem.*, **279**, 17038-17046.

Cornelissen, J. B. W. J., Gaasenbeek, C. P. H., Borgsteede, F. H. M., Holland, W. G., Harmsen, M. M. and Boersma, W. J. A. (2001). Early immunodiagnosis of fasciolosis in ruminants using recombinant *Fasciola hepatica* cathepsin L-like protease. *Int. J. Parasitol.*, **31**, 728-737.

Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S. and Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*, **15**, 5492-5503.

Cygler, M. and Mort, J. S. (1997). Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity. *Biochimie*, **79**, 645-652.

Dalton, J. P. and Heffernan, M. (1989). Thiol proteases released *in vitro* by *Fasciola hepatica*. *Mol. Biochem. Parasitol.*, **35**, 161-166.

Dalton, J. P., Clough, K. A., Jones, M. K. and Brindley, P. J. (1996a). Characterization of the cathepsin-like cysteine proteinases of *Schistosoma mansoni*. *Infect. Immun.*, **64**, 1328-1334.

Dalton, J. P., McGonigle, S., Rolph, T. P. and Andrews, S. J. (1996b). Induction of protective immunity in cattle against infection with *Fasciola hepatica* by vaccination with cathepsin L proteinase and with haemoglobin. *Infect. Immun.*, **64**, 5066-5074.

Dalton. J. P. and Mulcahy, G. (2001). Parasite vaccines – a reality? *Vet. Parasitol.*, **98**, 149-167.

Dalton, J. P., O'Neill, S. M., Stack, C. N., Collins, P. R., Walshe, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D., Khaznadji, E., Moiré, N., Brennan, G., Mousley, A., Kreshchenkoh, N., Maule, A. G. and Donnelly, S. M. (2003). *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *Int. J. Parasitol.*, **33**, 1173-1181.

Dalton, J. P. and Brindley, P. J. (2004). Cathepsin W. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1109-1112.

Dalton, J. P., McKerrow, J. H. and Brindley, P. J. (2004a). Trematode cysteine endopeptidases. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1176-1182.

Dalton, J. P., Skelly, P. and Halton, D. W. (2004b). Role of tegument and gut in nutrient uptake by parasitic platyhelminths. *Can. J. Zool.*, **82**, 211-232.

Dalton, J. P., Caffrey, C. R., Sajid, M., Stack, C., Donnelly, S., Loukas, A., Don, T., McKerrow, J., Halton, D. W. and Brindley, P. J. (2005). Proteases in trematode biology. In: *Protein Function, Metabolism and Physiology* (Maule, A, ed.). CAB International, Wallingford, Oxon, UK.

Day, S. R., Dalton, J. P., Clough, K. A., Leonardo, L., Tiu, W. U. and Brindley, P. J. (1995). Characterization and cloning of the cathepsin L proteinases of *Schistosoma japonicum*. *Biochem. Biophys. Res. Comm.*, **217**, 1-9.

Dowd, A. J., Smith, A. M., McGonigle, S. and Dalton, J. P. (1994). Purification and characterisation of a second cathepsin L proteinase secreted by the parasitic trematode *Fasciola hepatica*. *Eur. J. Biochem.*, **223**, 91-98.

Dowd, A. J., Tort, J., Roche, L., Ryan, T. and Dalton, J. P. (1997). Isolation of a cDNA encoding *Fasciola hepatica* cathepsin L2 and functional expression in *Saccharomyces cerevisiae*. *Mol. Biochem. Parasitol.*, **88**, 163-174.

Dowd, A. J., Dooley, M., Ó Fágáin, C. and Dalton, J. P. (2000). Stability studies on the cathespin L proteinase of the helminth parasite, *Fasciola hepstica*. *Enzyme Microb. Technol.*, **27**, 599-604.

Drake, F. H., Dodds, R. A., James, I. E., Connor, J. R., Debouck, C., Richardson, S., Lee-Rykaczewski, E., Coleman, L., Rieman, D., Barthlow, R., Hastings, G. and Gowen, M. (1996). Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. *J. Biol. Chem.*, **271**, 12511-12516.

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. and Wolthers, B. G. (1968). Structure of Papain. *Nature*, **218**, 929-932.

Fairweather, I., Threadgold, L. T. and Hanna, R. E. B. (1999). Development of *Fasciola hepatica* in the mammalian host. In: *Fasciolosis* (Dalton, J. P., ed.). CABI Publishers, Wallingford, UK, 47-111.

Felbor, U., Dreier, L., Bryant, R. A. R., Ploegh, H. L., Olsen, B. R. and Mothes, W. (2000). Secreted cathepsin L generates endostatin from collagen XVIII. *EMBO J.*, **19**, 1187-1194.

Frade, R. (1999). Structure and functions of proteases which cleave human C3 and are expressed on normal or tumor human cells: some are involved in tumorigenic and metastatic properties of human melanoma cells. *Immunopharmacology*, **42**, 39-45.

Friedrichs, B., Tepel, C., Reinheckel, T., Deussing, J., von Figura, K., Herzog, V., Peters, C., Saftig, P. and Brix, K. (2003). Thyroid functions of mouse cathepsins B, K, and L. *J. Clin. Invest.*, **111**, 1733-1745.

Godat, E., Lecaille, F., Desmazes, C., Duchêne, S., Weidauer, E., Saftig, P., Brömme, D., Vandier, C. and Lalmanach, G. (2004). Cathepsin K: a cysteine protease with unique kinin-degrading properties. *Biochem. J.*, **383**, 501-506.

Gorman, T., Aballay, J., Fredes, F., Silva, M., Aguillón, J. C. and Alcaíno, H. A. (1997). Immunodiagnosis of fasciolosis in horses and pigs using western blots. *Int. J. Parasitol.*, **27**, 1429-1432.

Grams, R., Vichasri-Grams, S., Sobhon, P., Suchart Upatham, E. and Viyanant, V. (2001). Molecular cloning and characterization of cathepsin L encoding genes from *Fasciola gigantica*. *Parasitol. Int.*, **50**, 105-114.

Grover, J. K., Vats, V., Uppal, G. and Yadav, S. (2001). Anthelmintics: a review. *Trop. Gastroenterol.*, **22**, 180-189.

Guillaume-Rousselet, N., Jean, D. and Frade, R. (2002). Cloning and charaterization of anti-cathepsin L single chain variable fragment whose expression inhibits procathepsin L secretion in human melanoma cells. *Biochem. J.*, **367**, 219-227.

Guo, Y. L., Kurz, U., Schultz, J. E., Lim, C. C., Wiederanders, W. and Schilling, K. (2000). The α1/2 helical backbone of the prodomain defines the intrinsic inhibitory specificity in the cathepsin L-like cysteine protease subfamily. *FEBS Lett.*, **469**, 203-207.

Halton, D. W. (1967). Observations on the nutrition of digenetic trematodes. *Parasitology*, **57**, 639-660.

Halton, D. W. (1997). Nutritional adaptations to parasitism within the Platyhelminthes. *Int. J. Parasitol.*, **27**, 693-704.

Hanahan, D. (1985). Techniques for transformation of *Escherichia coli*. In: *DNA Cloning, Volume 1, a practical approach* (Glover, D. M., ed.). IRL Press Ltd., London, UK., 109-135.

Harmsen, M. M., Cornelissen, J. B. W. J., Buijs, H. E. C. M., Boersma, W. J. A., Jeurissen, S. H. M. and van Milligen, F. J. (2004). Identification of a novel *Fasciola hepatica* cathepsin L protease containing protective epitopes within the propeptide. *Int. J. Parasitol.*, **34**, 675-682.

Hashmi, S., Britton, C., Liu, J., Guiliano, D. B., Oksov, Y. and Lustigman, S. (2002). Cathepsin L is essential for embryogenesis and development of *Caenorhabditis elegans*. *J. Biol. Chem.*, **277**, 3477-3486.

Hashmi, S., Zhang, J., Oksov, Y. and Lustigman, S. (2004). The *Caenorhabditis elegans* cathepsin Z-like cysteine protease, *Ce*-CPZ-1, has a multifunctional role during the worm's development. *J. Biol. Chem.*, **279**, 6035-6045.

Heussler, V. T. and Dobbeleaere, D. A. E. (1994). Cloning of a protease gene family of *Fasciola hepatica* by the polymerase chain reaction. *Mol. Biochem. Parasitol.*, **64**, 11-23.

Higgins, D. R. and Cregg, J. M. (1998). Introduction to *Pichia pastoris*. *Methods Mol. Biol.*, **103**, 1-15.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.

Hou, W.-S., Li, Z., Gordon, R. E., Chan, K., Klein, M. J., Levy, R., Keysser, M., Keyszer, G. and Brömme, D. (2001). Cathepsin K is a critical protease in synovial fibroblast-mediated collagen degradation. *Am. J. Pathol.*, **159**, 2167-2177.

Hou, W.-S., Li, Z., Büttner, F. H., Bartnik, E. and Brömme, D. (2003). Cleavage site specificity of cathepsin K toward cartilage proteoglycans and protease complex formation. *Biol. Chem.*, **384**, 891-897.

Hu, W., Brindley, P. J., McManus, D. P., Feng, Z. and Han, Z.-G. (2004). Schistosome transcriptomes: new insights into the parasite and schistosomiasis. *Trends Mol. Med.*, **10**, 217-225.

Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T. and Mort, J. S. (1997). Role of the occluding loop in cathepsin B activity. *J. Biol. Chem.*, **272**, 1197-1202.

Irving, J. A., Spithill, T. W., Pike, R. N., Whisstock, J. C. and Smooker, P. M. (2003). The evolution of enzyme specificity in *Fasciola spp. J. Mol. Evol.*, **57**, 1-15.

Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N. and Suzuki, K. (1987). Molecular cloning and sequencing of cDNA for rat cathepsin L. *FEBS Lett.*, **223**, 69-73.

Ishidoh, K. and Kominami, E. (1994). Multi-step processing of procathepsin L *in vitro*. *FEBS Lett.*, **352**, 281-284.

Ishidoh, K. and Kominami, E. (1994). Multi-step processing of procathepsin L *in vitro*. *FEBS Lett.*, **352**, 281-284.

Ishidoh, K. and Kominami, E. (1998). Gene regulation and extracellular functions of procathepsin L. *Biol. Chem.*, **379**, 131-135.

Jean, D., Rodrigues-Lima, F., Cassinat, B., Hermann, J., Cabane, J. and Frade, R. (1997). Co-expression and secretion of C3, the third component of complement and a C3-cleaving cysteine proteinase in a highly metastatic human melanoma cell line. *Imm. Lett.*, **58**, 107-112.

Jerala, R., Žerovnik, E., Kidrič, J. and Turk, V. (1998). pH-induced conformational transitions of the propeptide of human cathepsin L: a role for a molten globule state in zymogen activation. *J. Biol. Chem.*, **273**, 11498-11504.

Joseph, L. J., Chang, L. C., Stamenkovich, D. and Sukhatme, V. P. (1988). Complete nucleotide and deduced amino acid sequences of human and murine preprocathepsin L. An abundant transcript induced by transformation of fibroblasts. *J. Clin. Invest.*, **81**, 1621-1629.

Kafienah, W., Brömme, D., Buttle, D. J., Croucher, L. J. and Hollander, A. P. (1998). Human cathepsin K cleaves native type I and II collagens at the N-terminal end of the triple helix. *Biochem. J.*, **331**, 727-732.

Kang, T. H., Yun, D. H., Lee, E. H., Chung, Y. B., Bae, Y. A., Chung, J. Y., Kang, I., Kim, J., Cho, S. Y. and Kong, Y. (2004). A cathepsin F of adult *Clonorchis sinensis* and its phylogenetic conservation in trematodes. *Parasitology*, **128**, 195-207.

Karrer, K. M., Peiffer, S. L. and DiTomas, M. E. (1993). Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 3063-3067.

Khaznadji, E., Collins P. R., Dalton, J. P., Bigot, Y. and Moiré, N. (2005). A new multi-domain member of the cystatin superfamily expressed by *Fasciola hepatica*. *Int. J. Parasitol*, in press.

Kim, E.-A., Juhng, S.-K., Kim, H. W., Kim, G. D., Lee, Y. W., Cho, H. J. and Won, J. J. (2004). Imaging findings of hepatic paragonimiasis: a case report. *J. Kor. Med. Sci.*, **19**, 759-762.

King, S. and Scholz, T. (2001). Trematodes of the family Opisthorchiidae: a minireview. *Kor. J. Parasitol.*, **39**, 209-221.

Kirschke, H. (2004a). Cathepsin L. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1097-1102.

Kirschke, H. (2004b). Cathepsin S. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1104-1107.

Kofta, W., Mieszczanek, J., Płucienniczak, G. and Wędrychowicz, H. (2000). Successful DNA immunisation of rats against fasciolosis. *Vaccine*, **18**, 2985-2990.

Kreusch, S., Fehn, M., Maubach, G., Missler, K., Rommerskirch, W., Schilling, K., Weber, E., Wenz, I. and Wiederanders, B. (2000). An evolutionarily conserved tripartite tryptophan motif stabilizes the prodomains of cathepsin L-like cysteine proteases. *Eur. J. Biochem.*, **267**, 2965-2972.

Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. (2001). MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244-1245.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Law, R. H. P., Smooker, P. M., Irving, J. A., Piedrafita, D., Ponting, R., Kennedy, N. J., Whisstock, J. C., Pike, R. N. and Spithill, T. W. (2003). Cloning and expression of the major cathepsin B-like protein from juvenile *Fasciola hepatica* and analysis of immunogenicity following liver fluke infection. *Infect. Immun.*, **71**, 6921-6932.

Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S. and Brömme, D. (2002). Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S_2 subsite specificity. *Biochemistry*, **41**, 8447-8454.

Lecaille, F., Wiedauer, E., Juliano, M. A., Brömme, D. and Lalmanach, G. (2003). Probing cathepsin K activity with a selective substrate spanning its active site. *Biochem. J.*, **375**, 307-312.

Lee, J.-S., Lee, J., Park, S.-J. and Yong, T.-S. (2003). Analysis of the genes expressed in *Clonorchis sinensis* adults using the expressed sequence tag approach. *Parasitol. Res.*, **91**, 283-289.

Lemere, C. A., Munger, J. S., Shi, G.-P., Natkin, L., Haass, C., Chapman, H. A. and Selkoe, D. J. (1995). The lysosomal cysteine protease, cathepsin S, is increased in Alzheimer's disease and Down syndrome brain. An immunocytochemical study. *Am. J. Pathol.*, **146**, 848-860.

Li, R., Chen, X., Gong, B., Selzer, P. M., Li, Z., Davidson, E., Kurzban, G., Miller, R. E., Nuzum, E. O., McKerrow, J. H., Fletterick, R. J., Gillmor, S. A., Craik, C. S., Kuntz, I. D., Cohen, F. E. and Kenyon, G. L. (1996). Structure-based design of parasitic protease inhibitors. *Bioorg. Med. Chem.*, **4**, 1421-1427.

Li, Z., Hou, W.-S. and Brömme, D. (2000). Collagenolytic activity of cathepsin K is specifically modulated by cartilage-resident chondroitin sulfates. *Biochemistry*, **39**, 529-536.

Li, Z., Hou, W.-S., Escalante-Torres, C. R., Gelb, B. D. and Brömme, D. (2002). Collagenase activity of cathepsin K depends on complex formation with chondroitin sulphate. *J. Biol. Chem.*, **277**, 28669-28676.

Li, Z., Yasuda, Y., Li, W., Bogyo, M., Katz, N., Gordon, R. E., Fields, G. B. and Brömme, D. (2004). Regulation of collagenase activities of human cathepsins by glycosaminoglycans. *J. Biol. Chem.*, **279**, 5470-5479.

Linder, S., Schliwa, M. and Kube-Granderath, E. (1996). Direct PCR screening of *Pichia pastoris* clones. *Biotechniques*, **20**, 980-982.

Lipps, G., Füllkrug, R. and Beck, E. (1996). Cathepsin B of *Schistosoma mansoni*: purification and activation of the activation of the recombinant proenzyme secreted by *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **271**, 1717-1725.

Lorenzo, K., Ton, P., Clark, J. L., Coulibaly, S. and Mach, L. (2000). Invasive properties of murine squamous carcinoma cells: secretion of matrix-degrading cathepsins is attributable to a deficiency in the mannose 6-phosphate/insulin-like growth factor II receptor. *Cancer Res.*, **60**, 4070-4076.

McGrath, M. E., Eekin, A. E., Engel, J. C., McKerrow, J. H., Craik, C. S. and Fletterick, R. J. (1995). The crystal structure of cruzain: a therapeutic target for Chagas' Disease. *J. Mol. Biol.*, **247**, 251-259.

McGrath, M. E., Palmer, J. T., Brömme, D. and Somoza, J. R. (1998). Crystal structure of human cathepsin S. *Protein Sci.*, **7**, 1294-1302.

McKerrow, J. H., McGrath, M. E. and Engel, J. C. (1995). The cysteine protease of *Trypanosoma cruzi* as a model for antiparasite drug design. *Parasitol. Today*, **11**, 279-282.

McQueney, M. S., Amegadzie, B. Y., D'Alessio, K., Hanning, C. R., McLaughlin, M. M., McNulty, D., Carr, S. A., Ijames, C., Kurdyla, J. and Jones, C. S. (1997). Autocatalytic activation of human cathepsin K. *J. Biol. Chem.*, **272**, 13955-13960.

Mach, L., Mort, J. S. and Glössl, J. (1994a). Maturation of human procathepsin B: proenzyme activation and proteolytic processing of the precursor to the mature proteinase, *in vitro*, are primarily unimolecular processes. *J. Biol. Chem.*, **269**, 13030-13035.

Mach, L., Mort, J. S. and Glössl, J. (1994b). Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass dorms of the enzyme. *J. Biol. Chem.*, **269**, 13036-13040.

Marcet, R., Díaz, A., Arteaga, E., Finlay, C. M. and Sarracent, J. (2002). Passive protection against fasciolosis in mice by immunization with a monoclonal antibody (ES-78 MoAb). *Parasite Immunol.*, **24**, 103-108.

Mason, R. W. and Massey, S. D. (1992). Surface activation of pro-cathepsin L. *Biochem. Biophys. Res. Comm.*, **189**, 1659-1666.

Ménard, R., Carmona, E., Takebe, S., Dufour, É., Plouffe, C., Mason, P. and Mort, J. S. (1998). Autocatalytic processing of recombinant human procathepsin L. *J. Biol. Chem.*, **273**, 4478-4484.

Ménard, R. and Storer, A. C. (2004). Papain. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1125-1128.

Ménard, R. and Sulea, T. (2004). Cathepsin X. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1113-1116.

Merckelbach, A., Hasse, S., Dell, R., Eschlbeck, A. and Ruppel, A. (1994). cDNA sequences of *Schistosoma japonicum* coding for two cathepsin B-like proteins and Sj32. *Trop. Med. Parasitol.*, **45**, 193-198.

Michel, A., Ghoneim, H., Resto, M., Klinkert, M. Q. and Kunz, W. (1995). Sequence, characterization and localization of a cysteine proteinase cathepsin L in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.*, **73**, 7-18.

Mort, J. S. (2004). Cathepsin B. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1079-1086.

Mulcahy, G., O'Connor, F., McGonigle, S., Dowd, A. J., Clery, D. G., Andrews, S. J. and Dalton, J. P. (1998). Correlation of specific antibody titre and avidity with protection in cattle immunized against *Fasciola hepatica*. *Vaccine*, **16**, 932-939.

Munger, J. S., Haass, C., Lemere, C. A., Shi, G.-P., Wong, W. S., Teplow, D. B., Selkoe, D. J. and Chapman, H. A. (1995). Lysosomal processing of amyloid precursor protein to Aβ peptides: a distinct role for cathepsin S. *Biochem. J.*, **311**, 299-305.

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N. and Bode, W. (1991). The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.*, **10**, 2321-2330.

Nagano, I., Pei, F., Wu, Z., Wu, J., Cui, H., Boonmars, T. and Takahashi, Y. (2004). Molecular expression of a cysteine proteinase of *Clonorchis sinensis* and its application to an enzyme-linked immunosorbent assay for immunodiagnosis of clonorchiasis. *Clin. Diagn. Lab. Immunol.*, **11**, 411-416.

O'Neill, S. M., Parkinson, M., Strauss, W., Angles, R. and Dalton, J. P. (1998). Immunodiagnosis of *Fasciola hepatica* infection (fasciolasis) in a human population in the Bolivian Altiplano using purified cathepsin L cysteine proteinase. *Am. J. Trop. Med. Hyg.*, **58**, 417-423.

O'Neill, S. M., Brady, M. T., Callanan, J. J., Mulcahy, G., Joyce, P., Mills, K. H. G. and Dalton, J. P. (2000). *Fasciola hepatica* infection downregulates T_H1 responses in mice. *Parasite Immunol.*, **22**, 147-155.

O'Neill, S. M., Mills, K. H. G. and Dalton, J. P. (2001). *Fasciola hepatica* cathepsin L cysteine proteinase suppresses *Bordetella pertussis*-specific interferon-γ production *in vivo*. *Parasite Immunol.*, **23**, 541-547.

Ogino, T., Kaji, T., Kawabata, M., Satoh, K., Tomoo, K., Ishida, T., Yamazaki, H., Ishidoh, K. and Kominami, E. (1999). Function of the propeptide region in recombinant expression of active procathepsin L in *Escherichia coli*. *J. Biochem.*, **126**, 78-83.

Ondr, J. K. and Pham, C. T. N. (2004). Characterization of murine cathepsin W and its role in cell-mediated cytotoxicity. *J. Biol. Chem.*, **279**, 27525-27533.

Öörni, K., Sneck, M., Brömme, D., Pentikäinen, M. O., Lindstedt, K. A., Mäyränpää, M., Aitio, H. and Kovanen, P. T. (2004). Cysteine protease cathepsin F is expressed in human atherosclerotic lesions, is secreted by cultured macrophages, and modifies low density lipoprotein particles *in vitro*. *J. Biol. Chem.*, **279**, 34776-34784.

Park, H., Hong, K.-M., Sakanari, J. A., Choi, J.-H., Park, S.-K., Kim, K.-Y., Hwang, H.-A., Paik, M.-K., Yun, K.-J., Shin, C.-H., Lee, J.-B., Ryu, J.-S. and Min, D.-Y. (2001). *Paragonimus westermani*: cloning of a cathepsin F-like cysteine proteinase from the adult worm. *Exp. Parasitol.*, **98**, 223-227.

Park, H., Kim, S.-I., Hong, K.-M., Kim, M.-J., Shin, C.-H., Ryu, J.-S., Min, D.-Y., Lee, J.-B. and Hwang, U. W. (2002). Characterization and classification of five cysteine proteinases expressed by *Paragonimus westermani*. *Exp. Parasitol.*, **102**, 143-149.

Park, S. Y., Lee, K. H., Hwang, Y. B., Kim, K. Y., Park, S. K., Hwang, H. A., Sakanari, J. A., Hong, K. M., Kim, S.-I. and Park, H. (2001). Characterization and large-scale expression of the recombinant cysteine proteinase from adult *Clonorchis sinensis*. *J. Parasitol.*, **87**, 1454-1458.

Pearce, E. J. (2003). Progress towards a vaccine for shistosomiasis. *Acta Trop.*, **86**, 309-313.

Pezzella, A. T., Yu, H. S. and Kim, J. E. (1981). Surgical aspects of pulmonary paragonimiasis. *Cardiovasc. Dis.*, **8**, 187-194.

Podobnik, M., Kuhelj, R., Turk, V. and Turk, D. (1997). Crystal structure of the wild-type human procathepsin B at 2.5 Å resolution reveals the native active site of a papain-like cysteine protease zymogen. *J. Mol. Biol.*, **271**, 774-788.

Polgá, L. (2004). Catalytic mechanisms of cysteine peptidases. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1072-1079.

Poulin, R. and Cribb, T. H. (2002). Trematode life cycles: short is sweet? *Trends Parasitol.*, **18**, 176-183.

Quraishi, O., Nägler, D. K., Fox, T., Sivaraman, J., Cygler, M., Mort, J. S. and Storer, A. C. (1999). The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide. *Biochemistry*, **38**, 5017-5023.

Quraishi, O. and Storer, A. C. (2001). Identification of internal autoproteolytic cleavage sites within the prosegments of recombinant procathepsin B and procathepsin S: contribution of a plausible unimolecular autoproteolytic event for the processing of zymogens belonging to the papain family. *J. Biol. Chem.*, **276**, 8118-8124.

Rawlings, N. D. and Barrett, A. J. (1994). Families of cysteine peptidases. *Meth. Enzymol.*, **244**, 461-486.

Rawlings, N. D. and Barrett, A. J. (2004). Introduction: the clans and families of cysteine proteases. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1051-1071.

Roberts, L. S. and Janovy, J. Jr. (2000). *Foundations of Parasitology.* 6th ed. McGraw-Hill Higher Education, USA.

Robinson G. and Threadgold, L. T. (1975). Electron microscope studies of *Fasciola hepatica* – XII. The fine structure of the gastrodermis. *Exp. Parasitol.*, **37**

Roche, L., Dowd, A. J., Tort, J., McGonigle, S., McSweeney, A., Curley, G. P., Ryan, T. and Dalton, J. P. (1997). Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **245**, 373-380.

Roche, L., Tort, J. and Dalton, J. P. (1999). The propeptide of *Fasciola hepatica* cathepsin L is a potent and selective inhibitor of the mature enzyme. *Mol. Biochem. Parasitol.*, **98**, 271-277.

Rokni, M. B., Massoud, J., O'Neill, S. M., Parkinson, M. and Dalton, J.P. (2002). Diagnosis of human fasciolosis in the Gilan province of northern Iran. *Diag. Microbiol. Infect. Dis.*, **44**, 175-179.

Rozman, J., Stojan, J., Kuhelj, R., Turk, V. and Turk, B. (1999). Autocatalytic processing of recombinant human procathepsin B is a bimolecular process. *FEBS Lett.*, **459**, 358-362.

Sajid, M. and McKerrow, J. H. (2002). Cysteine proteases of parasitic organisms. *Mol. Biochem. Parasitol.*, **120**, 1-21.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory Press, New York, USA.

Santamaría, I., Velasco, G., Pendás, A. M., Fueyo, A. and López-Otín, C. (1998). Cathepsin Z, a novel human cysteine proteinase with a short propetide domain and a unique chromosomal location. *J. Biol. Chem.*, **273**, 16816-16823.

Santamaría, I., Velasco, G., Pendás, A. M., Paz, A. and López-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. *J. Biol. Chem.*, **274**, 13800-13809.

Schmitt, M. E., Brown, T. A. and Trumpower, B. L. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 3091.

Shi, G.-P., Chapman, H. A., Bhairi, S. M., DeLeeuw, C., Reddy, V. Y. and Weiss, S. J. (1995). Molecular cloning of human cathepsin O, a novel endoproteinase and homologue of rabbit OC2. *FEBS Lett.*, **357**, 129-134.

Shi, G.-P., Bryant, R. A. R., Riese, R., Verhelst, S., Driessen, C., Li, Z., Brömme, D., Ploegh, H. L. and Chapman, H. A. (2000). Role for cathepsin F in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages. *J. Exp. Med.*, **191**, 1177-1186.

Shin, M. H., Chung, Y.-B. and Kita, H. (2005). Degranulation of human eosinophils induced by *Paragonimus westermani*-secreted protease. *Kor. J. Parasitol.*, **43**, 33-37.

Shinde, U. P., Liu, J. J. and Inouye, M. (1997). Protein memory through altered folding mediated by intramolecular chaperones. *Nature*, **389**, 520-522.

Smith, A. M., Dowd, A. J., McGonigle, S., Keegan, P. S., Brennan, G., Trudgett, A. and Dalton, J. P. (1993a). Purification of a cathepsin L-like proteinase secreted by adult *Fasciola hepatica*. *Mol. Biochem. Parasitol.*, **62**, 1-8.

Smith, A. M., Dowd, A. J., Heffernan, M., Robertson, C. D. and Dalton, J. P. (1993b). *Fasciola hepatica*: a secreted cathepsin L-like proteinase cleaves host immunoglobulin. *Int. J. Parasitol.*, **23**, 977-983.

Smith, A. M., Carmona, C., Dowd, A. J., McGonigle, S., Acosta, D. and Dalton, J. P. (1994). Neutralisation of the activity of a *Fasciola hepatica* cathepsin L proteinase by anti-cathepsin L antibodies. *Parasite Immunol.*, **16**, 325-328.

Smooker, P. M., Whisstock, J. C., Irving, J. A., Siyaguna, S., Spithill, T. W. and Pike, R. N. (2000). A single amino acid substitution affects the substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Sci.*, **9**, 2567-2572.

Smyth, J. D. and Halton, D. W. (1983). *The Physiology of Trematodes*, 2nd ed., Cambridge University Press, Cambridge, UK.

Spithill, T. W. and Dalton, J. P. (1998). Progress in development of human liver fluke vaccines. *Parasitol. Today*, **14**, 224-228.

Tepel, C., Brömme, D., Herzog, V. and Brix, K. (2000). Cathepsin K in thyroid epithelial cells: sequence, localization and possible function in extracellular proteolysis of thyroglobulin. *J. Cell. Sci.*, **113**, 4487-4498.

Tkalcevic, J., Ashman, K. and Meeusen, E. (1995). *Fasciola hepatica*: rapid identification of newly excysted juvenile proteins. *Biochem. Biophys. Res. Comm.*, **213**, 169-174.

Tort, J., Brindley, P. J., Knox, D., Wolfe, K. H. and Dalton, J. P. (1999). Proteinases and associated genes of parasitic helminths. *Adv. Parasitol.*, **43**, 161-266.

Turk, B., Turk, D. and Turk, V. (2000). Lysosomal cysteine proteases: more than scavengers. *Biochim. Biophys. Acta*, **1477**, 98-111.

Turk, B., Turk, D., Dolenc, I. and Turk, V. (2004). Dipeptidyl-peptidase I. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1192-1196.

Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M. Turk, V. (1996). Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.*, **384**, 211-214.

Turk, D., Gunčar, G., Podobnik, M. and Turk, B. (1998). Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.*, **379**, 137-147.

Turk, D. and Gunčar, G. (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 203-213.

Turk, V., Turk, B. and Turk, D. (2001). Lysosomal cysteine proteases: facts and opportunities. *EMBO J.*, **20**, 4629-4633.

Velasco, G., Ferrando, A. A., Puente, X. S., Sánchez, L. M. and López-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. *J. Biol. Chem.*, **269**, 27136-27142.

Velasco, G. and López-Otín, C. (2004b). Cathepsin O. In: *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. Jr., eds.). 2nd ed. Elsevier Academic Press, London, 1102-1103.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M.-C., Storer, A. C. and Thomas, D. Y. (1995). Processing of the papain precursor: the ionization state of a conserved amino acid motif within the pro region participates in the regulation of intramolecular processing. *J. Biol. Chem.*, **270**, 10838-10846.

Wang, B., Shi, G.-P., Yao, P. M., Li, Z., Chapman, H. A. and Brömme, D. (1998). Human cathepsin F: molecular cloning, functional expression, tissue localization, and enzymatic characterization. *J. Biol. Chem.*, **273**, 32000-32008.

Wex, T., Bühling, F., Wex, H., Günther, D., Malfertheiner, P., Weber, E. and Brömme, D. (2001). Human cathepsin W, a cysteine protease predominantly expressed in NK cells, is mainly localized in the endoplasmic reticulum. *J. Immunol.*, **167**, 2172-2178.

Wiederanders, B. (2003). Structure-function relationships in class CA1 cysteine peptidase propeptides. *Acta Biochim. Pol.*, **50**, 691-713.

Wijffels, G. L., Panaccio, M., Salvatore, L., Wilson, L., Walker, I. D. and Spithill, T. W. (1994). The secreted cathepsin L-like proteinases of the trematode, *Fasciola hepatica*, contain 3-hydroxyproline residues. *Biochem. J.*, **299**, 781-790.

Wilson, L. R., Good, R. T., Panaccio, M., Wijffels, G. L., Sandeman, R. M. and Spithill, T. W. (1998). *Fasciola hepatica*: characterization and cloning of the major cathepsin B protease secreted by newly excysted juvenile liver fluke. *Exp. Parasitol.*, **88**, 85-94.

Yamasaki, H. and Aoki, T. (1993). Cloning and sequence analysis of the major cysteine protease expressed in the trematode parasite *Fasciola sp. Biochem. Mol. Biol. Int.*, **31**, 537-542.

Yamasaki, H., Mineki, R., Murayama, K., Ito, A. and Aoki, T. (2002). Characterisation and expression of the *Fasciola gigantica* cathepsin L gene. *Int. J. Parasitol.*, **32**, 1031-1042.

Zhang, R., Yoshida, A., Kumagai, T., Kawaguchi, H., Maruyama, H., Suzuki, T., Itoh, M., El-Malky, M. and Ohta, N. (2001). Vaccination with calpain induces a T_H1-biased protective immune response against *Schistosoma japonicum*. *Infect. Immun.*, **69**, 386-391.

Appendix A: Publications

Conference Presentations:

Khaznadji, E., Collins, P. R., Dalton, J. P. and Moiré, N. (2002). Expression and characterization of two cystatin domains from a new cysteine protease inhibitor expressed by *Fasciola hepatica*. Poster presented at 10th International Congress of Parasitology, Vancouver, BC, Canada, August 2002. Oral presentation based on this by E. Khaznadji.

Collins, P. R., Donnelly, S. M., Stack, C. N., O'Neill, S. M., Larkin, N., Ryan, T., Dalton, J. P., McIlwaine, A., Mousley, A., Maule, A. G. and Brennan, G. P. (2003). *Fasciola hepatica* cathepsin L-like proteases: biology, fnction, and potential in the development of first generation liver fluke vaccines. Oral presentation by P. R. Collins at The British Society for Parasitology Spring Meeting, Manchester, UK, 6th to 9th April 2003.

Collins, P. R., Stack, C. N., Donnelly, S. M., O'Neill, S. M., Mulcahy, G., Brinen, L. and Dalton, J. P. (2004). Functional expression and characterisation of *Fasciola hepatica* cathepsin L proteases in the yeast *Pichia pastoris*. Oral presentation by P. R. Collins at The Australian Society for Parasitology 46th Annual Scientific Meeting, Fremantle, WA, Australia, 26th to 30th September 2004.

Published papers:

Dalton, J. P., O'Neill, S. M., Stack, C. N., Collins, P. R., Walshe, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D., Khaznadji, E., Moiré, N., Brennan, G., Mousley, A., Kreshchenkoh, N., Maule, A. G. and Donnelly, S. M. (2003). *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *Int. J. Parasitol.*, **33**, 1173-1181.

Collins, P. R., Stack, C. M., O'Neill, S. M., Doyle, S., Ryan, T., Brennan, G. P., Mousley, A., Stewart, M., Maule, A. G., Dalton, J. P. and Donnelly, S. (2004). Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J. Biol. Chem.*, **279**, 17038-17046.

Khaznadji, E., Collins P. R., Dalton, J. P., Bigot, Y. and Moiré, N. (2005). A new multi-domain member of the cystatin superfamily expressed by *Fasciola hepatica*. *Int. J. Parasitol*, in press.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
International Journal for Parasitology 33 (2003) 1173–1181
www.parasitology-online.com

Invited review

Fasciola hepatica cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines

John P. Dalton[a,b,*], Sandra O. Neill[a,c], Colin Stack[a,d], Peter Collins[a], Alan Walshe[a], Mary Sekiya[a], Sean Doyle[d], Grace Mulcahy[b,e], Deborah Hoyle[f], Eric Khaznadji[g], Nathalie Moiré[g], Gerard Brennan[h], Angela Mousley[h], Natalia Kreshchenko[h], Aaron G. Maule[h], Sheila M. Donnelly[a]

[a]*Molecular Parasitology Unit, School of Biotechnology, Dublin 9, Ireland*
[b]*Ildana Biotech, INVENT, Dublin City University, Glasnevin, Dublin 9, Ireland*
[c]*School of Nursing, Dublin City University, Dublin 9, Ireland*
[d]*Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland*
[e]*Department of Veterinary Microbiology and Parasitology, Faculty of Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland*
[f]*Centre for Tropical Veterinary Medicine, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, Midlothian EH25 9RG, Scotland, UK*
[g]*INRA, UR86 "Bio-Agresseurs, Santé et Environnement", 37 380 Nouzilly, France*
[h]*School of Biology and Biochemistry, The Queen's University of Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast BT9 7BL, UK*

Received 11 November 2002; received in revised form 15 May 2003; accepted 16 June 2003

Abstract

Fasciola hepatica secretes cathepsin L proteases that facilitate the penetration of the parasite through the tissues of its host, and also participate in functions such as feeding and immune evasion. The major proteases, cathepsin L1 (FheCL1) and cathepsin L2 (FheCL2) are members of a lineage that gave rise to the human cathepsin Ls, Ks and Ss, but while they exhibit similarities in their substrate specificities to these enzymes they differ in having a wider pH range for activity and an enhanced stability at neutral pH. There are presently 13 *Fasciola* cathepsin L cDNAs deposited in the public databases representing a gene family of at least seven distinct members, although the temporal and spatial expression of each of these members in the developmental stage of *F. hepatica* remains unclear. Immunolocalisation and in situ hybridisation studies, using antibody and DNA probes, respectively, show that the vast majority of cathepsin L gene expression is carried out in the epithelial cells lining the parasite gut. Within these cells the enzyme is packaged into secretory vesicles that release their contents into the gut lumen for the purpose of degrading ingested host tissue and blood. Liver flukes also express a novel multi-domain cystatin that may be involved in the regulation of cathepsin L activity. Vaccine trials in both sheep and cattle with purified native FheCL1 and FheCL2 have shown that these enzymes can induce protection, ranging from 33 to 79%, to experimental challenge with metacercariae of *F. hepatica*, and very potent anti-embryonation/hatch rate effects that would block parasite transmission. In this article we review the vaccine trials carried out over the past 8 years, the role of antibody and T cell responses in mediating protection and discuss the prospects of the cathepsin Ls in the development of first generation recombinant liver fluke vaccines.

Keywords: Helminths; Trematodes; Parasites; Cathepsins; Proteases; Vaccines; Immunology; Biochemistry

1. Introduction

The trematode parasites *Fasciola hepatica* (temperate) and *Fasciola gigantica* (tropical) are the causative agents of liver fluke disease (fasciolosis) in cattle and sheep. Infection causes world-wide economic losses of approximately US$2,000 million annually to the agricultural sector (Spithill et al., 1999a). Fasciolosis is also an emerging pathogen of humans particularly in the South American countries Bolivia, Peru and Equador, in Egypt and Iran (O'Neill et al., 1998; Mas-Coma et al., 1999; Rokni et al., 2002). It is estimated that 2.4 million people are infected with liver fluke world-wide (Mas-Coma et al., 1999).

* Corresponding author. Tel.: +353-1-7005407; fax: +353-1-7005412.
E-mail address: john.dalton@dcu.ie (J.P. Dalton).

0020-7519/$30.00
doi:10.1016/S0020-7519(03)00171-1

Fasciolosis is acquired following the ingestion of vegetation or water contaminated with the encysted infectious liver fluke larvae, known as metacercariae. Metacercariae excyst in the intestines, burrow through the intestinal wall and migrate into the liver tissue where they spend about 8–12 weeks feeding on host tissue and blood and consequently causing extensive haemorrhaging and perforations. This acute stage of disease can result in death in highly infected sheep, but death is rarely seen in cattle. After this period the parasites move into the bile ducts where they complete their growth and maturation. The hermaphroditic adult liver flukes puncture the wall of the bile duct and feed on blood that provides the nutrient for the production of enormous numbers of eggs that are carried into the intestine with the bile fluids and are passed onto pasture with the faeces. An aquatic larval stage hatches from the eggs and infects an intermediate mud-snail host, such as *Lymnaea truncatula*. After several developmental and multiplication stages within the snail the parasites emerge and become encysted on vegetation and thus continue the cycle (Andrews, 1999).

A number of chemicals are commercially available for the treatment of animal fasciolosis, including closantal, clorsulon, rafoxanide, nitroxynil and triclabendazole (see Fairweather and Boray, 1999). Following dosing, the therapeutic concentration of these drugs in the blood is reached for approximately 1–2 days and hence only current parasitic infections are cleared. Animals on contaminated pastures are then susceptible to re-infection. Triclabendazole requires particular mention since this is the only drug effective against early stage parasites, i.e. the stages that migrate through the liver and induce acute fasciolosis. The drug was launched in Ireland in 1983 under the brand name Fasinex (Ciba-Geigy, now Novartis Animal Health) and has enjoyed substantial commercial success. However, in 1995 an anonymous letter to the *Irish Farmer's Journal* (Anon., 1995) reported the presence of triclabendazole-resistant liver flukes on a farm in Co. Sligo, an area where fasciolosis is endemic. If verified this provides an example of how quickly (within 12 years) drug resistance to anthelminthics can appear in the field. Triclabendazole resistance has now been reported in Australia (Overend and Bowen, 1995; Fairweather and Boray, 1999), Scotland (Mitchell et al., 1998) and the Netherlands (Moll et al., 2000; Gaasenbeek et al., 2001).

Apart from the appearance of drug resistant parasites, other pressures in the form of EU regulations, consumer concerns for 'greener' food and for animal welfare, and environmental issues regarding the passage of chemical residues onto pastures and into waterways will make the control of helminth diseases, including fasciolosis, with chemical drugs unsustainable. The future of liver fluke control will almost certainly depend on the development of a protective vaccine; the challenge will be to develop and commercialise such a product that will give a benefit that is visible to farmers (i.e. enhanced performance of animals, reduced deaths, etc.) at a competitive price (i.e. at a cost as low as drug treatments). It must be remembered, however, that vaccines provide 'protection' against infection and thereby have an advantage over drugs (particularly those that are effective against the adult liver fluke stages alone) that only provide a 'treatment'. Moreover, vaccines are considered safe and environmentally friendly.

Recent advances in molecular biology (proteomics and genomics) have made it relatively easy to identify and isolate potential vaccine components but in the case of liver flukes there is still a dearth of vaccine candidates that have been tried and tested in the target species, i.e. sheep and cattle. Research on a number of molecules including cathepsin Ls, glutathione *S*-transferase (GST), leucine aminpeptidase (LAP) and fatty acid binding proteins (FABP), however, has demonstrated the feasibility of inducing protective responses in laboratory and large animal models (reviewed by Spithill and Dalton, 1998; Spithill et al., 1999a). This article will review the work of our laboratory and others on liver fluke cathepsin L proteases and update the progress towards a vaccine.

2. Biochemical characteristics of cathepsin Ls

Fasciola hepatica cathepsin L proteases were the first of this class of enzyme to be described in helminths (Smith et al., 1993). It is now known that this enzyme class is expressed by all parasitic worms and in many cases they are secreted and hence are pivotal in the host–parasite interplay. However, they appear to be particularly prominent in trematodes as compared to nematodes where cathepsin B-like enzymes tend to be expressed in greater amounts (Tort et al., 1999).

Enzyme activity attributable to cysteine proteases was detected in our laboratory in extracts of *F. hepatica* parasites and ES products over a decade ago (Dalton and Heffernan, 1989). However, a more recent characterisation of these extracts using synthetic peptide substrates, such as Z-Phe-Arg-NHMec, demonstrated that the dominant activity was cathepsin L-like. Two distinct cathepsin L proteases have been purified in our laboratory from culture medium in which adult *F. hepatica* parasites were incubated and shown to be the predominant proteolytic activity secreted by the worms (Smith et al., 1993; Dowd et al., 1994). Subsequently, transcripts encoding these two major cathepsin L-like proteases, termed FheCL1 and FheCL2, have been isolated from cDNA libraries by immunoscreening with anti-sera prepared against the native enzymes (Dowd et al., 1997; Roche et al., 1997). These transcripts encode zymogens with structures that are similar to those of vertebrate cathepsin Ls in that they consist of a hydrophobic signal peptide (12–20 residues), a pro-segment (100 residues) and a mature enzyme (200 residues) (see Fig. 1). Unlike mammalian cathepsin Ls, FheCL1 and FheCL2 do not possess potential *N*-glycosylation sites in their mature

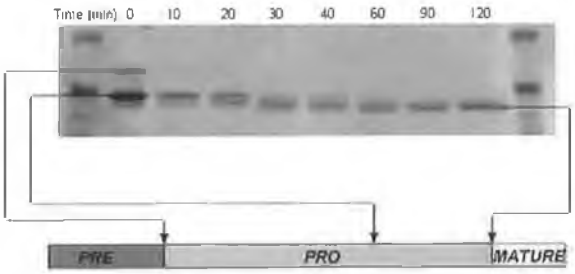


Fig. 1. Activation of Ni-NTA-agarose purified *P. pastoris*-expressed recombinant FheCL1 at pH 5.5. Purified recombinant FheCL1 resolves as two bands of 38 (upper) and 34 kDa (lower, arrowed) that represent the full pre-pro-mature and a truncated pro-mature form of the enzyme, respectively. While this preparation exhibits cathespin L activity it can be further activated to the fully processed mature enzyme (last lane on right) by incubating it for a period of 1–2 h in 0.1 M sodium acetate, pH 5.5. The positions of the major proteins resolved in the gel with respect to the cathepsin L pro-enzyme are shown schematically and were determined by N-terminal sequencing.

enzyme portion and electrophoretic/immunoblot analyses suggest that they are not glycosylated.

The substrate specificity of the cathepsin Ls was examined using synthetic fluorogenic peptide substrates that are classically used to characterise mammalian cathepsin Ls, Ss, Ks (Z-Phe-Arg-NHMec), cathepsin B (Z-Arg-Arg-NHMec) and cathepsin H (Z-Arg-NHMec) (Barrett and Kirschke, 1981). Both enzymes demonstrated high catalytic efficiency (k_{cat}/K_m) for Z-Phe-Arg-NHMec, but showed little or no activity against the latter two substrates. This substrate specificity is consistent with their grouping with the non-cathepsin B proteases; however, FheCL2 could also cleave substrates that contained proline in the P_2 position, and the substrate Z-Gly-Pro-Arg is used to distinguish between FheCL1 and FheCL2 activities (Dowd et al., 1994, 1997).

The *F. hepatica* cathepsin L-like proteases are active over a wide pH range (pH 3.0–8.0), and stability studies demonstrated that FheCL1 retains 100% of its activity when incubated at 37 °C for 24 h. In contrast, human cathepsin L is irreversibly inactivated above pH 7.0, a property that is thought to protect cells against uncontrolled proteolysis in the event of accidental leakage from the lysosomes (Dowd et al., 2000). The stability of the liver fluke enzymes may reflect the fact that the parasite employs them to generate a migratory path through a large tissue mass, the liver, on its way to the bile ducts.

3. The cathepsin L gene family

Early studies by Heussler and Dobbelaere (1994) indicated that *F. hepatica* expressed several cathepsin L genes, but since this report the number of cathespin L cDNA entries into the public databases has gradually grown so that at writing there are 13 known *F. hepatica* cathepsin genes. The number of *F. gigantica* cathepsin L entries in the databases has also grown mainly due to the efforts of Grams et al. (2001) and Yamasaki et al. (2002), and unsurprisingly many of these genes appear to be homologues of the *F. hepatica* genes. The family as it stands is presented in Fig. 2; some of the encoded polypeptides

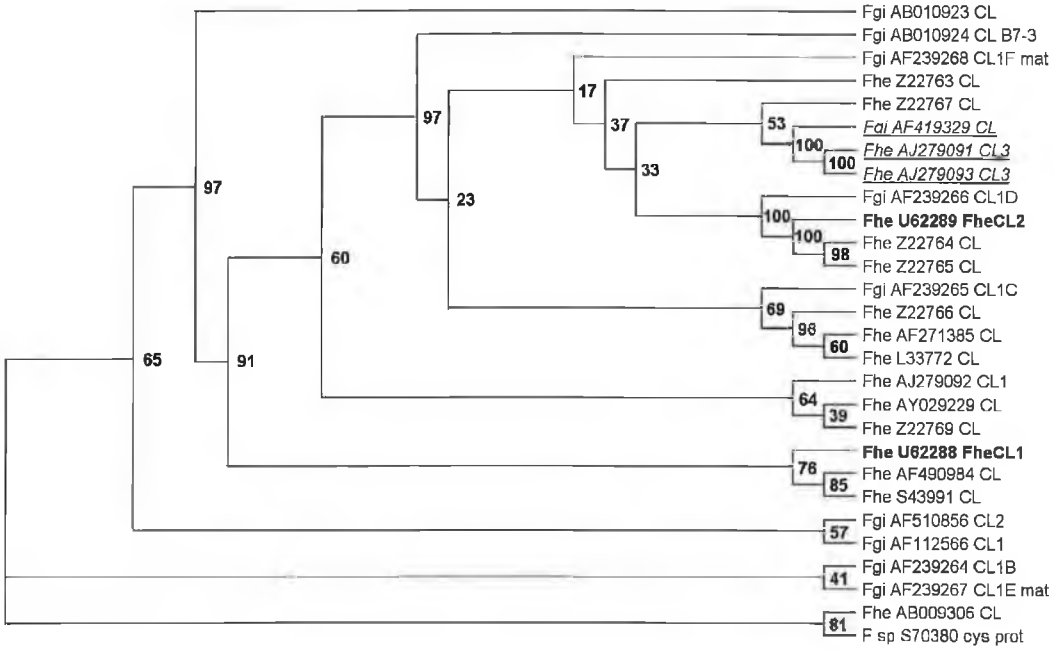


Fig. 2. Dendrogram of *F. hepatica* and *F. gigantica* cathepsin L gene family. Relationship of the cathepsin L cDNAs (public database accession numbers at end of branches). FheCL1 and FheCL2 are highlighted in bold. Sequences isolated from cDNA from newly excysted juveniles are italicised. Samples were analysed using 1,000 bootstrap replicates (numbers represent values obtained).

differ in only a few amino acids and may represent alleles rather than distinct genes. Nevertheless, it is clear that at least seven different cathepsin L genes are expressed by these trematodes. Phylogenetic studies of this gene family revealed that they belong to a clade that includes the mammalian cathepsin Ls, cathepsin Ss and cathepsin Ks (the parasite enzymes show 40–55% similarity with the mammalian homologues) (Tort et al., 1999) Of particular note is that the newly excysted juvenile (NEJ)-specific cathepsin L cDNAs (two from *F. hepatica* and one from *F. gigantica*) cluster together in the dendrogram shown in Fig. 2, which may suggest that cathepsin L expression in NEJ is limited to two genes whereas in the adult stages many more genes are expressed.

It is difficult to assess the temporal expression and localisation of each of the cathepsin L gene products because their close nucleic acid/amino acid sequences similarities preclude the development of specific PCR/antibody reagents. However, in situ hybridisation studies by Grams et al. (2001) with *F. gigantica* and by us with *F. hepatica* (Fig. 3) are consistent with immunolocalisation studies (Smith et al., 1993) and show that the vast majority of cathepsin L proteases are expressed in the cells lining the parasite's guts. It is possible that more than FheCL1 and FheCL2 proteases are secreted by the parasite, which would agree with the 2-D electrophoresis analysis by Wijffels et al. (1994) and Jefferies et al. (2001) that shows the presence of multiple cathepsin L proteases in adult *F. hepatica* ES products. The structure of one of the *F. gigantica* genes was recently determined by Yamasaki et al. (2002), and consists of four exons and three introns of similar size and arrangement to those found in crustacean and mammalian cathepsin Ls.

4. Regulation of cathepsin L activity

Enzymatic assays have shown that cathepsin L activity is expressed in the infective stage of the parasite, the NEJ, in the immature developmental stages that migrate through the liver parenchyma and in adults that reside in the bile ducts (Carmona et al., 1993; Smith et al., 1993). A proteomics approach to characterising NEJ proteins confirmed the presence of cathepsin L proteases in this stage (Tkalcevic et al., 1995) and now full-length *F. hepatica* and *F. gigantica* NEJ-specific cathepsin L sequences are available in the public databases (Fig. 2).

As mentioned above the proteases are synthesised in the epithelial cells that line the parasite gut (Fig. 3). Within these cells the proteases are packaged into secreted vesicles in high concentration ready for secretion. During trafficking to these secretory vesicles the activity (and stability) of the cathespin Ls is controlled by the presence of the pro-peptide on the protein (as has been demonstrated for mammalian cathepsin Ls) (Roche et al., 1997). It remains unclear whether the proteases are activated before they reach the vesicles or following their secretion by the NEJ and by the adult.

It is possible that other mechanisms exist to protect the parasite's cells from unwanted protease activity. Recently, Khaznadji, Collins, Dalton and Moire (unpublished) discovered a novel molecule that contains five cystatin

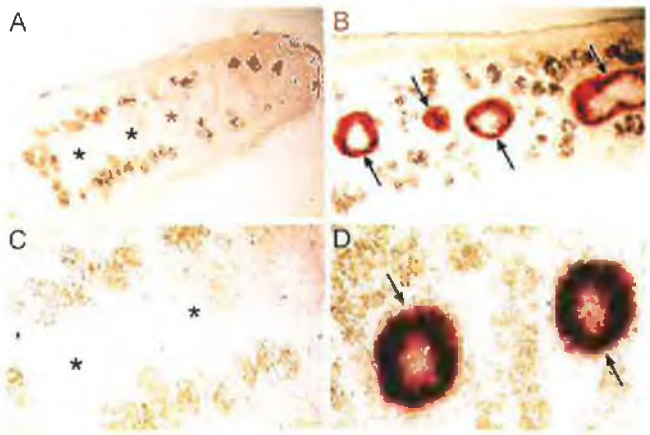


Fig. 3. Localisation of cathepsin L proteases by in situ hydridisation. Light micrographs showing the expression of cathepsin L protease in transverse sections (20 μM) of adult *F. hepatica*. In situ hybridisation was performed using digoxigenin-labelled DNA probes to cathepsin L protease gene. Specific sense [5′-tgtggctcctgttgggcattc-3′] and anti-sense [5′-ggattcggtttccaatcc-3′] primers were designed to amplify a region of 195 bp (including primers) from *F. hepatica* cathepsin L genes. Each probe (sense strand and anti-sense strand) was generated using asymmetrical PCR and digoxigenin-labelled di-nucleotide triphosphates (dNTPs) (Roche Diagnostics); cryostat sections were stained following overnight hybridisation at 50 °C. Gene expression can be seen as a dark red/brown chromogen formed by substrate cleavage by alkaline phosphatase conjugated to anti-digoxigenin antibodies bound to the probes. All sections were treated in the same way except that A and C were exposed to the sense strand probes (negative controls) whereas B and D were exposed to the anti-sense strand probes. Asterisks indicate gut lumen and arrows indicate positive staining/cathepsin L expression. (A) Low power micrograph showing unstained gut. Mag. × 20. (B) Low power micrograph showing positive staining in four gut. Mag. × 20. (C) High power micrograph showing several unstained gut. Mag. × 40. (D) High power micrograph showing strong positive staining in 2 gut. Mag. × 40.

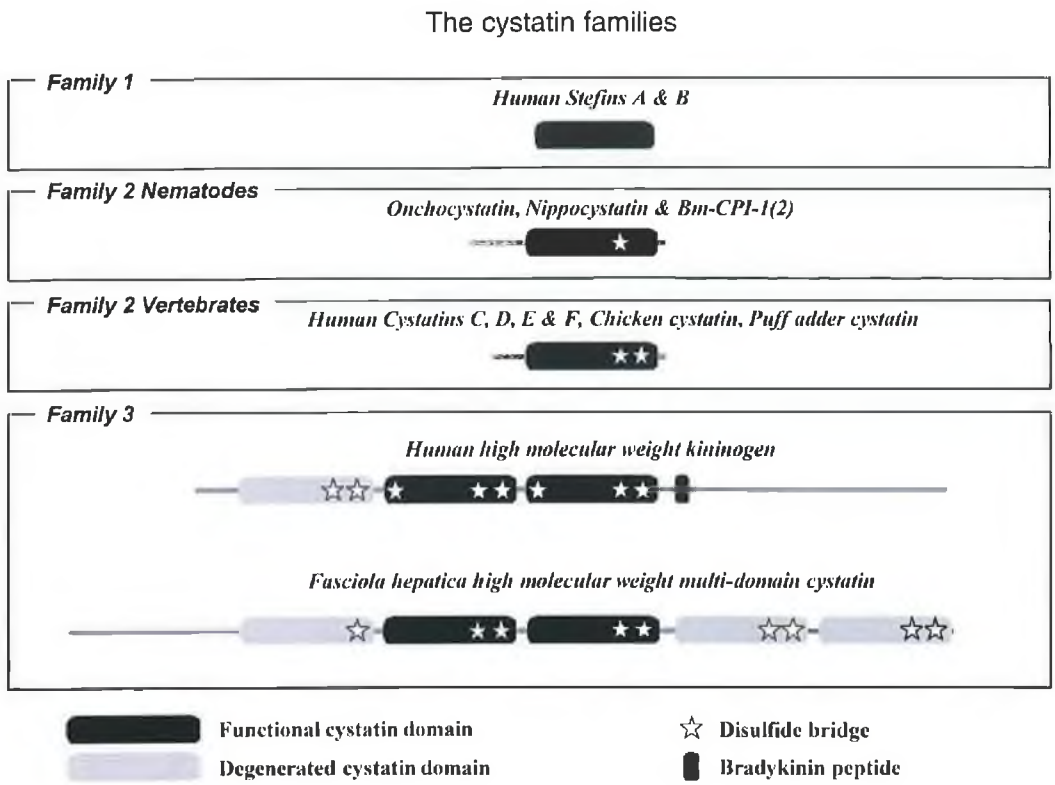


Fig. 4. The novel multi-domain *F. hepatica* cystatin. The Nematode cystatins-like Onchocystatin from *Onchocerca volvulus*, Nippocystatin from *Nippostrongylus brasiliensis* and Bm-CPI-1 from *Brugia malayi* are related to the cystatin family 2 of the cystatin superfamily (see Maizels et al., 2001). The *F. hepatica* multi-domain cystatin, however, is structurally related to the kininogen family (or cystatin family 3) but may be the first-described member of a new cystatin family.

domains, two of which are potent inhibitors of cathepsin L (the structure of the other three cystatin domains appears to be degenerate, Fig. 4). This multi-domain cystatin is expressed in NEJ and adult flukes but its expression is greater in the former stage. The cathepsin L protease may be required in large amounts by the NEJ to allow them quickly to invade the host intestinal wall following excystment, and a corresponding cystatin inhibitor may be necessary to protect the parasite from its own proteolytic activity during its storage or after secretion.

5. Functions of the cathepsin Ls

Cathepsin Ls that are secreted into the gut lumen following ingestion of host blood and liver tissue perform the task of digesting host tissues. The breakdown products of this endoproteolysis are then absorbed into the epithelial cells where they may undergo further degradation by dipeptidyl peptidases and aminopeptidases to amino acids that are then used in the anabolism of parasite proteins, or in the case of adults, in the production of eggs (Tort et al., 1999).

Because liver flukes possess a blind-ending intestine, the contents of the gut must be emptied by frequent regurgitation (approximately every 3 h). Accordingly, the protease must reach the outside tissue where they may carry out additional important functions for the parasite, including the excavation of a migratory tract. In this context, we have demonstrated that liver fluke cathepsin Ls can efficiently cleave interstitial matrix proteins such as fibronectin, laminin and native collagen (Berasain et al., 1997). We have also shown that the parasite cathepsin Ls can cleave immunoglobulins precisely in the hinge region (thus separating the Fab from the Fc regions) and prevent the antibody-mediated attachment of eosinophils to the parasite surface and hence may aid in protecting the parasite from immune attack (Carmona et al., 1993; Smith et al., 1993; Berasain et al., 1997).

More recently, however, we have proposed that the secreted cathepsin Ls may be involved in suppression and/or modulation of Th1 immune responses and induction of non-protective host Th2 responses. An analysis of cytokine production by antigen-stimulated spleen cells of *F. hepatica*-infected mice showed that these are predominantly of the Th2 type, i.e. production of interleukin (IL)-4, IL-5 and IL-10, but little or no IFN-γ (O'Neill et al., 2000). This is consistent with immunological observations in cattle which show that in the early stages of infections mixed Th1/Th2 responses are observed but as infection progresses, a Th2

Table 1
Summary of vaccine trials with *F. hepatica* cathepsin L proteases

Antigen	Host	Dose	Protection (%)	Reduction in viable eggs (%)
FheCL1[a]	Cattle	10–200 μg × 3	38.2–69.4	ND
FheCL1[a]	Cattle	200 μg × 3	43	40–65
FheCL1 + haem fraction[a]	Cattle	200 μg × 3	52	80
FheCL1 + haem fraction[a]	Cattle	200 μg × 3	72.4	98
FheCL1 + FheCL1[b]	Cattle	200 μg × 3	54	55
FheCL1[c]	Sheep	100 μg × 2	34	71
FheCL2[c]	Sheep	100 μg × 2	33	81
FheCL2 + FheCL2[c]	Sheep	100 μg each × 2	60	ND
FheCL2 + FheCL2 + LAP[c]	Sheep	100 μg each × 2	79	ND
FheCL proteases[d]	Sheep	120 + 90 μg	0	69.4

ND, not determined.
[a] Dalton et al. (1996).
[b] Mulcahy et al. (1998).
[c] Piacenza et al. (1999).
[d] Wijffels et al. (1994).

response predominates (Mulcahy et al., 1999). However, at high levels of infection in mice (approximately 15 metacercariae) a marked suppression of Th1 responses was observed (O'Neill et al., 2000). Moreover, when mice were concurrently infected with *F. hepatica* and *Bordetella pertussis*, the Th1-specific response to the bacterial pathogen (assessed by measuring IFN-γ) was suppressed. This suppression led to a delay in the clearance of the bacterium in the co-infected mice as compared to mice infected with *B. pertussis* alone (Brady et al., 1999). Liver fluke infection also suppressed the Th1 responses to a *B. pertussis* whole cell vaccine (WCV) suggesting that the parasite may secrete Th1 suppressive factors. Subsequently, O'Neill et al. (2001) demonstrated that injection of purified cathepsin L just prior to vaccination with *B. pertussis* WCV blocked the production of IFN-γ by WCV-stimulated spleen cells. It is possible, therefore, that secreted cathepsin Ls play some immunomodulatory/immunosuppressive function by blocking the development of protective Th1 responses in the host and thus aiding the progression of non-protective Th2 responses that would favour the longevity of the parasite.

6. Vaccine trials with cathepsin L proteases

The involvement in biological roles that are essential to the parasite's survival in the host makes the *F. hepatica* cathepsin L proteases likely targets at which novel vaccines could be directed. Native FheCL1 and FheCL2 proteases could be purified from the ES products of adult worms by a series of conventional gel permeation and ion exchange chromatography steps and produced in sufficient quantities to allow experimental vaccine trials in cattle. In our first vaccine trial a dose-ranging experiment demonstrated that significant protection against liver fluke infection (average 54%) could be obtained and that doses of as little as 10 μg of protein, given on three occasions, could induce protective responses (Dalton et al., 1996). Subsequent trials, performed with 200 μg per dose, confirmed the protective nature of the cathepsin L proteases (Table 1). When the cathepsin L proteases were administered in conjunction with other liver fluke antigens, such as a haem-binding protein, protection levels of 73% were achieved (Dalton et al., 1996, Table 1). A further vaccine trial in cattle showed that a combination of FheCL1 and FheCL2 elicited 53% protection against challenge, and trials in sheep induced protection levels from 33 to 60% (Mulcahy et al., 1998; Piacenza et al., 1999, Table 1). When FheCL1 and FheCL2 were combined with *F. hepatica* leucine aminopeptidase (LAP) protection levels were 79% in sheep (Piacenza et al., 1999).

Additional beneficial aspects of the protection induced by the cathepsin L vaccines were observed. First, in all vaccine trials the proportion of liver flukes that did not develop to maturity was greater in vaccinates than in non-vaccinated controls. Consequently, the damage to the host's liver during acute infection was significantly reduced. Secondly, vaccination of both sheep and cattle also elicited a highly significant reduction (50–98%) of the parasite's ability to produce eggs, and those eggs that were synthesised showed reduced 'hatch rates' (Dalton et al., 1996; Mulcahy et al., 1998; Table 1). The implications of these findings are that by reducing the parasite burden of the host and, at the same time, blocking the synthesis of viable eggs by those parasites that do survive in vaccinated animals, the vaccine would have a profound effect on pasture contamination and hence disease transmission.

7. What is the mechanism behind protection?

Studies examining the immune responses of cattle that were given either experimental or natural infections of *F. hepatica* metacercariae showed that animals generate IgG1 antibodies and little or no IgG2 antibodies

(Mulcahy et al., 1999; Hoyle et al., 2002). These observations suggest that the immune responses to liver fluke antigens are highly polarised to a Th2 type. On the other hand, antibody responses to the cathepsin L-based vaccines include high titres of both IgG1 and IgG2 indicating that protection is associated with the induction of a Th1 responses or a mixed Th1/Th2 response (Mulcahy et al., 1998, 1999).

A recent analysis of the cathepsin L-specific IgG1 antibody responses in infected cattle revealed that these antibodies do not appear in the serum until 4 or 5 weeks post-infection. Moreover, animals given a second infection 4 weeks after the primary infection did not exhibit any boosting of immune responses to the cathepsin L proteases (Hoyle et al., 2002; Fig. 5). It is clear, therefore, that cathepsin Ls are not highly immunogenic in the early stages of infection such that, at this stage, they may be considered 'hidden antigens'. This may be an important strategy used by the parasite to prevent antibodies being generated to critical molecules. Accordingly, vaccinating animals with cathepsin Ls would elicit a high titre of anti-cathepsin L antibodies that would upset this evasion strategy and, therefore, would be detrimental to the early migratory stages of the parasite. Indeed, when animals are vaccinated with cathepsin L a positive correlation was observed between protection and antibody titres (Mulcahy et al., 1998; Fig. 6).

The effects of the vaccine on egg production and 'hatch rate' may be mediated by antibodies that inhibit parasite feeding by blocking cathepsin L activity, thereby preventing the acquisition of amino acids for the synthesis of egg proteins. Additionally and/or alternatively, since cathepsin Ls have been immunolocalised to the reproductive organs of the mature parasite (Spithill et al., 1999a; Wijffels et al., 1994) and may play a role in egg production,

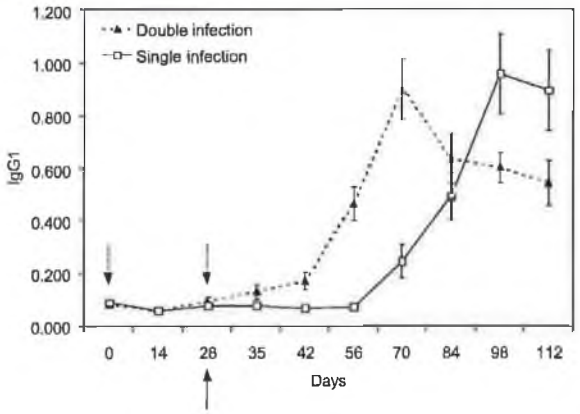


Fig. 5. IgG1 antibody response of experimental-infected cattle to FheCL1. Cattle were experimentally infected with metacercariae of *F. hepatica* either once (solid arrow) or twice (4 weeks apart, dotted arrows). Serum was collected weekly and assayed by ELISA for IgG1 antibodies to purified FheCL1 (IgG2 antibodies were either very low or undetectable). Antibody responses to FheCL1 in both experimental groups were observed 4–5 weeks after the primary infection.

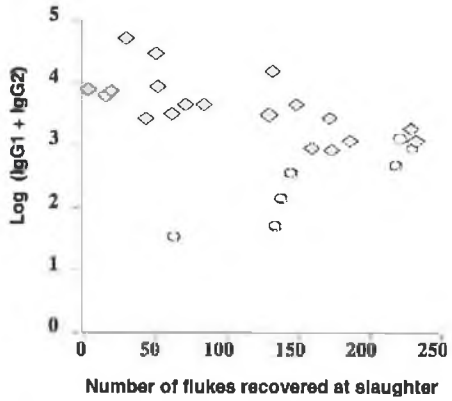


Fig. 6. Immune responses of FheCL1-vaccinated and non-vaccinated control cattle. Relationship between antibody titres and numbers of liver fluke recovered from livers of non-vaccinated animals and cattle vaccinated with cathepsin L proteases. In vaccinated animals, high total antibody titres (diamonds) correlate with lower fluke burdens (these antibodies consist of both IgG1 and IgG2). In contrast, in non-vaccinated control animals (circles) high antibody titres correlate with higher infections (these antibodies consist predominantly of IgG1, see text).

vaccine-induced antibodies may inhibit the enzyme and thereby have a direct effect on female reproductive structures.

While antibodies seem to be central to the protective effects of the cathepsin L vaccine, it is clear that the generation of a Th1 or Th1/Th2 type immune response is also critical. The induction of this type of response was dependant on the adjuvant employed in our studies i.e. Freund's Complete/Incomplete Adjuvant. However, it also seems likely that by blocking cathepsin L activity the vaccine prevents the parasite from suppressing the Th1 response of the host and from inducing the development of Th2 responses that are typical of natural infections.

8. Functional expression of cathepsin L proteases

While FheCL1 and FheCL2 could be purified from liver fluke ES products in quantities that allowed the detailed analysis of their biochemical and physico-chemical properties, and their assessment as vaccine candidates, it is obvious that for their future commercialisation a method of recombinant production is necessary. The production of cathepsin L proteases in a fully functional form requires expression of the pro-mature form of the enzyme since the pro-peptide region is required for correct folding of the molecule. Prokaryotic systems of expression, e.g. *E. coli*, are not suitable for the production of functionally active cathepsin proteases, although these systems have been employed to express non-active proteins, such as *F. gigantica* cathepsin L (Yamasaki et al., 2002), which were subsequently re-folded to active form. Unfortunately, re-folding methodologies are complex, cumbersome and expensive, and serve to obviate their application in commercial production schemes.

Yeast systems have proved more successful for the functional expression of many parasite proteases including the *F. hepatica* cathepsin Ls (Dalton et al., in press). We have reported that the yeast *Saccharomyces cerevisiae*, transformed with a full-length FheCL1 or FheCL2 cDNA, expressed and secreted functional enzyme into the culture medium from which homogeneous enzyme could be obtained by conventional purification techniques such as gel filtration and ion exchange chromatography (Dowd et al., 1997; Roche et al., 1997). While the functional expression of mammalian cathepsin proteases in *S. cerevisiae* has also been reported, these have necessitated the use of an α-factor pre- or pre-pro-signals. However, functional expression of FheCL1 and FheCL2 did not require such a fusion protein and N-terminal sequencing of expressed proteases revealed that they were processed in yeast to the mature enzyme by cleavage at the precise position as that observed for the native enzymes (Dowd et al., 1997; Roche et al., 1997). Notwithstanding these advances, the expression of cathepsin L protease was low (approximately 100 μg/l) and hence still not suitable for large-scale production. More recently, however, we have employed the yeast *Pichia pastoris* for protease expression using vectors that include the yeast α-factor pre-pro-signals (Invitrogen). High-level expression of FheCL1 has been achieved (800 mg/l) and the protease, which carries a His_6 tag, can be isolated by a one-step affinity chromatography procedure on Ni-NTA-agarose (Fig. 1). The protease is expressed as a pro-enzyme together with a semi-processed form, both of which can be activated at low pH to fully active mature enzyme (Fig. 1). The physico-chemical properties (substrate specificity, pH optima, stability, etc.) and molecular size of cathepsin L are similar to the native enzyme, and more importantly with respect to vaccines, no post-translational glycosylation occurs (unpublished data). These experiments demonstrate that liver fluke FheCL1 and FheCL2 could be produced at a commercial level using yeast as the heterologous expression system.

9. Concluding remarks

In our laboratory the cathepsin L proteases of *F. hepatica* have given consistent positive results when used as vaccines against fasciolosis in cattle and sheep and therefore have excellent prospects for the development of first generation anti-fluke vaccines. Vaccinations in sheep with a purified fraction of cysteine proteases of adult *F. hepatica* carried out by Wijffels et al. (1994), however, did not provide any protection against a challenge infection, although a >70% reduction of faecal eggs counts (FEC) was observed (Table 1). In addition, vaccination trials in cattle with *F. gigantica* cathepsin Ls did not protect animals against this fluke species (Spithill et al., 1999a). The discrepancies remain unresolved and may only be resolved if a comparative analysis of the immune responses to the vaccines was carried out. There are differences in vaccine formulations, doses and/or route of administration between the trials (see Spithill et al., 1999a,b) that may have induced a variety of effector mechanisms to the vaccines. Each of these factors will need to be investigated systematically during the developmental stages of a commercial vaccine.

Besides the cathepsin L proteases there are several other molecules that are potential vaccine candidates including GST, FABP (Spithill and Dalton, 1998) and LAP (Piacenza et al., 1999). While trials with these antigens have given some very variable results, no detailed analysis of immune responses to the vaccines has been carried out to determine immunological factors that correlate with protection/no protection. It is therefore still feasible that these molecules could be realised as a basis for commercial vaccines, either alone or in combination with other candidates, if the precise conditions for inducing protection were elucidated.

The commercialisation of a vaccine will require a system for production and downstream processing that is not so costly as to reduce the final product's competitiveness against current drug treatments. Both *E. coli* and yeast expression systems may be suitable for the production of vaccines for experimental trials, but before progressing into developmental phase the best/cheapest system for up-scale production/downstream processing will have to be considered for final commercialisation in an animal health market that already suffers from relatively low-valued/low profit-margin products (Dalton et al., in press). Nevertheless, the prospects for future vaccines are good, particularly considering the pressure on the industry to move away from chemical treatments.

References

Andrews, S.J., 1999. The life cycle of *Fasciola hepatica*. In: Dalton, J.P., (Ed.), Fasciolosis, CAB International, Wallingford, UK, pp. 1–30.

Anon., 18 March 1995. Resistance to fluke drug found on Sligo farm. Irish Farmer's J. 2.

Barrett, A.J., Kirschke, H., 1981. Cathespin B, cathepsin H and cathepsin L. Methods Enzymol. 80, 535–561.

Berasain, P., Goni, F., McGonigle, S., Dowd, A., Dalton, J.P., Frangione, B., Carmona, C., 1997. Proteinases secreted by *Fasciola hepatica* degrade extracellular matrix and basement membrane components. J. Parasitol. 83, 1–5.

Brady, M.T., O'Neill, S.M., Dalton, J.P., Mills, K.H.G., 1999. *Fasciola hepatica* suppresses protective Th1 responses against *Bordetella pertussis*. Infect. Immun. 67, 5372–5378.

Carmona, C., Dowd, A.J., Smith, A.M., Dalton, J.P., 1993. Cathespin L proteinase secreted by *Fasciola hepatica* in vitro prevents antibody-mediated eosinophil attachment to newly excysted juveniles. Mol. Biochem. Parasitol. 62, 9–18.

Dalton, J.P., Heffernan, M., 1989. Thiol proteases released *in* vitro by *Fasciola hepatica*. Mol. Biochem. Parasitol. 35, 161–166.

Dalton, J.P., McGonigle, S., Rolph, T.P., Andrews, S.J., 1996. Induction of protective immunity in cattle against infection with *Fasciola hepatica* by vaccination with cathepsin L proteinase and haemoglobin. Infect. Immun. 64, 5066–5074.

Dalton, J.P., Brindley, P.J., Knox, D.P., Hotez, P.J., Brady, C.P., Donnelly, S., O'Neill, S.M., Mulcahy, G., Loukas, A., 2003. Helminth vaccines: from mining genomic information for vaccine targets to systems used for protein expression. Int. J. Parasitol. 33, 621–640.

Dowd, A.J., Dooley, M., Fagain, C., Dalton, J.P., 2000. Stability studies on the cathepsin L proteinase of the helminth parasite, *Fasciola hepatica*. Enzyme Microb. Technol. 27, 599–604.

Dowd, A.J., Smith, A.M., McGongle, S., Dalton, J.P., 1994. Purification and characterisation of a second cathepsin L proteinase secreted by the trematode *Fasciola hepatica*. Eur. J. Biochem. 223, 91–98.

Dowd, A.J., Tort, J., Roche, L., Dalton, J.P., 1997. Isolation of a cDNA encoding cathepsin L2 and functional expression in *Saccharomyces cerevisiae*. Mol. Biochem. Parasitol. 88, 163–174.

Fairweather, I., Boray, J.C., 1999. Mechanism of fasciolicide action and drug resistance in *Fasciola hepatica*. In: Dalton, J.P., (Ed.), Fasciolosis, CAB International, Wallingford, UK, pp. 225–276.

Gaasenbeek, C.P.H., Moll, L., Cornelissen, J.B.W.J., Vellema, P., Borgsteede, F.H.M., 2001. An experimental study on triclabendazole resistance of *Fasciola hepatica* in sheep. Vet. Parasitol. 95, 37–43.

Grams, R., Vichasri-Grams, S., Sobhorn, P., Upatham, E.S., Viyanant, V., 2001. Molecular cloning and characterisation of cathepsin L encoding genes from *Fasciola gigantica*. Parasitol. Int. 50, 105–114.

Heussler, T.V., Dobbelaere, A.E.D., 1994. Cloning of a protease gene family of *Fasciola hepatica* by the polymerase chain reaction. Mol. Biochem. Parasitol. 64, 11–23.

Hoyle, D.V., Dalton, J.P., Chase-Topping, M., Taylor, D.W., 2002. Pre-exposure of cattle to drug-abbreviated *Fasciola hepatica* infections: the effect upon subsequent challenge infection and early immune response. Vet. Parasitol. 111, 65–82.

Jefferies, J.R., Campbell, A.M., van Rossum, A.J., Barrett, J., Brophy, P.M., 2001. Proteomic analysis of *Fasciola hepatica* excretory–secretory products. Proteomics 1, 1128–1132.

Mas-Coma, S., Bargues, M.D., Esteban, J.G., 1999. Human fasciolosis. In: Dalton, J.P., (Ed.), Fasciolosis, CAB International, Wallingford, UK, pp. 411–434.

Maizels, R.M., Gomez-Escobar, N., Gregory, W.F., Murray, J., Zang, X., 2001. Immune evasion genes from filarial nematodes. Int. J. Parasitol. 31, 889–898.

Mitchell, G.B.B., Maris, L., Bonniwell, M.A., 1998. Triclabendazole-resistant liver fluke in Scottish sheep. Vet. Rec. 143, 399.

Moll, L., Gassenbeek, C.P.H., Vellema, P., Borgsteede, F.H.M., 2000. Resistance of *Fasciola hepatica* against triclabendazole in cattle and sheep in the Netherlands. Vet. Parasitol. 91, 153–158.

Mulcahy, G., O'Connor, F., McGonigle, S., Dowd, A.J., Clery, D., Andrews, S.J., Dalton, J.P., 1998. Correlation of specific antibody titre and avidity with protection in cattle immunized against *Fasciola hepatica*. Vaccine 16, 932–939.

Mulcahy, G., Joyce, P., Dalton, J.P., 1999. Immunology of *Fasciola hepatica* infection. In: Dalton, J.P., (Ed.), Fasciolosis, CAB International, Wallingford, UK, pp. 341–376.

O'Neill, S.M., Parkinson, M., Strauss, W., Angles, R., Dalton, J.P., 1998. Immunodiagnosis of *Fasciola hepatica* infection (fascioliasis) in a human population in the Bolivian altiplano using purified cathepsin L cysteine protease. Am. J. Trop. Med. Hyg. 58, 417–423.

O'Neill, S.M., Brady, M.T., Callanan, J.J., Mills, K.H.G., Dalton, J.P., 2000. *Fasciola hepatica* infection down-regulates Th1 response in mice. Parasite Immunol. 22, 147–155.

O'Neill, S.M., Mills, K.H.G., Dalton, J.P., 2001. *Fasciola hepatica* cathepsin L cysteine proteinase suppresses *Bordetella pertussis*-specific interferon-γ production in vivo. Parasite Immunol. 23, 541–547.

Overend, D.J., Bowen, F.L., 1995. Resistance of *Fasciola hepatica* to triclabendazole. Aust. Vet. J. 72, 275–276.

Piacenza, L., Acosta, D., Basmadjian, I., Dalton, J.P., Carmoma, C., 1999. Vaccination with cathepsin L proteinases and with leucine aminopeptidase induces high levels of protection against fascioliasis in sheep. Infect. Immun. 67, 1954–1961.

Roche, L., Dowd, A.J., Tort, J., McGonigle, S., McSweeney, A., Curley, P., Ryan, T., Dalton, J.P., 1997. Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae*. Eur. J. Biochem. 245, 373–380.

Rokni, M.B., Massoud, J., O'Neill, S.M., Parkinson, M., Dalton, J.P., 2002. Diagnosis of human fasciolosis in the Gilan province of northern Iran: application of cathepsin L-ELISA. Diagn. Microbiol. Infect. Dis. 44, 175–179.

Smith, A.M., Dowd, A.J., McGonigle, S., Keegan, P.S., Brennan, G., Trudgett, A., Dalton, J.P., 1993. Purification of a cathepsin L proteinase secreted by adult *Fasciola hepatica*. Mol. Biochem. Parasitol. 62, 1–8.

Spithill, T.W., Dalton, J.P., 1998. Progress in the development of liver fluke vaccines. Parasitol. Today 14, 224–228.

Spithill, T.W., Smooker, P.M., Sexton, J.L., Bozas, E., Morrison, C.A., Creaney, J., Parsons, J.C., 1999a. Development of vaccines against *Fasciola hepatica*. In: Dalton, J.P., (Ed.), Fasciolosis, CAB International, Wallingford, UK, pp. 377–401.

Spithill, T.W., Smooker, P.M., Copeman, D.B., 1999b. *Fasciola gigantica*: epidemiology, control, immunology and molecular biology. In: Dalton, J.P., (Ed.), Fasciolosis, CAB International, Wallingford, UK, pp. 465–526.

Tkalcevic, J., Ashman, K., Meeusen, E.N.T., 1995. *Fasciola hepatica*: rapid identification of newly excysted juvenile proteins. Biochem. Biophys. Res. Commun. 213, 169–174.

Tort, J., Brindley, P.J., Knox, D., Wolfe, K.H., Dalton, J.P., 1999. Helminth proteinases and their associated genes. Adv. Parasitol. 43, 161–266.

Wijffels, G.L., Salvatore, L., Dosen, M., Waddington, J., Thompson, C., Campbell, N., Sexton, J., Wicker, J., Bowen, F., Friedel, T., Spithill, T.W., 1994. Vaccination of sheep with purified cysteine proteinases of *Fasciola hepatica* decreases worm fecundity. Exp. Parasitol. 78, 132–148.

Yamasaki, H., Mineki, R., Murayama, F., Ito, A., Aoki, T., 2002. Characterisation and expression of the *Fasciola gigantica* cathepsin L gene. Int. J. Parasitol. 32, 1031–1042.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 17, Issue of April 23, pp. 17038–17046, 2004

Printed in U.S.A.

Cathepsin L1, the Major Protease Involved in Liver Fluke (*Fasciola hepatica*) Virulence

PROPEPTIDE CLEAVAGE SITES AND AUTOACTIVATION OF THE ZYMOGEN SECRETED FROM GASTRODERMAL CELLS*

Received for publication, August 11, 2003, and in revised form, January 29, 2004
Published, JBC Papers in Press, January 30, 2004, DOI 10.1074/jbc.M308831200

Peter R. Collins‡§, Colin M. Stack‡¶‖, Sandra M. O'Neill‡, Sean Doyle¶, Thecla Ryan‡, Gerard P. Brennan, Angela Mousley**‡‡, Michael Stewart**, Aaron G. Maule**, John P. Dalton‡§§¶¶, and Sheila Donnelly‡‖‖**

*From the ‡School of Biotechnology, Dublin City University, Dublin 9, Republic of Ireland, the ¶Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Republic of Ireland, the **Parasitology Research Group, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast BT9 7BL, United Kingdom, and the §§Institute for the Biotechnology of Infectious Diseases (IBID), University of Technology, Sydney (UTS), Westbourne Street, Gore Hill, Sydney, New South Wales 2065, Australia*

The secretion and activation of the major cathepsin L1 cysteine protease involved in the virulence of the helminth pathogen *Fasciola hepatica* was investigated. Only the fully processed and active mature enzyme can be detected in medium in which adult *F. hepatica* are cultured. However, immunocytochemical studies revealed that the inactive procathepsin L1 is packaged in secretory vesicles of epithelial cells that line the parasite gut. These observations suggest that processing and activation of procathepsin L1 occurs following secretion from these cells into the acidic gut lumen. Expression of the 37-kDa procathepsin L1 in *Pichia pastoris* showed that an intermolecular processing event within a conserved GXNXFXD motif in the propeptide generates an active 30-kDa intermediate form. Further activation of the enzyme was initiated by decreasing the pH to 5.0 and involved the progressive processing of the 37 and 30-kDa forms to other intermediates and finally to a fully mature 24.5 kDa cathepsin L with an additional 1 or 2 amino acids. An active site mutant procathepsin L, constructed by replacing the Cys^{26} with Gly^{26}, failed to autoprocess. However, [Gly^{26}]procathepsin L was processed by exogenous wild-type cathepsin L to a mature enzyme plus 10 amino acids attached to the N terminus. This exogenous processing occurred without the formation of a 30-kDa intermediate form. The results indicate that activation of procathepsin L1 by removal of the propeptide can occur by different pathways, and that this takes place within the parasite gut where the protease functions in food digestion and from where it is liberated as an active enzyme for additional extracorporeal roles.

Fasciola hepatica is a helminth parasite that causes liver fluke disease in cattle and sheep worldwide and has recently emerged as an important pathogen of humans (1). Cathepsin L1, a major cysteine protease secreted by the parasite plays a pivotal role in various aspects of its pathogenicity. For example, the enzyme takes part in nutrient acquisition by catabolizing host proteins to absorbable peptides (2), facilitates the migration of the parasite through the host intestine and liver by cleaving interstitial matrix proteins such as fibronectin, laminin, and native collagen (3), and is implicated in the inactivation of host immune defenses by cleaving immunoglobulins (4, 5). Further, cathepsin L1 has recently been shown to suppress Th1 immune responses in infected laboratory animals making them susceptible to concurrent bacterial infections (6, 7, 8). Accordingly, the protease has been recognized as an important target at which parasite intervention strategies should be directed (9). In this regard, we have shown that the induction of anti-cathepsin L immune responses prior to a challenge infection of *F. hepatica* larvae elicits high levels of protection in cattle against disease (10, 11, 12).

Phylogenetic studies have shown that the *F. hepatica* cathepsin L1 belongs to an enzyme lineage that eventually gave rise to the mammalian cathepsin Ls, Ks, and Ss (2). The isolation of a cDNA encoding *F. hepatica* cathepsin L revealed that the enzyme, like its mammalian homologs, is synthesized as an inactive preproenzyme consisting of a prepeptide, a propeptide and mature enzyme region (13). By analogy with the mammalian proteases we assume that the prepeptide is removed following translocation into the endoplasmic reticulum and that the N-terminal propeptide extension is involved in several functions including intracellular targeting of the enzyme (14–16), correct folding of the mature enzyme by acting as an intramolecular chaperone (17–19) and prevention of uncontrolled proteolysis by binding to the enzyme substrate cleft in a reverse, non-productive direction (14). As has been demonstrated for the propeptide of human cathepsin L and other papain-like cysteine proteases (14, 17, 19, 20) the free propeptide of the *F. hepatica* procathepsin L is a specific and potent inhibitor of the cognate mature enzyme at neutral pH but does not bind to the enzyme at pH 5.5–3.5 (21).

Human lysosomal procathepsin L is stable at high pH because the propeptide protects the protein from the denaturing effects of the alkali (15). Removal of the propeptide to generate the fully active mature enzyme occurs at the lower lysosomal

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by a grant received from Enterprise Ireland and Ildana Biotech.

‖ Funded by the Health Research Board (HRB), Ireland.

‡‡ Funded by a joint North-South Cooperation grant from the HRB (Ireland) and the Research and Development Office (Northern Ireland).

¶¶ To whom correspondence should be addressed: Institute for the Biotechnology of Infectious Diseases (IBID), University of Technology, Sydney (UTS), Westbourne Street, Gore Hill, Sydney, NSW 2065, Australia. Tel.: 61-2-9514-4142; Fax: 61-2-9514-4201; E-mail: john.dalton@uts.edu.au.

‖‖ Funded by The Wellcome Trust.

pH of 5.5 (14, 15, 22, 23). It has been widely reported that cleavage of the proregion of cysteine proteases can occur autocatalytically *in vitro* under acidic conditions (15, 24, 25, 26). As the stability of the propeptide-protease complex is dependent on electrostatic interactions, reduction of the environmental pH, weakens the bond between the propeptide and the catalytic site. As a consequence, the proenzyme possibly adopts a looser conformation, in which the propeptide is bound less tightly into the active site making it more susceptible to proteolysis (16, 27).

The precise mechanism of proteolytic conversion from proenzyme to mature enzyme is still actively debated. The three-dimensional structure of procathepsin L (14) and procathepsin B (28, 29, 30) show that the N terminus of the mature enzyme is quite removed from the active site thus making it difficult to visualize an autocatalytic cleavage event. Moreover, circular dichroism studies reveal that activation does not involve significant conformational changes in the structure of procathepsin L (15) or procathepsin B (16). Therefore, the initial event in autocatalysis may involve an active proenzyme, possibly created by the reduced pH, that cleaves another proenzyme in the vicinity of the N terminus and sets off a chain reaction (15, 16). However, more recent studies on procathepsin B and procathepsin S identified autoproteolytic intermediates of processing when cystatin was included in *in vitro* activation reactions, which supported the view that the segment of the propeptide that binds the active site cleft is susceptible to cleavage (31). Accordingly, it was suggested that an initial slow intramolecular cleavage event within this segment of the propeptide triggers a more rapid cascade of intermolecular cleavages at the N terminus (31).

Earlier studies on the processing of yeast-expressed recombinant papain identified a conserved heptapeptide (Gly-Xaa-Asn-Xaa-Phe-Xaa-Asp) motif located between residues −42 and −36 in the propeptide that may be a site of initial cleavage in the pH-dependent autoactivation of the enzyme (32). It was suggested that the lowering of the pH perturbed the negative charge of Asp^{-36} resulting in a conformational change that switched on the processing events by allowing proteolysis to occur at the Ala^{-37}/Asp^{-36} bond. Following this primary cleavage further removal of the remaining amino acids of the propeptide may result from the proteolytic activity of the intermediate species, resulting in fully active mature protease (32).

Given the importance of cathepsin L proteases in the virulence of *F. hepatica* and other helminth pathogens (2) it is important to understand the mechanisms of their synthesis, processing and activation. We have previously shown that cathepsin L1 is secreted by this parasite and therefore functions extracellularly (4, 33) similar to the mammalian cathepsin Ls involved in the processing of the hormone thyroglobulin (34–36). In the present study we have employed confocal laser and electron immunocytochemistry to show that the enzyme is stored in secretory vesicles of the gastrodermal epithelial cells in its inactive proenzyme form. We investigate the autoactivation of *F. hepatica* cathepsin L1 using the wild-type proenzyme expressed in *Pichia pastoris*. Intermolecular processing was studied by generating an inactive procathepsin L mutant, where the catalytic Cys residue at position 26 was substituted with a Gly, which was incapable of autoactivation. Our results show that *P. pastoris*-expressed cathepsin L can autoactivate at low pH to an active intermediate form by an initial cleavage within the conserved heptapeptide (Gly-Xaa-Asn-Xaa-Phe-Xaa-Asp), as previously described for papain (32). Further activation of the enzyme occurs by cleavage in the vicinity of the N terminus. In contrast, intermolecular processing of the mutant inactive procathepsin L to a mature product by exogenous mature cathepsin L did not involve the formation of the above intermediate but was initiated by a direct cleavage close to the N terminus. These data provide a mechanism by which *F. hepatica* synthesizes and secretes a fully activated protease that is essential to its existence as a parasite.

EXPERIMENTAL PROCEDURES

Materials—Z-Phe-Arg-NHMec[1] and Z-Phe-Ala-CHN_2 were obtained from Bachem (St. Helens, UK). DTT and EDTA were obtained from Sigma (Dorset, Poole, UK). Prestained molecular weight markers and the AvrII and SnaBI restriction enzymes were obtained from New England Biolabs (UK) Ltd. (Hitchin, UK). Primers were obtained from Sigma-Genosys (Pampisford, UK). The pPIC9K vector and *Pichia pastoris* strain GS115 were obtained from Invitrogen Corp. (San Diego, CA). Ni-NTA agarose and columns were obtained from Qiagen (Crawley, UK).

In Vitro Cultivation of Parasites—Adult *F. hepatica* were obtained from infected cattle at a local abattoir and cultured *in vitro* in RPMI 1640 containing 30 mM HEPES, 1% glucose and 25 mg/ml gentamycin as described by Dalton and Heffernan (37). The medium was collected after 6 h, cleared by centrifugation at 14,000 × *g* for 30 min at 4 °C and stored at −20 °C.

Preparation of Antipropeptide and Antimature Cathepsin L1 Antisera—Native mature cathepsin L1 (nFheCL1) was purified from excretory-secretory (ES) products of adult *F. hepatica* by gel permeation and ion exchange chromatography and antiserum prepared in rabbits as previously described (4, 38). Recombinant cathepsin L1 propeptide was generated and purified as described by Roche *et al.* (21) and antiserum prepared by immunizing New Zealand White rabbits five times with 20 μg of protein formulated in Freund's Complete and Incomplete Adjuvant.

Immunofluorescence and Immunoelectron Microscopy—Adult *F. hepatica* were recovered from infected cattle at a local abattoir, washed, and transported to the laboratory in mammal saline (0.9% NaCl) at 37 °C. Parasites were rinsed in mammal saline and allowed to regurgitate their gut contents before being flat-fixed in 4% paraformaldehyde (PFA) in PBS (pH 7.2) for 4 h. They were washed in antibody diluent (ABD: PBS with, 0.1% bovine serum albumin; 0.3% Triton X-100; 0.1% sodium azide) for 24 h before being incubated for 48 h at 4 °C in antiserum prepared against purified mature cathepsin L1 (diluted 1:3000) and subsequently washed in ABD (24 h, 4 °C). Swine anti-rabbit tetramethyl rhodamine isothiocyanate (TRITC; 1:100; Dako Ltd.) was used to visualize bound primary antibody before the worms were washed in ABD (24 h, 4 °C) and mounted on glass microscope slides in PBS/glycerol (1:9) containing 2.5% 2,4-diazabicylco 2.2.2 octane. Specimens were viewed using a Leica TCS-NT confocal scanning laser microscope.

For electron microscopy worms were washed in mammal saline and fixed for 1 h in 2% double-distilled glutaraldehyde (GTA) (Agar Scientific) in 0.1 M sodium cacodylate buffer (pH 7.2) containing 3% sucrose at 4 °C. Following thorough washing in buffer, specimens were dehydrated through graded ethanol to propylene oxide, infiltrated and embedded in Agar 100 resin (Agar Scientific). Ultrathin sections (80–90 nm) were cut on a Reichert Ultracut E ultramicrotome, collected on bare 200-mesh nickel grids and dried at room temperature. For immunogold labeling sections were etched with 10% hydrogen peroxide for 5 min and rinsed thoroughly with 20 mM Tris-HCl buffer (pH 8.2) containing 0.1% bovine serum albumin and Tween 20 (1:40 dilution). Grids were incubated in normal goat serum (1:20 dilution) for 30 min and then transferred to primary antibody diluted to 1:20,000 with 0.1% bovine serum albumin/Tris-HCl buffer for 12–18 h. Grids were then washed in bovine serum albumin/Tris-HCl and transferred to a 20 μl drop of 10 nm gold-conjugated goat anti-rabbit IgG (Bio Cell International) for 2 h at room temperature. Following another buffer wash, grids were lightly fixed with 2% double-distilled GTA for 3 min, and finally washed with buffer and rinsed with distilled water. Grids were double stained with uranyl acetate (5 min) and lead citrate (3 min) and examined in a FEI (Philips) CM100 transmission electron microscope, operating at 100 keV.

Controls consisted of (i) incubation of whole-mounts/sections with secondary antibody in the absence of primary antibody and (ii) incubation with preimmune serum followed by the secondary antiserum.

[1] The abbreviations used are: Z-Phe-Arg-NHMec, benzyloxycarbonyl-L-phenylalanyl-L-arginine 4-methylcoumarinyl-7-amide; Z-Phe-Ala-CHN_2, benzyloxycarbonyl-L-phenylalanyl-L-alanine-diazomethylketone; DTT, dithiothreitol; PBS, phosphate-buffered saline; NTA, nitrilotriacetic acid.

Construction of Expression Vectors Encoding cDNA for Wild-type Procathepsin L1 (rFheproCL1) and Gly^{26} Procathepsin L1 (rmutFheproCL1) and Transformation into Pichia pastoris—The *F. hepatica* procathepsin L (FheproCL1) was amplified by PCR from the pAAH5 *Saccharomyces cerevisiae* expression vector into which the full-length cDNA had been previously cloned in our laboratory (13). Primers were used (see primers A and D below) to incorporate a SnaBI restriction site at the 5′-end of the gene and an AvrII restriction site and His_6 tag sequence at the 3′-end. The 980-bp fragment was inserted into the AvrII/SnaB1 site of *P. pastoris* expression vector pPIC9K (Invitrogen).

Mutants were generated from this construct by a PCR-based site-directed mutagenesis method known as gene splicing by overlap extension (SOEing) (39) using the pPIC9K-FheproCL1 DNA as a template. The construction of the inactive FheproCL1 mutant involved changing the active site cysteine (Cys^{26}) residue to a glycine in a two-step PCR process. The primers used were as follows: primer A, 5′-GCGGCTACGTATCGAATGATGATTTGTGGCAT-3′; primer B, 5′-GAATGCCCAACCGGAGCCAC-3′; primer C, 5′-GTGGCTCCGGTTGGGCATTC-3′; primer D, 5′-GCGCCTAGGTCAGTGGTGGTGGTGGTGGTGGGCCC-3′.

The underlined nucleotides indicate the replacements introduced. Each reaction used one flanking primer that hybridized at one end of the target sequence (primer A or D) and one overlapping internal primer that hybridizes at the site of the mutation and contains the mismatched base (primer B or C). In the first round of amplification two sections of the cDNA were amplified using primer combinations A+B and C+D. These two PCR products, with an overlap of 21 bp at one end of each fragment, were then combined in a second PCR to amplify the entire rmutFheproCL1 cDNA. Primers for this reaction were the two outside primers used in each of the two first round reactions (primers A and D). For all PCRs, high-fidelity Taq polymerase was used (25 cycles at 94 °C for 30 s, 55 °C for 1 min, and 72 °C for 2 min). The rmutFheproCL1 cDNA was then inserted into the AvrII/SnaB1 site of expression vector pPIC9K. Both the wild-type (rFheproCL1) and mutant (rmutFheproCL1) plasmid insert were sequenced to verify the presence of correct gene sequence and mutation.

The rFheCLI or rmutFheCL1 plasmids were linearized by digestion with SalI and introduced to *P. pastoris* GSII5 cells by spheroplasting (40). Transformants were selected for their ability to grow on histidine-deficient agar plates and on agar plates containing minimal media and methanol. Insertion of rFheCLI or rmutFheCL1 into *P. pastoris* was confirmed by PCR using primers specific to the yeast genome (41).

Expression and Purification of Procathepsin L1 and Gly^{26} Procathepsin L1—Yeast transformants were cultured in 250 ml of BMGY broth, buffered to pH 6.0, in 1 liter of baffled flasks at 30 °C until an OD_{600} of 2–6 was reached. Cells were harvested by centrifugation at 2000 rpm for 5 min, and protein expression induced by resuspending in 50 ml of BMMY broth, buffered at pH 6.0, 7.0, or 8.0, containing 1% methanol (41). The cultures were grown at 30 °C with shaking at 225 rpm for 3 days, 1 ml samples were removed daily and then filter-sterilized methanol was added to maintain a final concentration of 1%. To assess the effect of the cathepsin L cysteine inhibitor Z-Phe-Ala-diazomethylketone (-CHN_2) on the production of procathepsin L1 by yeast the inhibitor was added at the time of induction to a final concentration of 25 μM and then twice daily at the same concentration over the subsequent 3 days.

Recombinant proteins were purified from yeast medium by affinity chromatography using Ni-NTA-agarose (Qiagen). Briefly, a column prepared with 1 ml of resin was equilibrated by passing through 10 ml of 50 mM sodium phosphate buffer, pH 8.0, containing 300 mM NaCl and 10 mM imidazole. 10 ml of yeast media supernatant was mixed with 40 ml of the same buffer and applied to the column. The column was washed with 15 ml of 50 mM sodium phosphate buffer, pH 8.0, containing 300 mM NaCl and 20 mM imidazole, and bound protein eluted using 50 mM sodium phosphate buffer, pH 7.0, containing 300 mM NaCl and 250 mM imidazole. Purified recombinant proteases were dialysed against PBS and stored at −20 °C. Samples of yeast medium supernatants (10 μl) and purified proteins (~1–5 μg) were analyzed by 12% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting (see below).

Fluorometric Substrate Assay for Cathepsin L1 Activity—Cathepsin L1 activity was determined by the fluorometric substrate assay described by Dowd *et al.* (38). Yeast culture supernatants or purified recombinant protease were assayed in a total volume of 1 ml of substrate/buffer mix (2.5 mM EDTA, 2 mM DTT, 0.1 M sodium phosphate buffer, pH 6.0, and 10 μM Z-Phe-Arg-NHMec). The reaction was incubated at 37 °C and stopped after 30 min by the addition of 200 μl of 10% acetic acid. Fluorescence was recorded at an excitation wavelength of 370 nm and an emission wavelength of 440 nm. The activity of the samples were calculated from a standard curve of NIIMec ranging from 0 to 10 μM, and presented as nanomoles of NHMec min^{-1} ml^{-1}.

In Vitro Processing of Procathepsin L1 and Gly^{26} Procathepsin L1—Purified recombinant procathepsin L1 (~5 μg) was incubated in activation buffer (0.1 M sodium citrate buffer, pH 5.0, 1 mM DTT, and 1.25 mM EDTA) at 37 °C. Samples were taken at various time points up to 2 h and the proteolytic cleavage of the propeptide visualized by 12% SDS-PAGE. Aliquots were also assayed for enzyme activity with the specific substrate Z-Phe-Arg-NHMec as described above.

Exogenous processing of Gly^{26} procathepsin L1 was carried out by mixing 5 μg of the purified mutant enzyme with 0.5 μg of wild-type procathepsin L that had been activated as described above. Processing of the mutant protein over a period of 3 h at 37 °C was analyzed by 12% SDS-PAGE. To confirm that [Gly^{26}]procathepsin L1 was expressed as a correctly folded protein a similar experiment was performed in parallel with mutant enzyme preparations that had been unfolded by heating at 95 °C for 3 min.

Identification of Cleavage Products by N-terminal Sequencing—Following 12% SDS-PAGE, proteins were transferred to polyvinylidene difluoride membrane using a semidry transfer cell at 15V for 20 min. The membrane was washed with dH_2O and stained with 0.025% Coomassie Brilliant Blue R-250 in 40% methanol (40). Protein bands of interest were subjected to N-terminal sequencing at Genosphere Biotechnologies (Paris, France).

Sequence Analysis—The *F. hepatica* cathepsin L1 protein sequence was aligned with several related cathepsin sequences using ClustalX 1.81. Protein sequences used included *Carica papaya*, *Fasciola hepatica* cathepsin L1, *F. hepatica* cathepsin L2, *F. gigantica* cathepsin L1, *Schistosoma mansoni* cathepsin L2, *Caenorhabditis elegans* CPL-1, mouse cathepsin L, rat cathepsin L, and human cathepsin L.[2] Sequences were numbered according to the papain numbering used by Vernet *et al.* (32) where the propeptide residues are recorded as a negative beginning from the cleavage site between propeptide and mature enzyme.

SDS-PAGE and Immunoblotting—Protein samples were separated by 12% SDS-PAGE and gels stained with a 0.1% w/v solution of Coomassie Brilliant Blue R-250 in 40% methanol/10% acetic acid (42). Immunoblots were prepared by transferring proteins to nitrocellulose membranes, presoaked in a transfer buffer (50 mM Tris, 384 mM glycine, 20% methanol), and then blocking these for 1 h at room temperature with 5% milk in PBS/0.1% Tween 20. The nitrocellulose membranes were probed with polyclonal rabbit antimature cathepsin L1 serum, rabbit anti-recombinant propeptide and preimmunized control rabbit serum diluted 1/1,000 in 1% milk/PBS/0.1% Tween-20 for 45 min at room temperature. They were then washed three times for 5 min each with 1% milk/PBS/0.1% Tween-20, followed by incubation for 45 min at room temperature with a 1/10,000 dilution of secondary antibody (goat anti-rabbit IgG-peroxidase conjugate) prepared in 1% milk/PBS/0.1% Tween-20. Blots were washed as before and bound antibody visualized with 3,3′-diaminobenzidine (DAB, Sigma).

RESULTS AND DISCUSSION

Secretion of Cathepsin L Proteases by F. hepatica Parasites—When adult *F. hepatica* are maintained *in vitro* they secrete two major proteases that were previously characterized as 27.5 cathepsin L1 and 29 kDa cathepsin L2 in our laboratory (4, 38). N-terminal sequencing revealed a sequence AVPDK for both enzymes that correlated with the N terminus of fully processed mature cathepsin L enzymes. In addition, immunoblotting experiments demonstrated that while they are both reactive with sera prepared against purified mature cathepsin L1, neither reacts with sera prepared specifically against a recombinant propeptide of cathepsin L1 (Fig. 1, see also Fig. 3*B*). Isolation of cDNAs encoding the two enzymes showed that they exhibit a high level of identity (77%) at the amino acid level that explains their cross-reactivity with antiserum prepared against either enzyme (13, 38).

The two proteases are members of a phylogenetic lineage

[2] Protein sequences and accession numbers from GenBank™ are as follows: *Carica papaya* papain (P00784), *F. hepatica* cathepsin L1 (U62288), *F. hepatica* cathepsin L2 (U62289), *F. gigantica* cathepsin L1 (AF112566), *S. mansoni* cathepsin L2 (Z32529), *C. elegans* CPL-1 (NP_507199), mouse cathepsin L (P06797), rat cathepsin L (KHRTL), and human cathepsin L (M20496).

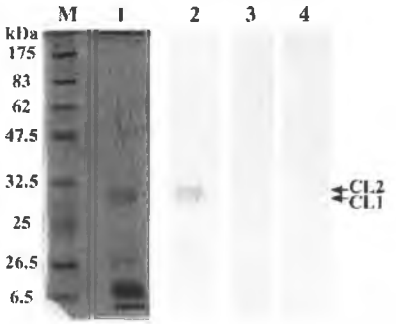


FIG. 1. **SDS-PAGE and immunoblot analysis of *F. hepatica* excretory-secretory (ES) products.** ES products obtained from medium in which *F. hepatica* were cultured was analyzed by SDS-PAGE and immunoblotting. *Lane M*, molecular size markers; *lane 1*, Coomassie Blue-stained 12% SDS-PAGE gel of ES products; *lane 2*, ES products probed with antiserum prepared in rabbits against native mature cathepsin L1; *lane 3*, ES products probed with antiserum prepared in rabbits against recombinant propeptide of procathepsin L1; *lane 4*, ES products probed with normal rabbit serum. *Arrows* indicate position of major secreted and fully processed cathepsin L1 and cathepsin L2.

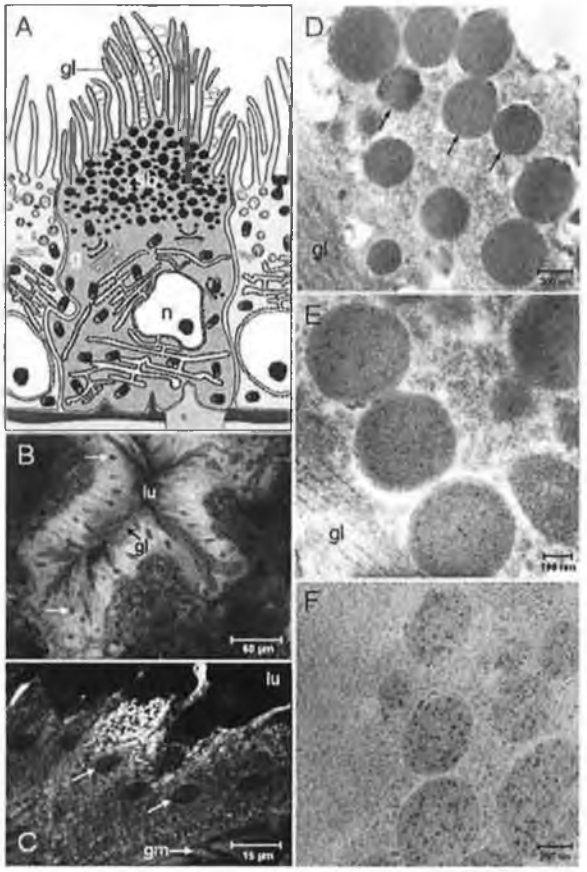


FIG. 2. **Immunolocalization of procathepsin L1 in *F. hepatica* gastrodermal epithelial cells.** *A*, schematic representation of gastrodermal epithelial cells in *F. hepatica* (after Smyth and Halton, Ref. 46). *g*, Golgi; *gl*, gut lamellae; *n*, nuclei; *sb*, secretory bodies. *B*, confocal scanning laser micrograph showing immunoreactivity for mature cathepsin L1 within the gastrodermis of adult *F. hepatica*; the nuclei (*arrows*) of the gastrodermal cells are clearly visible. *gl*, gut lamellae; *lu*, gut lumen. *C*, confocal scanning laser micrograph showing immunoreactivity for mature cathepsin L1 within the gastrodermal cells of adult *F. hepatica*; the nuclei (*arrows*) of the gastrodermal cells are clearly visible. Immunoreactivity appears as a punctuate pattern at the apical end of the cells where the secretory vesicles are located. *gm*, gut muscle; *lu*, gut lumen. *D*, electron micrograph showing non-reactivity of secretory bodies (*arrows*) within the epithelial cells of the gastrodermis with control preimmunized rabbit serum. *gl*, gut lamellae. *E*, electron micrograph showing the localization *F. hepatica* procathepsin L1 to secretory bodies of the gastrodermal epithelial cells with antiserum prepared against the recombinant propeptide of procathepsin L1. *gl*, gut lamellae. *F*, electron micrograph showing the localization of *F. hepatica* cathepsin L1 to secretory bodies of the gastrodermal epithelial cells with antiserum prepared against purified mature portion of cathepsin L1. Note that the labeling is confined to the contents of the secretory bodies.

that includes the mammalian cathepsin Ls and have substrate specificities that are typical of this group, *i.e.* a preference for positively charged residues, such as arginine, in the P1 position and hydrophobic residues, such as phenylalanine and leucine, in the P2 position (14). In contrast to mammalian cathepsin Ls, however, the enzymes exhibit activity over a wide pH range (4.0–8.5) and are remarkably stable at neutral pH; they retain most of their activity following a 24-h incubation at pH 7.0 and 37 °C whereas human cathepsin L is completely inactivated in less than 20 min under the same conditions (43). These enhanced stability properties were suggested to be important in host protein digestion in the parasite gut and in facilitating the migration of the parasite through the host intestine and liver (33).

Immunolocalization and *in situ* hybridization studies previously carried out by our laboratories had shown that the cathepsin L proteases are synthesized within the gastrodermis that lines the parasite gut (4, 33). The gastrodermis is composed of epithelial cells that are columnar in shape and contain numerous dense vesicles situated at the apical end (see Fig. 2*A*). Although secretory in function, these vesicles have been described as lysosomal-like and for almost thirty years have been suspected to contain proteases (44–47). Confocal laser immunocytochemistry using antiserum prepared against the native mature cathepsin L localized this protease to the gastrodermal epithelial cells (Fig. 2*B*) throughout the gut of adult *F. hepatica*. The pattern of immunoreactive labeling appeared punctate and was more dense toward the apical end of the gastrodermal cells indicating that the cathepsin L protease was associated with the secretory vesicles (Fig. 2*C*). Electron microscopy revealed that the enzyme is indeed stored in secretory vesicles within these cells (Fig. 2, *D–F*). The secretory vesicles were strongly immunoreactive with an antiserum prepared against the recombinant-expressed propeptide (Fig. 2*E*) and to the mature portion of cathepsin L1 (Fig. 2*F*) indicating that the enzymes are present in these vesicles as the inactive proforms, procathepsin L. (The specificity of the sera for the propeptide and mature regions of the procathepsin L1 is demonstrated in Fig. 3*B*.)

The above observations provide important insights into the regulation of protease activity within the gastrodermis of these parasites. First, in agreement with the observation by Halton and co-workers (44, 46, 47) it appears that the function of the vesicles is to package proteolytic enzymes for delivery into the gut lumen. Second, the accumulation of procathepsin L1, the inactive precursor, in the secretory vesicles rather than active mature protease makes good sense given the high abundance of these vesicles in the epithelial cells and hence the possibility of protease leakage into the cytoplasm (see Fig. 2*A*). Third, the data suggest that activation of the procathepsin L1 takes place following secretion of the proteases into the gut lumen. Electron microscopy has shown that the contents of the secretory vesicles are extruded into the gut lumen where they mix with the ingested blood and tissue meal between the long extruding lamellae (Refs. 47 and 48; see Fig. 2*A*). Enzyme activation is likely to take place here as the pH is estimated to be ~5.5 or slightly lower (47, 48). Once activated, the protease would be in direct contact with the bloodmeal and could immediately begin its function in protein catabolism. Since these helminth para-

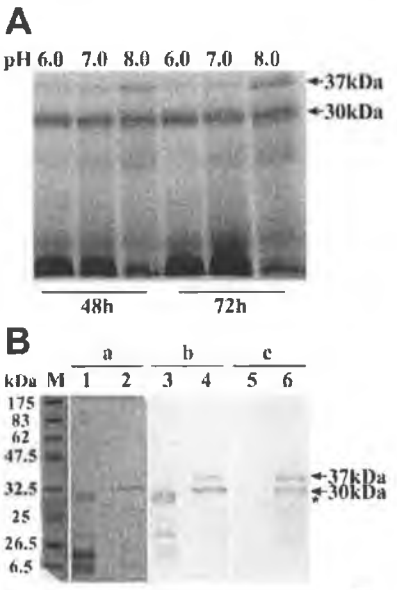


FIG. 3. **SDS-PAGE and immunoblot analysis of *F. hepatica* procathepsin L1 expression from *P. pastoris*.** *A*, cultures of *P. pastoris* transformed with a cDNA encoding procathepsin L1 were induced with 1% methanol. Fermentations were carried out at 30 °C in media buffered to pH 6.0, 7.0, and 8.0. Aliquots (10 μl) removed at 48 and 72 h after induction were analyzed by 12% SDS-PAGE. The 37 and 30 kDa components are indicated. *B*, comparison of *P. pastoris*-expressed procathepsin L with native *F. hepatica*-secreted mature cathepsin L1. Aliquots of adult *F. hepatica* ES products (*lanes 1*, *3*, and *5*) were compared with samples taken from *P. pastoris* medium after induction for 72 h at pH 8.0 (*lanes 2*, *4*, and *6*). *Lane M*, molecular size markers; *panel a*, Coomassie Blue-stained 12% SDS-PAGE gel; *panel b*, samples probed with antiserum prepared in rabbits against mature portion of cathepsin L1; *panel c*, samples probed with antiserum prepared in rabbits against recombinant propeptide portion of procathepsin L1. Note, the antipropeptide serum reacts with the 37 and 30 kDa *P. pastoris*-expressed components (indicated with *arrows*) but not with parasite-produced mature cathepsin L1 or cathepsin L2 (indicated by an *asterisk*). Control preimmunized rabbit serum was not reactive with any proteins (not shown).

sites have a blind-ending gut, removal of undigested products and accumulated hematin is achieved by simply voiding the contents approximately every 3 h (46, 47). In this way active mature cathepsin L would be regularly liberated from the parasite into the surrounding intestinal or liver tissues where it could perform its function in tissue degradation and thus aid the movement of the parasite through these tissues (47, 49).

The secretion of *F. hepatica* procathepsin contrasts with the intracellular trafficking of mammalian procathepsin L through the Golgi apparatus to acidic lysosomes where the enzyme is processed to the mature form by removal and degradation of the propeptide (14, 15). However, a low level of procathepsin L is also secreted from normal cells and is thought to play an extracellular processing function (50), while high levels of secretion has been correlated with the metastatic activity of transformed cells (51, 52). A comparison may also be made with human cathepsin K, which is secreted by osteoclasts and functions in bone resorption (25, 53, 54). Following adherence of osteoclasts to the bone surface an extracellular compartment known as an absorption pit is created between cell and bone into which cathepsin K is secreted. The pH within the pit is slightly acidic and is considered important not only for the dissolution of the bone but also for the activation of procathepsin K to its active proteolytic form (53, 54). Cathepsin B, L, and K have also been shown to be secreted into the extracellular lumen of thyroid follicles where they play a role in the proteolytic processing of covalently cross-linked soluble thyroglobulin to liberate the thyroid hormone thyroxine (T_4) (34–36).

Expression and Processing of Wild-type Procathepsin L1—To study the mechanism by which the *F. hepatica* procathepsin L1 is processed to an active mature enzyme the recombinant proenzyme was expressed in the yeast *P. pastoris*. The transformed yeast cells were induced by 1% methanol in BMMY medium buffered at pH 6.0, 7.0, or 8.0 and then the recombinant protein secreted into the medium analyzed by SDS-PAGE. Two major proteins migrating at 37 and 30 kDa were visualized (Fig. 3*A*). These components migrate higher than the mature cathepsin L1 and cathepsin L2 secreted by the adult *F. hepatica* parasites in 12% SDS-PAGE (Fig. 3*B*). Immunoblotting experiments showed that polyclonal antiserum prepared against the propeptide portion and against the mature portion of procathepsin L both reacted with the *P. pastoris*-expressed 37 and 30 kDa components confirming that these are two forms of the procathepsin L. By contrast, the fully processed enzymes secreted by the parasite are reactive with antiserum prepared against the mature portion of the procathepsin L1 but are not reactive with antiserum prepared against the propeptide portion (Fig. 3*B*).

The 37 and 30 kDa components were purified from the pH 8.0 yeast medium by Ni-NTA agarose affinity chromatography and subjected to N-terminal sequence analysis. This showed that the 37 kDa protein represented the procathepsin L (plus 5 amino acids of the yeast α-factor secretory signal at the N terminus) while the 30-kDa component represented an intermediate processed form beginning at residue Leu^{-41} (papain numbering) within the propeptide (see Fig. 5).

The amount of the recombinant proteins in the yeast medium increased over the 3 days following induction with methanol; this increase was observed at pH 6.0, 7.0, and 8.0 (Fig. 3). However, the relative ratio of the 37 and 30 kDa proteins was dependent on the buffering pH of the medium, with the proportion of 37 kDa increasing with increasing pH. A pH profile of activity for the recombinant *F. hepatica* cathepsin L showed that it was similar to that of the native mature enzyme (38, 43), *i.e.* the recombinant enzyme was active over the pH range 4.0 to 8.5 but was most active at pH 6.0.[3] Therefore, by increasing the pH of the yeast medium above the pH optimum for activity of the *F. hepatica* cathepsin L the level of processing of the 37-kDa form to the 30-kDa intermediate form was reduced.

To address this issue further, yeast cells were cultured at pH 8.0 in the presence and absence of the cathepsin L-specific inhibitor Z-Phe-Ala-CHN_2 (Fig. 4, *A* and *B*). Although the inhibitor was added twice daily at a concentration of 25 μM, which is 250-fold greater than its K_i for cathepsin L (38, 40), samples of medium taken daily showed that cathepsin L activity was not completely inhibited (it is possible that the inhibitor is not stable in the medium) (see Fig. 4*B*). Nevertheless, the proportion of 37-kDa component in culture supernatants containing inhibitor was greater than that in cultures without inhibitor, although the 30-kDa protein was still evident. These data support the idea that procathepsin L1 secreted from *P. pastoris* autoprocesses to an intermediate active form and that this autoprocessing is prevented by the cathepsin L protease inhibitor Z-Phe-Ala-CHN_2. Estimations of yeast cell viability showed that at the concentrations used the inhibitor did not affect the growth of the yeast cells.[3]

To examine whether further intermolecular processing could take place at a pH lower than 6.0, the 37 kDa procathepsin L1 and 30-kDa intermediate forms were first purified from pH 8.0 yeast culture supernatant by affinity chromatography on Ni-NTA agarose and dialyzed against PBS (pH 7.3). The pH of

[3] P. R. Collins and J. P. Dalton, unpublished data.

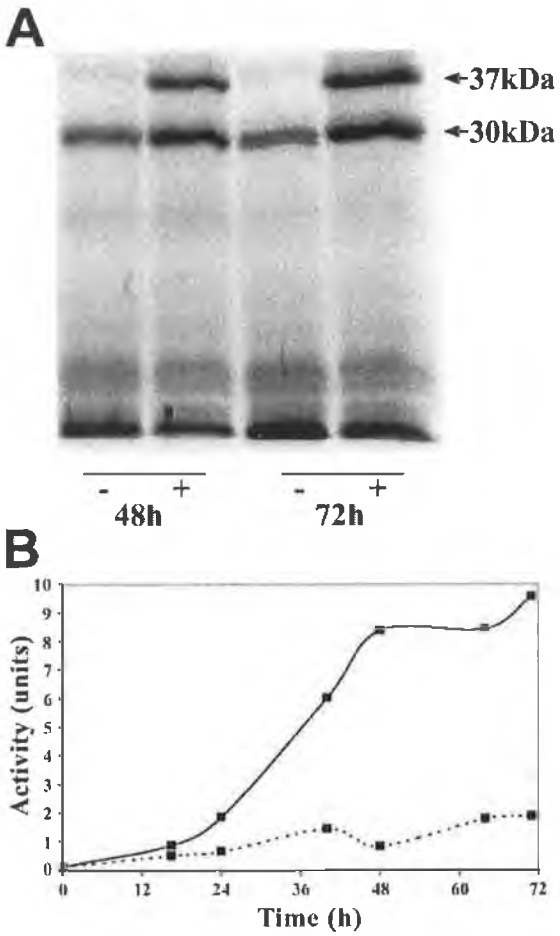


FIG. 4. **Production of procathepsin L1 transformed *P. pastoris* in the presence and absence of Z-Phe-Ala-diazomethylketone.** *A*, transformed *P. pastoris* were cultured at 30 °C and pH 8.0 in the presence (+) or absence (−) of the cysteine protease inhibitor Z-Phe-Ala-CHN_2 (25 μM). Aliquots of culture supernatant, removed after 48 and 72 h, were analyzed by SDS-PAGE. The 37 and 30 kDa bands are indicated. *B*, recombinant yeast cultured in the presence (*dotted line*) and absence (*solid line*) of 25 μM Z-Phe-Ala-CHN_2. Cathepsin L activity in the culture supernatant was measured with the fluorogenic substrate Z-Phe-Arg-NHMec. Activity units are presented as nmol of NHMec released min^{-1} ml^{-1}.

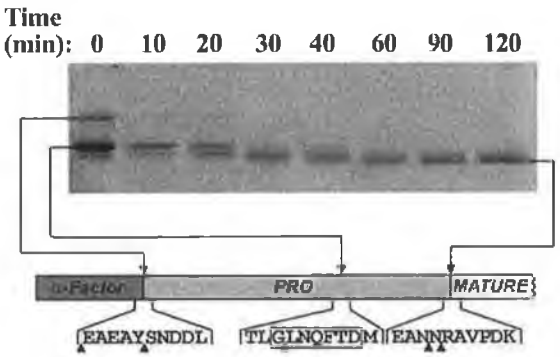


FIG. 5. **Autocatalytic processing of procathepsin L1 at pH 5.0.** Purified procathepsin L1 (5.0 μg) was incubated at 37 °C for 2 h in 0.1 M sodium citrate buffer, pH 5.0, containing 2 mM DTT and 2.5 mM EDTA. Samples removed at various time points were analyzed by SDS-PAGE. The predominant proteins of 37, 30, and 24.5 kDa were subjected to N-terminal sequencing. The schematic indicates the amino acid sequences surrounding the intermolecular processing sites (*arrowed*) that generated these components and their location within the propeptide of procathepsin L1. The *box* indicates the conserved motif GXNXFXD.

samples taken from this preparation were then reduced to 5.0 by the addition of 0.1 M sodium citrate containing 1 mM DTT and 1.25 mM EDTA, and then incubated at 37 °C (Fig. 5). Analysis of samples removed at various time points over a period of 2 h revealed that the 37 kDa procathepsin L is processed to the 30-kDa intermediate form and then through various intermediates that eventually give rise to a single peptide migrating at 25.4 kDa. Corresponding with this processing was the gradual 5-fold increase in the activity of cathepsin L by the end of the 2-h incubation period (data not shown). N-terminal sequence analysis of the 25.4 kDa product revealed two sequences, NRAVP and RAVPD, which are just 2 and 1 amino acids from the mature protein processing site, respectively (see Fig. 5). This intermolecular processing was completely inhibited by adding 10 μM Z-Phe-Ala-CHN_2 to the starting mix and therefore can be attributed to the action of cathepsin L (data not shown).

The cleavage site between Gly^{-42}-Leu^{-41} that generated the 30 kDa intermediate form observed by SDS-PAGE analysis (29.6 kDa from amino acid sequence) was of particular interest as it occurred within the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} motif that was previously identified as pivotal in the intermolecular processing of papain expressed in the yeast *S. cerevisiae*. By random mutagenesis studies, and by sequence comparisons of various papain-like proteases, Vernet *et al.* (32) showed that the Gly^{-42} and Asp^{-36} residues within this motif were the most constrained and proposed that two cleavage sites, Gly^{-42}-Leu^{-41} and Ala^{-37}-Asp^{-36} were important for proenzyme processing. The latter was considered more significant because cleavage at this site was predicted to perturb the negative charge of the Asp^{-36} residue (32). The Asp^{-36} residue is absolutely conserved in non-cathepsin B-like cysteine proteases of helminth parasites, plants and mammals (including cathepsin L, cathepsin S, and cathepsin K) and crystallographic studies on human cathepsin L showed that it participates in an important salt bridge between propeptide and mature enzyme (14).

On inspection of the *F. hepatica* procathepsin L propeptide sequence we found that the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} motif is completely conserved in this parasitic helminth and in others including *F. gigantica* and *S. mansoni*, and in the free-living helminth *C. elegans* (Fig. 6). Moreover, the Leu^{-41} residue is also conserved in all of these sequences thus preserving the Gly^{-42}-Leu^{-41} bond and implicating its importance in propeptide function in these helminths. However, these two residues are not conserved in the human, mouse, and rat cathepsin L sequences where Gly^{-42} is replaced by Glu or Ala and Leu^{-41} is replaced by Met (Fig. 6). While it is tempting to speculate that the changes at these positions explains why an intermediate 30-kDa form was not observed in studies on the intermolecular processing of human procathepsin L (15), McQueney *et al.* (25) found that the intermolecular processing of an inactive mutant form of procathepsin K by activated cathepsin K did involve a cleavage between Ala^{-42} and Met^{-41}. As pointed out by Coulombe *et al.* (14), while there may be a direct involvement of the Asp^{-36} in the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} motif in the pH-dependant processing of proforms of cysteine proteases, cleavage at other peptide bonds within this motif could influence the local charge state that then may trigger processing.

Expression and Exogenous Processing of Gly^{26} Procathepsin L1—To further study the intermolecular processing of *F. hepatica* procathepsin L1 we expressed an inactive form of the protein by replacing the active site Cys^{26} with Gly^{26}. Expression of this recombinant mutant in medium buffered at pH 6.0

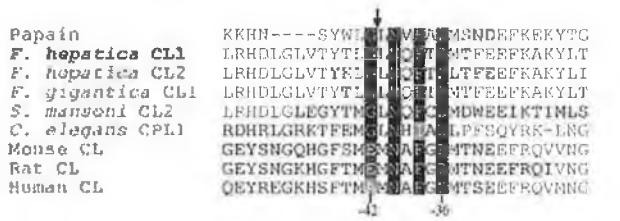


FIG. 6. **Comparison of the GXNXFXD motif between papain and cathepsin L proteases of helminths and mammals.** The GXNXFXD motif sequence of papain is identical to that of *F. hepatica* cathepsin L1 and cathepsin L2 and to the cathepsin Ls of related the trematodes *F. gigantica* and *S. mansoni* (identical residues highlighted in *black*). The *arrow* indicates the position of the intermolecular cleavage site, between the Gly^{-42}-Leu^{-41} bond (papain numbering), that generates the intermediate 30 kDa component of the *F. hepatica* procathepsin L when it is expressed in *P. pastoris*. The Gly^{-42} is substituted for Glu in the rat and mouse cathepsin L sequences, and for Ala in the human cathepsin L sequence (highlighted in *gray*, papain numbering).

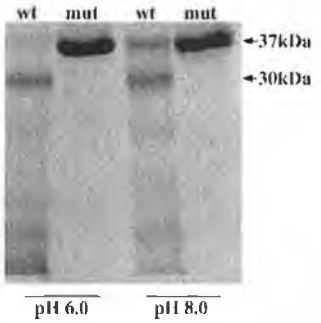


FIG. 7. **Secretion of the mutant procathepsin L1 from *P. pastoris*.** *P. pastoris* cells transformed with wild-type procathepsin L1 (*wt*) or the procathepsin L1 that had the active site Cys^{26} replaced by Gly (*mut*) were cultured at 30 °C in media buffered to pH 6.0 and 8.0. After 48 h of induction with 1% (v/v) methanol samples of culture supernatant were analyzed by SDS-PAGE. The 30-kDa intermediate form of procathepsin L was observed only in the wild-type cultures.

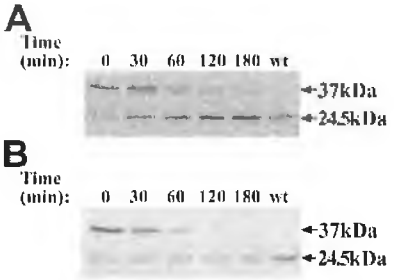


FIG. 8. **Intermolecular processing of procathepsin L1 Cys^{26}-Gly^{26} mutant by exogenously added active cathepsin L1.** *A*, purified procathepsin Cys^{26}-Gly^{26} mutant (5.0 μg) was incubated for up to 3 h in 0.1 M sodium citrate buffer, pH 5.0, containing 1 mM DTT and 1.25 mM EDTA at 37 °C in the presence of 0.5 μg of wild-type cathepsin L1 that was activated according to procedures shown in Fig. 5. The position of the mature wild-type cathepsin L at 24.5 kDa (2.5 μg, loading) is indicated in the lane labeled *wt* with the *arrow*. *B*, parallel experiments were carried as above but in this case purified procathepsin Cys^{26}-Gly^{26} mutant (5.0 μg) was unfolded by heating at 95 °C for 3 min prior to the addition of activated wild-type cathepsin L1.

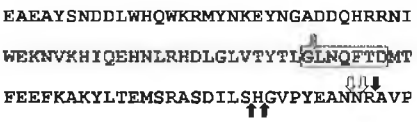


FIG. 9. **Amino acid sequence of the propeptide of *F. hepatica* cathepsin L1 with cleavage sites that lead to activation indicated.** The *downward solid arrow* indicates the cleavage site observed when native mature cathepsin L1 is secreted *in vitro* by adult *F. hepatica*. The *first open arrow* indicates the site of cleavage within the GXNXFXD motif (*boxed*) that gives rise to the 30 kDa intermediate component when procathepsin L1 is expressed in *P. pastoris*, and the second and third *open arrows* show the sites of cleavage when this is further activated *in vitro* at pH 5.0 (see Fig. 5). The upward *solid arrows* show the sites of cleavage of procathepsin Cys^{26}-Gly^{26} mutant by exogenously added mature cathepsin L1.

and 8.0 showed that it migrates as a single band at 37 kDa in SDS-PAGE (Fig. 7). N-terminal sequencing confirmed that the 37 kDa protein represented the complete procathepsin L1, and enzyme assays showed that it lacks activity against the substrate Z-Phe-Arg-NHMec.[3] The production of the mutant procathepsin L1 is almost 4-fold greater than that of the combined 37 and 30 kDa components observed in parallel cultures of wild-type procathepsin L1 (see Fig. 7). Furthermore, the mutant enzyme did not process to lower molecular size forms when incubated at pH 5.0.[3]

Addition of recombinant wild-type cathepsin L1, which was activated as described above, to the mutant procathepsin L, at pH 5.0, resulted in the progressive appearance of a minor band at ~35 kDa over an incubation period of 3 h and a second major band at 24.5 kDa that co-migrates with fully active cathepsin L (the intensity of the 24.5-band increases over the reaction period, Fig. 8). N-terminal sequencing attempts were unsuccessful for the 35 kDa protein but a sequence of HGVPY (with less than 10% GVPYE) was obtained for the 24.5 kDa protein; the cleavage site is therefore located 10 amino acids prior to the N terminus of mature cathepsin L1. Experiments in which activated wild-type cathepsin L1 was added to heat-denatured mutant procathepsin L showed that the latter was sensitive to complete degradation by active cathepsin L1 and confirms that the expressed mutant is correctly folded (intensity of the 24.5 does not increase, compare Fig. 8, *A* and *B*).

An overview of the cleavage sites observed by autoactivation by the wild-type enzyme and by intermolecular processing of mutant enzyme by exogenously added wild-type cathepsin L1 is shown in Fig. 9. The cleavage sites determined for the native mature cathepsin secreted by *F. hepatica* parasites *in vitro* are also indicated. It is clear that autoactivation and intermolecular activation occur by cleavages at different sites and thus processing of procathepsin L1 to an active enzyme can be achieved by more than one pathway. Of particular note is the absence of a cleavage site within the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} motif that would generate the 30-kDa intermediate component by exogenous activation. The dominance of this cleavage site in yeast-expressed wild-type cathepsin L and the strict conservation of the four relevant residues suggests an important role for this motif in the processing of *F. hepatica* procathepsin L1. The importance of this motif in the processing of papain was demonstrated by Vernet *et al.* (32) using various mutants in Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} sequence; certain residue replacements gave rise to some non-functional incorrectly folded enzymes, while others cause the accumulation of inactive or partly active precursors.

Further intermolecular processing and activation of the 30-kDa intermediate component and direct activation of procathepsin L1 by exogenous enzyme took place in the C-terminal portion of the propeptide. Elucidation of the three-dimensional structure of human procathepsin L (14) and cathepsin B (29, 30) revealed that this part of the propeptide is less structured and is readily accessible to proteases. Testament to this fact is the collective data showing that papain (32) cathepsin L (14), cathepsin B (29, 30), cathepsin S (31), and cathepsin K (25) can all be activated to mature enzymes either by *cis-* or *trans*-cleavage at various bonds within this segment. Cathepsin B of the helminth parasite *S. mansoni* can also be activated by

trans-cleavage at this segment (55, 56). However, the sites of cleavage of the wild-type and mutant procathepsin L1 differed suggesting that prior cleavage within the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} motif of the wild-type enzyme may have altered the accessibility of the C terminus to protease action. It is also evident that both processing pathways leave additional amino acids at the N terminus of the mature enzyme, whereas cathepsin L1 recovered from the medium in which parasites were cultured was fully processed to the characteristic N-terminal residue. This suggests that additional protease action, possibly by an exopeptidase, may be required to remove these additional amino acids.

In the present study we show that the secretion and activation of *F. hepatica* cathepsin L1 has similarities with that of their mammalian counterparts. Most importantly, we have elucidated that cathepsin Ls are secreted into the parasite lumen from vesicles synthesized within the gastrodermal epithelial cells and that it is possible for these to become active by autocatalysis within the slightly acidic environment of the gut. The cathepsin L proteases are synthesized in copious amounts by *F. hepatica* parasites (0.5–1.0 μg per adult parasite per hour) and early autoradiographic studies by Hanna (57) estimated that the turnover rate from the synthesis of the contents of the secretory vesicles to their liberation into the gut lumen was rapid (~1 h). The means by which these enzymes are trafficked to the secretory vesicles and the possible involvement of the Gly^{-42}-Xaa-Asn-Xaa-Phe-Xaa-Asp^{-36} motif (32) or a pH-dependent interaction with a membrane-bound receptor (58) requires investigation. A role for carbohydrate moieties may be excluded since the *F. hepatica* procathepsin L studied here contains no potential *N*-glycosylation sites, and studies have indicated that the native enzyme is not glycosylated (38, 42), although homologs from other helminths, such a *S. mansoni*, are glycosylated (59).

Mammalian cathepsins play a number of important biological functions such as protein turnover, antigen processing and tissue remodeling (50). However, they have also been implicated in various pathological conditions including tumor invasion and metastasis (50, 60, 61), osteoporosis (53, 54, 62), and chronic inflammatory disease (63, 64). In parasites, cysteine proteases perform a number of pivotal functions such as feeding, tissue penetration and immunomodulation that allow them to infect and establish disease in the host (2, 56). A number of studies have demonstrated that these enzymes represent targets at which new chemotherapeutic and immunoprophylactic strategies can be directed (2, 12, 56, 65). While the goal of improving our knowledge on the activation mechanisms of mammalian cathepsin protease is to provide new ways of treating the diseases they cause, similar studies on the parasite homologs would allow comparative analyses for the purpose of designing parasite-specific treatments. The feasibility of this approach was demonstrated in our laboratory by showing that propeptides of *F. hepatica* procathepsin Ls are potent inhibitors of their cognate enzymes but exhibit little or no activity against their mammalian homologs.

Acknowledgments—We thank Nial Larkin and John P. Dalton for the antipropeptide serum.

REFERENCES

1. MasComa, S., Bargues, M. D., and Esteban J. (1999) in *Fasciolosis* (Dalton, J. P., ed) pp. 411–435, CABI Publishers, Wallingford, UK
2. Tort, J., Brindley, P. J., Knox, D., Wolfe, K. H., and Dalton, J. P. (1999) *Adv. Parasitol.* **43,** 161–266
3. Berasain, P., Goni, F., McGonigle, S., Dowd, A., Dalton, J. P., Frangione, B., and Carmona, C. (1997) *J. Parasitol.* **83,** 1–5
4. Smith, A., Dowd, A., McGonigle, S., Keegan, P. S., Brennan, G., Trudgett, A., and Dalton, J. P. (1993) *Mol. Biochem. Parasitol.* **62,** 1–8
5. Carmona, C., Dowd, A., Smith, A., and Dalton J. P. (1993) *Mol. Biochem. Parasitol.* **62,** 9–18
6. Brady, M. T., O'Neill S. M., Dalton J. P., and Mills K. H. G. (1999) *Infect. Immun.* **67,** 5372–5378
7. O'Neill, S. M., Brady, M. T., Callanan, J. J., Mulcahy, G., Joyce, P., Mills, K. H., and Dalton, J. P (2000). *Parasite Immunol.* **22,** 147–155
8. O'Neill, S. M., Mills, K. H., and Dalton, J. P. (2001) *Parasite Immunol.* **23,** 541–547
9. Spithill, T. W., Smooker, P. M., Sexton, J. L., Bozas, E., Morrison, C. A., Creaney, J., and Parsons, J. C. (1999) in *Fasciolosis* (Dalton, J. P., ed) pp. 377–410, CABI Publishers, Wallingford, UK
10. Dalton, J. P., McGonigle, S., Rolph, T. P., and Andrews, S. J. (1996) *Infect. Immun.* **64,** 5066–5074
11. Mulcahy, G., O'Connor, F., McGonigle, S., Dowd, A., Clery, D. G., Andrews, S. J., and Dalton, J. P. (1998) *Vaccine* **16,** 932–939
12. Dalton, J. P., and Mulcahy, G. (2001) *Vet. Parasitol.* **98,** 149–167
13. Roche, L., Dowd, A. J., Tort, J., McGonigle, S., McSweeney, A., Curley, G. P., Ryan, T., and Dalton, J. P. (1997) *Eur. J. Biochem.* **245,** 373–380
14. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J.* **15,** 5492–5503
15. Menard, R., Carmona, E., Takebe, S., Dufour, E., Plouffe, C., Mason, P., and Mort, J. S. (1998) *J. Biol. Chem.* **273,** 4478–4484
16. Rozman, J., Stojan, J., Kuhelj, R., Turk, V., and Turk B. (1999) *FEBS Lett.* **459,** 358–362
17. Baker, D., Silen, J. L., and Agard, D. A. (1992) *Protein Struct. Funct. Genet.* **12,** 339–344
18. Shinde, U. P., Liu. J. J., and Inouye, M. (1997) *Nature* **389,** 520–522
19. Cappetta, M., Roth, I., Díaz, A., Tort, J., and Roche, L. (2002) *Biol. Chem.* **383,** 1215–1221
20. Ogino, T., Kaji, T., Kawabata, M., Satoh, K., Tomoo, K., Ishida, T., Yamazaki, H., Ishidoh, K., and Kominami, E. (1999) *J. Biochem.* **126,** 78–83
21. Roche, L., Tort, J., and Dalton, J. P (1999) *Mol. Biochem. Parasitol.* **98,** 271–277
22. Mason, R. W., and Massey, S. D. (1992) *Biochem. Biophys Res. Comm.* **189,** 1659–1666
23. Ishidoh, K., and Kominami, E. (1994) *FEBS Lett.* **352,** 281–284
24. Mach, L., Mort, J. S., and Glossl, J. (1994) *J. Biol. Chem.* **269,** 13036–13040
25. McQueney, M. S., Amegadzie, B. Y., D'Alessio, K., Hanning, C. R., McLaughlin, M. M., McNulty, D., Carr, S. A., Ijames, C., Kurdyla, J., and Jones, C. S. (1997) *J. Biol. Chem.* **272,** 13955–13960
26. Yamasaki, H., Mineki, R., Murayama, K., Ito, A., and Aoki, T. (2002) *Int. J. Parasitol.* **32,** 1031–1042
27. Jerala, R., Zerovnik, E., Kidre, J., and Turk, V. (1998) *J. Biol. Chem.* **273,** 11498–11504
28. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., and Katunuma, N. (1991) *EMBO J.* **10,** 2321–2330
29. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) *FEBS Lett.* **384,** 211–214
30. Podobnik, M., Kuhelj, R., Turk, V., and Turk, D. (1997) *J. Mol. Biol.* **271,** 774–788
31. Quraish, O., and Storer, A. C. (2001) *J. Biol. Chem.* **276,** 8118–8124
32. Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M. C., Storer, A. C., and Thomas, D. Y. (1995) *J. Biol. Chem.* **270,** 10838–10846
33. Dalton, J. P., O'Neill, S. M., Stack, C., Collins, P., Walsh, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D., Khaznadji, E., Moire, N., Brennan, G., Mousley, A., Kreshchenko, Maule, A, and Donnelly, S. (2003) *Int. J. Parasitol.* **33,** 621–640
34. Tepel, C., Bromme, D., Herzog, V., and Brix, K. (2000) *J. Cell Sci.* **113,** 4487–4498
35. Brix, K., Linke, M., Tepel, C., and Herzog, V. (2001) *Biol. Chem.* **382, 717–725**
36. Friedrichs, B., Tepel, C., Reinheckel, T., Deussing, J., von Figura, K., Herzog, V., Peters, C., Saftig, P., and Brix, K. (2003) *J. Clin. Investig.* **111,** 1733–1745
37. Dalton, J. P., and Heffernan, M. (1989) *Mol. Biochem. Parasitol.* **35,** 161–166
38. Dowd, A. J., McGonigle, S., and Dalton, J. P. (1994) *Eur. J. Biochem.* **223,** 91–98
39. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene (Amst.)* **77,** 51–59
40. Higgins D. R., and Cregg, J. M. (1998) *Methods Mol. Biol.* **103,** 1–15
41. Linder, S., Schliwa, M., and Kube-Granderath, E (1996) *BioTechniques* **20,** 980–982
42. Dowd, A. J., Tort, J., Roche, L., Ryan, T., and Dalton, J. P. (1997) *Mol. Biochem. Parasitol.* **88,** 163–174
43. Dowd, A. J., Dooley, M., O'Fagain, C., and Dalton, J. P. (2000) *Enz. Microb. Tech.* **27,** 599–604
44. Halton, D. W. (1967) *Parasitology* **57,** 639–660
45. Robinson G. and Threadgold, L. T. (1975) *Exp. Parasitology* **37,** 20–36
46. Smyth, J. D. and Halton, D. W. (1983) *The Physiology of Trematodes*, Second Edition, Cambridge University Press, Cambridge, UK
47. Halton (1997) *Int. J. Parasitol.* **57,** 693–704
48. Senft (1976) in *Biochemistry of Parasites and Host-Parasite Relationship* (H. Vanden Bossche, ed) pp. 335–342, North Holland Publishing Company, Amsterdam
49. Fairweather, I., Threadgold, L., and Hanna, R. E. B. (1999) in *Fasciolosis* (Dalton, J. P., ed) pp. 47–112, CABI Publishers, Wallingford, UK
50. Ishidon, K., and Kominami E. (1998) *Biol. Chem.* **379,** 131–135
51. Lorenzo, K., Ton, P., Clark, J. L., Coulibaly, S., and Mach, L. (2000) *Cancer Res.* **60,** 4070–4076
52. Guillaume-Rousselet, N., Jean, D., and Frade, R. (2002) *Biochem. J.* **367,** 219–227
53. Shi, G. P., Chapman, H. A., Bhairi, S. M., Delecuw, C., and Reddy, V. Y., and Weiss, S. J. (1995) *FEBS Lett.* **357,** 129–134

54. Drake, F. H., Dodds, R. A., James, I. E., Connor, J. R., Debouck, C., Richardson, S., Lac-Rykaczewski, E., Coleman, L., Rieman, D., Barthlow, R., Hastings, G., and Gowen, M. (1996) *J. Biol. Chem.* **271,** 12511–12516
55. Lipps, G., Fullkrug, R., and Beck, E. (1996) *J. Biol. Chem.* **271,** 1717–1725
56. Sajid, M., and McKerrow, J. H. (2002) *Mol. Biochem. Parasitol.* **120,** 1–21
57. Hanna, R. E. B. (1976) *Exp. Parasitol.* **38,** 167–180
58. McIntyre, G. F., Godbold, G. D., and Erickson, A. H. (1994) *J. Biol. Chem.* **269,** 567–572
59. Brady, C. P., Brindley, P. J., Dowd, A. J., and Dalton, J. P. (2000) *Exp. Parasitol.* **94,** 75–83
60. Yagel, S., Warner, A. H., Nellans, H. N., Lala, P. K., Waghorne, C., and Denhardt, D. T. (1989) *Cancer Res.* **49,** 3553–3557
61. Sheahan, K., Shuja, S., and Murnane, M. J. (1989) *Cancer Res.* **49,** 3809–3814
62. Kakegawa, H., Nikawa, T., Tagami, K., Kamioka, H., Sumitani, K., Kawata, T., Drobnic-Kosork, M., Lenarcic, B., Turk, V., and Katunuma (1993) *FEBS Lett.* **321,** 247–250
63. Esser, R. E., Angelo, R. A., Murphey, M. D., Watts, L. M., Thornburg, L. P., Palmer, J. T., Talhouk, J. W., and Smith, R. E. (1994) *Arthritis Rheum.* **37,** 236–247
64. Trabandt, A., Aicher, W. K., Gay, R. E., Sukhatme, V. P., Nilson-Hamilton, M., Hamilton, R. T., McGhee, J. R., Fassbender, H. G., and Gay, S. (1990) *Matrix* **10,** 349–361
65. Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S., Wong, J. Y. M., Smythe, M. L., and McManus, D. P. (1997) *Mol. Biochem. Parasitol.* **89,** 1–9

A new multi-domain member of the cystatin superfamily expressed by *Fasciola hepatica*[☆].

Eric Khaznadji[a,b], Peter Collins[b], John P Dalton[b], Yves Bigot[d] and Nathalie Moiré[c,a]*

a. INRA, UR86 "Bio-Agresseurs, Santé et Environnement". 37 380 Nouzilly. FRANCE

b. Institute for the Biotechnology of Infectious Diseases, Westbourne Street, Gore Hill, Sydney, NSW 2065, AUSTRALIA.

c. UMR 483 Immunologie Parasitaire et Vaccinologie, Université François Rabelais, UFR de Pharmacie, Parc Grandmont, 37200 Tours. 37200 Tours - FRANCE

d. Laboratoire d'Etude des Parasites Génétiques, EA3868, Université François Rabelais, UFR des Sciences et Techniques, Parc Grandmont, 37200 Tours - FRANCE

*Corresponding Author, Tel: +33 2 47 36 73 30, Fax: +33 2 47 36 72 52

Email address: nathalie.moire@univ-tours.fr

☆Note: The nucleotide sequence data reported in this paper has been submitted into the EMBL Databases under the accession number **AJ312374**.

Abstract

Cystatins are cysteine protease inhibitors that are widespread in the plant and animal kingdoms. Cystatins are expressed by helminth parasites that may employ these proteins to regulate parasite cysteine protease activity and to modulate host immune responses. Here, we describe the cloning of a cDNA encoding a high molecular weight protein of *Fasciola hepatica* that contains two domains with significant identity to the cardinal cystatin signatures and four domains with degenerated cystatin signatures. This is the first report of a multi-domain cystatin in an invertebrate species. While cystatins are divided into three evolutionary related families, our phylogenic analysis shows that all cystatin domains within this protein, like several other helminth cystatins, belong to the cystatin family 2. The DNA region encoding the domain 4 that is the best conserved at level of its cystatin signatures was expressed in *Drosophila* cells and recombinant protein was produced and purified. This protein was a potent inhibitor of the papain and of the major cysteine protease of *F. hepatica*, the cathepsin L1.

Keywords: Cystatin; Cysteine protease inhibitor; Proteases; Cathepsin L1; Parasites, *Fasciola hepatica*; Newly excysted juvenile.

1. Introduction

Fasciolosis is an infection caused by the parasitic trematode *Fasciola hepatica*. This parasite is widespread and infects a wide range of wild and domestic mammals including the sheep and bovine (Torgerson and Claxton, 1999). The disease causes poor grow performances in sheep and cattle that lead to major annual economic looses in agricultural communities in both the developed and developing world. Moreover, human fasciolosis is a public health problem in several regions including the Peru, Bolivian altiplano and the Nile delta of Egypt (Mas-Coma *et al.,* 1999a; Mas-Coma *et al.*, 1999b). Identification of *F. hepatica* antigens as candidates for vaccines and the study of host-parasite interactions has become a major research focus since the parasites resistant to chemical treatments have emerged in several countries including the United Kingdom, Ireland, the Netherlands and Australia (Gaasenbeek *et al.,* 2001). Several vaccination experiments have been performed using proteases, anti-oxidants or fatty acid binding proteins from adult and immature flukes and elicited significant protection rates (30 to 80%) after challenge infection (for review see Spithill and Dalton, 1998).

There is little data concerning proteins from newly excysted juvenile (NEJ) flukes due to the minute amount of material that can be obtained from these stages. In efforts to identify potential targets for vaccination in this developmental stage, we screened a subtractive library and identified a large number of mRNA fragments specifically or over-expressed by NEJ. One of the specific mRNAs was found to possess similarities with motifs that are signatures for cystatins. Cystatins are tight-binding inhibitors of papain-like cysteine proteases and are widespread in plants and animals. Members of the cystatin super-family are classified into three evolutionary related families (Rawlings and Barrett, 1990; Turk and Bode, 1991). Members of family 1, termed Stefins, are low molecular weight, single domain cystatins that

do not contain disulphide bridges. Members of family 2 also possess a single cystatin domain but their structure contains at least two disulphide bridges. In contrast, the members of family 3 contain three cystatin domains and are typically represented by the mammalian blood plasma kininogens. Cystatins are expressed by filarial nematodes such as *Acanthocheilonema vitae*, *Brugia malayi* (Maizels *et al.*, 2001) and *Onchocerca volvulus* (Lustigman *et al.*, 1992) and intestinal nematodes such as *Nippostrongylus brasiliensis* (Dainichi *et al.,* 2001). Recent work also shows that *Schistosoma mansoni* express a cystatin related to the family 1 (Morales *et al.*, 2004). Filarial cystatins are pathogenicity factors and are thought to play a key role in balancing the host-parasite immune relationship (Shierack *et al.*, 2003). Here, we describe the characterisation of a cDNA encoding a protein that contains six cystatin-like domains, two of which are well conserved while four are degenerate. One of the conserved domains, domain 4, suspected to be an active protease inhibitor, was expressed as a recombinant protein in *Drosophila* Schneider 2 cells and shown to be a potent inhibitor of papain and of the major cysteine protease, cathepsin L1, of *F. hepatica.* Phylogenic analyses revealed that the cystatin domains all belong to a lineage of cystatins that consist mainly of helminth cystatins that belong to the family 2. This protein is the first multi-domain cystatin to be described in invertebrates.

2. Material and methods

2.1. Parasites

Fasciola hepatica metacercaria were obtained from in vitro-infected snails *Lymnaea truncatula* maintained in our laboratory. Excystment of juvenile flukes was performed as

previously described (Wilson *et al.*, 1998). Adult *F. hepatica* flukes were recovered from naturally infected bovine livers at a local abattoir.

2.2. Construction of cDNA library from newly excysted juvenile

Total RNA was isolated from 3 adults and from 5000 NEJ *F. hepatica* with a modified caesium chloride purification method of Sambrook (Sambrook *et al.*, 1989). Reverse transcription was made with 1 μg of adult total RNA or 500 ng of juvenile worm total RNA with the Smart™ PCR cDNA synthesis kit (Clontech) and the Super Script II MMLV reverse transcriptase (Invitrogen). NEJ cDNAs (10 μg) were digested with proteinase K (Promega), purified, blunted with T4 DNA polymerase (Promega) and size fractionated through a Chromaspin 400 column (Macherey Nagel). The cDNAs of more than 500 bp were collected, precipitated with ethanol and dissolved in water. The resulting cDNAs were immediately used for ligation with *Eco R*I adaptators (Promega). The reaction was phosphorylated with T4 polynucleotide kinase (Promega) and excess adaptator was removed by passage through a Chromaspin 400 column. The fractions containing cDNAs were treated with ethanol and the resulting precipitated cDNAs were dissolved in water. Two hundred ng of cDNA were used for ligation with 500 ng of λTriplEx vector (Clontech) that had been previously digested with *Eco R*I. The ligation was mixed with 50 μl of packaging extract (Promega) and incubated for 3h at 22°C to obtain the phage library. The resulting library was plated and titrated using standard protocol.

A subtractive cDNA library enriched with specific juvenile cDNAs was made using the PCR Select™ Subtraction kit (Clontech). The subtraction products were subcloned into the pGEM-T easy vector and the recombinant vectors transfected into competent cells (Promega). One hundred and fifty colonies were individually collected and each of the

corresponding DNA was used to probe adult and juvenile total cDNAs in Southern blot. A 472 bp clone that hybridised cDNA from juvenile worms and not from adult worms was selected, sequenced and studied further.

2.3 Screening of the NEJ library

The 472 bp DNA (100 ng) was labelled with 25 μCi dCTP [αP^{32}] (ICN France) using random priming kit (Promega) and used to screen the NEJ cDNA library prepared in λTriplEx vector and also used to probe cDNAs from adult or NEJ in Southern blot. A total of 4000 colonies were plated and transferred to nylon membranes (Porablot NY plus, Macherey Nagel). These membranes were incubated overnight at 55°C in hybridization solution (6X SSC, 1% SDS, 5X Denhardt's and 50% formamide containing 100 μg/ml denatured herring sperm DNA) to which the radiolabelled 472 bp probe was added. The membranes were then washed twice for 30 min in 3X SSC/1% SDS at 55°C and twice for 30 min in 1.5X SSC/1% SDS at 55°C and then placed in contact with X-OMAT AR5 Film (Kodak) for 48h. Following a second round of screening, as described above, a 2359 bp cDNA clone was isolated and its sequence determined.

2.4 Southern blot analysis

Genomic DNA (gDNA) was extracted from adults using the Aquapure genomic DNA isolation kit (Bio-Rad) and 3μg were digested with *EcoR*I and *Rsa*I and the products were separated on a 1% agarose gel. Electrophoresis of the cDNAs was performed using the same conditions except that 250 ng sample of cDNA from adults and NEJ *F. hepatica* were used.

The DNA was denatured by incubating the gel 10 min in a 0.2N HCl solution followed by an incubation of 20 min in a denaturing solution (1.5M NaCl, 0.5N NaOH). The DNA was then transferred to a nylon membrane in 20X SSC, 3M NaCl, 300 mM sodium citrate (pH 7) (Sambrook *et al.*, 1989) and hybridization was performed as described above (section 2.3).

The 472 bp fragment (see section 2.2) was used to probe the membrane with the cDNAs and a 336 bp fragment corresponding to cystatin domain 3 was used to probe the membrane with the genomic DNA.

2. 5. Semi-quantitation of mRNA levels by PCR

PCR was performed by using 1 ng adult and NEJ cDNAs obtained as described above (section 2.2). The primers used were as followed: 5' CTG ATG TTG GGT TCC TTG G 3' for the forward primer and 5' TTA ACA GCT TAT AGA CAA ATC ATG A 3' for the reverse primer corresponding to a fragment of 1368 bp spanning from the beginning of the cystatin domain 3 to the stop codon. PCR conditions were: 1 min at 94°C, 30 s at 53°C, 60 s at 72°C. A final 10 min extension was performed at 72°C.

2.6. Expression and purification of the recombinant cystatin domain in Schneider 2 cells

The fragment of 336 bp encoding the cystatin domain 4, was generated by PCR with the forward (5' CCA TGG TCT CTC ACC GGC GCT CCT C 3' containing the *Nco* I restriction site (underlined)) and the reverse (5' CTC GAG TTA GAA GCG ACG ACA ATT ATC C 3' containing the *Xho* I restriction site (underlined) followed by a stop codon) primers. PCR conditions were : 1 min at 94°C, 30 s at 55°C, 60 s at 72°C for 25 cycles. A final 10 min extension was performed at 72°C.

The DNA fragment was first inserted into the pGEM-T easy vector (Promega) and digested with the appropriate restriction enzymes and then cloned into the expression vector pMT/BiP/V5-his (InVitrogen). The recombinant vector was sequenced to ensure that no substitutions or alterations occurred in the reading frame. The transfection into *Drosophila* Schneider cells was performed with the *Drosophila* Expression System (DES) purchased from Invitrogen as described (Khaznadji *et al.,* 2003). Transfected cell lines were grown in DES culture medium and protein synthesis was induced by the addition of copper sulphate (500 μM final) for 48 h. The supernatant was harvested and dialysed overnight at 4°C against Tris/HCl 25 mM, pH 8.0. The supernatant was loaded onto an anion exchange column (Poros 2OHQ PerSeptive Biosystems) equilibrated in the same buffer. Cystatin domain 4 was collected in the column run-through. The cystatin domain 4-containing fractions were centrifuged through a 50 kDa filter (Amicon). The cystatin domain 4 was recovered in the <50 kDa fraction, concentrated using ultrafiltration units (Centriprep, Millipore) and dialysed against 10 mM sodium acetate, pH 5.0. Total protein concentration was determined by BCA assay (Pierce, Interchim, France) and purification steps were followed in SDS-PAGE with silver staining (Khaznadji *et al.*, 2003). After purification, a faint contaminating band with a molecular mass of 50 kDa was observed and cystatin concentration was estimated by comparison with human cystatin C in SDS-PAGE after staining with Coomassie blue.

2.7. Enzymatic assays and analysis on non-denaturared polyacrylamide gel with gelatin.

The enzymes used were papain (EC3.4.22.2; Sigma) and the recombinant *F. hepatica* cathepsin L1 purified from yeast (Collins *et al.*, 2004). Human cystatin C was purchased from Calbiochem and Z-Phe-Ala-diazomethylketone (Z-Phe-Ala-CHN2) was purchased from Sigma. The colorimetric substrate benzoyl-DL-Arg-p-nitroanilide (Z-DL-Arg-p-nitroanilide)

and the fluorogenic substrate benzyloxycarbonyl-Phe-Arg-7-amido-4-methylcoumarin (Z-Phe-Arg-AMC) used to assay papain and cathepsin L1, respectively, were from Bachem. The enzymatic assay buffer was 100 mM sodium phosphate, containing 1 mM dithiothreitol and 2 mM EDTA, adjusted to pH 6.5 for papain and to pH 6.0 for cathepsin L1. The reactions were performed at room temperature for papain and at 37°C for cathepsin L1. The reactions were allowed to proceed for 30 min and then stopped by the addition of 200 μl of 10 M acetic acid. Released p-nitroanilide was measured by spectrophotometry at 405 nm. Z-Phe-Arg-AMC hydrolysis was measured by fluorometry with an excitation at 370 nm and emission at 440 nm.

The inhibition of *F. hepatica* cathepsin L1 was also visualised on non-denaturated polyacrylamide gel containing gelatin substrate. Recombinant cathepsin L1 was activated in sodium acetate, pH 5.5, for 2 h and dialysed against PBS, pH 7.3 (Collins *et al.*, 2004). The activated enzyme (5μM) was incubated in PBS, pH 7.3, with an equal amount of recombinant cystatin domain 4 for 30 min. Samples of cathepsin L1, cystatin and the mixture were then applied to 8% polyacrylamide gels that had been set with or without 0.1% (w/v) gelatin (Roche *et al.*, 1997). After electrophoresis, the gel without gelatin was stained with Coomassie Blue while the other was incubated overnight in sodium acetate buffer, pH 5.5, and then stained.

2.8. Databank searches and sequence analyses

The Infobiogen facilities were used for database searches (Genbank release 132, updated 11/16/2002, Swissprot release 40 and TrEMBL 21, both updated 12/06/2002), sequence alignments and calculations. First, 97 sequences (Table 1) identified from BLAST searches and from the PROSITE database were aligned using CLUSTAL W (Thompson *et*

al., 1994). The beginning of the sequences was the conserved G residue and the end the last conserved C. At each step, the sequences were manually adjusted to facilitate the quality of alignment. Phylogenic analyses were performed using the parsimony and Neighbour Joining software of the PHYLIP package, version 3.5c (Felsenstein, 1993). Due to the important size of the data file, consensus trees were calculated from a sampling of 100 bootstraps.

3. Results

3.1.Cloning of a cDNA encoding a protein containing six putative cystatin domains

Screening of a subtractive library enriched in NEJ cDNAs of *F. hepatica* allowed the isolation of a 472 bp cDNA fragment that contained a sequence encoding a complete cystatin domain. This fragment was then used to isolate a 2359 bp cDNA from a second library that contained complete NEJ cDNAs of *F. hepatica*. The 472 bp cDNA fragment was also used as probe to screen adult and NEJ cDNAs on Southern blot (Fig. 1); a cDNA fragment of about 2359 bp was detected in the NEJ sample but not in adults. The cDNAs from adults and NEJ were also used in PCR to compare the cystatin expression in both stages (Fig. 2). In this case, a 1368 bp transcript was detectable in the 2 samples. However, semi-quantification by PCR showed that while the transcript is detectable after 20 amplification cycles in the NEJ sample it was only detectable in adults after 25 cycles. Collectively, this data confirms that the mRNAs for the cystatin are more abundant in NEJ stage than in adult stage. To confirm the presence of the gene in the parasite, Southern blot was performed on genomic DNA (Fig.3). Adult genomic DNA was digested and probed with a 336 bp fragment. The probe corresponds to the *EcoR* I-*Rsa* I restriction fragment of the cDNA and recognized a major fragment of

about 374 bp as expected. A fragment of about 250 bp hybridized less intensively. This fragment may correspond to a region with weaker homology.

Analysis of the 2359 bp cDNA isolated from the complete NEJ library, revealed the presence of several potential open reading frames (ORF) and one of 2072 bp that encodes a putative protein of 690 amino acids. This ORF is flanked at its 3' extremity by a 287 bp untranslated region that contained a typical polyadenylation signal located 23 bp upstream to the polyA tail. The 5' region is not complete and lacks the typical consensus sequence for initiation of transcription and the start codon.

Search for similarities between the amino acid sequence of the protein encoded by the 2072 bp ORF and the databases revealed that the protein contained two putative cystatin domains of 106 amino acids from amino acid 238 to 343 (domain 3) and 355 to 460 (domain 4), respectively. The best hits found with these two domains were observed with the mouse kininogen (25% identity) and for the ovarian cystatin from the common carp (29% identity) for domains 3 and 4, respectively. Search for similarities within the amino acid sequence of the protein encoded by the 2072 bp ORF revealed the presence of six tandemly repeated domains that spanned along all the protein length (Fig. 4). The six domains exhibited low sequence similarities to each other (15.5 to 29%) and their extremities were difficult to define since no significant similarities were found between the residues located at their C and N-termini. While these observations indicate that these six domains may have originated from a single ancestral cystatin domain, domain 4 was the only to contain the three cardinal signatures of a cystatin domain (Fig. 4): the N-terminal glycine (referred as G^{355} for domain 4), the highly conserved central motif QxVxG that was located in the first hairpin loop of the protein and corresponded to the papain-binding site, and the second C-terminal hairpin-loop containing the PW motif. This domain also contained four cysteine residues. All other domains lacked one or several of these features. Domain 3 lacked the proline residue of the

PW motif and showed a R/Q substitution within the QxVxG motif. The domains 2, 5 and 6 lacked the QxVxG motif and the conserved G in their N-termini. Only domain 2 showed the C-terminal loop PW, the two others lacking the conserved W residue. Domain 1 was incomplete at its N-terminus and only contained the conserved motif QxVxG. Finally, the domains 1 and 2 had two and three cysteine residues whereas the four others had four cysteine residues.

3.2. Expression and purification of the cystatin domain 4

Since the cystatin domain 4 has conserved papain-binding sites we suspected that it might be a potent inhibitor of cysteine proteases. To demonstrate this property the region of cDNA that encodes this domain was transformed into *Drosophila* S2 cells using a vector that directs expressed products into the culture medium. Recombinant domain 4 was secreted in the S2 cells media at levels of approximately 10 mg/L. The cystatin domain 4 was purified by a single anion exchange chromatography step followed by ultra-centrifugation through a 50 kDa cut-off filter. The molecular mass of the recombinant purified protein was around 19 kDa as shown in Fig. 5.

3.3. Inhibitory activity of cystatin domain 4 against papain and F. hepatica cathepsin L1

The purified *F. hepatica* cystatin domain 4 was assayed for inhibitory activity against the cysteine protease papain and the major *F. hepatica* cathepsin L1 protease. Five μM of cystatin domain 4 completely inhibits (>98% inhibition) 1 μM of papain as measured by Z-DL-Arg-p-nitroanilide hydrolysis. In a similar assay with the recombinant *F. hepatica* cathepsin L1, the recombinant cystatin inhibits completely the Z-Phe-Arg-AMC hydrolysis.

Figure 6A shows the inhibition of gelatinolytic cathepsin L1 activity by domain 4 in polyacrylamide gel under non-denaturating conditions. In these gels, cathepsin L migrates as a smear and a corresponding broad band of gelatinolytic activity in polyacrylamide gel containing gelatin (Fig. 6A, compare lanes 1). Cystatin remains at the top of the gel (Fig. 6A, Panel a, lane 3) but when mixed at equimolar ratios with cathepsin L1 it complexes with the enzyme and completely inhibits its gelanolytic activity (Fig. 6A, compare lanes 2).

The inhibition profile of domain 4 in comparison with human cystatin C and the cathepsin L-specific inhibitor Z-Phe-Ala-CHN2 is shown in Fig. 6B. The decrease of cathepsin L1 activity with cystatin concentration followed the hyperbolic inhibition curve of the tight-binding inhibitors. This inhibition profile demonstrated that domain 4 has a similar potency to these inhibitors, i.e. an inhibition of 95% at an equimolar ratios, suggesting that the recombinant domain 4 is produced as a fully functional correctly-folded protein. Similar studies were carried out with domain 3; however, its expression in insect cells was low making purification difficult. A partially purified fraction obtained after anion exchange chromatography exhibited inhibition of papain and cathepsin L but kinetic data could not be established (data not shown).

3.4. Phylogenic relationship with other cystatins

In an attempt to further characterize each of the *F. hepatica* cystatin domains, phylogenic analyses were developed using sequences available in databanks. Overall, our calculations were done with a data comprising of 97 cystatin sequences identified in the genomes of plants, insects, nematodes, trematodes and cold- and warm-blood vertebrates. Phytocystatins were used as outgroup to root the trees as previously described (Margis *et al.*, 1998).

Consensus trees obtained using the parsimony and Neighbour-Joining methods were similar in topology and in the bootstrap values at nodes.

The consensus tree developed with the parsimony method and a sampling of 100 bootstraps is presented in Fig. 7. It revealed four animal cystatin lineages that were named family 1 to 4. Their existence is sustained by significant bootstrap values ranging from 62 to 97% at nodes located on the root of each of these 4 families. Family 1 to 3 correspond to those previously described (Rawlings and Barrett, 1990; Turk and Bode, 1991) and our analyses demonstrated that each of them is composed of two sub-lineages comprising cystatins from invertebrate or vertebrate origin, respectively. Surprisingly, our data also revealed a fourth family so far never described and that is presently only composed of proteins from invertebrate origin. Concerning the six domains of the *F. hepatica* polycystatin, we observed that all belong to the invertebrate sub-lineage of the Family 2. Their clustering within a single branch indicates that each domain originated from a single ancestral cystatin by duplication.

4. Discussion

The present study describes the sequence of a novel multi-domain cystatin that is over-expressed in the infective NEJ stage of *F. hepatica.* The only other molecules with multiple cystatin domains are the mammalian kininogens, although these are members of the family 3 cystatins. Kininogens are multi-functional proteins containing three cystatin domains and biologically active bradykinin domains that are potent vasodilators (Kaufmann *et al*, 1993). However, while two of the cystatin domains of kininogen are potent inhibitors of cysteine proteases the other is degenerate and lacks inhibitory activity (Salvesen *et al.*, 1986). *Fasciola hepatica* multi-cystatin shows similar structural features of the kininogens with functional and

degenerated domains, however our phylogenic analysis suggests that it is more related to family 2 cystatins. The family 2 cystatins represented by human cystatin C and chicken egg cystatin are secretory proteins with one cystatin domain and two characteristic disulphide bridges. While human cystatins E and F also belong to the family 2 cystatins they exhibit a closer relationship with domain 3 of kininogen and, contrary to most cystatins, they have a restricted pattern of expression (Ni *et al.*, 1998). Nematode cystatins have also been classified in the family 2 cystatin but only contain one disulphide bridge. A new subgroup of the family 2 cystatins, cystatin-related epididimal spermatogenic (CRES) proteins, that have tissue-specific expression (reproductive tract), was recently described by Cornwall and Hsia (2003). These cystatins contain only the PW site but lack the conserved QxVxG motif. Therefore, it is becoming clear that family 2 cystatins now consists of several groups and sub-groups. In the phylogenic study, we show that *F. hepatica* cystatins and the nematode cystatins are included in the invertebrate sub-group of the family 2.

Cystatins have a papain-binding site that binds to the catalytic site of papain-like proteases and inhibits them reversibly (Abrahamson, 1994). This site is created by several conserved regions of the protein coming together including, a glycine in the N-terminal region, a central motif QxVxG in the first hairpin loop of the protein and a PW motif in the second hairpin loop in the C terminal region (Abrahamson *et al.*, 1987; Bjork *et al.*, 1996; Hall *et al.*, 1993). The *F. hepatica* cystatin domains 3 and 4 contain a glycine in the N-terminal part, and the highly conserved tryptophan in the C-terminal region. However, domain 3 lacks the proline residue of the PW motif of the second hairpin loop but this clearly does not affect its inhibitory activity. The domain II of human kininogen also lacks this proline residue but is, nevertheless, a potent inhibitor of cysteine proteases (Salvesen *et al.*, 1986). A second tryptophan residue is present in the *F. hepatica* cystatins three amino acids before the PW motif where a tyrosine residue is usually conserved within the cystatin super

family. However all the nematode cystatins and the human cystatin F possess a tryptophan and hence this substitution might be associated with a specific function (Lustigman *et al.*, 1992; Ni *et al.*, 1998). The central motif QxVxG is conserved for the domain 4, however, the domain 3 contains a positively charged arginine (RIVAG) in this motif which is unusual but similarly found in the human cystatin F (Ni *et al.*, 1998), and may affect the target specificity. The *F hepatica* cystatin domains 3 and 4 both possess the four highly conserved cysteine residues toward the C-terminal end, corresponding to those that form disulphide bridges within the cystatin members of the families 2 and 3 (Rawlings and Barrett, 1990). Therefore we assume that these bridges are intact in these proteins. The *F. hepatica* cystatin domains 1, 2, 5, and 6 lack one or several features of the papain-binding site. Domains 1 and 2 possess, respectively, the conserved central and the C-terminal motif. Additionally, they contain only one conserved disulphide bridge like nematode cystatins (Maizels *et al.*, 2001). Cystatin domains 5 and 6 lack the 3 conserved regions and only show the 2 putative conserved disulphide bridges. A similar configuration is found in the human kininogen; the cystatin domain 1 of this molecule lacks all the three parts of the papain-binding site but have two conserved disulphide bridges. This degenerated domain does not exhibit inhibitory effect on cysteine proteases (Salvesen *et al.*, 1986).

Given that we have shown that two cystatin domains of the NEJ multi-domain protein described here exhibit the ability to inhibit in vitro cathepsin L activity it is natural to suspect that the multi-cystatin is involved in regulating cysteine protease activity in this stage of parasite. The NEJ is the invasive stage of the parasite but remains in a protease-free environment until it is ingested into the intestines. NEJ of *F. hepatica* express copious amounts of cathepsin B (Wilson *et al.*, 1998) and cathepsin L proteases (Harmsen *et al.*, 2004). Within a short period after excystment the expression of both cathepsin B and cathepsin L cysteine proteases are switched on to allow the rapid penetration of the host

intestinal wall (Wilson *et al.*, 1998; Mulcahy *et al.*, 1999); hence cystatins may be required to protect the NEJ from its own proteases.

An additional function of the cystatin may be the modulation of host immune responses. Parasite nematode cystatins are involved in immune evasion such as antigen presentation inhibition (Dainichi *et al.*, 2001; Manoury *et al.*, 2001), polyclonal activation and induction of a Th2 immune response (Schönemeyer *et al.*, 2001). A similar function for at least the two functional cystatin domains of the multi-cystatin may be expected. However, the inhibition of host antigen presentation by nematode cystatins is mediated by the legumain inhibitory motif within the cystatin. The legumain inhibitory motif is distinct from the papain binding motif and compounded by three essential amino acids: serine, asparagine and aspartic acid (Alvarez-Fernandez *et al.*, 1999). Since the *F. hepatica* cystatin domains are devoid of any legumain-binding motif, we suggest that the *F. hepatica* multi-cystatin is not involved in the inhibition of host antigen presentation.

Fasciola hepatica has been shown to polarize the host cytokine response to a Th2 profile. Analysis of cytokine production in infected mice (O'Neill *et al.*, 2000) and rats (Tliba *et al.*, 2002) have shown that only Th2 cytokines (IL4, IL10) and little or no IFNγ are produced. Tliba *et al.* (2002) have studied the very early response in the liver where only juvenile and immature flukes are present and showed that an increase in IL10 production and a reduced proliferative response of liver mononuclear cells occurs soon after infection. Filarial cystatins were shown to induce the massive production of IL10. This production of IL10 has been involved in the induction of an anti-inflammatory environment with a strong inhibition of cellular proliferation, which favours the survival of the worms (Hartmann and Lucius, 2003). *Fasciola hepatica* cystatins produced by NEJ may have such an effect and this immunoregulatory function may, as described for filarial cystatins (Schierack *et al.*, 2003), be a specific adaptation to the parasitic life style.

Acknowledgements :

We are grateful to Florence Carreras and Katia Courvoisier for maintaining the *Fasciola hepatica* life cycle and Michèle Péloille for DNA sequencing. We wish to thank Dr Mary Sekiya for providing the fluke genomic DNA. This work was supported by a grant from the Région Centre (France).

References

Abrahamson, M., 1994. Cystatins. Methods Enzymol. 244, 685-700.

Abrahamson, M., Ritonja, A., Brown, M. A., Grubb, A., Machleidt, W. Barrett, A. J., 1987. Identification of the probable inhibitory reactive sites of the cysteine proteinase inhibitors human cystatin C and chicken cystatin. J. Biol. Chem. 262, 9688-9694.

Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., Abrahamson, M., 1999. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. J. Biol. Chem. 274, 19195-19203.

Bjork, I., Brieditis, I., Raub-Segall, E., Pol, E., Hakansson, K., Abrahamson, M., 1996. The importance of the second hairpin loop of cystatin C for proteinase binding. Characterization of the interaction of Trp-106 variants of the inhibitor with cysteine proteinases. Biochemistry. 35, 10720-10726.

Collins, P.R., Stack, C.N., O'Neill, S.M., Doyle, S., Ryan, T., Brennan, G.P., Mousley, A., Stewart, M., Maule, A.G., Dalton, J. P., Donnelly, S., 2004. Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. J. Biol. Chem. 279, 17038-17046.

Cornwall, G. A., Hsia, N., 2003 A new subgroup of the family 2 cystatins. Mol Cell Endocrinol. 28, 1-8.

Dainichi, T., Maekawa, Y., Ishii, K., Zhang, T., Nashed, B. F., Sakai, T., Takashima, M., Himeno, K., 2001. Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response. Infect. Immun. 69, 7380-7386.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) version 3.5.c University of Washington, Seattle.

Gaasenbeek, C. P., Moll, L., Cornelissen, J. B., Vellema, P., Borgsteede, F. H., 2001. An experimental study on triclabendazole resistance of *Fasciola hepatica* in sheep. Vet. Parasitol. 95, 37-43.

Hall, A., Dalboge, H., Grubb, A., Abrahamson, M., 1993. Importance of the evolutionarily conserved glycine residue in the N-terminal region of human cystatin C (Gly-11) for cysteine endopeptidase inhibition. Biochem. J. 291, 123-129.

Harmsen, M. M., Cornelissen, J. B. W. J., Buijs, E. C. M., Boersma, W. J. A., Jeurissen, S. H. M., van Milligen, F. J., 2004. Identification of a novel *Fasciola hepatica* cathepsin L protease containing protective epitopes within the propeptide. Int. J. Parasitol. 34, 675-682.

Kaufmann, J., Haasemann, M., Modrow, S., Muller-Esterl, W., 1993. Structural dissection of the multidomain kininogens. Fine mapping of the target epitopes of antibodies interfering with their functional properties. J. Biol. Chem. 268, 9079-9091.

Khaznadji, E., Boulard, C., Moiré, N., 2003. Expression of functional hypodermin A, a serine protease from *Hypoderma lineatum* (Diptera, Oestridae), in Schneider 2 cells. Exp Parasitol. 104, 33-39.

Lustigman, S., Brotman, B., Huima, T., Prince, A. M., McKerrow, J. H., 1992. Molecular cloning and characterization of onchocystatin, a cysteine proteinase inhibitor of *Onchocerca volvulus*. J. Biol. Chem. 267, 17339-17346.

Maizels, R. M., Gomez-Escobar, N., Gregory, W. F., Murray, J., Zang, X., 2001. Immune evasion genes from filarial nematodes. Int. J. Parasitol. 31, 889-898.

Manoury, B., Gregory, W. F., Maizels, R.M., Watts, C., 2001. Bm-CPI-2, a cystatin homolog secreted by the filarial parasite *Brugia malayi*, inhibits class II MHC-restricted antigen processing. Curr. Biol. 11, 447-451.

Margis, R., Reis, E.M., Villeret, V., 1998. Structural and phylogenetic relationships among plant and animal cystatins. Arch. Biochem. Biophysics. 359, 24-30.

Mas-Coma, M.S., Esteban, J.G., Bargues, M.D., 1999a. Epidemiology of human fascioliosis: a review and proposed new classification. Bull. World Health Organ. 77, 340-346.

Mas-Coma, S., Angles, R., Esteban, J.G., Bargues, M.D., Buchon, P., Franken, M., Strauss, W., 1999b. The Northern Bolivian Altiplano: a region highly endemic for human fascioliosis. Trop. Med. and Int. Health 4, 454-467.

Morales F.C., Furtado D.R., Rumjanek F.D., 2004. The N-terminus moiety of the cystatin SmCys from *Schistosoma mansoni* regulates its inhibitory activity *in vitro* and *in vivo*. Mol. Biochem. Parasitol. 134, 65-73.

Mulcahy, G., Joyce, P., Dalton, J.P., 1999. Immunology of *Fasciola hepatica*. *In* Fasciolosis, (ed. J.P. Dalton). Wallingford, UK.: CAB International. pp. 314-376.

Ni, J., Fernandez, M.A., Danielsson, L., Chillakuru, R.A., Zhang, J., Grubb, A., Su, J., Gentz, R., Abrahamson, M., 1998. Cystatin F is a glycosylated human low molecular weight cysteine proteinase inhibitor. J. Biol. Chem. 273, 24797-27804.

O'Neill, S.M., Brady, M.T., Callanan, J.J., Mulcahy, G., Joyce, P., Mills, K.H., Dalton, J.P., 2000. *Fasciola hepatica* infection downregulates Th1 responses in mice. Parasite Immunol. 22, 147-155.

Rawlings, N. D, Barrett, A.J., 1990. Evolution of proteins of the cystatin superfamily. J. Mol. Evol. 30, 60-71.

Roche, L., Dowd, A.J., Tort, J., McGonigle, S., McSweeney, A., Curley, G.P., Ryan, T., Dalton, J.P., 1997. Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae*. Eur. J. Bioch. 245, 373-380.

Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A., Barrett, A.J., 1986. Human low-Mr kininogen contains three copies of a cystatin sequence that are divergent in structure and in inhibitory activity for cysteine proteinases. Biochem. J. 234, 429-434.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. A laboratory Manual. New York: Cold Spring Harbor.

Schönemeyer, A., Lucius, R., Sonnenburg, B., Brattig, N., Sabat, R., Schilling, K., Bradley, J., Hartmann, S., 2001. Modulation of human T cell responses and macrophage functions by onchocystatin, a secreted protein of the filarial nematode *Onchocerca volvulus*. J. Immunol. 167, 3207-3215.

Shierack, P., Lucius, R., Sonneburg, B., Schilling, K., Hartmann, S., 2003 Parasite-specific immunomodulatory functions of filarial cystatin. Infect. Immun. 71, 2422-2429.

Spithill, T.W., Dalton, J.P., 1998. Progress in development of liver fluke vaccines. Parasitol. Today. 14, 224-228.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: Improving the sensitivity of progressive multiple sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Tliba, O., Moiré, N., Le Vern, Y., Boulard, C., Chauvin, A., Sibille, P., 2002. Early hepatic immune response in rats infected with *Fasciola hepatica*. Vet. Res. 33, 261-70.

Torgerson, P., Claxton, J., 1999. Epidemiology and Control. *In* Fasciolosis, (ed. J.P. Dalton). Wallingford, UK.: CAB International. pp. 113-149.

Tort, J., Brindley, P.J., Knox, D., Wolfe, K.H., Dalton, J.P., 1999. Proteinases and associated genes of parasitic helminths. Adv. Parasitol. 43, 161-266.

Turk, V., Bode, W., 1991. The cystatins: protein inhibitors of cysteine proteinases. FEBS Lett. 285, 213-219.

Wilson, L.R., Good, R.T., Panaccio, M., Wijffels, G.L., Sandeman, R.M., Spithill, T.W., 1998. *Fasciola hepatica*: characterization and cloning of the major cathepsin B protease secreted by newly excysted juvenile liver fluke. Exp. Parasitol. 88, 85-94.

Table 1. Cystatins used for the phylogenic study.

Protein Name	Species	Acc. N°.	Tree name
Cysteine proteinase inhibitor	*Zea mays*	Q41897	Cyst1ZeaM
Cystatin II	*Zea mays*	Q41825	Cyst2ZeaM
Cysteine proteinase inhibitor	*Zea mays*	P93627	Cyst3ZeaM
Cystatin I precursor	*Zea mays*	P31726	Cyst4 ZeaM
Cystatin A	*Helianthus annuus*	Q10992	Cyst-HelA
Multicystatin precursor	*Solanum tuberosum*	P37842	Cyst1to Cyst8 SolT
Putative cysteine proteinase inhibitor B	*Arabidopsis thaliana*	O22202	Cyst1-AraT
Cysteine proteinase inhibitor	*Arabidopsis thaliana*	Q41906	Cyst2-AraT
Cysteine proteinase inhibitor like	*Arabidopsis thaliana*	O23494	Cyst3-AraT
Cysteine proteinase inhibitor	*Arabidopsis thaliana*	Q41916	Cyst4-AraT
Cysteine proteinase inhibitor	*Sorghum vulgare*	Q41294	Cyst-SorV
Cysteine protease inhibitor	*Pyrus communis*	O24462	Cyst-PyrC
Cysteine proteinase inhibitor	*Brassica campestris*	Q39268	Cyst1-BraC
Cysteine proteinase inhibitor	*Brassica campestris*	Q39270	Cyst2-BraC
Cysteine proteinase inhibitor	*Brassica campestris*	Q42380	Cyst3-BraC
Cysteine proteinase inhibitor	*Glycine max*	O04720	Cyst1-GlyM
Cysteine proteinase inhibitor	*Glycine max*	Q39840	Cyst2-GlyM
Cysteine proteinase inhibitor	*Glycine max*	Q39841	Cyst3-GlyM
Cysteine proteinase inhibitor	*Glycine max*	Q39842	Cyst4-GlyM
Cysteine proteinase inhibitor	*Ambrosia artemisiifolia*	Q38678	Cyst1-AmbA
Cysteine proteinase inhibitor	*Ricinus communis*	Q43635	Cyst-RicC
Cysteine proteinase inhibitor	*Carica papaya*	Q39561	Cyst-CarP
Oryzacystatin-I	*Oryza sativa*	P09229	Cyst1-Ory-S
Oryzacystatin-II	*Oryza sativa*	P20907	Cyst2-Ory-S
Cysteine proteinase inhibitor	*Vigna unguiculata*	Q06445	Cyst-VigU
Cystatin precursor	*Ornithodoros moubata*	AAS01021	Cyst-OrnM
Putative secreted cystatin	*Ixodes scapularis*	Q8MVB6	Cyst-IxoS
Putative cystatin	*Ixodes ricinus*	CAD68002	Cyst-IxoR
Cystatin	*Lepydoglyphus destructor*	Q8WQ46	Cyst-LepD
Cystatin	*Theromyson tessulatum*	Q8IT43	CystB-TheT
Cystatin-like protein	*Drosophila melanogaster*	P23779	CystL-DroM
Cystatin	*Drosophila melanogaster*	1702209A	Cyst-DroM
Sarcocystatin A precursor	*Sarcophaga peregrina*	P31727	CystA-SarP
Sarcocystatin B	*Sarcophaga crassipalpis*	Q8TYOY1	Cyst-B-SarC
Sarcocystatin A	*Sarcophaga crassipalpis*	BAB88880	Cyst-A-SarC
Cystatin precursor	*Tachypleus tridentatus*	JC4536	Cyst-TacT
Cysteine protease inhibitor like (cli-1)	*Caenorhabditis elegans*	O61973	Cyst-CaeE-cli1
Cysteine protease inhibitor like (cli-2)	*Caenorhabditis elegans*	Q9TYY2	Cyst-CaeE-cli2
Cystatin-1	*Haemonchus contortus*	O44396	Cyst-HaeC
Nippocystatin	*Nippostrongylus brasiliensis*	Q966W0	Cyst-NipB
Ls-cystatin precursor	*Litomosoides sigmodontis*	Q9NH95	Cyst-LitS
CPI-1	*Brugia malayi*	P90698	Cyst1-BruM
CPI-2	*Brugia malayi*	O16159	Cyst2-BruM
Cystatin	*Acanthocheilonema vitae*	Q17108	Cyst-AcaV
CPI-1	*Onchocerca volvulus*	Q9UA1	CPI1-OneV
cystatin	*Onchocerca volvulus*	Q25620	Cyst-OncV
Onchocystatin precursor	*Onchocerca volvulus*	P22085	CystX-OncV
Cysteine protease inhibitor	*Schistosoma mansoni*	AAQ16180	Cyst1-SchM
Cystatin homolog	*Schistosoma mansoni*	A48570	Cyst2-SchM
Dopamine receptor	*Fugu rubripes*	Q90517	Cyst-FugR
Cystatin	*Acipenser sinensis*	AAK16731	Cyst-AciS
Cystatin precursor	*Cyprinus carpio*	P35481	Cyst-CypC
Cystatin	*Bitis ariensis*	P08935	Cyst-BitA
Cystatin	*Bitis gabonica*	AAR24527	Cyst-BitG

Protein Name	Species	Acc. N°.	Tree name
Family-2 cystatin precursor	*Onchorhynchus keta*	Q98967	Cyst-OncK
Cystatin precursor	*Onchorhynchus mykiss*	Q91195	Cyst-OncM
Cystatin	*Gallus gallus*	P01038	Cyst-GalG
Cystatin 8	*Mus musculus*	AAH49753	CRES-MusM
Hypothetical cysteine proteases inhibitor containing protein	*Mus musculus*	Q9DAP1	CRES-1-MusM
Stefin 1	*Mus musculus*	P35173	Cyst1-MusM
Stefin 2	*Mus musculus*	P35174	Cyst2-MusM
Stefin 3	*Mus musculus*	P35175	Cyst3-MusM
Stefin B	*Mus musculus*	Q62426	CystB-MusM
Cystatin C	*Mus musculus*	P21460	CystC-MusM
Leukocystatin	*Mus musculus*	O89098	CystF-MusM
Kininogen domain 2	*Mus musculus*	O08677	KNG2-MusM
Kininogen domain 3	*Mus musculus*	O08677	KNG3-MusM
Cystatin A	*Rattus norvegicus*	P01039	CystA-RatN
Stefin B	*Rattus norvegicus*	P01041	CystB-RatN
Cystatin C	*Rattus norvegicus*	P14841	CystC-RatN
Cystatin S	*Rattus norvegicus*	P19313	CystS-RatN
Kininogen domain 2	*Rattus norvegicus*	P08934	KNG2-RatN
Kininogen domain 3	*Rattus norvegicus*	P08934	KNG3-RatN
Rat T-kininogen	*Rattus norvegicus*	P01048	T-KNG2-RatN T-KNG3-RatN
Major acute phase alpha-1 protein	*Rattus norvegicus*	P70517	MAAP1-RatN MAAP2-RatN
Cystatin C	*Oryctolagus cuniculus*	O97862	Cyst-OryC
Stefin A1	*Sus scrofa*	Q28988	Cyst1-SusS
Stefin A5	*Sus scrofa*	Q28986	Cyst5-SusS
Stefin A8	*Sus scrofa*	Q28987	Cyst8-SusS
Stefin B	*Sus scrofa*	Q29290	CystB-SusS
Stefin D1	*Sus scrofa*	P35479	Cysti-SusS
Stefin B	*Ovis aries*	O10994	CystB-OviA
Stefin A	*Bos taurus*	P80416	CystA-Bos
Stefin B	*Bos taurus*	P25417	CystB1-Bos
Cystatin C	*Bos taurus*	P01035	CystB2-Bos
Stefin C	*Bos taurus*	P35478	CystO-BosT
Kininogen	*Bos taurus* *Bos taurus*	P01044	KNG2-BosT KNG3-BosT
Cystatin C	*Saimiri sciureus*	O19093	CytC-SaiS
Cystatin C	*Macaca mulatta*	O19092	CytC-MacM
Stefin A	*Homo sapiens*	P01040	CystA-HomS
Stefin B	*Homo sapiens*	P04080	CystB-HomS
Cystatin C	*Homo sapiens*	P01034	CystC-HomS
Cystatin D	*Homo sapiens*	P28325	CystD-HomS
Cystatin SN precursor	*Homo sapiens*	P01037	CystN-HomS
Cystatin E	*Homo sapiens*	Q15828	CystM-HomS
Cystatin S precursor	*Homo sapiens*	P01036	CystS-HomS
Cystatin SA precursor	*Homo sapiens*	P09228	CystT-HomS
Leukocystatin	*Homo sapiens*	O76096	Cyst7-HomS
Cystatin 8	*Homo sapiens*	O60676	Cyst8-HomS
CS13	*Homo sapiens*	Q8WXU6	Cyst11-HomS
Kininogen	*Homo sapiens*	P01042	KNG2-HomS KNG3-HomS

Legends to figures

Fig.1. Southern blot analysis of adult and juvenile cDNA. cDNAs from adult and juvenile parasite were separated by agarose gel electrophoresis, transferred onto nylon membrane and probed with radiolabelled 472 bp probe.

Fig.2. PCR amplification on adult and juvenile cDNA. cDNAs (1 ng) from adult (lanes 1,3 5, 7) and juvenile (lanes 2, 4, 6, 8) parasite were submitted to 15 (lanes 1, 2), 20 (lanes 3, 4), 25 (lanes 5, 6) and 30 cycles (lanes 7, 8, 9) of PCR. Lane 9 is the control without DNA. PCR products were electrophoresed on agarose gel and stained with ethidium bromide. Molecular sizes of the DNA ladder (M) are in bp.

Fig.3. Southern blot analysis of genomic DNA. Genomic DNA was digested and the digestion products were electrophoresed on agarose gel and stained with ethidium bromide (lane 1) and transferred onto nylon membrane and probed with radiolabelled 336 bp probe and exposed for 24h (lane 2). Molecular sizes of the DNA ladder (M) are in bp.

Fig. 4. Alignment of F. hepatica cystatin domains. The six domains were aligned using ClustalW and adjusted manually. The three parts of the papain binding domain are framed. Disulphide bridges are shown with brackets and cysteine residues are in pale grey. Numbering corresponds to the amino acids of each domain.

Fig.5. Expression and purification of the recombinant F. hepatica cystatin domain 4. A 15% SDS-polyacrylamide gel was loaded with *Drosophila* supernatants before induction (lane 1) *Drosophila* supernatants after induction with copper sulphate (lane 2), fractions eluted from the anion exchange chromatography (lane 3), and fractions recovered from the 50 kDa filter (lane 4). Molecular mass markers (M) are indicated in kDa. Gels were stained with silver nitrate.

Fig.6. Inhibitory activity of recombinant cystatin. A. Purified and activated recombinant cathepsin L1 (lanes 1) was complexed with an equal amount (5μM) of cystatin (lanes 2) and analysed by non denaturated polyacrylamide gels prepared in the absence (panel a) and presence (panel b) of gelatin. Gels were stained immediately after electrophoresis (panel a) or were incubated overnight before staining to observe gelatinolytic activity (panel b). Lanes 3 show recombinant cystatin alone. **B.** 1μM of recombinant *F. hepatica* cathepsin L1 was incubated 30 min with increasing concentrations of recombinant *F. hepatica* cystatin domain 4 (●), human cystatin C (■) and cathepsin L inhibitor Z-Phe-Ala-CHN2 (▲). Residual activities were monitored by Z-Phe-Arg-AMC hydrolysis.

Fig.7. Phylogenic relationship of F. hepatica within the 3 cystatin families. Consensus tree illustrating the phylogenic relationship of *F. hepatica* domains (*Cyst1-6-FasH*) with other known cystatin proteins. The tree is based on amino acid sequences of the mature proteins beginning at the conserved G at the C-terminal end of the protein and ending at the conserved C at the N-terminal end of the protein (this corresponds to amino acid 37 to 142 of human cystatin C). Only bootstrap values at nodes higher than 40% are represented. The references for each protein are given in the Table 1.

Fig.1

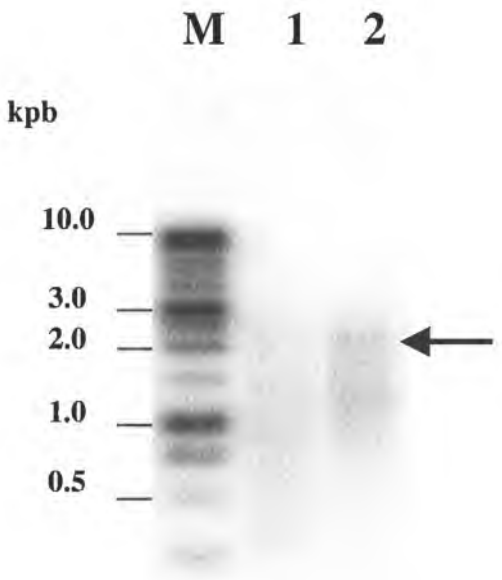
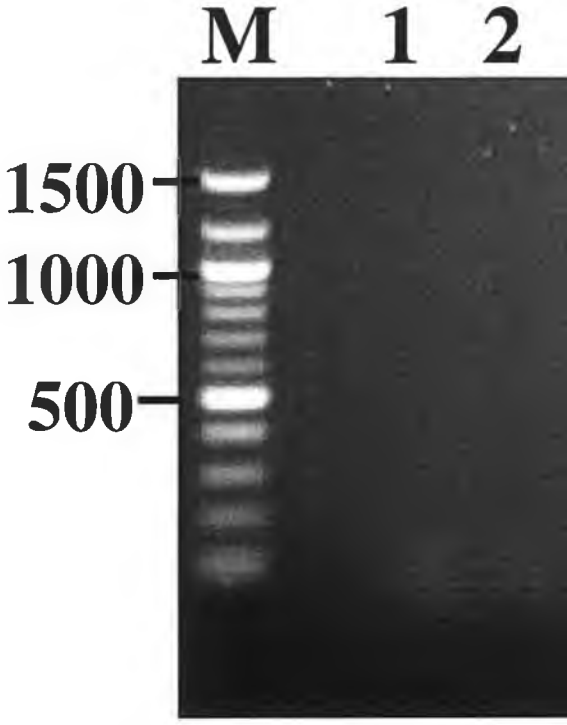


Fig.2



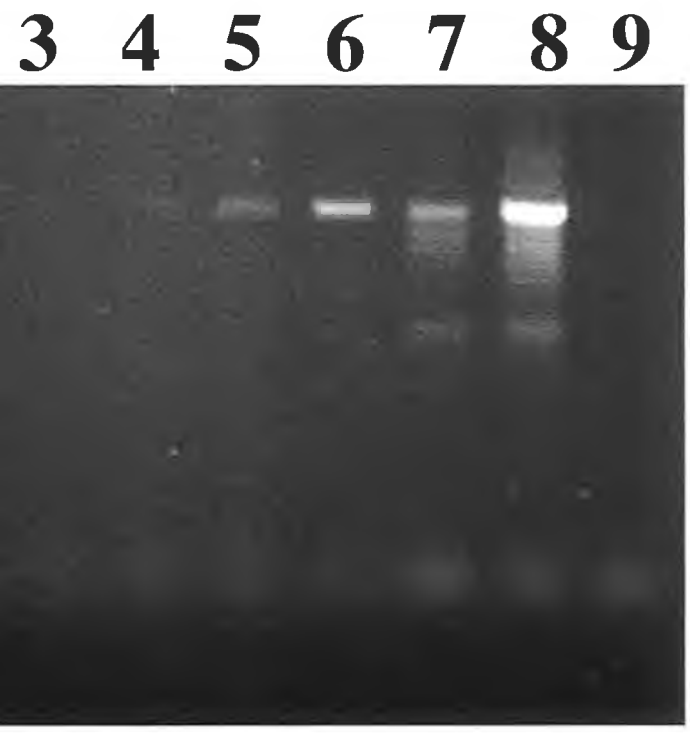

3 4 5 6 7 8 9

Fig. 3

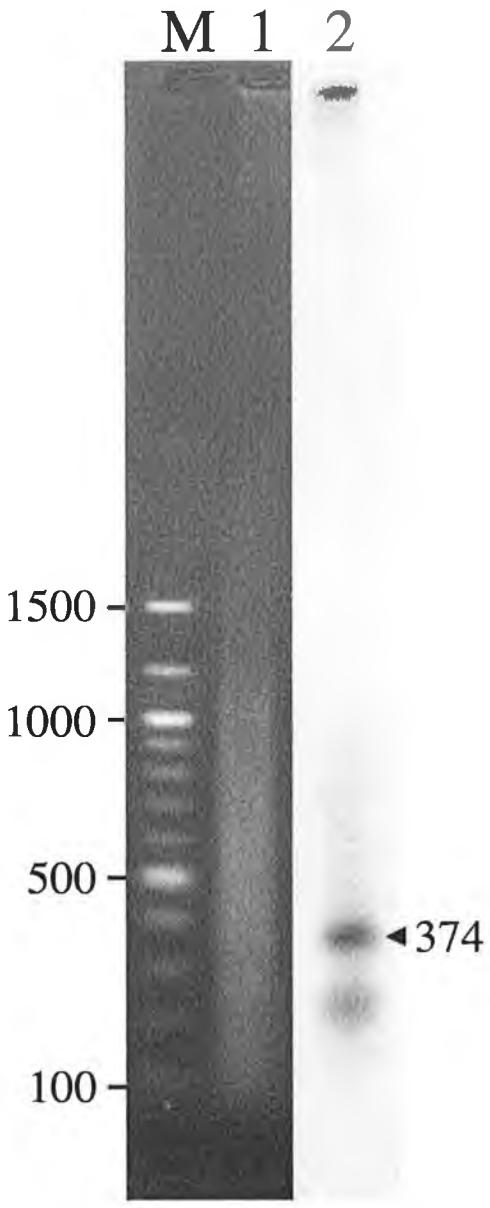


Fig. 4

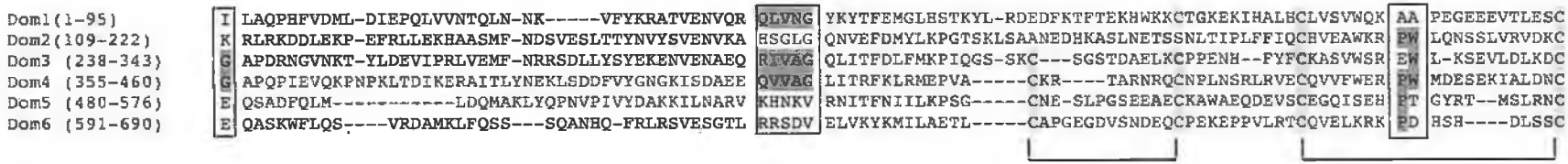

Dom1(1-95) I LAQPHFVDML-DIEPQLVVNTQLN-NK-----VFYKRATVENVQR QLVNG YKYTFEMGLHSTKYL-RDEDFKTFTEKHWKKCTGKEKIHALHCLVSVWQK AA PEGEEEVTLESC
Dom2(109-222) K RLRKDDLEKP-EFRLLEKHAASMF-NDSVESLTTYNVYSVENVKA HSGLG QNVEFDMYLKPGTSKLSAANEDHKASLNETSSNLTIPLFFIQCHVEAWKR PW LQNSSLVRVDKC
Dom3 (238-343) G APDRNGVNKT-YLDEVIPRLVEMF-NRRSDLLYSYEKENVENAEQ RLVAG QLITFDLFMKPIQGS-SKC---SGSTDAELKCPPENH--FYFCKASVWSR EW L-KSEVLDLKDC
Dom4 (355-460) G APQPIEVQKPNPKLTDIKERAITLYNEKLSDDFVYGNGKISDAEE QVVAG LITRFKLRMEPVA-----CKR----TARNRQCNPLNSRLRVECQVVFWER PW MDESEKIALDNC
Dom5 (480-576) E QSADFQLM----------LDQMAKLYQPNVPIVYDAKKILNARV KHNKV RNITFNIILKPSG-----CNE-SLPGSEEAECKAWAEQDEVSCEGQISEH PT GYRT--MSLRNC
Dom6 (591-690) E QASKWFLQS----VRDAMKLFQSS---SQANHQ-FRLRSVESGTL RRSDV ELVKYKMILAETL-----CAPGEGDVSNDEQCPEKEPPVLRTCQVELKRK PD HSH----DLSSC

Fig.5

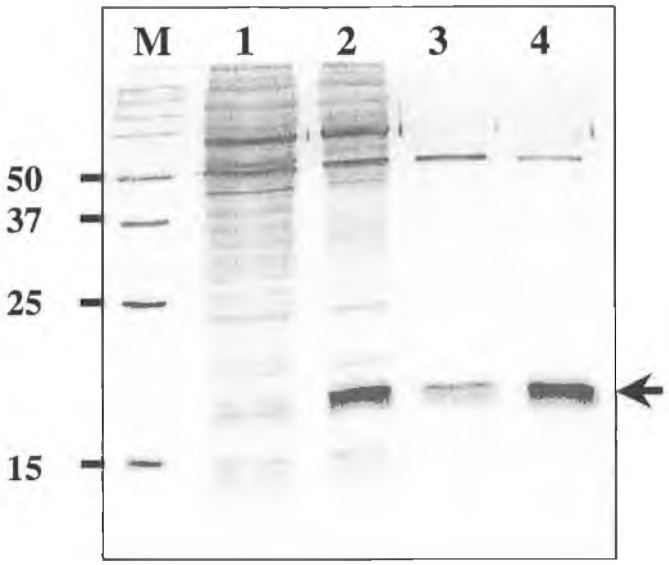


Fig.6A

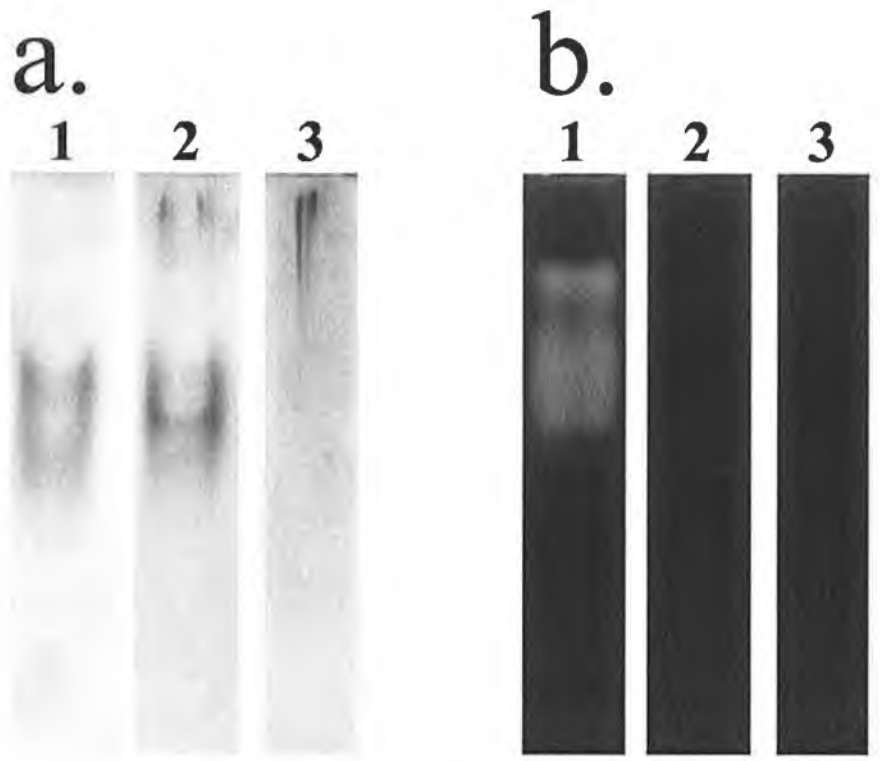


Fig. 6B

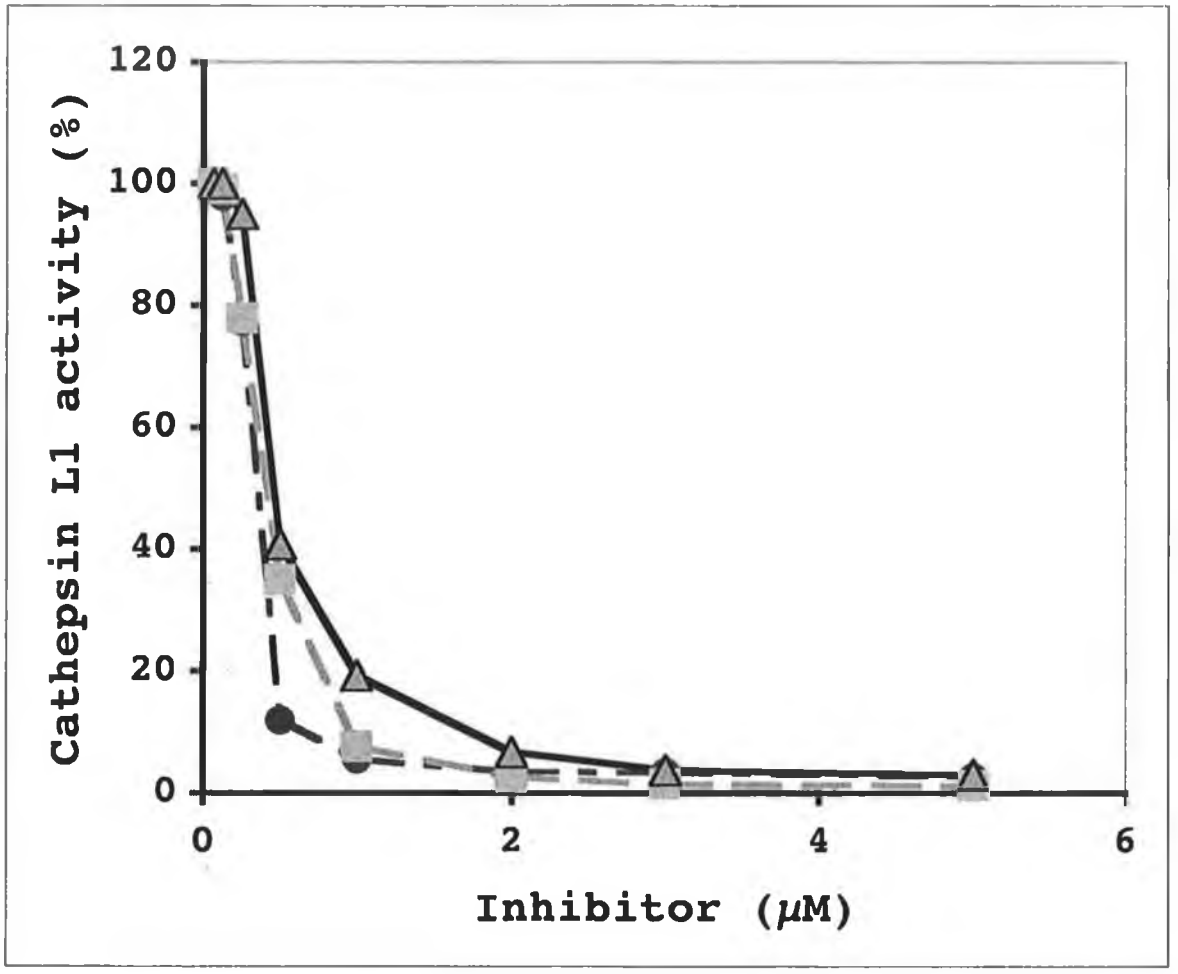
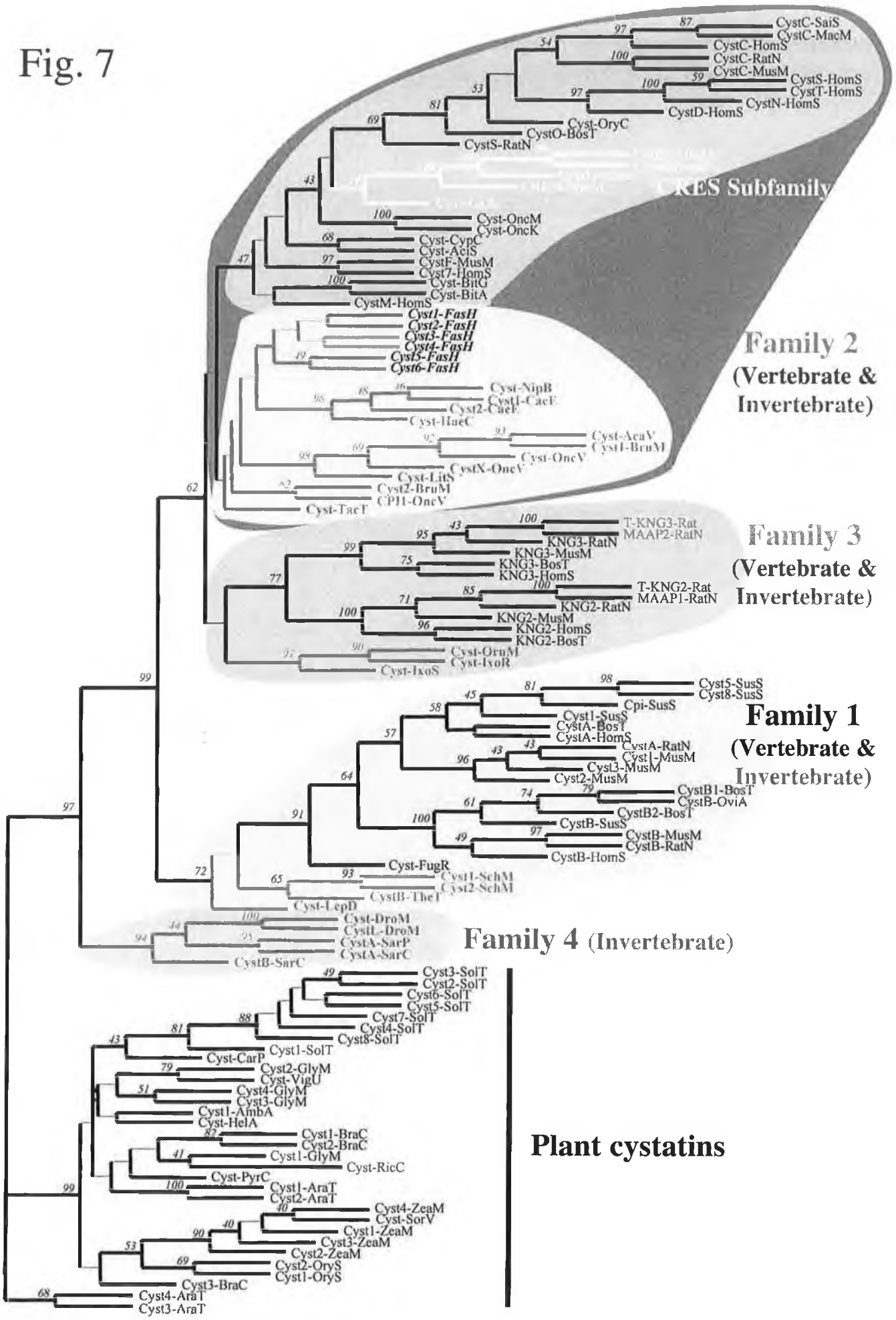


Fig. 7


CystC-SaiS
CystC-MacM
CystC-HomS
CystC-RatN
CystC-MusM
CystS-HomS
CystT-HomS
CystN-HomS
CystD-HomS
Cyst-OryC
CystO-BosT
CystS-RatN
CRES Subfamily
Cyst-OncM
Cyst-OncK
Cyst-CypC
Cyst-AciS
CystF-MusM
Cyst7-HomS
Cyst-BitG
Cyst-BitA
CystM-HomS
Cyst1-FasH
Cyst2-FasH
Cyst3-FasH
Cyst4-FasH
Cyst5-FasH
Cyst6-FasH
Family 2
(Vertebrate &
Invertebrate)
Cyst-NipB
Cyst1-CacF
Cyst2-CacF
Cyst-HaeC
Cyst-AcaV
Cyst1-BruM
Cyst-OncV
CystX-OncV
Cyst-LitS
Cyst2-BruM
CPH-OncV
Cyst-TacT
T-KNG3-Rat
MAAP2-RatN
KNG3-RatN
KNG3-MusM
KNG3-BosT
KNG3-HomS
Family 3
(Vertebrate &
Invertebrate)
T-KNG2-Rat
MAAP1-RatN
KNG2-RatN
KNG2-MusM
KNG2-HomS
KNG2-BosT
Cyst-OrnM
Cyst-IxoR
Cyst-IxoS
Cyst5-SusS
Cyst8-SusS
Cpi-SusS
Cyst1-SusS
CystA-BosT
CystA-HomS
CystA-RatN
Cyst1-MusM
Cyst3-MusM
Cyst2-MusM
CystB1-BosT
CystB-OviA
CystB2-BosT
CystB-SusS
CystB-MusM
CystB-RatN
CystB-HomS
Family 1
(Vertebrate &
Invertebrate)
Cyst-FugR
Cyst1-SchM
Cyst2-SchM
CystB-TheT
Cyst-LepD
Cyst-DroM
CystL-DroM
CystA-SarP
CystA-SarC
CystB-SarC
Family 4 (Invertebrate)
Cyst3-SolT
Cyst2-SolT
Cyst6-SolT
Cyst5-SolT
Cyst7-SolT
Cyst4-SolT
Cyst8-SolT
Cyst1-SolT
Cyst-CarP
Cyst2-GlyM
Cyst-VigU
Cyst4-GlyM
Cyst3-GlyM
Cyst1-AmbA
Cyst-HelA
Cyst1-BraC
Cyst2-BraC
Cyst1-GlyM
Cyst-RicC
Cyst-PyrC
Cyst1-AraT
Cyst2-AraT
Cyst4-ZeaM
Cyst-SorV
Cyst1-ZeaM
Cyst3-ZeaM
Cyst2-ZeaM
Cyst2-OryS
Cyst1-OryS
Cyst3-BraC
Cyst4-AraT
Cyst3-AraT
Plant cystatins