Studies on cathepsin B of *Eimeria tenella* and pyroglutamyl peptidase of *Leishmania major*

Marie Schaeffer

Division of Infection and Immunity

Institute of Biomedical and Life Sciences

This thesis is submitted for the degree of Doctor of Philosophy

Faculty of Biomedical and Life Sciences

University of Glasgow

March 2005

Abstract

Proteolytic enzymes play important roles in parasitic protozoa, including the organisms responsible for malaria, leishmaniasis and trypanosomiasis. Cysteine peptidases have attracted particular attention over recent years as some have been reported to play key roles in host-parasite interactions and virulence, and may constitute promising targets for new selective inhibitors and anti-parasite agents.

The obligate intracellular protozoan parasite *Eimeria* (mainly *E. tenella*) belonging to the subclass of Coccidia (with other parasites such as *Toxoplasma* and *Cryptosporidium*), is responsible for coccidiosis, which inflicts a serious economical burden to the poultry industry. Invasion of host cells by *Eimeria* involves the regulated release from specialised secretory organelles forming the apical complex, whose proteins' maturation, trafficking and secretion rely on proteolytic processing. A sequence encoding a cathepsin B-like cysteine peptidase of the clan CA, family C1 was identified in the genome database of *E. tenella*. The sequence corresponded to a single copy gene, and did not carry any introns.

The *E. tenella* enzyme, sharing 42% identity with the *Toxoplasma gondii* toxopain-1, which is reportedly involved in the invasion of host cells, was expressed as a soluble inactive zymogen in *E. coli*. Recombinant cathepsin B of *E. tenella* was functionally expressed as a mature enzyme in the *Pichia pastoris* inducible system as a glycosylated protein, but the glycosylations did not apparently affect the enzyme's activity. The biochemical characteristics of the recombinant enzyme were consistent with what has been reported in the literature for cathepsin Bs. The cathepsin B of *E. tenella* was detected in oocyst and sporozoite extracts, and, more specifically, in microneme preparations. In the sexual stages, the enzyme localised to mature macrogametes in discrete granules, the size and location of which suggest that they are wall-forming bodies, organelles involved in the oocyst wall formation in *Eimeria*. These data suggest that the cathepsin B may play roles in both cell invasion by sporozoites and the formation of the protective wall of the oocyst.

The genus *Leishmania* belongs to the family *Trypanosomatidae*, which includes the *Trypanosoma* species, and is the cause of leishmaniasis, which affects 2 million people and causes about 60000 deaths a year. Many peptidases encoded by the *Leishmania* genome are cysteine peptidases, some of which have been identified as important virulence factors, and therefore potential drug targets. Pyroglutamyl peptidases I (PPI) are cysteine peptidases of the clan CF, family C15, which hydrolyse N-terminal L-pyroglutamate (pGlu) residues. The pGlu modification is a post-transcriptional modification catalysed by the glutaminyl cyclase in human that confers relative aminopeptidase resistance and, in some cases, is essential to the modified peptides' activity. PPIs have been identified in a variety of organisms but no definitive biological function has been attributed to them.

A single copy gene encoding a PPI was identified in the genome database of *L. major*. Active recombinant enzyme was successfully produced in *E. coli*, and its biochemical properties coincided with those of mammalian PPIs. The active site catalytic triad E101, C210, and H234 (*L. major* PPI numbering) was confirmed by mutagenesis. The PPI activity was detected in *L. major* promastigotes, and the enzyme localised to the parasite cytoplasm. *PPI* knock-out mutants were generated, and knock-out of the PPI activity did not seem to induce a phenotype in *L. major*. The parasites retained the properties *in vivo* and *in vitro* of *L. major* wild type cells. The over-expression of the active PPI in *L. major* promastigotes seemed to impair metacyclogenesis and *in vivo* infectivity, while the over-expression of the C210A mutant did not have any detrimental effect. Susceptibility to a natural pGlu modified anti-microbial peptide, gomesin, was tested on the different cell lines, which were all equally susceptible. The data suggest that pGlu-modified peptides in *L. major* might be required for metacyclogenesis and modulation of the immune response in the host.

Acknowledgements

First, I would like to thank my supervisors Prof Graham Coombs and Prof Jeremy Mottram, for giving me the chance to do a PhD course here, and for their support and advice throughout these 3 years. Thanks also to my assessor, Dr Brian Shiels, for advice and guidance.

Then, I would like to thank Intervet, Schwabenhein, Germany, for funding of the *Eimeria* project, and providing with the *Eimeria* oocysts.

I would also like to thank Dr Fiona Tomley and all her group at the Institute for Animal Health (IAH) in Compton, for hosting me for 2 weeks, and for useful discussions and advice. Specially, thanks to Dr Liz Bromley for providing the microneme and rhoptry samples, I know they are precious!

Thanks to Prof David Ferguson at the University of Oxford for the immunolocalisation work on *Eimeria*; and Dr Antonio de Miranda (Sao Paolo, Brazil) for providing with the gomesin and gomesin analogue peptides. Thanks also to everybody at Biological Services (Glasgow), especially Maurice Dixon, for the *in vivo* experiments in mice.

I would also like to thank everybody in the North lab, for making it a very nice atmosphere to work in. It has been most enjoyable to work with you all. Special thanks to Dr Helen Denton, for the gel filtration work and for priceless advice on enzyme assays; to Dr Gareth Westrop and Dr Roderick Williams, for all their help and advice; to Dorothy Armstrong, for being so helpful, to Susan Baillie, for extraction of the parasites from mice lesions, and to Alan Scott, for running some of the BioCad purifications. Thanks to Dr Sébastien Besteiro, and Audrey Ambit at the Anderson College, for providing me with protocols, reagents, advice and fun times!

Muchas gracias a Marcelo, para compartir todos las buenas y malas épocas, para estar aquí para mí, siempre de apoyo, para ayudarme a ser mas centrádo en mí trabájo, y a ápagar la máquina también! Vos ha sído una gran fuente de la motivatión.

Je voudrais enfin remercier ma sœur et mes parents, pour avoir toujours soutenu mes choix sans toujours les comprendre. Merci de m'avoir toujours fait sentir que vous étiez fiers de moi.

Table of contents

Chapter 1: General introduction

Chapter 2: Material and Methods

Chapter 3: Expression and characterisation of the recombinant cathepsin B of *E. tenella*

Chapter 4: Expression and characterisation of recombinant pyroglutamyl peptidase I of *L. major*

Chapter 5: Functional study of the pyroglutamyl peptidase I of *L. major*

List of figures

Chapter 1. General introduction

Chapter 3: Expression and characterisation of the recombinant cathepsin B of *E. tenella*

Chapter 4: Expression and characterisation of recombinant pyroglutamyl peptidase I of *L. major*

Chapter 5: Functional study of the pyroglutamyl peptidase I of *L. major*

List of tables

Chapter 1: General introduction

Chapter 2: Material and Methods

Chapter 3: Expression and characterisation of the recombinant cathepsin B of *E. tenella*

Chapter 4: Expression and characterisation of recombinant pyroglutamyl peptidase I of *L. major*

Chapter 5: Functional study of the pyroglutamyl peptidase I of *L. major*

List of abbreviations

AMC	7-amino-4-methylcoumarin
β-NA	β-naphtylamide
bp	base pair
cDNA	complementary DNA
cm	centimeter
°C	degree Celsius
DEPC	diethylpyrocarbonate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E64	(2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl}carbamoyl)oxirane-2-carboxylic acid
EDTA	ethylenediamine tetraacetic acid
g	gram
gDNA	genomic DNA
h	hour
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HIFCS	heat inactivated foetal calf serum
HRP	horseradish peroxidase
Ig	immunoglobulin
IPTG	isopropylthio-β-D-galactoside
kb	kilobase
kDa	kilodalton
K_m	Michaelis constant
kV	kilovolt
l	litre
LB	Luria-Bertani
μF	microfarad
μg	microgram
μl	microlitre
μM	micromolar
M	molar
Mb	megabase
mg	milligram
min	minute
ml	millilitre
mM	millimolar
mm	millimetre
mRNA	messenger RNA
nm	nanometer
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Pfu	*Pyrococcus furiosis*
pNA	paranitro-anilide
Ω	ohm

OD	optical density
PMSF	phenylmethylsulfonyl fluoride
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
rpm	revolutions per minute
RT-PCR	reverse transcriptase PCR
s	second
SD	standard deviation
SDS	sodium dodecylsulfate
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SE	standard error
SL	splice leader
Taq	*Thermus aquaticus*
TCA	trichloro acetic acid
TCEP	Tris(2-carboxyethyl)phosphorine hydrochloride
TM	transmembrane
Tm	melting temperature
U	unit
UV	ultraviolet
V	volt
v	volume
w	weight
X-Gal	5-bromo-4-chloro-3-indol-β-D-galactopyranoside

À Emile Keller,
mon grand-père

Chapter 1
General introduction

Proteolytic enzymes seem to play important roles in parasitic protozoa, including the organisms responsible for malaria, leishmaniasis and trypanosomiasis. Cysteine peptidases have attracted particular attention over recent years and have been the subject of extensive studies because of their importance in parasite survival, interaction with the host cells and pathogenicity (McKerrow, 1993; Rosenthal, 1999; Klemba and Goldberg, 2002). Despite the progress in combating infectious diseases and the constant search for new drugs or treatments, some parasites still inflict a dramatic social and economical burden to tropical or sub-tropical regions of the world (human and animal infections), as well as to developed societies (in particular with the development of intensive husbandry of domestic animals enhancing transmission of parasites).

The protozoan parasite *Eimeria* (mainly *E. tenella*), belonging to the subclass of Coccidia (with other parasites such as *Toxoplasma* and *Cryptosporidium*), is responsible for coccidiosis, which is a serious problem in the poultry industry. Worldwide expenditures just for anti-coccidiodial drugs added to feed is estimated to be $250 to $300 millions annually and there is a rising problem of parasites becoming resistant to these coccidiostats (www.antecint.co.uk/main/oocide.htm). Also, the intensive use of drugs in farm animals raises public health concerns, as about 4% of eggs and 10% of chicken liver tested in the UK in 2003 contained residues of coccidiostats (http://research.utu.fi/residues/). In December 1998, the European Union banned the use of four, widely used growth-promoting antibiotics used in aviary production, due to public concern that antibiotic residues in meat could lead to antibiotic resistance in humans. The ban may be broadened to more in-feed drugs within the next 5 years. Coccidiosis is not limited to the poultry industry and some *Eimeria* species can also affect cattle, sheeps and pigs, which constitutes a potential threat for other intensive livestock industries. There is, therefore, an increasing economic need to develop efficient new methods of prevention and treatment of coccidiosis. The interest in cysteine peptidases has been supported by the fact that some of them seem to be promising targets for new selective inhibitors and anti-parasite agents (Selzer *et al.*, 1997; Rosenthal, 1999).

The species of *Leishmania* parasites, belonging to the family for *Trypanosomatidae* (together with the species of *Trypanosoma*), are responsible for leishmaniasis diseases, varying in severity of symptoms depending on the species contracted. It affects a variety of mammalian hosts, and it is believed that 12 million people are affected in the world, and potentially 350 million people are at risk of being infected. In developed countries, leishmaniasis has been on the increase as an opportunistic disease in immuno-deficient patients, infected with the human immunodeficiency virus (HIV). Treatment of the disease mainly relies upon the use of pentavalent antimonials, drugs that have been used since the 1940's. Because of the length of time the drugs have been used and sometimes inappropriate dosage, parasites have been developing resistance to these drugs, making them less and less effective (Ouellette and Papadopoulou, 1993). In the state of Bihar in northwestern India, unresponsiveness to pentavalent antimonials has increased from 34 to 64% between 1994 and 1997 (Thakur *et al.*, 1998). Though other means of treatment exist, most of them are very expensive and have severe side effects. Despite constant efforts to develop new drugs, very few are getting to an advanced stage in chemotherapy trials, and complexity of the immune response generated by *Leishmania* infections has meant that no vaccine has been developed so far (Croft and Coombs, 2003). Available chemotherapy being inadequate and expensive, there is an urgent need to develop new and better drugs. Interest has been brought to cysteine peptidases of *Leishmania* as potential drug targets, as some of them have been shown to be important virulence factors (Mottram *et al.*, 2004).

1.1. Peptidases: generalities

Peptidases (also termed proteases or proteinases) catalyse the cleavage of peptide bonds in proteins or oligomeric peptides, either within the polypeptide chain (endopeptidase activity) or from amino or carboxyl ends (exopeptidase activity). Proteolytic enzymes are numerous and diverse; they represent about 2% of all gene products (Rawlings and Barrett, 1999) and range from monomers of 10 kDa to multimeric complexes of several hundred kDa. The location of the scissile bond in the peptidic chain and the preferred amino acid sequence around it differ between individual peptidases, but this could not be used to establish a comprehensive classification system (Rawlings and Barrett, 1999). In 1960, Hartley suggested dividing peptidases into groups depending on the catalytic mechanism used (McDonald, 1995). Five main catalytic types could be recognised based on the active site residue that provides the nucleophilic attack during the catalytic reaction. They are now known as serine, aspartic, metallo-, cysteine, and threonine peptidases. Serine, threonine and cysteine peptidases effect catalysis after formation of a tetrahedral transition intermediate, whereas aspartate peptidases and metallopeptidases require the intervention of a molecule of water for cleavage. In the early 1990s, with the increase of data on primary sequences of proteins and three-dimensional structures of peptidases, Rawlings and Barrett began to bring together sets of peptidases that were similar both in molecular structures and in evolutionary origin to create the MEROPS database (Rawlings *et al.*, 2002; www.merops.co.uk). This emphasises the growing interest in peptidases and the wealth of resources now available.

Serine peptidases have a catalytic triad involving serine, aspartate and histidine (Rawlings and Barrett, 1994). They are divided into two groups in eukaryotes: regulatory peptidases with strict specificity and digestive enzymes with broader specificity, such as trypsin and chymotrypsin for example. In aspartic peptidases, 2 aspartates are involved in the catalytic process (Rawlings and Barrett, 1995b). Among eukaryotic aspartic peptidases, digestive enzymes and regulatory peptidases can be found as well as intracellular acid peptidases. Metallo-peptidases constitute a vast group of enzymes involving a metallic ion (zinc in most cases) in the catalytic process (Rawlings and Barrett, 1995a). Examples of metallo-peptidases are matrix degradation enzymes such as elastases, and regulatory peptidases. Some enzymes utilize a catalytic threonine for cleavage (Seemueller *et al.*, 1995) and other cryptic proteases may also exist (Sajid and McKerrow, 2002). Based on the fact that the majority of peptidases described so far in parasitic protozoa are of the cysteine class, the main focus of this review will be put on this type of proteolytic enzyme and they will be treated independently.

Although a single peptide bond is cleaved during one catalytic reaction, the flanking amino acids in the substrate protein play a role in determining peptidase specificity, as do the sequences flanking the catalytic amino acid residues on the peptidase. The convention used to identify substrate residues and binding pocket amino acids on the peptidases was elaborated by Schlechter and Berger in 1967. The substrate residues are designated as P, and the corresponding sites on the enzyme, as S. Prime residues are located on the C-terminal side of the cleavage site, whereas non-prime residues are on the N-terminal side. Thus, a five residue substrate site cleaved between the third and the fourth amino acid would be P3-P2-P1-P'1-P'2. Sequences that directly flank the active site residues are often very conserved and play a role in the enzymatic specificity and substrate binding, along with helping to determine the chemical environment. Specificity strictness varies from one enzyme another; usually several peptide bonds are hydrolysed by the same peptidase, but sometimes only one specific bond is cleaved, provided they fulfil certain sequence or conformational requirements. The substrate specificity of an enzyme often gives good indication on the enzyme's biological role and importance, although it can vary between *in vitro* and *in vivo* conditions.

1.1.1. Cysteine peptidases

1.1.1.1. Classification

The first cysteine peptidase was purified in 1879 from the papaya fruit *Carica papaya* and named papain. The first crystallographic structure of a cysteine peptidase was also obtained using papain (Drenth *et al.*, 1968). Various cysteine peptidases have been identified since from a wide variety of eukaryotic organisms and the major mammalian peptidase have been designated cathepsins B, H, K, L, and S. All of the mammalian lysosomal cysteine peptidases are known as cathepsins, the major ones being cathepsins B and L, but not all cathepsins are cysteine peptidases. For instance, cathepsins D and E are aspartic peptidases, and cathepsins A and G are serine peptidases (Barrett and McDonald, 1985). Cysteine peptidases are grouped into clans (Barrett, 1994), further divided into families depending on sequence similarities and substrate specificities (Fig. 1.1). The main focus of my studies will be on clan CA, family C1 cysteine peptidases.

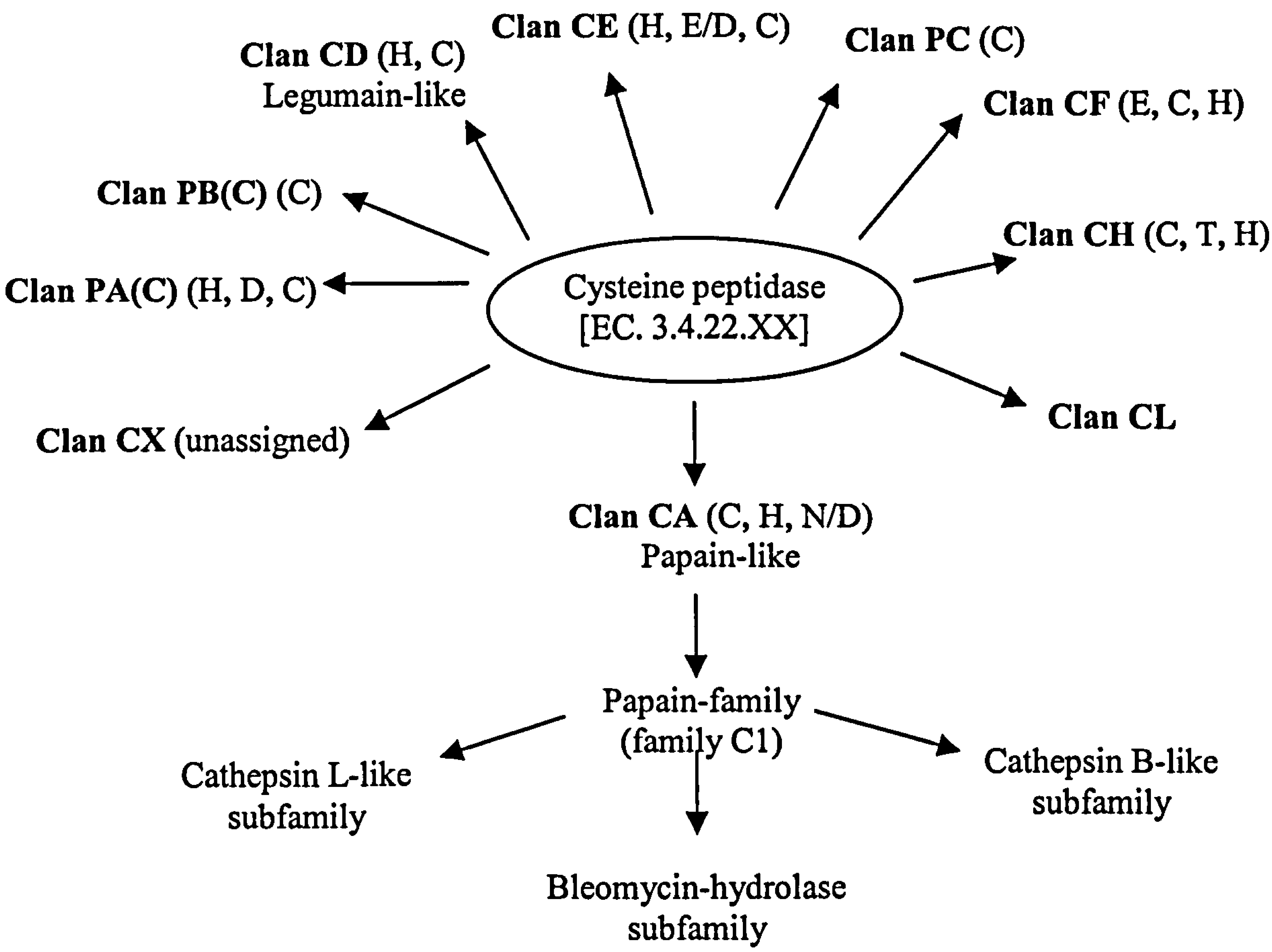


FIG. 1.1. **Cysteine peptidase superfamily.** Classification in subfamilies was based on amino acid sequence around the active site (see Sajid and McKerrow, 2002). Clans are indicated in bold and the amino acids (in parentheses) show the order of the catalytic residues in the linear polypeptide sequence (Mottram *et al.*, 2003).

Although cysteine peptidase's main role is in protein degradation, the multiplicity and apparent redundancy of the clan CA enzymes makes the discovery of their individual functions rather difficult. This is less the case for clan CD enzymes, whose specificity of functions seem much greater (Mottram *et al.*, 2003). Nevertheless, the recent progresses in genetic knockouts have enabled a better insight into the function of cysteine peptidases and proved that they are less redundant than was thought. Cathepsin K in mammals, for example, has been demonstrated to be essential for bone remodeling (Chapman *et al.*, 1997) and linked to the genetic disorder pycnodysostosis (Gelb *et al.*, 1996). Some other cysteine peptidases are implicated in tumour invasion (Duffy, 1992) and arthritis (Trabandt *et al.*, 1990). Therefore, they have attracted increasing interest in the recent years as targets for human therapy, in particular in terms of research for new anti-cancer drugs.

1.1.1.2. Catalytic mechanism (Clan CA, family C1)

Cysteine peptidases have also been referred as thiol or sulfhydryl peptidases. The catalytic mechanism of cysteine peptidases is close to those of serine peptidases with a nucleophile property conferred by the thiol group. This nucleophilic behaviour is further enhanced by the presence of an active site histidine, which acts as a proton acceptor. The '-SH' group of the cysteine side chain forms a thiolate-imidazolium $-S^{-}\ {}^{+}HN{=}$ group with the imidazole group of the histidine (Polgar, 2004). The delocalised electron cloud over the thiolate-imidazolium group constitutes a charge relay diad, which allows an enzymatic activity within a pH range from around 4.0 (pK_a of cysteine) to 8.5 (pK_a of histidine), and is often stabilised by a highly conserved asparagine and the chemical environment of the catalytic reaction (pH, ionic force). After formation of the transient catalytic thiol-ester tetrahedral intermediate between the enzyme and the substrate, the peptidase returns to an active enzyme state. As many as one million peptide bonds per second can be degraded with such a mechanism.

1.1.1.3. Structure and specificity

The majority of family C1 cysteine peptidases are synthesised as immature precursors or zymogens, with a pro-domain and, sometimes, a carboxy-terminal extension. These are cleaved in an endogenous or exogenous manner to lead to the mature catalytic enzyme. The activation is a regulated process involving the proteolytic cleavage of the N-terminal pro-peptide in a pre-lysosomal compartment and release of this pro-region, which acts as self-inhibitor. The pro-regions are tightly bound to the catalytic cleft and block access to the active site (Cygler *et al.*, 1996) to maintain the peptidase in an inactive form during trafficking to the lysosome. The pro-region can also contain folding information for the enzyme (Pandey *et al.*, 2004). Final activation can by facilitated by decreasing the pH (Turk *et al.*, 2000) and this is favoured in the acidic pH environment of the lysosome. Most of the mature forms are monomeric (with the exception of the tetrameric cathepsin C) with a molecular mass ranging from 20 to 30 kDa (without the pro-region of approximately 60 to 100 amino acids). Most of the enzymes are endopeptidases, although cathepsins B and H also exhibit exopeptidase activity (Barrett and Rawlings, 2001). Although most cysteine peptidases are located in lysosomes or lysosome-like organelles, some can also be located on the cell surface or secreted (Robertson *et al.*, 1996). Trafficking of the mammalian peptidases requires glycosylation and mannose-6-phosphate residue recognition by the mannose-6-phosphate receptor (Kornfeld and Mellman, 1989). An alternative mechanism involves transmembrane receptors (McIntyre and Erickson, 1993).

Some detailed analysis of the substrate binding sites have been performed, in particular on papain (Turk *et al.*, 1998), and shown that some residues have a role of specificity determinants. Preferences for certain residues at particular positions and recognition of specific motifs can vary from one peptidase to the other, although similarities are found among the same subfamily. Natural or synthesised low-molecular-weight inhibitors are often useful for determining binding pockets structures of cysteine peptidases and elucidating the relative importance of different binding determinants by co-crystallisation with the enzyme (Otto and Schirmeister, 1997). This avoids, in particular, the oxidation of the catalytic cysteine to sulfinic or sulfonic acid. Among the natural small inhibitors of papain-like cysteine peptidases, cystatins, stefins, kininogens and serpin can be cited. They have a low selectivity and are involved in humans in the control of cysteine peptidase activity and in the protection against self peptidases inappropriately escaping from lysosomes, as well as for exogenous organisms. A generic synthetic cysteine inhibitor widely used is E64 (L-*trans*-epoxysuccinyl-leucyl-amido (4-guanidino) butane) and inhibits, in particular, clan CA, family C1 peptidases.

On this basis, specificity determinants for most clans of peptidases could be determined. The legumain-like family of cysteine peptidases (clan CD), for instance, possess a very strict specificity at the S1 site (Barrett and Rawlings, 2001; Dickinson, 2002; Mottram *et al.*, 2003), are insensitive to E64 inhibition and present specificities at the P2 and P3 positions (Mathieu *et al.*, 2002). Papain-like peptidases are sensitive to E64 and possess specificity at the S2 position. Cathepsin B activity can be discriminated over cathepsin L activity using substrate peptides with, respectively, glutamate or alanine (Alves *et al.*, 2001). The different amino acid composition of the active site, thus the different chemical environment, gives a different S2 specificity to cathepsins B and L. Structural differences could also be elucidated for these two subclasses of peptidases. For instance, cathepsin L possesses a conserved motif in the pro-region named ERFNIN (sequence Glu-X-Arg-X-(Ile/Val)-Phe-X-Asn-X-Ile-X-Asn, X being any amino acid), which cathepsin B does not have (Karrer *et al.*, 1993). Furthermore, cathepsin B is characterised by the presence of an insertion loop of 20 amino acids, named the occluding loop because of its positioning in the active site, which seems to favour dipeptidyl carboxypeptidase activity (exopeptidase activity) (Naegler *et al.*, 1997).

1.1.2. Cysteine peptidases in parasitic protozoa

1.1.2.1. Generalities

A large number of the peptidases described to date for parasitic protozoa are cysteine peptidases of clan CA, family C1. Frequently they are abundant and stage-regulated, and most are closely related to mammalian lysosomal cathepsin L. Many of the cathepsin L-like cysteine peptidases could be detected in lysosome-like organelles by immunoelectron microscopy. For example, *Leishmania mexicana* cathepsin L-like enzymes were localised in large organelles known as megasomes (large lysosomes) in the amastigote form of the parasite (Pupkis *et al.*, 1986). In some cases they occur on the surface or are secreted. Trafficking of cysteine peptidases in some parasitic protozoa differs from that of mammals and some parasites, including the trypanosomatids, lack the mannose-6-phosphate pathway (Huete-Perez *et al.*, 1999) and may use a way involving a peptide loop in the pro-domain. Furthermore, to achieve proper targeting to an intracellular compartment or secretion, the N-terminal extremity of the enzyme is essential (Brooks *et al.*, 2000), and a hydrophobic signal peptide of about 15 to 22 amino acids upstream of the pro-region can be found. In some cases, a C-terminal extension may also be present, like for example for the CPB of *L. mexicana* (Mottram *et al.*, 1997), but this sequence is not required for correct trafficking or activation of the enzyme (Mottram *et al.*, 1997).

The major cysteine peptidase of *Trypanosoma cruzi*, cruzipain, could be detected on the surface of epimastigotes in addition to its main lysosomal location (Parussini *et al.*, 1998). Protozoan parasites such as *Entamoeba histolytica* are known to secrete cysteine peptidases. The parasites occur on mucosal surfaces in their hosts and this secretion of cysteine peptidase may confer to them an advantage for tissue invasion (Que *et al.*, 2003; Pertuz Belloso *et al.*, 2004). Due to these extra-lysosomal locations, some parasite cysteine peptidases are adapted to a wide range of pH.

Some structural differences can be found between mammalian and parasite peptidases. The ERFNIN amino acid motif in the pro-region of cathepsin-L is highly conserved in mammals, which is not the case in parasitic protozoa (Sajid and McKerrow, 2002). Furthermore, in some exceptional cases, cathepsin B-like enzymes lack the occluding loop or it is modified (Ward *et al.*, 1997).

Phylogenetic analysis showed that cysteine peptidases from diverse parasitic organisms are not closely related (Hughes, 1994), with the exception of the cysteine peptidases from *Leishmania* and *Trypanosoma.* Interestingly, the diversification of cysteine peptidases of *Haemonchus contortus* coincides with the emergence of mammals in evolution (Hughes, 1994). This might suggest that parasite cysteine peptidase diversification occurred as an adaptive response to mammalian hosts. Furthermore, isoenzymes encoded by multiple copies of a gene, for example the *CPB* genes coding for isoenzymes of cathepsin L-like enzymes in *Leishmania mexicana,* may have distinct roles in the interaction between the parasite and its host and reflect a highly adaptive process (Mottram *et al.*, 1997).

1.1.2.2. Identified protozoan cysteine peptidases and their roles

Members of all major peptidase classes (mainly cysteine peptidases but also several serine peptidases and fewer metallo-peptidases and aspartic peptidases (McKerrow *et al.*, 1993)) have been reported and characterised in parasitic protozoa, and extensive studies have been performed on cysteine peptidases. Detailed biochemical data are still restricted to a few parasites and a few enzymes, and our understanding of the physiological roles of most of the enzymes studied is still limited. However, there has been in recent years an increasing amount of data suggesting that cysteine peptidases play various indispensable roles in the biology of parasites. Beside their catabolic and protein processing functions, they have been reported to play key roles in host-parasite interactions and virulence in various parasitic organisms. Only a few examples will be mentioned here, describing some of the cysteine peptidases best characterised so far.

1.1.2.2.1. Cysteine peptidases of Apicomplexa

1.1.2.2.1.1. *Plasmodium*

Species of the genus *Plasmodium* can parasitise a wide range of hosts. *P. falciparum* is the agent responsible for the most virulent form of human malaria. Recent estimates suggest that the number of infected humans exceeds 500,000,000, and that 1-2 million persons die each year. The life cycle is split between a vertebrate host and an insect vector, which is the Anopheline mosquito. In the human host the parasite is found primarily inside red blood cells. The parasite reproduces asexually until the red blood cells breaks open and a large number of merozoites is released, which will infect more red blood cells. The characteristic "chill and fever" (paroxysm) associated with malaria occurs periodically. The different malarias produce fevers of different frequency, depending on how long it takes to complete shizogony in erythrocytes (48 h for *P. falciparum*). Severe and complicated malaria is usually caused by delay in successfully treating an uncomplicated infection with *P. falciparum.*

Three *P. falciparum* cathepsin-L like peptidases have been identified (falcipains-1, -2 and -3) and the main cysteine peptidase activity in the food vacuoles appears to be falcipain-2 (Singh and Rosenthal, 2001). The food vacuole also contains the aspartyl peptidases known as plasmepsins I and II (Coombs *et al.*, 2001), which have been more widely characterised than cysteine peptidases. Nevertheless, there is clear evidence that cysteine peptidases are involved in rupturing of red blood cells after schizogony and are therefore involved in parasite release (Sijwali and Rosenthal, 2004). Furthermore, all three peptidases are expressed by trophozoites and hydrolyse haemoglobin, suggesting roles in this process (Rosenthal *et al.*, 2002; Sijwali and Rosenthal, 2004). Falcipain-2 and falcipain-3 are food vacuole hemoglobinases, and falcipain-2, the principal hemoglobinase of *P. falciparum*, also hydrolyses ankyrin, suggesting additional activity against erythrocyte cytoskeletal targets (Hanspal *et al.*, 2002). The definite function of falcipain-1 is still uncertain, as various roles have been attributed to the enzyme. Disruption of the gene encoding falcipain-1 showed the enzyme was not essential in the erythrocyte stage of the parasite or for erythrocyte invasion (Sijwali *et al.*, 2004), though apparently specific inhibition of falcipain-1 had previously been shown to inhibit erythrocytes invasion (Greenbaum *et al.*, 2002). Another study suggested falcipain-1 may have an important role during parasite development in the insect vector (Eksi *et al.*, 2004). There is an apparent synergy between plasmepsin and falcipain in the degradation of haemoglobin (Gluzman *et al.*, 1994; Sijwali and Rosenthal, 2004) and these enzymes may be considered as potential joint targets for anti-malarial drugs. Multiple orthologs of the falcipains exist and have been identified in other plasmodial species and the predicted mature forms of the peptidases of the different species are highly conserved (Rosenthal, 1996). For example, in *P. vivax*, vivapain-2 and -3 resemble falcipain-2 and -3 and are potentially appropriate therapeutic targets too (Na *et al.*, 2004).

1.1.2.2.1.2. *Toxoplasma*

Toxoplasma gondii is responsible for toxoplasmosis and belongs to the subclass Coccidia along with other parasites including *Cryptosporidium* and *Eimeria*. *Toxoplasma* infection is common, virtually all warm-blooded animals, including humans can become infected and it is often transmitted to humans via domestic and feral cats. Infection is usually asymptomatic, but acute infection can cause lymphadenopathy and in case of immunodepression (for example, AIDS patients) leads to severe complications including encephalitis, pneumonitis and retinitis. Infection with *T. gondii* of pregnant women not-previously exposed to the parasite and therefore not immune to it may increase the probability of having a child born prematurely or even stillborn, especially if the infection occurs during the first three months of pregnancy.

In *T. gondii,* a cysteine peptidase gene has been identified and encodes an enzyme closely related to cathepsin B, named Toxopain-1 (Que *et al.*, 2002). Toxopain-1 is encoded by a single copy gene (*TGCP1*) and was reported to localise to rhoptries, secretory organelles required for apicomplexan parasite invasion. Evidence has been provided that it may be involved in cell invasion and rhoptry protein processing (Que *et al.*, 2002), as well as being critical for cell invasion *in vivo* (Que *et al.*, 2004). Upon release onto the host cell surface, a number of microneme proteins have been shown to be cleaved, suggesting a role for peptidases at that point (Kim, 2004), though most of the process may involve subtilisin-like serine peptidases.

1.1.2.2.1.3. *Cryptosporidium*

Cryptosporidium is an intracellular parasite that infects the small intestine of an unusually wide range of mammals, including humans. The two species responsible for human cryptosporidiosis are *C. parvum* and *C. hominis,* the latter being restricted to human, and the former being able to infect various mammalian hosts (Xu *et al.*, 2004). Cryptosporidiosis is a self-limiting diarrhoeal illness in immunocompetent humans, lasting 3 to 20 days. In immunodeficient patients (for example, infected with HIV) the symptoms can be a lot more severe, sometimes leading to death. Some species of *Cryptosporidium* have a major economic significance, as cattle can be infected. The parasites are transmitted mainly via contaminated water or food. *Cryptosporidium* infects the mucosal epithelium of the digestive and respiratory tracts. Infection is initiated when sporozoites excyst from an ingested sporulated oocyst in the small intestine of the host. Sporozoites invade host's intestinal cell lines, activate caspases and induce apoptosis (Ojcius *et al.*, 1999).

Cysteine peptidase activity has been detected in partially excysted oocysts but its role still has to be evaluated (Forney *et al.*, 1996). Nevertheless, there is an indication that cysteine peptidase might be involved in mucus penetration by the parasites and in sporozoite and merozoite invasion of host cells (Brown S. M. A., PhD thesis).

1.1.2.2.1.4. *Eimeria*

The diseases caused, transmission, pathogenicity, as well as the characteristic biochemical and molecular features of the parasite, will be discussed in 1.2.

Cysteine peptidase activity has been detected in *Eimeria* and using specific inhibitors it has been suggested that cysteine peptidases might be involved in invasion by sporozoites and merozoites of the host cells (Adams & Bushell, 1988; Coombs *et al.*, 1997), as well as in the penetration of mucus layers (Brown S. M. A., unpublished data). However very little is known about cysteine peptidases of coccidian parasites and Toxopain-1 of *Toxoplasma gondii* is the only one characterised so far (Que *et al.*, 2002).

Little is known about other eimerian peptidases and few have been reported so far. Study of *E. tenella* oocysts revealed that proteins in crude oocyst homogenates were degraded in the absence of peptidase inhibitors, and that the degradation could be prevented fully by using both typical serine peptidase and cysteine peptidase inhibitors. This is strongly suggestive of the presence of these two types of peptidases (Michalski *et al.*, 1994). Furthermore, *in vitro* incubation of sporozoites with serine peptidase and cysteine peptidase inhibitors significantly decreased host cell invasion, suggesting that these peptidases may be involved in invasion by *Eimeria* sporozoites (Adams & Bushell, 1988). Furthermore, a 46 kDa cysteine peptidase could be detected on the surface of the sporozoites of *E. tenella* (Brown S. M. A., unpublished data). Recent studies highlighted the expression of an aspartyl peptidase, named eimepsin, in the sporulated oocyst of *E. tenella* (Jean *et al.*, 2001). Eimepsin has been found in the refractile body organelles (possibly a storage location) of resting sporozoites (Laurent *et al.*, 1993) and within the apical tips of merozoites and invading sporozoites (Jean *et al.*, 2000). This gives support to the possibility that eimepsin has a function in host cell invasion, as the apical tip of apicomplexan parasites is known to be involved in both attachment to and penetration of host cells. But the specific function of eimepsin in the parasite is still unknown. Metallo-peptidase activity could also be detected in sporulated oocysts using specific inhibitors (Kaga *et al.*, 1998), but still very little is known about these enzymes.

1.1.2.2.2. Cysteine peptidases of trypanosomatids

1.1.2.2.2.1. *Leishmania*

Leishmania species are responsible for various diseases in humans. The diseases caused, transmission, pathogenicity, as well as the characteristic biochemical and molecular features of the parasite, will be discussed in 1.3.

A total of 65 putative cysteine peptidases, grouped into clans and 13 families, have been revealed in the *L. major* database (Mottram *et al.*, 2004). Most of them belong to the same clan as papain (clan CA). A review of cysteine peptidases in *Leishmania* is presented in Table 1.1. (from Mottram *et al.*, 2004).

Table 1.1. **Cysteine peptidases and cysteine peptidase inhibitors of *Leishmania major*.** Nomenclature for Clans and Families are described in the MEROPS database (http://www.merops.sanger.ac.uk), (from Mottram *et al.*, 2004).

Clan	Family	Number	Gene	Peptidase/inhibitor	Features	Reference
CA	C1	1	*CPA*	CPA	Cathepsin L-like, lysosomal	Mottram *et al.*, 1998
	C1	8	*CPB*	CPB	Cathepsin L-like, lysosomal	Mottram *et al.*, 1998
	C1	1	*CPC*	CPC	Cathepsin B-like, lysosomal	Mottram *et al.*, 1998
	C2	27	_	Contain calpain-like domain	Some calcium-dependant	
	C12	2	_	Ubiquitin C-terminal hydrolase	Ubiquitin pathway	
	C19	15	_	Ubiquitin hydrolase	Ubiquitin pathway	
	C48	1	_	SUMO-like	Ubiquitin pathway	
	C51	2	_	D-alanyl-glycyl endopeptidase-like	CHAP domain. Function unknown	
	C54	2	*ATG4.1* *ATG4.2*	ATG4	Autophagy	Williams, Mottram and Coombs, unpublished
	C65	1	_	Obutain	Ubiquitin pathway	
CD	C13	1	*GPI8*	GPI:protein transmamidase	GPI protein anchor biosynthesis	Hilley *et al.*, 2000; Mottram *et al.*, 2003
	C14	1	*MCA5*	Metacaspase	Caspase-like protein, function unknown	Arnoult *et al.*, 2002
	C50	1	_	Separase	Thought to control sister chromatid separation during mitosis	Arnoult *et al.*, 2002
CF	C15	1	*PGP*	Pyroglutamyl-peptidase I	Hydrolyses N-terminal glutamyl residues of proteins	Schaeffer, Mottram and Coombs, unpublished
PC(C)	C56	1	*PFPI*	PfPI	Similar to intracellular protease of *Pyrococcus*, function unknown	Eschenlauer, Coombs and Mottram, unpublished
I-	I42	1	*ICP*	Inhibitor of cysteine peptidases	Inhibits Clan CA, family C1 cysteine peptidases	Sanderson *et al.*, 2003

Different peptidase activities have been detected in *Leishmania* and the proteins have mainly been located in the amastigote form of the parasites in the megasomes (Pupkis *et al.*, 1986), which are large lysosomes. In the promastigote form of the parasite, major activities are thought to be located in the multi-vesicular tubule-lysosome (Mullin *et al.*, 2001).

The family C1 of clan CA comprises 3 identified and characterised enzymes. There are 2 cathepsin L-like enzymes; the peptidase CPA (single copy gene (Souza *et al.*, 1992)), and the peptidase CPB (multiple copy gene located on a single tandem array of 19 non-identical copies in *L. mexicana* (Mottram *et al.*, 1997), and 8 in *L. major* (Mottram *et al.*, 2004)). These enzymes possess an unusual carboxy-terminal extension of unknown function. The cathepsin B-like enzyme CPC is encoded by a single copy gene (Bart *et al.*, 1997). Analysis of the peptidase activity showed a stage-specific expression of these different enzymes (Robertson and Coombs, 1994). Targeted gene disruption of all *CPB* genes resulted in a decrease in virulence compared to wild-type parasites, a reduction of infectivity to macrophages (Mottram *et al.*, 1996) and a shift in the immune response to infection in mice from a predominantly Th-2 to a Th-1 response (Alexander *et al.*, 1998; Buxbaum *et al.*, 2003). Re-expression of multiple *CPBs* of the array in the null mutant restored most of the infectivity in mice, showing that the peptidase is a virulence factor (Mottram *et al.*, 1996; Denise *et al.*, 2003). This suggests that these cysteine peptidases promote the parasite's resistance to macrophages' microbial activity and the host's immune response.

Both *L. major* and *L. pifanoi* possess clan CA family C1 cathepsin L-like cysteine peptidase genes quite similar to the ones from *L. mexicana* (Sakanari *et al.*, 1997; Rafati *et al.*, 2001), *L. major* possessing as well a cathepsin B-like gene (Sakanari *et al.*, 1997). Another example of immuno-evasion promoted by cysteine peptidase activity was described in Souza Leao *et al.*, 1995. The degradation of host MHC class II molecules internalised by amastigotes forms of *L. amazonensis* was inhibited by cysteine peptidase inhibitors.

There are other clan CA cysteine peptidases in *Leishmania*; notably 27 family C2 peptidases with calpain-like domains have been identified in the *L. major* genome database. Calpains are involved in signal transduction pathways, and remodelling of cystoskeleton or membrane attachments. Other genes for clan CA peptidases have also been identified, and, among them, there are 19 genes for cysteine peptidases of the families C12, C19, C48, and C65 involved in the ubiquitin pathway.

Among the other clans of cysteine peptidases, clan CD peptidases, including glycosylphosphatidylinositol (GPI):protein transamidase (Hilley *et al.*, 2000) and the family of caspase-like peptidases, the metacaspases (Arnoult *et al.*, 2002), have attracted particular attention. The GPI:protein transamidase is involved in the biosynthesis of GPI-anchored proteins, and among the GPI-anchored proteins in *Leishmania*, the GP63 metallo-peptidase, thought to be involved in resistance to the host's immune response (Ellis *et al.*, 2002; Yao *et al.*, 2003), can be found. Caspase-like peptidases are involved in apoptosis in many organisms. Programmed cell death in *Leishmania* is believed to be caspase-independent (Zangger *et al.*, 2002), and the reason for the presence of metacaspase remains unclear, and its function remains unknown. A gene encoding a pyroglutamyl peptidase I has also been identified (clan CF, family C15) and work on this enzyme will be presented in Chapters 4 and 5 of this thesis.

1.1.2.2.2.2. *Trypanosoma*

Trypanosoma cruzi causes Chagas disease, also known as South American trypanosomiasis. The early stage of the disease is characterised by a mild febrile state, followed by severe heart and gastrointestinal problems some years later. The parasite invades muscle cells and as cardiac muscle is often infected, the illness may lead to death through heart failure.

A major *T. cruzi* clan CA, family C1 cysteine peptidase is known as cruzain (or cruzipain, or gp57/51). It is encoded by multicopy genes present in tandem arrays on more than one chromosome (Eakin *et al.*, 1992), and is expressed in all life-cycle stages, though the highest expression rate is in the replicative insect epimastigote stage. Cruzain has a similar carboxy-terminal extension to that of *L. mexicana* CPB cysteine peptidase (Eakin *et al.*, 1992) and its amino acid sequence shows high similarity with the cathepsin L-like enzymes of other *Trypanosoma* and relatively high similarity with *Leishmania* sequences (Sakanari *et al.*, 1997). Cruzain, unlike most cathepsin L-like enzymes, also shows a carboxypeptidase activity, which is a characteristic of cathepsin B enzymes

(Judice *et al.*, 2004). The enzyme's specific function remains uncertain though it has been linked to differentiation from epimastigotes to the metacyclic stage (Harth *et al.*, 1993; Tomas *et al.*, 1997), and it has been described to have a kininogenase activity (Del Nery *et al.*, 1997). Kininogens are multifunctional glycoproteins present in mammalian plasma and are potent inhibitors of papain-like cysteine peptidases. Cleavage of kininogens with cruzain leads to the production of a potent pro-inflammatory peptide and thus cruzain may be an important virulence factor (Del Nery *et al.*, 1997). Furthermore, substrate cleavage analysis of the different isoenzymes of cruzain showed that they might have different substrate specificities (Del Nery *et al.*, 1997), as is the case for the cathepsin L-like isoenzymes of *L. mexicana* (Mottram *et al.*, 1997). A more recent study also linked cruzain to host cell invasion through a kinin-independent pathway (Aparicio *et al.*, 2004). There is some evidence as well that *T. cruzi* releases cysteine peptidase activity and this is implicated in the chronic heart disease (Morrot *et al.*, 1997).

African trypanosomes (including the human infective *T. brucei rhodesiense* and *T. brucei gambiense,* and the non-human infective *T. brucei brucei* causing Nagana in cattle) are responsible for the usually fatal African trypanosomiasis. The number of human cases occurring each year is estimated at more than 40 000 (http://www.who.int/en/). Less is known about the peptidases of these African parasites compared to the American forms. The major cysteine peptidase of *Trypanosoma brucei* is named trypanopain (Troeberg *et al.*, 1996) (clan CA, family C1). Its precise function is still unknown, though it may be involved in parasite differentiation during the life cycle (Pamer *et al.*, 1989). Cysteine peptidases also seem to be released into the bloodstream, which may cause platelet aggregation and complication of the infection (Troeberg *et al.*, 1996), though these enzymes may only be active for a short period of time before they are inhibited by natural hosts peptidase inhibitors in the serum (Troeberg *et al.*, 1996). By using RNA interference to selectively knock down the cathepsin B activity of *T. brucei,* it was shown the cathepsin B activity, and not the trypanopain activity as previously thought, was essential to the parasite in culture and played a major role in host serum protein degradation by the bloodstream parasite (Mackey *et al.*, 2004).

There are at least two families of cysteine peptidases in *T. congolese*, the CP1 (Fish *et al.*, 1995) and CP2 families, which are 90% identical in their amino acid sequences and differ in their N-terminal sequences (Boulange *et al.*, 2001). Congopain belongs to the CP2-type of enzymes. Though the two types of enzymes are very similar in sequence, they differ in substrate specificity and pH optimum for activity (Boulange *et al.*, 2001), suggesting distinct roles *in vivo*, as well as different effects on the tolerance they can elicit in cattle. The exact role for these enzymes in the parasite remains unknown, but it is thought they may have potential as anti-disease vaccines (Authie *et al.*, 2001).

During *T. congolese* infection, it was observed that a cysteine peptidase of the parasite (congopain) elicited an IgG1 antibody production in cattle which showed a degree of resistance to disease during experimental infections (Authie *et al.*, 1993). Congopain is a 33-kDa lysosomal cysteine peptidase, inhibited by E64 (Authie *et al.*, 1992), which shares similarities with the cruzain from *T. cruzi* and the mammalian cathepsin L and may be an attractive target for anti-trypanosome drugs (Lalmanach *et al.*, 2002). It is a major antigen in infected cattle and antibodies against this antigen inhibit the enzyme's activity (Authie *et al.*, 2001). Unlike the related mammalian cathepsins B and L, congopain accommodates a prolyl residue at the P2' site (Chagas *et al.*, 1997).

1.1.3. Why are parasite cysteine peptidases good drug targets?

Cysteine peptidase activity has been demonstrated in most parasitic protozoa and some of the cysteine peptidases well characterised show common features, for example the similar substrate specificity and similar genetic organisation of the cathepsin L-like cysteine peptidase of *Leishmania* and *T. cruzi*. Therefore, there is good indication that in aiming to obtain novel anti-parasite drugs, it would be useful to target a series of cysteine peptidases of different organisms. Furthermore, some cysteine peptidases involved in human diseases are well characterised and already targeted by a library of compounds. Thirteen papain-like cysteine peptidases are encoded in the human genome and qualify as pharmaceutical targets for the treatment of osteoporosis, arthritis (Taubert *et al.*, 2002), asthma, auto-immune diseases, and potentially for certain forms of cancer (Bromme and Kaleta, 2002). Recent advances in the design of mammalian cysteine peptidase inhibitors were reviewed by Bromme and Kaleta, 2002. To give an example, CLIK-148, a cathepsin-L specific inhibitor designed from analysis of the substrate-binding pocket, computer graphics and X-ray crystallographic data, has been demonstrated to be an effective protective drug *in vivo* against tumour-induced osteoporosis and bone metastasis of cancer cells due to bone collagen degradation and is without side effects (Katunuma *et al.*, 2002). Therefore, a great methodological knowledge has already been developed to validate cysteine peptidases as drug targets, and a large number of inhibitors are already available to direct against often homologous parasite cysteine peptidases.

Although a variety of homologous cysteine peptidases are present in the parasites' hosts, a number of structural and biochemical differences between the two counterparts have been elucidated. This constitutes a clear advantage in drug design, as anti-parasite compounds should not inadvertently target a host cysteine peptidase. Firstly, many parasitic cathepsin-B like peptidases cleave preferably substrates with arginine at the P2 position whereas cathepsin-L like peptidases cleave substrates with a hydrophobic amino acid at P2 (Rawlings & Barrett, 1994). It has been demonstrated that cathepsin L of *Trypanosoma cruzi* and *Leishmania mexicana* prefer hydrophobic residues at the P2 position of substrate, whereas cathepsin B of the same organisms can also accept basic residues (Del Nery *et al.*, 1997; St Hilaire *et al.*, 2000). For human cathepsin-B and cathepsin-L, the preferred residues are glutamate and alanine, respectively. Furthermore, some cysteine peptidases of parasites are difficult to classify, as some predicted cathepsin B-like enzymes based on sequence and structure homology show cathepsin L-like substrate specificity due to single amino-acid substitutions in the catalytic site (i.e. *L. major* cathespin B-like enzyme; Chan *et al.*, 1999). Mammalian cathepsin-L have a highly conserved ERFNIN amino acid motif in the pro-region which is not in the parasite homologues and cathepsins B have an inserted "occluding" loop in the catalytic domain. All vertebrate cathepsins B have maintained the loop whereas some parasite enzymes have lost it in the evolution process, like *Giardia*, or modified the loop sequence or the glycosylation content of the loop (Sajid & McKerrow, 2002). Therefore, parasite cysteine peptidases do not conform to the general properties described for vertebrate cysteine peptidases. As parasites' cysteine peptidases have both lysosomal and extra-lysosomal functions, they are adapted to more diverse chemical environments than their host homologues and are more stable over a wider pH-range than mammalian cysteine peptidases. The reduced stability of mammalian cysteine peptidases at pHs higher than 4.0 (lysosomal pH) may constitute a protection mechanism avoiding inappropriate proteolysis. Parasite cysteine peptidases are not so tightly restricted and different isoforms of the same enzyme can be located in different cell compartments (Parussini *et al.*, 1998). Furthermore, at least some parasite cysteine peptidases are trafficked in the cells through the primitive mannose-6-phosphate-independent pathway and N-glycosylation is not necessarily required. For example, the major route of trafficking of *L. mexicana* cysteine peptidases to lysosomes may occur via the flagellar pocket, as treatment with cysteine peptidase inhibitors leads to a build up of inactive cathepsin B-like cysteine peptidase in the flagellar pocket (Selzer *et al.*, 1999) and experiments with mutated inactive cathepsin L-like cysteine peptidase led to build up of unprocessed precursor of the enzyme in the flagellar pocket as well (Brooks *et al.*, 2000). The trafficking of cysteine peptidases in lower eukaryotic organisms may therefore differ significantly from that in mammals, which is mainly based on binding to specific receptors of mannose-6-phosphate residues on the surface of the enzyme (Kornfeld and Mellman, 1989).

All of these features constitute potential advantages in targeting the cysteine peptidase family to develop new anti-parasite chemotherapy.

1.1.4. Cysteine peptidase inhibitors as potential anti-protozoa drugs

Over recent years, there has been a general increase of resistance of parasites against available drugs, and therefore there is an increasing need to develop new therapeutics and identify new drug targets. For example, *Leishmania* strains resistant to antimonials, currently the frontline drugs of chemotherapy against leishmaniasis, are becoming frequent (Werbovetz, 2002). Furthermore, antimonials show high toxicity. The same problem is encountered with anti-malarial agents (White, 1998) and anti-trypanosomiasis agents. The differences between parasite cysteine peptidases and their homologues in mammals have permitted the development of inhibitors and the use of parasite cysteine peptidases as chemotherapeutical targets. It has now been clearly demonstrated that cysteine peptidase inhibitors can be used as anti-parasite agents (Robertson, 1999; Selzer *et al.*, 1997 and 1999; Fujii *et al.*, 2005) and as tools for the study of enzyme function (Greenbaum *et al.*, 2004). As cysteine peptidases are often numerous in parasites, key issue that remain are identification of cysteine peptidases that are not redundant in function and play vital roles and to find good and specific inhibitors, often by chemical synthesis, screening of banks of molecules and rational drug design based on structural information. Through construction of homology-based enzyme models, structures of inhibitors can be predicted, then they can be synthesised and tested (Selzer *et al.*, 1997; Scheidt *et al.*, 1998). Target validation can be achieved in different ways; firstly by demonstrating that the target is essential through genetic manipulation, secondly by showing that the parasite lines resistant to a specific inhibitor possess a mutated version of the target enzyme, and thirdly by proving there is a good correlation between inhibition of the target activity and the inhibition of the parasite's growth (Coombs and Mottram, 1997).

Following this type of scheme, both reversible and irreversible inhibitors of cysteine peptidases have been identified for different parasites. In *Leishmania*, cathepsin B- and L-like enzymes are attractive new chemotherapy targets as they are required for parasite growth and are demonstrated virulence factors (Mottram *et al.*, 1996; Mottram *et al.*, 1998). Using models of both enzymes, inhibitors could be identified (Selzer *et al.*, 1997) and it is now clear that inhibition of both types of activities is required for parasite clearance (Mottram *et al.*, 1997). Progression of *L. major* infection *in vivo* could be reduced using these inhibitors without toxicity (Selzer *et al.*, 1999). This brings good hope that new therapeutics against leishmaniasis will be commercially available in the future. Similar studies have been undertaken on *Trypanosoma cruzi* and *Plasmodium falciparum*. In *T. cruzi*, there are as many as 100 genes encoding cruzain. This may be a problem in the attempt of designing specific inhibitor. Nevertheless, an effective inhibitor has been identified and named K11777. It is a derivative of vinyl-sulfone compounds and may be used in human clinical trails for the treatment of Chagas disease in the near future (Jacobsen *et al.*, 2000). In *P. falciparum* strains, falcipain-2 is highly conserved (Singh and Rosenthal, 2001) and a series of peptidyl vinyl sulfones have been demonstrated to be good falcipain-2 and -3 inhibitors of potential therapeutic use in the treatment of malaria (Shenai *et al.*, 2003). Safety analysis for these different potent drugs suggest that the selectivity may be due to lack of redundancy of parasite peptidases, higher concentration of the host peptidases in intracellular compartments, and differential uptake of inhibitors by parasites (McKerrow *et al.*, 1999; Selzer *et al.*, 1999). So it seems that the development of ideal reversible, tight binding and very specific inhibitors (or less specific inhibitors that still work as cysteine peptidase' role in the parasite differs from those in the host) of cysteine peptidases for these parasites is on the way and there is hope that some will be commercially available in the future and provide new chemotherapies against major parasitic diseases. Exciting findings acquired in this area could also be applicable to other human diseases involving cysteine peptidase activities, such a cancer.

1.2. *Eimeria*

The genus *Eimeria* (Apicomplexa phylum, subclass Coccidia) includes more than 1700 species of obligate intracellular parasites that infect all classes of vertebrates and some invertebrates. An eimerian species was most probably the first protozoon ever visualised, by Antony van Leeuwenhoek in 1674 in the bile of a rabbit. As oocysts constitute the stage that leaves the host, usually in faeces, 98% of all *Eimeria* species are known only from this life-cycle stage. The diseases caused by these parasites are referred to collectively as coccidiosis, and they vary tremendously in virulence. Some species cause diseases that result in mild symptoms that might go unnoticed and eventually disappear, while other species cause highly virulent infections that are rapidly fatal.

Eimeria species are both site- and host-specific. Most species develop in the cells of the host's gastrointestinal tract, but some exceptions can be found. *Eimeria* species seem to be limited to specific locations in specific cells in a limited organ. Most species show high host specificity too (*E. tenella* only infects chickens), but some can cross genetic boundaries and infect several hosts.

There has not been any evidence that *Eimeria* infects humans, though closely related genera, like *Plasmodium*, *Cryptosporidium* and *Toxoplasma* for example, can do so. *Eimeria* infections in wild and domesticated animals are frequent and widespread, but harmless in most cases. It is only when many animals are confined together in a restricted area, as it is the case for chicken flocks infected with the species *E. tenella*, that the parasite becomes a problem as animals are continually re-infected and eventually develop coccidiosis. It is particularly a problem for young birds, as adults may become immune.

Coccidiosis is the most important death-causing agent in the poultry industry and is a major economical issue. At least 11 species of *Eimeria* can infect chickens, but the most pathogenic species is *E. tenella*. *E. necatrix* (responsible for intestinal coccidiosis), *E. acervulina* (responsible for upper digestive tract coccidiosis) and *E. maxima* (responsible for middle digestive tract coccidiosis). All cause chronic intestinal coccidiosis and are responsible for disease outbreaks, but are less significant than *E. tenella*. *E. necatrix* produces fewer oocysts, therefore the development of a sufficient level of environmental contamination requires more time, whereas *E. acervulina* and *E. maxima* are less pathogenic and mainly affect older birds. *E. tenella* develop in the cells of the caeca (two blind sacs near the end of the intestine). After three days of infection, the host chicken stops feeding, droops and by the fourth day, blood appears in the droppings. By the eight or ninth day, the bird is either dead due to excessive blood loss, or on its way to recovery. This type of coccidiosis is the most common in young chickens and is therefore a major problem. Furthermore, with the rising problem of resistance of eimerian strains to most chemicals used as anti-coccidial agents, there is a need to identify new drug targets and develop new drugs.

1.2.1. Life cycle, transmission and pathogenicity

Eimeria species are homoxenous; they have a direct life cycle that takes place within the same host. The infective stage is the sporulated oocyst, the environmentally-resistant form of the parasite which contains 4 sporocysts (Fig. 1.2, b, d) containing 2 sporozoites each. Sporulated oocysts (Fig. 1.2, a, c) can survive in a litter for many months.

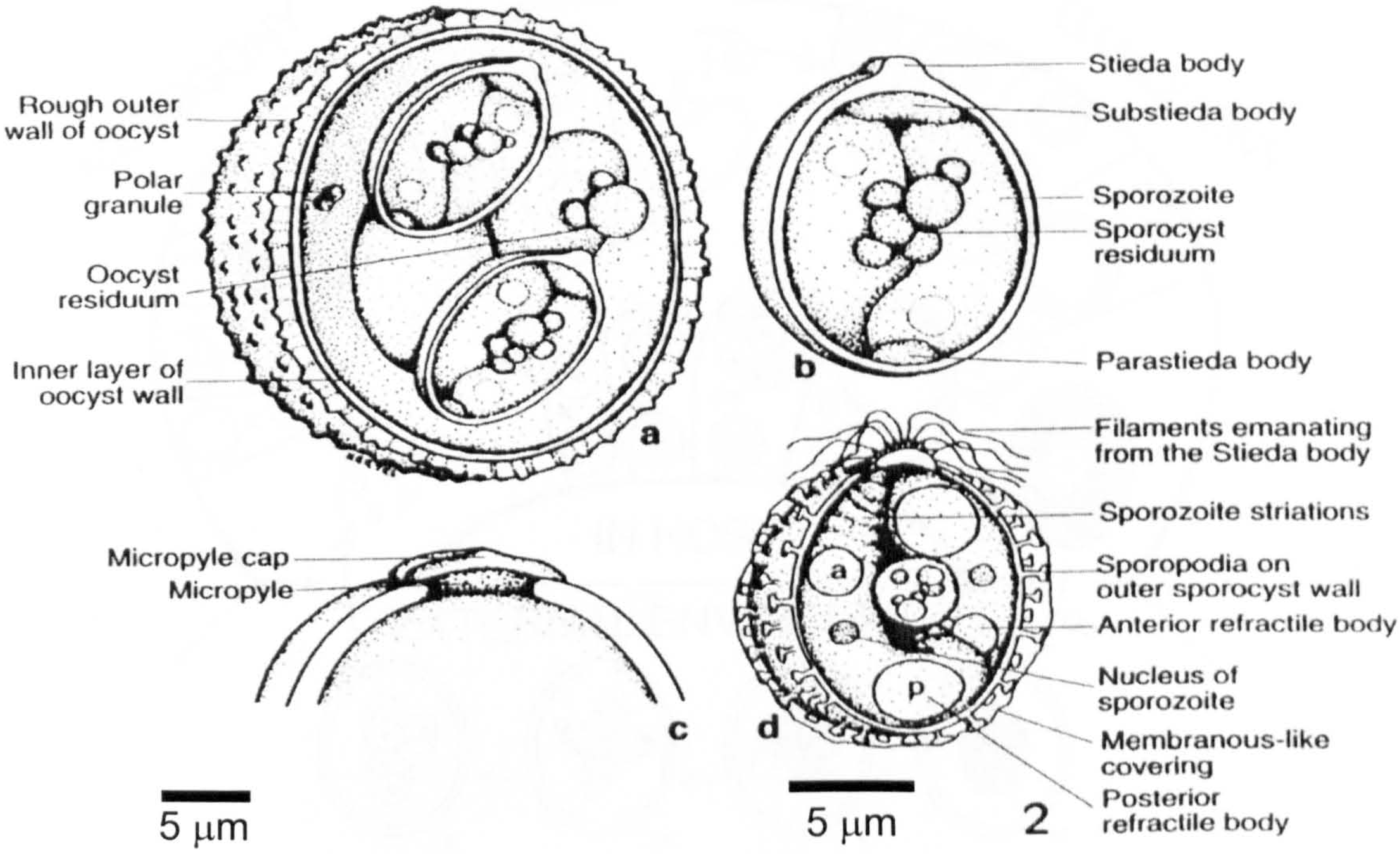


FIG. 1.2. ***Eimeria* sporulated oocyst structure**. Line drawings of the parts of hypothetical sporulated oocysts/sporocysts of *Eimeria*: a, completely sporulated oocyst showing major structural features with four sporocysts each with two sporozoites; b, sporulated sporocyst showing major features, including two sporozoites; c, end of an oocyst showing other possible structures, a micropyle and micropyle cap, present in some oocysts, especially those of ruminants; d, another sporulated sporocyst showing a variety of structural features, some of which may be present on the sporocysts of different *Eimeria* species (a = anterior, p = posterior refractile bodies of sporozoite) (Duszynski, 2001).

Unsporulated oocysts are released in the faeces of an infected host (Fig. 1.3, u). Under the appropriate combination of environmental factors (conditions of moisture, temperature and direct exposure to sunlight) often optimal in a litter, the oocyst sporulates, which means that sporocysts and sporozoites develop. This is an aerobic process involving high oxygen consumption and an energy source from stored mannitol. Once sporulation is completed the metabolism is more lipid based. The diploid sporoplasm undergoes a meiosis (Fig. 1.3, v) followed by mitosis to form 4 sporoblasts (Fig. 1.3, w). Each sporoblast forms a sporocyst containing 2 sporozoites each (Fig. 1.3, x). This process takes about 2 days at room temperature.

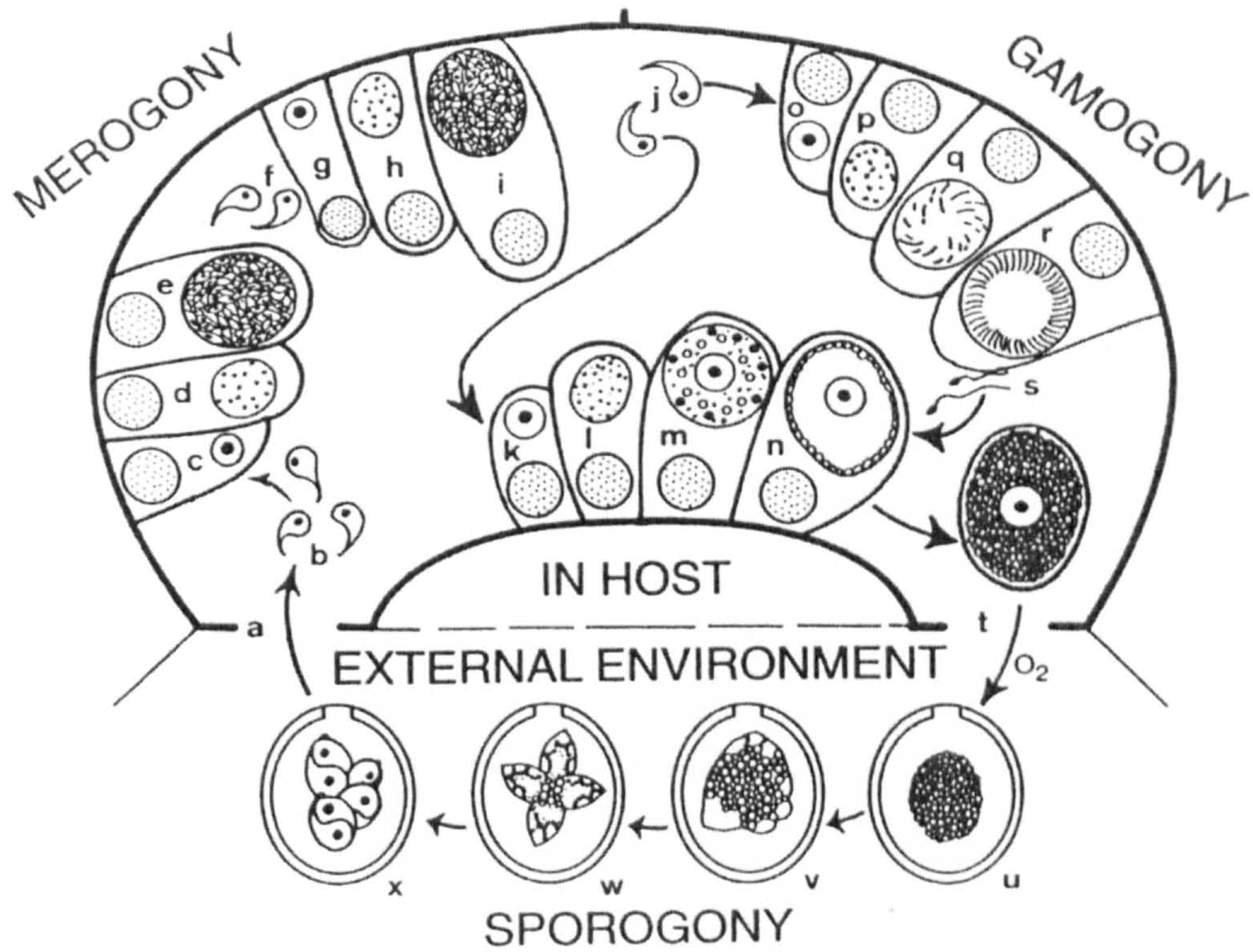


FIG. 1.3. **Complete life cycle of *Eimeria*.**
(http://biology.unm.edu/biology/coccidia/eimeriabiol.html).

A host is infected when it ingests sporulated oocysts from contaminated food or water (Fig. 1.3, a). The oocyst wall is mechanically (or/and enzymatically) broken in the stomach/gut of the host and the biochemical environment is favourable to the active process of excystation. Sporocysts are set free and exposed to enzymes of the bile, which leads to release of the sporozoites (Fig. 1.3, b). The sporozoites move actively and enter epithelial cells for their further development (Fig. 1.3, c). In *E. tenella*, the sporozoites penetrate the villus epithelium of the caecum at the tip of the villi. In the cell, the parasite is engulfed in a parasitophorous vacuole where uninucleate sporozoites develop to multinucleate schizonts (Fig. 1.3, d) and schizonts undergo cytoplasm division to produce merozoites (around 900) (Fig. 1.3, e). The merozoites break out of the epithelium, disrupting the host cell and are released back into the intestinal lumen where they re-infect other epithelial cells (Fig. 1.3, f). Several cycles of asexual multiplication (schizogony or merogony) occur and lead to the formation of a large number of merozoites (Fig. 1.3, g-j). Merozoites of different generations differ in size and quantity. *Eimeria* infections are self-limiting as asexual generations are limited to 2 to 4 before sexual generation occurs. Then, instead of giving a further generation of schizonts, merozoites re-enter the epithelium and begin the process of gametogony by which macrogametocytes (female gamete) (Fig. 1.3, k-o) or microgametocytes (male) (Fig. 1.3, o-r) are formed. The mechanism that triggers this developmental switch is still unknown. The microgametocyte buds off to produce many flagellated microgametes that leave the epithelial cells in which they were produced to enter cells containing macrogametocytes (Fig. 1.3, s). The process of fertilisation by which a microgamete actively enters a macrogamete, leads to the formation of an intracellular zygote and the granules present within the macrogamete coalesce to form the outer cyst wall. The oocyst disrupts the host's cell membrane, is released in the intestinal lumen and is discharged in the faeces (Fig. 1.3, t). The entire cycle from the infection to the production of oocyst lasts 6 days.

A picture of a cell-invading merozoite is shown on Fig. 1.4. and a diagram of a sporozoite is provided in Fig. 1.5. A typical structural feature of apicomplexan parasites is the presence of a unique apical complex, composed of polar rings, rhoptries, micronemes, often a conoid, and other subcellular organelles that differ from species to species (Tomley, 1997). The apical complex is involved in cell invasion (Soldati *et al.*, 2001; Tomley and Soldati, 2001). The refractile bodies, typical of the sporozoite stage, are composed of lipid components that might be incorporated into the parasitophorous vacuole (Entzeroth *et al.*, 1998).

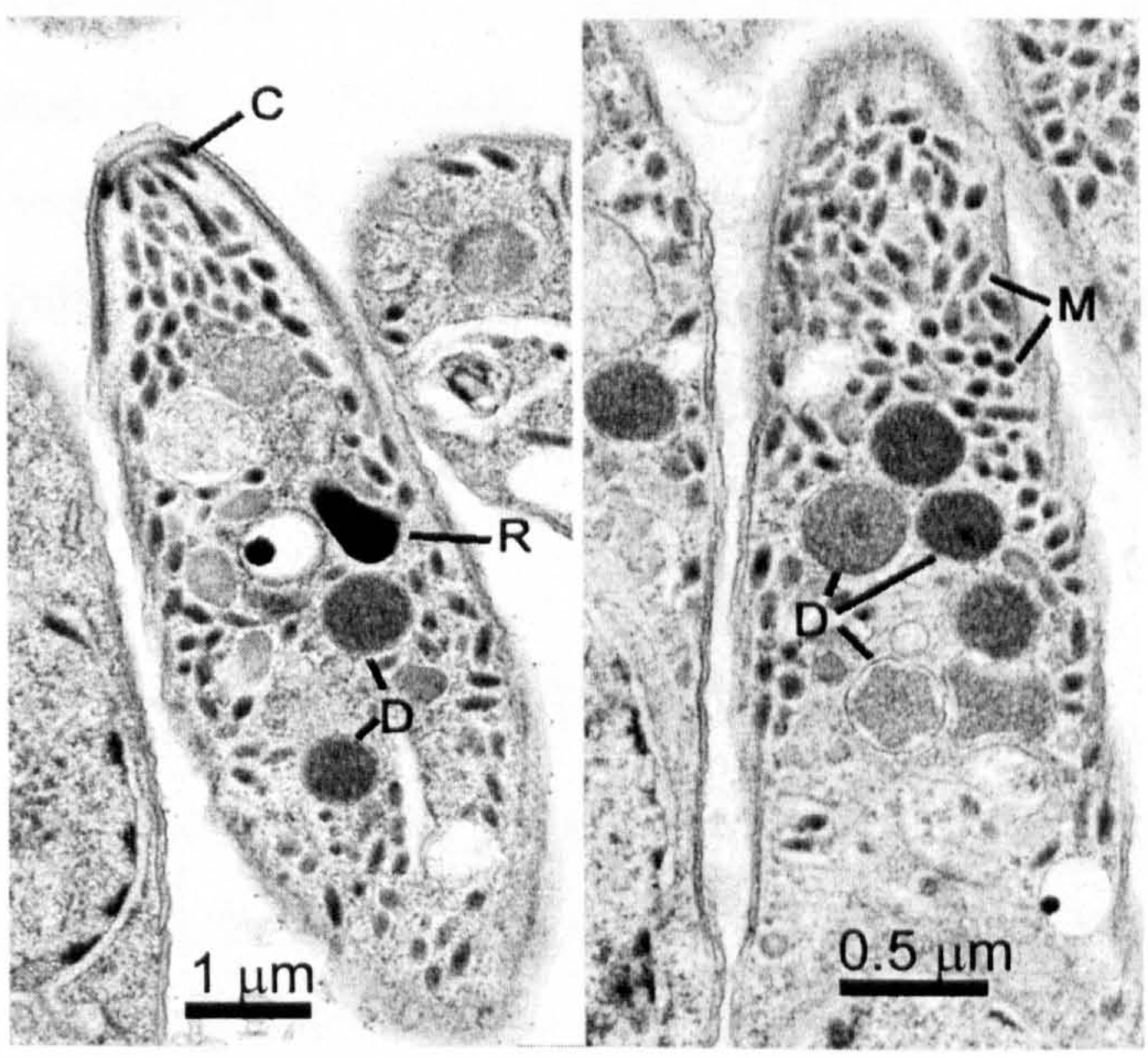


FIG. 1.4. **Electron-micrograph of second generation merozoites of *E. tenella*.** (Courtesy of Prof D. Ferguson, Oxford). C: conoid, D: dense granules, M: micronemes, R: rhoptry.

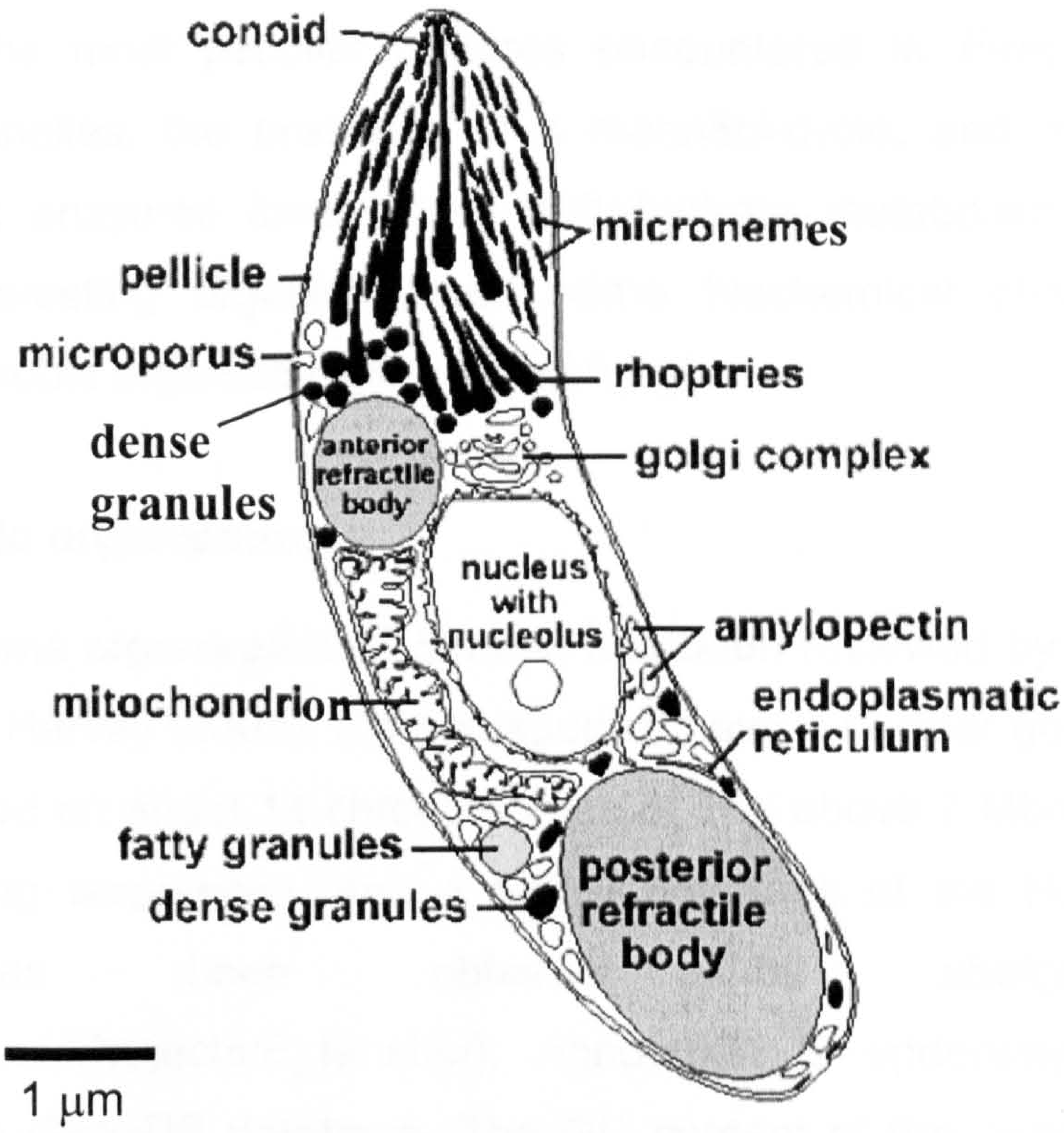


FIG. 1.5. **Structure of a sporozoite of *Eimeria* species** (www.saxonet.de).

The release of the merozoites induces rupture of host's epithelial cells, destroying tissue and causing haemorrhages in epithelial capillaries. This will lead to villous atrophy in the intestine resulting in malabsorption and accumulation of cell debris, blood clots, and necrotic materials in the caecum of infected birds with *E. tenella* causes necrosis of the organ, sometimes leading to death. Not all species of *Eimeria* cause such dramatic symptoms. The pathogenicity will very much depend on the type of species ingested, the number of sporulated oocysts ingested, the age of the host (old chicken for example may be immune against *E. tenella*), and the environmental conditions (such as crowding effect; Williams, 2001).

1.2.2. Biochemical and molecular biology features

Studies of the biochemistry of *Eimeria* have been made difficult by its intracellular location. It is therefore difficult to obtain enough material for analysis and parasites cannot be grown outside of the host cells. It is mainly oocysts and sporozoites, the extracellular stages of the life cycle of the parasite, that have been biochemically characterised.

Among the most peculiar features encountered in *Eimeria*, there are plastid-like organelles, the presence of a mannitol-cycle, and pyrophosphate-linked glycolytic enzymes involved in carbohydrate metabolism. This makes *Eimeria* an interesting organism, with some biochemical characteristics of mammals, anaerobic organisms, plants and fungi.

1.2.2.1. Genomic organisation

The genome organisation of *Eimeria* has been reviewed by Shirley (2000) and Shirley and Harvey (2000). *Eimeria* species have a nuclear genome of about 60 Mbp contained on about 14 chromosomes of 1 to above 7 Mbp. The genome is currently being sequenced, and a 5-time coverage of the Houghton strain genome has been obtained by shotgun reads (www.sanger.ac.uk/Projects/E_tenella/). Annotation is underway and will be available on the GeneDB database. The GC content of the genome has been determined for a few species and is around 50% (53% for *E. tenella*), there is an abundance of the tri-nucleotide repeat GCA (Shirley, 2000), the codon usage has been determined for nine genes of *E. tenella* (Ellis *et al.*, 1994) and bias was associated with the over-representation of G- or C-rich codons. Additional to the nuclear genome, a 6 kb mitochondrial DNA molecule and a circular plastid DNA molecule of about 35 kb could be detected (Williamson *et al.*, 1994). The plastid DNA molecule is contained in a plastid-like organelle, and it has been suggested that it could have a green algae origin (Kohler *et al.*, 1997). The position of the plastid-like organelle in the cells and its full biological function are unknown, though it has been suggested for other apicomplexan parasites (*T. gondii*, *P. falciparum*, *Sarcosystis muris*, and *Babesia ovis*) that it contains some components of the respiratory chain (Hackstein *et al.*, 1995), proteins involved in haem synthesis, in Type II fatty acid biosynthesis (Vollmer *et al.*, 2001), tRNAs and transcription factors (Wilson and Williamson, 1997), and proteins for isoprenoid biosynthesis (important for protein prenylation and signal transduction regulation) (Vial, 2000). Furthermore, spherical virus-like particles could be revealed on electron microscopy and viral-like double-stranded RNA molecules could be detected (Ellis and Revets, 1990), but the role of these RNA viruses in the biochemistry of *Eimeria* is unknown so far.

Further analysis of the function of the genes requires good molecular biology tools. There is hope that transfection technologies will provide a good method of study, as transient transfection in sporozoites of *E. tenella* could be performed (Kelleher and Tomley, 1998). But no gene knock-out or RNA interference experiments have been successful so far, as studies are mainly limited to the extracellular non-replicative forms of the parasite and *in-vitro* propagation of the life-cycle is very inefficient (Tierney and Mulcahy, 2003).

1.2.2.2. Energy metabolism

The major energy sources for *Eimeria* through its life cycle are carbohydrate compounds (routinely mannitol, and glycogen and amylopectin as storage forms), though lipids might be metabolised too (Coombs *et al.*, 1997). The presence of anaerobic-specific enzymes (like PPi-PFK, phosphofructokinase, Muller *et al.*, 2001) in *Eimeria* suggests that the parasite may be adapted to this metabolic mode (Denton *et al.*, 1994). Sporozoites for example may be facultative anaerobes. A surprising feature is the presence of mannitol at high level in unsporulated oocysts (Schmatz *et al.*, 1989). This is a carbohydrate previously detected only in fungi, with mannitol-cycle associated enzymes. The mannitol cycle occurs in stages other than the oocyst, such as sporozoites (Michalski *et al.*, 1992). The function though is still uncertain, though there is good evidence that it may be required during sporulation (Allocco *et al.*, 1999). Respiratory chain and other catabolic pathways are similar to the typical eukaroytic pathways and were reviewed in detail by Coombs *et al.* (1997).

Interestingly, the shikimate pathway, necessary for synthesis of aromatic amino acids and absent from mammals, is present in coccidia (Roberts *et al.*, 1998). This provides an additional potential target for chemotherapy and some herbicide-like compounds have been found to have some efficacy against enzymes of the shikimate pathway in apicomplexan parasites, like in *Plasmodium falciparum* (McConkey, 1999).

1.2.3. Treatment and control

Avian coccidiosis is considered to be one of the most important veterinary diseases of domestic animals. Over the years, many anti-coccidial compounds have been shown to have efficacy and have been used as feed additives to control the disease. The major drugs currently used against *E. tenella* are polyether ionophores (Dutton *et al.*, 1995). A list of commercially available anti-coccidial drugs and their mechanism of action is given in Table 1.2.

Table 1.2. **Mechanism of action of commercially available avian anticoccidial drugs.** (From Coombs *et al.*, 1997).

Compound class	Examples	Mechanism of action
Polyether ionophore	Salinomycin Lasalocid Maduramicin Semduramicin	Perturb ion gradients
Carbanilide/pyrimidine	Nicarbazin	Oxidative phosphorylation uncoupler?
Febrifugine	Halofuginone	Not known
Triazine	Diclazuril Toltrazuril	Pyrimidine metabolism? Mitochondrial respiration? Chlorophyll α–D1 complex?
Quinolone	Decoquinate Buquinolate	Mitochondrial respiration/electron transport Topoisomerase?
Pyridinol	Clopidol	Mitochondrial respiration/electron transport
Thiamine analogue	Amprolium	Thiamine uptake and utilisation
Nitrobenzamide	Zoalene Nitromide	Nicotinamide antagonist?
Guanidine	Robenidine	Oxidative phosphorylation uncoupler?
Benzylpurine	Arprinocid	Purine salvage? Ion chelation (N-oxide active metabolite?)
Organic arsenical	Roxarsone	Binds protein sulphydryl groups?
Polyketide	Oxytetracycline Chlortetracycline	Protein synthesis?
Sulphonamide	Sulphadimethoxine Sulfaquinoxiline	Dihydropteroate synthetase
Aminopyrimidine and sulphonamide	Ormethoprim and sulphadimethoxine	Dihydrofolate reductase and Dihydropteroate synthetase

However, resistance to these drugs has occurred (Augustine *et al.*, 1996) and effort has been put in attempts to understand better the parasite's biology and the host immune response to the parasite to find alternative ways of control, like the development of vaccines. The basis of this approach is the observation that it is possible to acquire protective immunity to *Eimeria* species, which can be boosted by increasing the exposure to the parasite's antigens. Current vaccines are composed of oocysts (usually attenuated to avoid averse effect) of various species of *Eimeria*. Commercial live vaccines are already available, like Coccivac® (1952), Immunocox® (1987), Paracox® (Shirley, 1989), Livacox® (1993). Vaccine efficacy has been reviewed by Chapman *et al.* (2002). Vaccination is the only alternative to chemotherapy at the moment, and in practice the administration of drugs remains the easiest, most cost effective way and probably the most effective way of controlling the disease (Waldenstedt *et al.*, 1999). But the intensive use of drugs in farm animals raises public health concerns as residues of coccidiostats can be found in the meat and this might lead to antibiotic resistance in humans. Rotating treatments involving an alternation of the administration of drugs and vaccines, or of different vaccines, or of different drugs can be used. Therefore, there are already methods of control available but none seems ideal and absolutely efficient; therefore there is a constant need for identifying new potential drug targets, developing new drugs or new control strategies. Increasing knowledge of the parasite's biology in concordance with genome sequencing projects should bring useful information for the identification and validation of better anti-coccidial agents.

1.3. *Leishmania*

The genus *Leishmania* belongs to the family of *Trypanosomatidae*, of the order Kinetoplastida. They are flagellated parasites that occur as intracellular amastigotes in vertebrate hosts and promastigotes in invertebrate vectors. About 21 species have been identified to date, most of them having a specific mammalian reservoir (mainly rodents), and some of them being human pathogens. The parasite is endemic in areas of the tropics and subtropics, in a total of 80 countries in South and Central America, Southern Europe, Asia, the Middle East, and Africa, with sporadic cases elsewhere (Fig. 1.6).

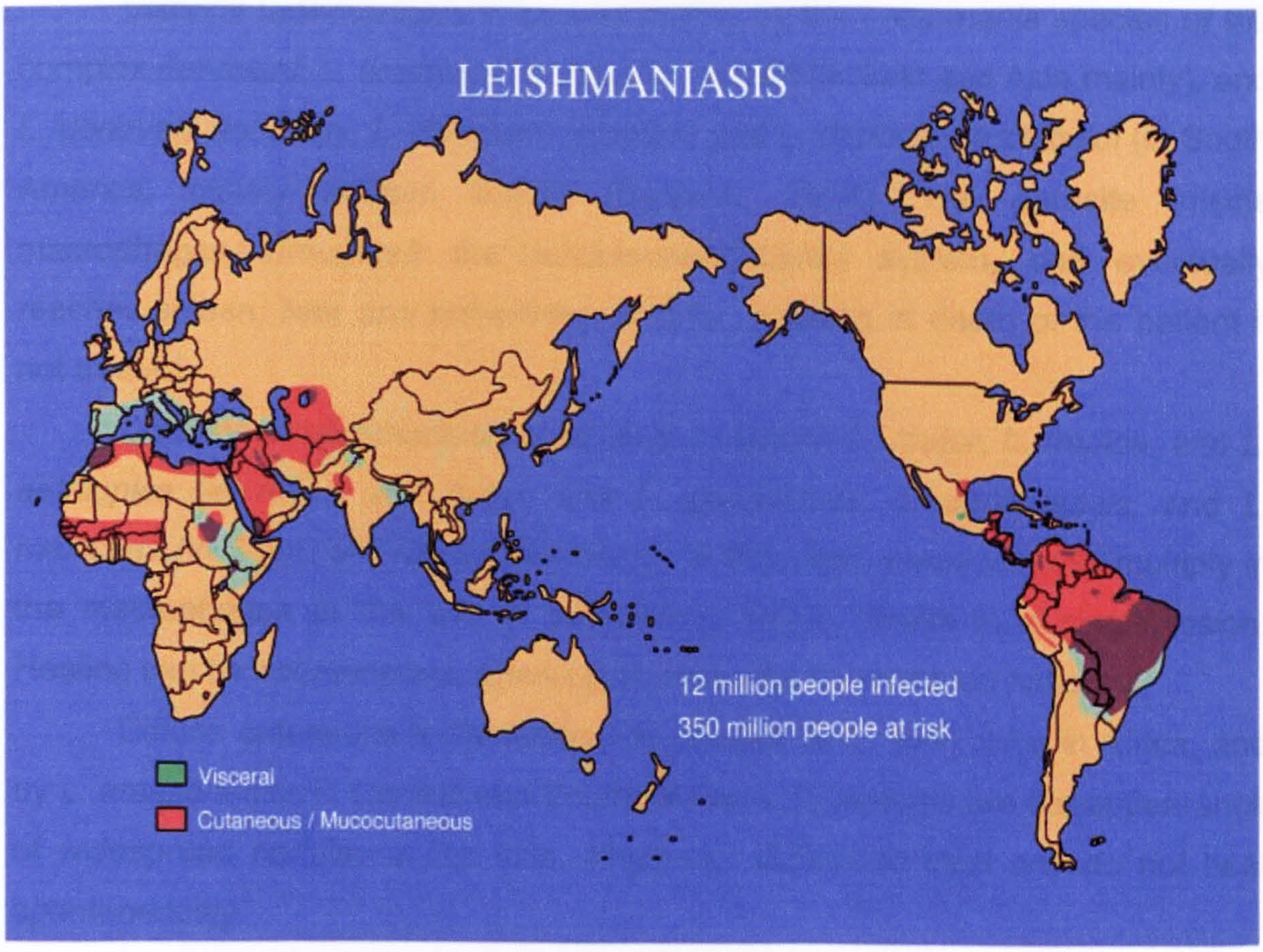


FIG. 1.6. **Leishmaniasis distribution in the world.**
(srs.wehi.edu.au/media/images/handman/world_map.jpg)

Leishmania affects 2 million people and causes about 60000 deaths a year (Croft *et al.*, 2002; Reithinger *et al.*, 2002), and there has been an ever increasing number of cases reported every year over the last 20 years. The last 10 years have also seen an emergence of leishmaniasis as an important opportunistic infection in immuno-depressed patients; in particular those infected with the human immunodeficiency virus (HIV) (Ambroise-Thomas, 2001). Infection with *Leishmania* is also a frequent cause of clinical disease in the dog, especially in tropical, sub-tropical and temperate areas.

Different *Leishmania* strains are responsible for different forms of the disease in humans, with different severity of symptoms. The four types of diseases are: visceral, cutaneous, diffuse cutaneous and mucocutaneous leishmaniasis. According to the human clinical criteria, canine leishmaniasis is classified as visceral, although the term "generalized canine leishmaniasis" might be more appropriate because it involves visceral and cutaneous tissues (Alvar *et al.*, 2004).

Visceral leishmaniasis is caused mainly by the *Leishmania* species of the complex *donovani*; *L. donovani chagasi* (in the Middle East and Asia mainly), and *L. donovani donovani*, *L. donovani infantum* and *L. donovani archibaldi* (in South America, mainly eastern Brazil) (Desjeux, 2004). The parasite infects macrophages throughout the reticuloenendothelial system, and eventually reaches spleen, liver and bone-marrow cells, resulting in death of the patient if not treated.

Cutaneous leishmaniasis is caused mainly by *L. major*, *L. tropica*, and *L. aethiopica* (in Africa and Asia); and *L. guyanensis*, *L. panamensis*, and *L. mexicana* (in South and Central America). In this case, parasites only multiply in the macrophages at the site of inoculation, which results in a single lesion. Healing can be spontaneous, resulting in immunity, in untreated patients.

Diffuse cutaneous leishmaniasis in caused by *L. aethiopica* in Africa, and by *L. amazonensis* in Central and South America. Symptoms are the appearance of widespread nodules in the skin, which are difficult to treat and do not heal spontaneously.

Mucocutaneous leishmaniasis is caused by *L. braziliensis* and *L. panamensis* in South America only. Parasites are inoculated by a sand fly bite at any location on the body, which forms a small lesion that heals spontaneously. But the parasites invade and erode cartilaginous tissues (nose and palate) causing disfigurement and eventually death from secondary infections if untreated.

Leishmaniasis is therefore an important human disease, in constant progression and emergence of resistance to available drugs is an increasing problem. It is essential new drugs are identified in order to increase chances of controlling the disease.

1.3.1. Life cycle and transmission

The life cycle of *Leishmania* alternates between a vertebrate and an insect host (Fig. 1.7). The natural reservoir hosts, besides humans, include dogs and various wild animals. The vectors are small Dipteran insects (sandflies) of the genus *Phlebotomus* and *Lutzomyia*.

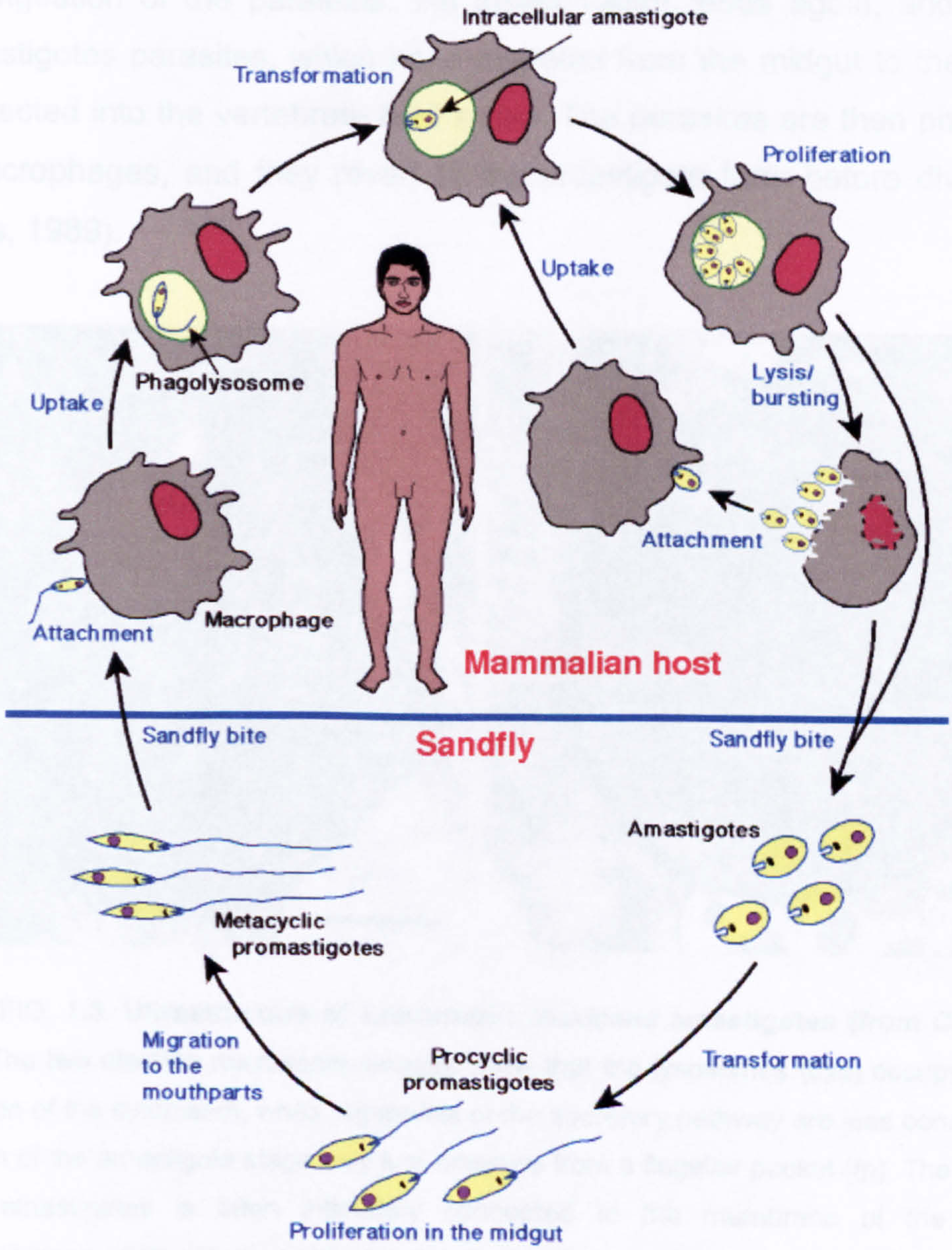


FIG. 1.7. **Life cycle of *Leishmania* species.** (www.wehi.edu.au/media/images/ leishmania_cycle.gif).

Amastigotes (in the vertebrate host) have virtually no flagellum (Fig. 1 8). They are ingested by macrophages by phagocytosis. The parasites, resistant to lysosomal enzymes released into the phagolysosome, multiply by binary fission within the macrophages (Pearson *et al.*, 1981). This makes the macrophage die eventually and release the parasites that can infect new macrophages. The sandfly vector (female only) ingests the infected macrophages from the blood or skin, while having a blood meal. The parasites then transform into procyclic promastigotes in the midgut of the insect (Walters, 1993). Promastigotes are more elongated, flagellated forms (15-20 μm in size), with the kinetoplast near the anterior end, which multiply by binary fission too (Fig. 1.9). About 10 days after ingestion of the parasites, the insect vector feeds again, and metacyclic promastigotes parasites, which have migrated from the midgut to the proboscis, are injected into the vertebrate host's skin. The parasites are then phagocytosed by macrophages, and they revert to the amastigote form before dividing again (Sacks, 1989).

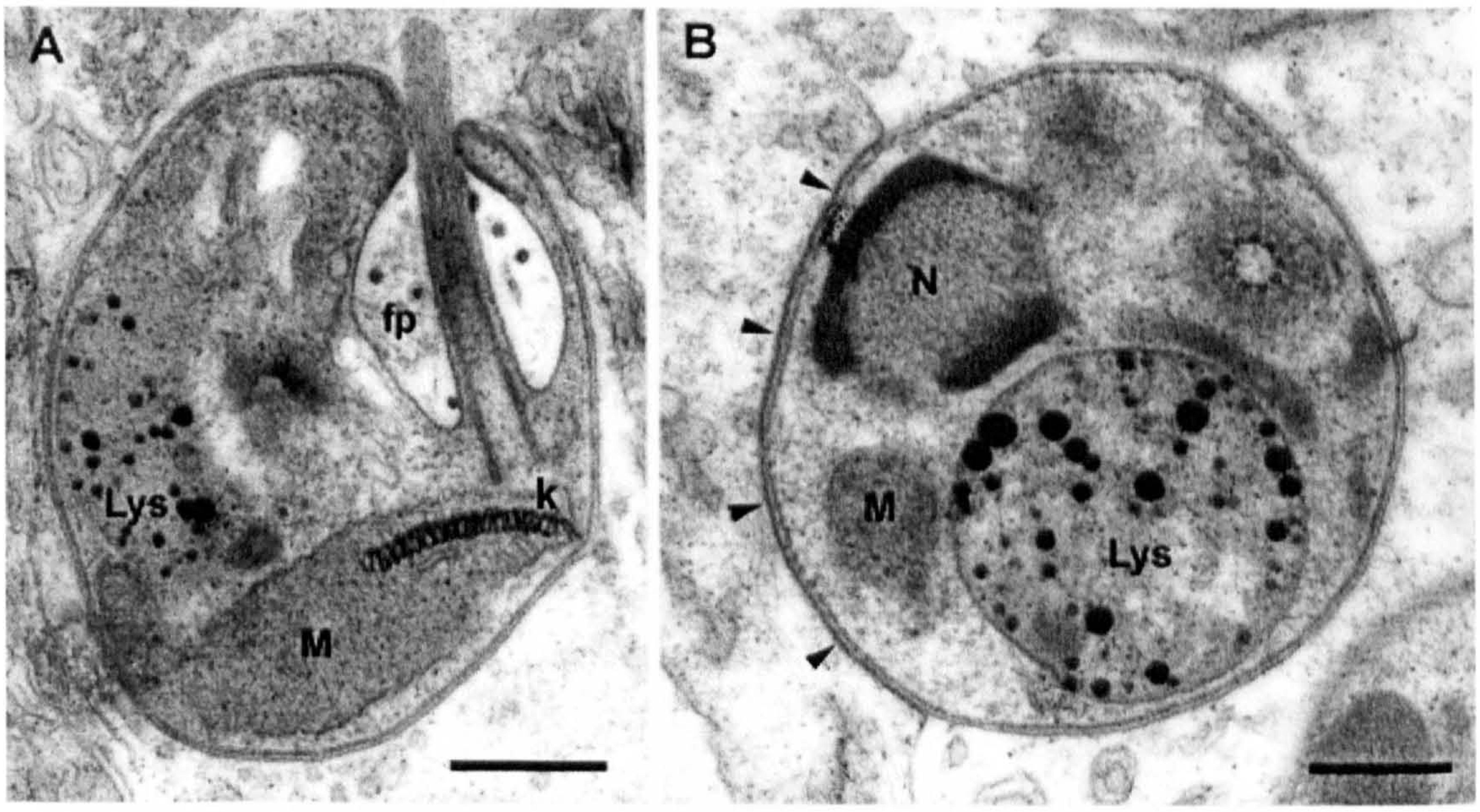


FIG. 1.8. **Ultrastructure of *Leishmania mexicana* amastigotes (from Coombs *et al*, 1986)**. The two electron microscopy images show that the lysosomes (Lys) occupy a significant proportion of the cytoplasm, while organelles of the secretory pathway are less conspicuous. The flagellum of the amastigote stage only just emerges from a flagellar pocket (fp). The posterior end of the amastigotes is often intimately connected to the membrane of the macrophage phagolysosome compartment (arrowheads). N, nucleus; M, mitochondrion; k, kinetoplast. Scale bars: 500 nm.

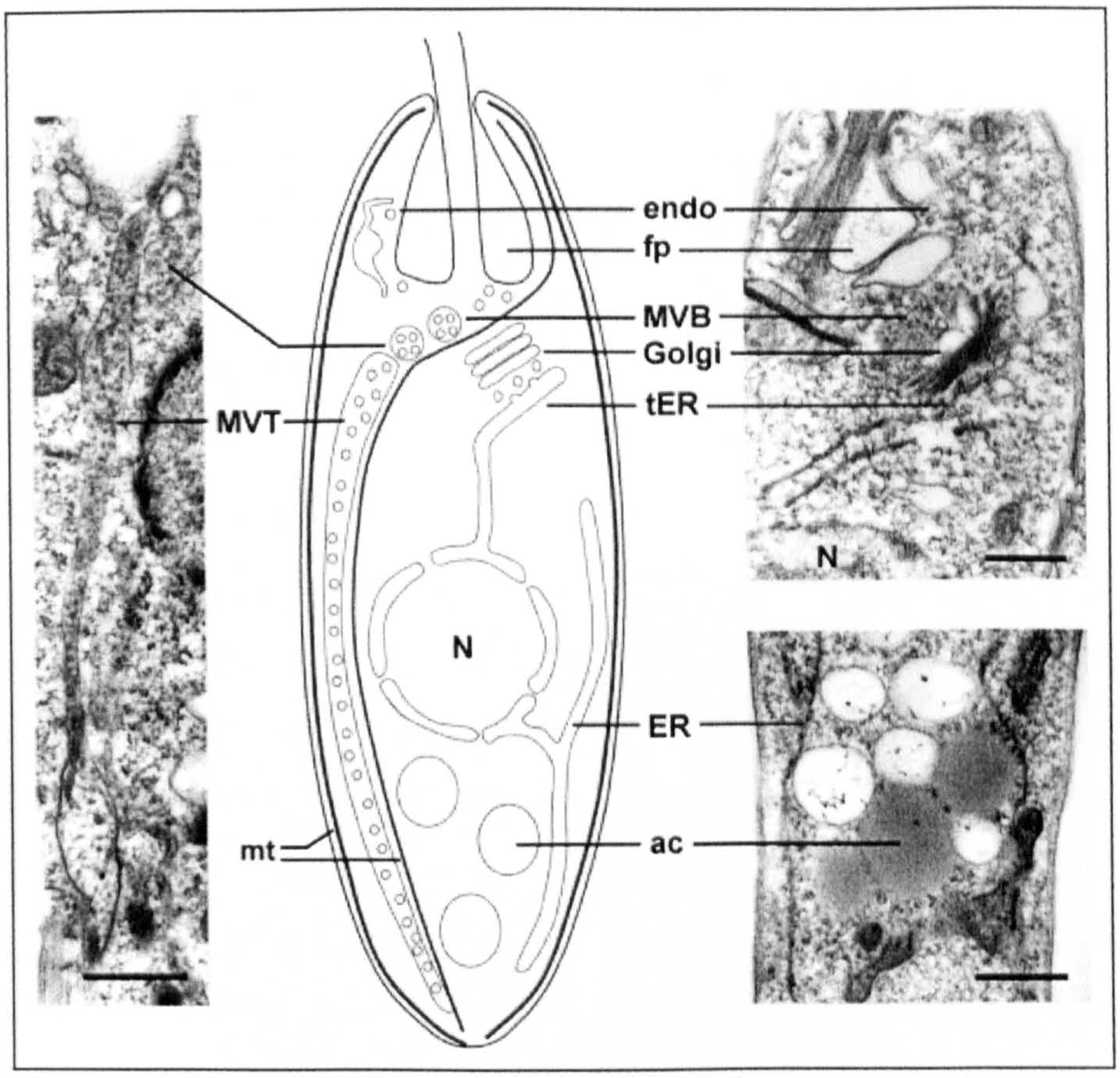


FIG. 1.9. **Ultrastructure of *Leishmania mexicana* promastigotes (from Waller and McConville, 2002).** Tubular 'early' endosomes (endo) are closely associated with the flagellar pocket (fp). A single Golgi apparatus is also proximal to the flagellar pocket and is flanked by a transitional endoplasmic reticulum (tER). Multivesicular bodies (MVBs) can be seen arising. The lysosome compartment, termed the multivesicular tubule (MVT) extends from near the flagellar pocket, in close proximity to MVBs, to the posterior end of the cell. One or two microtubules (mt) run along the length of the MVT-lysosome. Acidocalcisomes (ac) cluster principally in the posterior of the cell (note that glutaraldehyde and osmium fixation extracts the contents of some of the acidocalcisomes). N, nucleus. Scale bars: 500 nm.

1.3.2. Biochemical and molecular biology features

Studies of the biochemistry of *Leishmania* have been possible due to the relatively easy maintenance of the promastigote form of the parasite in culture. The promastigotes multiply by binary fission and differentiate into metacyclic promastigotes when reaching the stationary phase of growth. This has made genetic manipulations, and studies of particular gene's functions possible.

Among the most peculiar features present in *Leishmania*, is the kinetoplast, a unique cellular structure located within the single mitochondrion and possessing its genetic information. The kinetoplast is located at the base of the flagellum near the anterior end (Fig. 1.8). At the metabolic level, glycolysis is a regulated event that takes place in the glycosomes, which are specific organelles related to peroxisomes (Parsons, 2004), and oxidative stress is regulated using trypanothione, an antioxidant absent from the mammalian host, which utilises glutathione instead.

1.3.2.1. Genomic organisation

Leishmania is a diploid organism. The haploid genome of *Leishmania major* Friedlin (the subject of this study) is composed of 36 chromosomes, that vary in size from 326.9 to 2821.3 kb, with a total genome size of 34.7 Mb (Zhou *et al.*, 2004). In terms of codon usage, the G+C content at the third codon position represents the main source of codon usage variation (Alvarez *et al.*, 1994). The genome is divided into transcription units (Myler *et al.*, 2000), which are transcribed into polycistronic mRNAs. The control of gene expression is mainly post-transcriptional, using trans-splicing and polyadenylation, which require the addition of a 5' cap and a 3' poly A tail. At the 5' end of all mRNAs, a splice leader sequence (SL) is added by trans-splicing (Graham, 1995). The SL sequence is composed of a non-translated 39-41 nucleotide sequence, which is well conserved between trypanosomatid species (Gibson *et al*, 2000). At the 3' end no specific signal for the addition of the poly A tail has yet been described, although polypyrimidine tracts have been identified.

Additional to the nuclear genome, the kinetoplast contains a catenated network of DNA (kDNA), which can be divided into 2 groups: maxicircles and minicircles, where maxicircles are found at a copy number of 20-50 per kinetoplast, and minicircles are present at about 10000 per kinetoplast. Maxicircles are homogeneous circular DNA molecules, 20-35 kb in size, which have many of the characteristics of conventional mitochondrial DNA, encoding proteins involved in energy production. Minicircles are a heterogeneous group and 0.5-1.5 kb in size, which differ in sequence number and types from species to species. They do not encode proteins, but encode for guide RNAs (gRNAs), which have a role in the maturation of the mitochondrial mRNAs, a process known as RNA editing, a process typical of trypanosomatids (Schneider, 2001).

After transcription, U nucleotides are inserted, and create open reading frames from non-sense sequences at the RNA level (Benne *et al.*, 1986). U deletions can also occur, but at lower frequencies. The small RNA molecules transcribed from the minicircles and maxicircles contain the editing information (Simpson *et al.*, 2004). RNA editing is used as a mechanism of regulation of mitochondrial gene expression (Worthey *et al.*, 2003).

1.3.2.2. Energy metabolism and trypanothione

Leishmania species present a number of particular biochemical features. First, they present a unique compartmentation of glycolysis within unusual peroxisomes called glycosomes. Like peroxisomes, glycosomes are bounded by a single-membrane. In *Leishmania* amastigotes, glycosomes comprise 1% of the total cell volume (Coombs *et al.*, 1986). Some enzymes commonly found in peroxisomes are also found in the glycosomes (Wiemer *et al.*, 1996). So the glycosomes might have evolved from peroxisomes, but might also reflect recruitment of proteins from a photosynthetic organism through lateral gene transfer by endosymbiosis (Hannaert *et al.*, 2003) or phagocytosis (Waller *et al.*, 2004). This photosynthetic organism is thought to have been an alga, as some pathways are related to those of algae, plants and chloroplasts (Hannaert *et al.*, 2003), and some plastid-like proteins are found in the glycosome. Glycolysis occurs as a result of a cooperation between the cytoplasm, the mitochondrion and the glycosome, and purines, which the parasite cannot synthesise *de novo*, are also salvaged using enzymes associated with glycosomes. Although some enzymes of glycolysis are similar to those of the mammalian host, the glycosome possesses unique enzymes of algal origin, which may be interesting for drug design (Parsons, 2004).

Another interesting characteristic is that *Leishmania* parasites regenerate thiols using trypanothione (a glutathione-spermidine conjugate), rather than glutathione (Henderson and Fairlamb, 1987). This is the anti-oxidant used to reduce hydrogen peroxide and other oxidants. Trypanothione and the enzymes involved in its synthesis and recycling are absent from the mammalian host and so have been proposed as suitable targets against which to develop drugs (Fairlamb and Cerami, 1992; Chibale and Musonda, 2003).

1.3.3. Treatment and control

As a first strategy of control of the disease, the elimination of natural reservoirs such as dogs, rodents and insect vectors has been used. Sandfly population control is an ongoing process and continued surveillance is required to keep the population low. However, this is possible only under certain environmental conditions. Therapy against leishmaniasis mainly relies upon pentavalent antimonials, which have been used since the 1940's, and include sodium stibogluconate or meglumine antimonite; a prolonged course being necessary for treatment. The mode of action is unclear, but the retention in macrophages and conversion to more toxic trivalent derivative are important factors (Shaked-Mishan *et al.*, 2001). Pentamidine is used as an alternative to antimonials, in case the latter are not effective (Amato *et al.*, 1998). But pentamidine resistance has been described, relying on the decrease in mitochondrial membrane potential, which prevents the accumulation of the drug in the mitochondrion and thus reduces its toxicity (Basselin *et al.*, 2002). Amphotericin B, an antibiotic, is also used, where resistance to other drugs occurs (Olliaro and Bryceson, 1993). Many of these treatments present severe side effects, or are too expensive for extensive use in developing countries. Antimonials have a cumulative toxicity and parasites are increasingly becoming resistant to these drugs (Croft *et al.*, 2002). Pentamidine is expensive and more toxic; amphotericin B has lower toxicity but is even more expensive. Vaccines have not been successful so far due to the complexity of the immune response caused by the parasite (Croft and Coombs, 2004). Recently a new compound inhibiting various enzymes of cell signalling pathways, miltefosine, has been registered for use against leishmaniasis (Seifert *et al.*, 2003), but there is a need for new drug discovery and validation of new treatments.

1.4. Aims of this study

The overall aim of this project was to identify and characterise cysteine peptidases as possible drug targets in the parasitic protozoa *Eimeria* and *Leishmania*.

This project was divided into 2 parts:

- Identification of cathepsin B-like peptidases (clan CA, family C1) in *Eimeria tenella*, production of active recombinant enzyme for biochemical characterisation and drug screening, and production of specific antibodies to study the localisation of the enzyme in the parasite.
- Identification and characterisation of genes encoding a pyroglutamyl peptidase I (cysteine peptidase of the clan CF, never described before in parasites) in *Leishmania major*, production of active recombinant enzyme for biochemical characterisation, and study of its role by gene replacement and over-expression in the parasite.

Chapter 2

Material and Methods

2.1. Parasites

2.1.1. *Eimeria tenella* sporozoite preparation

A suspension of sporulated oocysts of *E. tenella* (H strain) at a concentration of 10^7 cells/ml in 5% (w/v) sodium hypochlorite, 1% (w/v) amphotericin B was kindly provided by Intervet, Schwabenheim, Germany. About 10^8 sporulated oocysts were washed in phosphate-buffered saline (PBS) (20 mM sodium phosphate buffer, 150 mM sodium chloride, pH 7.4), with 3 min centrifugation at 13000 *g* at room temperature to sediment them between each wash. The oocysts were resuspended in 5 ml PBS and 5 ml 3 mm sterile glass beads (Sigma, Poole) were added. The sample was vortexed for a few minutes, checking regularly by phase contrast microscopy for release of sporocysts. The sporocysts suspended in the liquid phase were transferred into a fresh tube, the beads were washed with PBS and the liquid phase containing sporocysts was added to the first batch of sporocysts. The combined sporocyst sample was centrifuged at 800 *g* for 5 min at room temperature. The cloudy supernatant was removed and pelleted sporocysts were resuspended in 1% (w/v) taurodeoxycholic acid containing 0.25% (v/v) trypsin (Sigma, Poole, ref. T4549) and placed in a shaking water bath at 44°C for 1 h. Subsequently, the excystation mixture was filtered through a non-absorbent cotton wool column and the ~95% pure sporozoites were collected in the flow through (contaminated by some cell debris). The sporozoites were sedimented at 3000 *g* for 5 min at room temperature and used immediately or stored at -20°C. The efficiency of excystation of 6 to 12 months old sporulated oocysts was about 50%. The sporulated oocysts used were not older than 12 months. The overall efficiency of sporozoite purification was 30 to 35%.

2.1.2. *Leishmania major* culture

Leishmania major (MHOM/IL/80/Friedlin) promastigotes were cultured in HOMEM medium (Berens *et al.*, 1976) (GibcoBRL, Paisley) with 10% (v/v) heat inactivated foetal calf serum (HIFCS) at 25°C with air as the gas phase. Cultures were inoculated at ~10^5 cells/ ml and cells were sub-passaged when stationary phase was reached (1-2 x 10^7 cells/ml) (after about one week).

L. major metacyclic promastigotes were purified from stationary phase cultures as described by Da Silva and Sacks, 1987. Briefly, promastigotes were pelleted at 1300 *g* for 10 min at 4°C, washed in PBS and resuspended in PBS at 10^8 cells/ml. Peanut agglutinin (Vector Laboratories, Burlingame, CA, US) was added at 50 μg/ml (w/v) and the sample was incubated 25 min at room temperature. The supernatant contained the non-agglutinated metacyclic promastigotes.

L. major amastigotes were purified from infected BALB/c mice by Susan Baillie (University of Glasgow) as described (Hart *et al.*, 1981) and incubated at 25°C in presence of 30 μg/ml gentamicin (Sigma, Poole).

The densities of all cultures were determined using an improved Neubauer haemacytometer (Weber Scientific, Hamilton, NJ, US).

2.1.3. *L. major* harvest and lysis

Parasites were harvested by centrifugation at 1300 *g* for 10 min at 4°C, washed twice in PBS, and stored at -80°C until use. Parasites were lysed by resuspension to a concentration in parasite equivalents of 2 x 10^9 cells/ml in lysis buffer, composed of 50 mM Tris/HCl pH 8.0, 0.25% (v/v) Triton X-100, 20% (v/v) glycerol, and, unless stated otherwise, a cocktail of peptidase inhibitors (10 μM (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl) carbamoyl]3-methylbutyl}carbamoyl) oxirane-2-carboxylic acid [E-64], 2 mM 1,10-phenanthroline, 4 μM pepstatin A, 1 mM phenylmethylsulfonyl fluoride [PMSF]). The mixture was incubated for 10 min on ice and centrifuged at 13000 *g* for 15 min at 4°C. The supernatant (designated as the soluble fraction) was separated from the pellet, and the pellet (designated as the membrane-bound fraction) was resuspended in the same volume as before in lysis buffer. Both samples were either used immediately or stored at -20°C until use.

2.1.4. *L. major* cryo-preservation

Stationary phase *L. major* promastigotes were pelleted at 1300 *g* for 10 min and resuspended at 10^8 cells/ml in HOMEM medium with 10% (v/v) HIFCS. One ml of cell suspension was mixed with 1 ml of HIFCS containing 30% (v/v) glycerol in a cryotube vial (Nunc, Roskilde, Denmark), and placed overnight at -70°C in an isopropyl alcohol bath (for slow decrease of temperature), before transfer to liquid nitrogen (-196°C) for long-term preservation up to several years.

2.1.5. Bioassay for leishmanicidal activity

Promastigote viability was measured using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay as described in Silva *et al.*, 2000. The assay is based on the loss of mitochondrial activity associated with cell death, and therefore loss of the capacity of reduction of MTT (yellow) into formazan (blue). *L. major* promastigotes were pelleted at 1300 *g* for 10 min and resuspended at 5.5×10^7 cells/ml in PBS. 5×10^6 *L. major* cells were incubated in a microtiter plate with gomesin and gomesin analogues at various concentrations in a total volume of 100 µl. Gomesin and gomesin analogues were kindly provided by Dr Antonio de Miranda (Sao Paolo, Brazil). The primary amino acid structure of gomesin is ZCRRLCYKQRCVTYCRGR* (using the single letter amino acid code, were Z stands for pyroglutamate and the asterisk indicates an α-amide) (Silva *et al.*, 2000). The gomesin analogues were [Ala^1]-gomesin, where the N-terminal pyroglutamate was replaced by an alanine, and Ac-[Ala^1]-gomesin, where the N-terminal pyroglutamate was replaced by an acetylated alanine. After an incubation of 1 h at 25°C, 100 µl of 2 mg/ml MTT dissolved in PBS was added (final concentration of 1 mg/ml), the mixture was incubated for 1 h at 25°C, and then the absorbance at 620 nm was read using a microtiter plate reader.

2.2. Molecular biology techniques

2.2.1. Isolation of genomic DNA from *E. tenella* and *L. major*

1 x 10^8 *E. tenella* sporozoites or *L. major* promastigotes were pelleted by centrifugation at 1300 *g* for 10 min at 4°C and the cells were resuspended in 150 µl of TELT buffer (50 mM Tris-HCl pH 8.0, 62.5 mM ethylenediamine tetraacetic acid [EDTA], 2.5 M LiCl, 4% (v/v) Triton X-100) (Medina-Acosta and Cross, 1993). The suspension was incubated for 5 min at room temperature, before addition of 150 µl of phenol-chloroform (1:1, v/v). The sample was gently mixed by inversion and centrifuged at 13000 *g* for 5 min at 4°C. The aqueous phase was transferred to a new microfuge tube and 0.1 volume of 3 M sodium acetate and 2 volumes of 100% ethanol were added to precipitate the DNA. The mixture was incubated 5 min on ice before centrifugation at 13000 *g* for 5 min at room temperature. The DNA pellet was washed with 70% (v/v) ethanol, air dried briefly and resuspended in 50 µl of TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). After RNAse H (Sigma, Poole) treatment (1 mg/ml) for 1 h at 37°C, the sample was stored at 4°C.

2.2.2. Isolation of genomic DNA from *Pichia pastoris*

A single colony was used to inoculated 10 ml of YPD medium (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose) and grown overnight at 30°C, 300 rpm shaking. Cells were pelleted at 1500 *g* for 10 min, washed with 10 ml of sterile water and resuspended in 2 ml of SCED solution (1 M sorbitol, 10 mM sodium citrate, pH 7.5, 10 mM EDTA, 10 mM dithiothreitol [DTT]). To disrupt the yeast cell wall, 0.3 mg of lyticase (227 U/mg, Sigma, Poole) was added and incubated without shaking at 37°C for 30 min. Then 2 ml of 1% (w/v) sodium dodecylsulfate (SDS) were added, and the mixture was incubated for 5 min on ice, before addition of 1.5 ml of 5 M potassium acetate, pH 8.9, and centrifugation at 10000 *g* for 10 min at 4°C. Two volumes of ethanol were added to the supernatant and the mixture was incubated for 15 min at room temperature, before centrifugation at 10000 *g* for 20 min at 4°C. The pellet was resuspended in 0.7 ml of TE buffer and 1 volume of phenol/chloroform (1:1) was added, before centrifugation at 10000 *g* for 10 min at 4°C. One volume of chloroform was added to the aqueous phase, and centrifuged as before. Then 0.5 volume of 7.5 M ammonium acetate, pH 7.5, and 2 volumes of ethanol were added to the aqueous phase, and the mixture was incubated for 10 min on dry ice. The sample was centrifuged at 13000 *g* for 20 min at 4°C, the pellet was washed with 1 ml of 70% (v/v) ethanol, air dried and resuspended in 100 µl of TE buffer. The DNA was stored at 4°C.

2.2.3. Isolation of total RNA

2.2.3.1. *E. tenella*

Total RNA from *E. tenella* sporulated oocysts (H strain) was isolated using the method described in Johnston *et al.*, 1998. The RNA obtained was further DNase treated for 30 min at 37°C using 1 U/µg of RNA isolated (DNase I, RNase-free, Sigma, Poole) before storage at –70°C.

2.2.3.2. *L. major*

1 x 10^8 *L. major* promastigotes were pelleted at 1500 *g* for 10 min at 4°C and resuspended in 1 ml of TRIzol® reagent (GibcoBRL, Paisley). The suspension was incubated at room temperature for 5 min, 0.2 ml of chloroform was added and the suspension was incubated at room temperature for 2 min, before being centrifuged at 13000 *g* for 15 min at 4°C. The upper aqueous phase was transferred to a fresh tube and 0.5 ml of isopropyl alcohol was added. The suspension was mixed vigorously and incubated at room temperature for 10 min, before being centrifuged at 12000 *g* for 10 min at 4°C. The supernatant was discarded and 1 ml of 70% (v/v) ethanol was added to the RNA pellet, mixed and centrifuged at 7500 *g* for 5 min at 4°C. This step was repeated, before air-drying of the pellet, and resuspension in 100 µl of double distilled water (ddH_2O). The RNA obtained was further Dnase-treated for 30 min at 37°C using 1 U/µg of RNA isolated (DNase I, RNase-free, Sigma, Poole) before storage at –70°C. All equipment and reagents were made RNAse-free by treatment with diethylpyrocarbonate (DEPC) at 0.01% (v/v).

2.2.4. Polymerase chain reaction (PCR)

PCR was used to amplify fragments of DNA situated between two known regions. All oligonucleotides (or primers) were synthesised by MWG-Biotech, Ebersberg, Germany. The two primers used in each amplification had sequences complementary to those flanking the region to amplify. The annealing temperature and elongation time for the PCR cycles where optimised for each reaction. A summary of all the primers used, their sequences and melting temperatures (Tm) is presented in Table 2.1. The machine used was the GeneAmp PCR system 2400 (Perkin Elmer, Beaconsfield). If Tm was high, the maximum temperature used for annealing was 65°C.

For reactions using the *Taq* polymerase system (Promega, Southampton), the final concentrations of the PCR reagents were 1X Thermophilic DNA buffer (provided with the enzyme), 1.5 mM $MgCl_2$, 0.5 mM dNTPs mix, 100 pmol of each primer, 50-100 ng of cDNA or gDNA, 1 unit of *Taq*, in a total volume of 50 μl. PCR cycles were: 94°C for 2 min, 1 cycle; 94°C for 1 min, Tm-5°C for 1 min, 72°C for 1 x (number of kb of fragment to amplify) min, 30 cycles; 72°C for 7 min.

For reactions using the High Fidelity PCR system (Roche, Lewes), the final concentrations of the PCR reagents were 1X PCR buffer 2 (with 1.5 mM $MgCl_2$ final concentration, provided by the manufacturer), 0.5 mM dNTPs mix, 100 pmol of each primer, 50-100 ng of cDNA or gDNA, 2.5 units of High Fidelity enzyme, in a total volume of 100 μl. PCR cycles were: 80°C for 5 min, 1 cycle; 94°C for 2 min, 1 cycle; 94°C for 15 s, Tm-5°C for 30 s, 72°C for 1.5 x (number of kb of fragment to amplify) min, 30 cycles; 72°C for 7 min.

For reactions using the Long Template PCR system (Roche, Lewes), the final concentrations of the PCR reagents were 1X PCR buffer 2 (with 2.75 mM $MgCl_2$ final concentration), 0.5 mM dNTPs mix, 100 pmol of each primer, 50-100 ng of gDNA, 3.75 units of polymerase enzyme mix, in a total volume of 50 μl. PCR cycles were: 94°C for 2 min, 1 cycle; 94°C for 10 s, Tm-5°C for 30 s, 68°C for 1 x (number of kb of fragment to amplify) min, 25 cycles; 68°C for 7 min.

Table 2.1. **Summary of the primers used, their sequences and Tm (melting temperature in °C).** A, adenine; T, thymine; C, cytosine; G, guanine; I, inosine.

RACE Primers	Sequence (5'-3')	Tm
CatBGSP	AGCAGGTACGGCAGCAACTC	61.4
CatBGSP1	TGCCGACGGAAGTGATCCCGC	65.7
CatBGSP2	ACAACCTCATGGCCTCCTGG	61.4
AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	>75
AUAP	GGCCACGCGTCGACTAGTAC	54.7
SL	TAACGCTATATAAGTATCAGTTTC	65
PPrace1	GGGTGGCGCTTCAGGCCCTCCG	71.4
PPrace2	GTGATGAACACGACGATGCCCG	64
QCrace1	GCAGGCTCGAGTGGGGTGTTG	65.7
QCrace2	GATGAGACAGAGCAGGATGCCCATG	66.3
PCR Primers	**Sequence (5'-3')**	**Tm**
pETNdeCatBfor	GTTGTCATATGCCCTCCGATGATTTGGG	66.6
PETXhoCatBrev	GCCTCTCGAGTTATAGGTCCTGCGCTGACGGCAG	>75
pETCatB31for	AAGGCATATGGGTACTTACCTTTCTTTCTATTCAG	66
pETCatB31rev	GCCTCTCGAGTCATAGGTCCTGCGCTGACG	73.6
pGAPCatBfor	ATCACAGCAGCCTCGAGAAGCGGGAGGCCGAAGCTATGCCCTCCGATG	>75
pGAPCatBrev	GGTGCCGCGGTAGGTCCTGCGCTGACGG	>75
pGAPCatBrev2	GGTGCCGCGGTTATAGGTCCTGCGCTGACGG	>75
pPIC9CatBfor	GCGGCTACGTAATGCCCTCCGATGATTTGGG	72.1
pPIC9CatBrev	GGTGCCTAGGTTATAGGTCCTGCGCTGACGG	73.5
3'AOX1	GCAAATGGCATTCTGACATCC	57.9
5'AOX1	GACTGGTTCCAATTGACAAGC	57.9
α-factor	TACTATTGCCAGCATTGCTGC	57.9
LmPPfor	CGCTTCATATGCTGCACTCCAAGGCG	68
LmPPrev	CGATGGTCGACTCACATGGTCATCAAAG	66.6
OEforLmPP	GGAGAGATATCATGGGCAGCCATCATCATCATC	71.9
OErevLmPP	GCTTGGATCCTCACATGGTCATCAAAGACAGCAGG	71.7
Hind5'for	CTGCTAAGCTTGCTGCTACTCGCCTACTTGCCTCG	74.2
Sal5'rev	AGTGGTCGACTTTAAGCGGTGCGGAACTGCGAGG	74.3
Sma3'for	GACCCCCGGGCGAAGACATCGTTCGCAGCGAAGG	>75
Bgl3'rev	CTCCAGATCTTACCCGTTTCACGTTACACGGCTG	71.9
CheckKOPPFor	CTGCGGCAGCTTGTCCTGCTACTAGTG	69.5
CheckKOPPRev	GCACCGCTCCTTCGCTGCGAACG	69.6
QCfor	TCTGCATATGCTCACCTGGAAGAGCAACC	68.1
QCrev	TTGTAAAGCTTTCATTCTTTACGCGTCCAG	64
QCshortfor	GCGGCATATGCTGAAGCCGAGTCTGTCCGCG	74.8
Mutagenesis Primers	**Sequence (5'-3')**	**Tm**
mutCatBXafor	GGTGTTTGCGGAGATCGAGGGGCGTCTGGAAACTGACAAG	>75
mutCatBXarev	CTTGTCAGTTTCCAGACGCCCCTCGATCTCCGCAAACACC	>75
mutCatBTEVfor	CTGCCTGTGAAGGAGATTGTGGACTCGCAGCGGGTTCTGGAAACTG	>75
mutCatBTEVrev	CAGTTTCCAGAACCCGCTGCGAGTCCACAATCTCCTTCACAGGCAG	>75
mutCPPfor	CAGGGCGCTACCTGGCCAACTGCGCCCTC	>75
mutCPPrev	GAGGGCGCAGTTGGCCAGGTAGCGCCCTG	>75
mutHPPfor	GGCATCTTCGTTAGCGTTGTCGATCCG	68
mutHPPrev	CGGATCGACAACGCTAACGAAGATGCC	68
mutE1PPfor	GGCTCATCTGCGTTCAGGTGCAGGGCTGC	73.7
mutE1PPrev	GCAGCCCTGCACCTGAACGCAGATGAGCC	73.7
mutE2PPfor	CAGGGCTGCAACCAGCTCTTTTCTAGCG	69.5
mutE2PPrev	CGCTAGAAAAGAGCTGGTTGCAGCCCTG	69.5

2.2.5. Reverse transcription (RT)-PCR

2.2.5.1. cDNA synthesis

Five µg of total RNA from *E. tenella* sporozoites or *L. major* promastigotes were added to 10 pmol of oligo-dT primer in a total volume of 11 µl. The mix was incubated for 10 min at 70°C and placed on ice for 1 min. Four µl of 5X First Strand buffer (provided with the enzyme), 2 µl of 0.1 M DTT, 1 µl of 10 mM dNTP mix and 40 units RNase OUT (GibcoBRL, Paisley) were added and the mix was equilibrated at 42°C for 5 min before addition of 200 units of SUPERScript II (GibcoBRL, Paisley). The sample was incubated for 50 min at 42°C, then 15 min at 70°C to terminate the reaction. After incubation on ice for a few minutes, 2 units of RNase H (Sigma, Poole) were added and the mix was incubated for 20 min at 37°C to eliminate the template RNA, before storage at -20°C. As a negative control and to rule out that any subsequent PCR may produce fragments from amplification on remaining gDNA, a sample of RNA was treated as described above, except that the addition of SUPERScript II enzyme was omitted. This sample was used as template for negative control PCRs on cDNA.

2.2.5.2. Rapid amplification of cDNA ends (RACE)

2.2.5.2.1. Gene encoding a cathepsin B-like enzyme in *E. tenella*

To confirm the chosen start methionine, an experiment of rapid amplification of cDNA ends was performed at the 5' end (5' RACE system kit, GibcoBRL, Paisley). Total mRNA from sporozoites was transcribed into single-stranded cDNA using SUPERScript II reverse transcriptase and the specific primer CatBGSP. Excess dNTPs and primer were removed from cDNA and a homopolymeric tail of dCs was added to the end of the single-stranded cDNA using terminal deoxynucleotidyl transferase, following the manufacturer's instructions. The 5' end was then amplified from dC-tailed cDNA using the abridged anchor primer (AAP) and the specific nested primer CatBGSP1 with the *Taq* polymerase, followed by a secondary amplification using the abridged universal anchor primer (AUAP) and the catBGSP2 primer with the *Taq* polymerase.

2.2.5.2.2. Gene encoding a pyroglutamyl peptidase I-like enzyme in *L. major*

PCR on total cDNA was performed using a primer with a sequence complementary to the splice leader (SL primer) and the PPrace1 primer, with the *Taq* polymerase. PCR product was diluted 1:50 (v/v) in ddH_2O and 2 µl were used as a template for a semi-nested PCR using the SL and the PPrace2 primers, with the *Taq* polymerase.

2.2.5.2.3. Gene encoding a glutaminyl cyclase-like enzyme in *L. major*

The same method as in 2.2.5.2.2. was used, with the SL and QCrace1 primers for the first PCR, and the SL and QCrace2 primers for the second PCR.

2.2.6. Site-directed mutagenesis

Site-directed mutagenesis was used to introduce site-specific mutation in the gene encoding the cathepsin B-like enzyme of *E. tenella* in order to introduce Factor Xa (with the primers mutCatBXafor and mutCatBXarev, 7 point-mutations required) or Tabacco etch virus (TEV) protease NIa (with the primers mutCatBTEVfor and mutCatBTEVrev, 12 point-mutations required) cutting sites into the translated protein. The pET28a(+) construct with the sequence for the N-terminal His-tagged *E. tenella* cathepsin B was used as a template.

The method was also used for generating active site mutants of the pyroglutamyl peptidase I (PPI) of *L. major*. Based on the work of Le Saux and co-workers (Le Saux *et al.*, 1996), the choice was made to mutate cysteine-210 to alanine (using the primers mutCPPfor and mutCPPrev), histidine-234 to serine (using the primers mutHPPfor and mutHPPrev), glutamic acid-101 to glutamine (using the primers mutE1PPfor and mutE1PPrev) and glutamic acid-107 to glutamine (using the primers mutE2PPfor and mutE2PPrev) (*L. major* PPI numbering). The most likely glutamic acid residue to be part of the catalytic triad was Glu-101, but the very close Glu-107 was mutated too. One or 2 point-mutations were required, and the pET28a(+) construct with the sequence for the N-terminal His-tagged *L. major* PPI was used as a template.

The QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, US) was used. The reactions contained 1X reaction buffer (provided with the kit), 100 ng of plasmid DNA (expression vector carrying the gene to mutate), 2 μM of each forward and reverse primer, 1 μl of dNTP mix, 2.5 units of *Pfu Turbo* DNA polymerase. Cycles were: 95°C for 30 s, 1 cycle; 95°C for 30 s, 55°C for 1 min, 68°C for 7 min, 18 cycles. Then 10 units of *DpnI* enzyme (provided with the kit) were added and incubated at 37°C for 2 h, to remove any residual methylated template. Fifty μl of *E. coli* XL-1 Blue Supercompetent cells (provided with the kit) were transformed with 1 μl of PCR using the heat shock method (see 2.2.7.5.1.) and 250 μl were plated on Luria-Bertani (LB) agar (1% tryptone (w/v), 0.5% yeast extract (w/v), 1% NaCl (w/v), 20% (w/v) agar) plates. The plates were incubated overnight at 37°C. Plasmid was extracted from colonies, sequenced and re-transformed in *E. coli* BL21(DE3) for expression.

2.2.7. DNA fragment cloning

Cloning of DNA fragments was done using 2 different strategies. The first one involved digestion by restriction enzymes of both DNA insert and vector recipient prior to ligation of the 2 fragments together. The second involved cloning of PCR products by amplification with the *Taq* polymerase. *Taq* polymerase adds a single A-nucleotide to the 3' ends of the PCR product, which can then be inserted in the pGEM-T Easy vector (Promega, Southampton), containing a 3' single T-nucleotide overhang.

2.2.7.1. Digestion of DNA with restriction enzymes

Typically, restriction digests used 3 units of restriction enzyme in a total volume of 20 μl per 500 ng of DNA to digest, and reactions contained 1X buffer (final concentration). The buffer used in each case was provided with the enzyme by the manufacturer (Promega, Southampton). For multiple digests, compatible buffer systems were chosen. The reactions were incubated 1-4 h at 37°C or 25°C (depending on the enzyme).

2.2.7.2. Isolation of DNA fragments

PCR products or DNA fragments obtained by restriction digests were separated on 1% (w/v) agarose gels containing 0.5 μg/ml of ethidium bromide, in 0.5X TBE (45 mM Tris-borate, 1 mM EDTA). The DNA ladder 1 kb Plus (Promega, Southampton) was used. DNA was visualized using a transilluminator (UVP Laboratory Products, Cambridge) at 312 nm. Fragments of DNA were purified from the gel using the QIAquick gel extraction kit (Qiagen, Crawley).

2.2.7.3. Ligations

Ligations of fragments into the pGEM-T Easy vector (Promega, Southampton) used 50 ng of vector, 150 ng of insert, 1X buffer and 3 units of T4 DNA ligase. The reactions were incubated 16 h at 4°C. A positive control (using the control insert DNA) and a background control (vector without insert DNA) were used.

Ligations of gel-purified digested DNA fragments and expression vector, or *L. major* transfection vectors, used 500 ng of insert, 250 ng of vector, 1X buffer and 3 U of T4 DNA ligase (Promega, Southampton). Reactions were incubated for 16 h at 16°C. The only *E. coli* expression vector used was the pET28a(+) vector (Novagen, Nottingham) and *L. major* transfection vectors were derivatives of the pXG vector, developed by Beverley and co-workers (Ha *et al.*, 1996).

2.2.7.4. Competent cells

2.2.7.4.1. *E. coli*

E. coli DH5α competent cells were from Promega, Southampton. BL21(DE3), JM109(DE3) and HMS174(DE3) cells were made competent the day of use using the calcium chloride method described in Sambrook *et al.*, 1989. A single colony was picked from LB agar plate and grown overnight at 37°C in 5 ml of LB broth. This culture was used to inoculate 100 ml of LB, and the cells were grown at 37°C until the cell density reached an optical density (OD) at 600 nm of 0.4-0.6. Cells were then pelleted at 2500 *g* for 10 min at 4°C, resuspended in 10 ml of ice-cold sterile 0.1 M $CaCl_2$, pelleted again as before and resuspended in 2 ml ice-cold 0.1 M $CaCl_2$. Cells were stored on ice until use.

2.2.7.4.2. *Pichia pastoris*

P. pastoris X-33 and KM71 cells were made electro-competent the day of use using the sorbitol method described in the Invitrogen (Paisley) "*Pichia* Expression Kit" manual (Version F, catalog no. K1710-01), with some modifications. A single colony was used to inoculate 5 ml of YPD medium and grown overnight at 30°C. One ml of this culture was used to inoculate 500 ml of YPD and grown to OD_{600} of 1.3-1.5. Cells were centrifuged at 1500 *g* for 5 min at 4°C, resuspended in 100 ml YPD medium, 20 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid (HEPES) pH 8, and 2.5 ml of 1 M DTT was added. The mixture was incubated at 30°C for 15 min without shaking. The suspension was brought to 500 ml with ice-cold sterile water. The cells were sedimented as before, resuspended in 250 ml ice-cold sterile water, sedimented again as before, resuspended in 20 ml ice-cold 1 M sorbitol, sedimented as before and resuspended in 1 ml of ice-cold 1 M sorbitol. The cells were stored at 4°C until use.

2.2.7.5. Transformation of competent cells

2.2.7.5.1. *E. coli*

E coli cells were transformed with the ligation products using a heat shock method (Sambrook *et al.*, 1989). Ten ng of DNA was incubated with 200 µl of competent cell suspension on ice for 20 min, heat shocked for 50 s at 42°C, and incubated on ice again for 2 min. One ml of LB was added and the mixture was incubated at 37°C for 1 h before plating of 100 µl on a LB agar plate containing the appropriate antibiotic for selection. In the case of pGEM-T Easy ligation product transformation in DH5α, a white/blue selection was used, by adding isopropylthio-β-D-galactoside (IPTG) (40 µg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (40 µg/ml). White colonies correspond to positive clones containing a DNA insert in the β-galactosidase gene, which leads to the expression of an inactive β-galactosidase not able to act on the chromogenic substrate X-gal.

2.2.7.5.2. *Pichia pastoris*

pPIC9 and plasmid derivatives were linearized with *Sac*I, and pGAPZαA and plasmid derivatives were linearized with *Avr*II. Linearized DNA was transformed by electroporation as described in the Invitrogen (Paisley) "*Pichia* Expression Kit" manual (Version F, catalog no. K1710-01) using the Bio-Rad (Hemel Hempstead) GenePulser at a charging voltage of 1500 V, capacitance of 25 μF and resistance of 200 Ω. Eighty μl of competent cell suspension were incubated with 10 μg of linearized DNA in an ice-cold 0.2 cm electroporation cuvette (BioRad, Hemel Hempstead) for 5 min on ice. Reactions were pulsed once and 1 ml of ice-cold 1 M sorbitol was immediately added to the reaction. The reaction was transferred to a new tube and either 300 μl were directly plated on RBD plates (1 M sorbitol, 2% (w/v) glucose, 1.34% (w/v) Yeast Nitrogen Base with Ammonium Sulfate without amino acids [GibcoBRL, Paisley], 4 x 10^{-5}% (w/v) biotin, 0.005% (w/v) amino acids (L-glutamic acid, L-methionine, L-lysine, L-leucine, L-isoleucine [Sigma, Poole]), 20% (w/v) agar), in the case of transformation of the KM71 cells with the pPIC9 and plasmid derivatives, or the reaction was incubated for 2 h at 30°C before plating of 300 μl on YPB plates containing 100 μg/ml of Zeocin, in case of transformation of the X-33 cells with the pGAPZαA and plasmid derivatives. Plates were incubated at 30°C for 3-4 days.

2.2.7.6. Plasmid DNA purification

The isolation of plasmid DNA from *E. coli* was performed using the QIAprep Miniprep kit (Qiagen, Crawley). The presence of the correct insert was checked by restriction digests, which were separated by agarose gel electrophoresis.

2.2.7.7. DNA sequencing

Plasmid DNA was sent to GRI Genomics (Braintree, UK) for sequencing. Sequences were analysed using the Vector NTI program (Informax, Bethesda, MA, US).

2.2.8. *L. major* promastigotes transfection

2.2.8.1. Strategy for *L. major* pyroglutamyl peptidase I (PPI) gene cloning for over-expression and knockout

For over-expression of the native and cysteine-210-mutated PPIs in *L. major*, the vector pGL102 was chosen (derivative of pXG vector [Cruz *et al.*, 1991a]). The restriction sites *Sma*I (5' end) and *Bam*HI (3' end) were chosen on the vector. As another *Sma*I site exists in the gene, the engineered site was the one for *Eco*RV, cohesive end, which will religate into the *Sma*I site on the vector after digestion. The restriction sites were added to the ends of the fragment by PCR on ~100 ng of either the pETLmPP plasmid, or the pET28a(+) construct with the sequence for the cysteine-210-mutated N-terminal His-tagged *L. major* PPI, using the primers OEforLmPP and OErevLmPP.

The 834 bp excised bands were gel-purified, digested with *Eco*RV and *Bam*HI, re-gel purified and cloned into pGL102 previously digested with *Sma*I and *Bam*HI to give the final plasmids pGL102LmPPI and pGL102LmPPI*. The sequence was confirmed using the PCR primers for sequencing.

For gene knockout, the vector pGL345 was chosen (derivative of pXG vector, hygromycin selection). Flanks at the 5' (~ 500 bp) and 3' (~ 1 kb) ends of the PPI gene were selected in the *L. major* database. Restriction sites for *Hin*dIII (5' end) and *Sal*I (3' end), or *Sma*I (5' end) and *Bgl*II (3' end) were, respectively, introduced by PCR at the ends of the 5' flank and the 3' flank. PCR used the primers Hind5'for and Sal5'rev for the 5' flank, and the primers Sma3'for and Bgl3'rev for the 3' flank.

The excised 5' flank band was gel-purified, digested with *Hin*dIII and *Sal*I, re-purified and cloned into pGL345 previously digested with the same enzymes. The excised 3' flank band was gel-purified, digested with *Sma*I and *Bgl*II, re-gel purified and sub-cloned in the pGL345 plasmid containing the 5' flank previously digested with the same enzymes to give the pGLLmPPKO. The hygromycin selection marker was replaced by the blasticidin resistance marker (cut from the pGL842 vector) on the plasmid by digestion with *Spe*I and *Bam*HI.

2.2.8.2. DNA preparation

In the case of episomal constructs for over-expression, 20 μg of circular plasmid was ethanol-precipitated (Sambrook *et al.*, 1989), resuspended in 20 μl of sterile water, and stored at -20°C.

In the case of integrative constructs for gene replacement, 40 μg of plasmid was linearised by digestion with *Hin*dIII and *Bgl*II. The digested DNA fragments were separated on ethidium bromide-free agarose gel, fragments were revealed by gel immersion in a 6.5 x 10^{-4}% (w/v) methylene blue solution, gel extracted with the QIAquick gel extraction kit (Qiagen, Crawley), ethanol precipitated (Sambrook *et al.*, 1989), resuspended in 20 μl of sterile water, and stored at -20°C.

2.2.8.3. Transfection

L. major promastigotes from log-phase cultures at a density of about 0.5-1 x 10^7 cells/ml were preferentially used. About 10^8 cells were used per transfection. The cells were pelleted at 1300 *g* for 10 min at 4°C, washed twice in sterile ice-cold electroporation buffer (120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4, 25 mM HEPES, 2 mM EDTA, and 5 mM $MgCl_2$, pH 7.6), and resuspended in electroporation buffer to a concentration of about 2 x 10^8 cells/ml. Cells (0.5 ml per transfection) were put in contact with the DNA to transfect (20 μg) and incubated on ice for 15 min. A no DNA control, to check efficiency of drug selection, and empty vector control to compare phenotypes (for episomal expression only), were always included. The mixture was placed into an ice-cold 4 mm Gene Pulse cuvette (BioRad, Hemel Hempstead) and pulsed twice at 25 μF, 1500 V, (3.75 kV/cm) on the BioRad Gene Pulser II apparatus. At least 10 s separated the 2 pulses. The cells were then incubated on ice for 20 min before being transferred to 5 ml HOMEM medium with 10% HIFCS and incubated at 25°C overnight. The next day, the cells were placed in medium containing the appropriate selection drug (depending on the construct transfected). All selection drugs were from Calbiochem, Nottingham. Hygromycin B was used at a final concentration of 50 μg/ml; blasticidin S hydrochloride at 15 μg/ml; neomycin (G418) at 50 μg/ml; and puromycin at 40 μg/ml. Clonal dilution (from 1:2 (v/v) to 1:64 (v/v), serial dilution) was performed on transfected cells in 96-well plates, in order to isolate clonal populations. Genomic DNA was extracted from drug-resistant clones and correct integration was confirmed using the CheckKOPPFor and CheckKOPPRev primers.

2.2.9. Southern-blot analysis

2.2.9.1. Southern blotting of DNA fragments

Southern blotting allows detection of specific sequences within the genomic DNA. Between 2 and 10 μg of gDNA were digested to completion using specific restriction enzymes, and the DNA fragments were separated on a 1% (w/v) agarose gel in 1x TBE buffer. The DNA was blotted according to the Southern method (Sambrook *et al.*, 1989). After electrophoresis for 8 h at 50 V, the gel was incubated in 125 mM HCl for 30 min under gentle agitation to allow the depurination of the DNA. The gel was then incubated for 30 min in 1.5 M NaCl, 0.5 M NaOH (for DNA denaturation), and 30 min in 1 M Tris pH 7.5, 1.5 M NaCl (for neutralisation). Each wash was interspersed by a 10 min wash in ddH_2O. The DNA was transferred by capillarity onto a Hybond-N+ nylon membrane (Amersham) in 20X SSC (300 mM tri-sodium citrate, 3 M NaCl, pH 7) using standard methods (Sambrook *et al.*, 1989). The DNA was then fixed onto the membrane by ultraviolet (UV) cross-linking (Spectrolinker XL-1000 UV linker, Spectronics Corporation, Westbury, NY, US).

2.2.9.2. Southern hybridisation with nucleic acid probes

Nylon membrane was pre-incubated at 65°C for 4 h in 20 ml of Church-Gilbert hybridisation solution (340 mM Na_2HPO_4, 158 mM NaH_2PO_4, 240 mM SDS, 1 mM EDTA), supplemented with 0.2 mg/ml of denatured salmon sperm DNA (GibcoBRL, Paisley). Thirty ng of nucleic acid probe was prepared from agarose gel-purified restriction endonuclease fragments using the Prime-It II

Random Primer kit (Stratagene, La Jolla, CA, US) according to the manufacturer's instruction. Labeled A nucleotides were introduced using 50 μCi of α^{32}dATP (Perkin Elmer, Beaconsfield) per probe, and the dATP buffer from the kit. Probe was purified on a Microspin S-200 HR column (Amersham, Chalfont St.Giles) and boiled for 5 min before incubation on ice for 2 min. The membrane was hybridised by incubation with the labeled probe in 20 ml of Church-Gilbert solution at 65°C overnight. Membranes were then washed under high stringency 2 times for 5 min in 2X SSC, 0.1% (w/v) SDS, followed by 15 min in 1X SSC, 0.1% SDS, and 2 times for 10 min in 0.1X SSC, 0.1% SDS. All wash solutions were pre-warmed at 65°C. Membranes were sealed in polythene and exposed to X-ray film (Konica Medical Film) for 1 to 10 days. Hybridisation signals were detected with a film processor (X-Ograph imaging system Compact X4).

2.3. Biochemical methods

2.3.1. Recombinant protein expression

2.3.1.1. In *E. coli*

For all expression experiments in *E. coli* the pET28a(+) vector (Invitrogen, Pailsey) was used.

For expression in *E.coli* of the pro-mature cathepsin B of *E. tenella* with a N-terminal His-tag, the cloning was performed between the *Nde*I site and the *Xho*I site. The recognition sequences for these restriction enzymes were firstly added to the sequence by PCR, with the primers pETNdeCatBfor and pETXhoCatBrev. The full-length product was cloned into pGEMT-easy and the resulting plasmid was cut with *Nde*I and *Eco*RI, and *Hin*dIII and *Xho*I. The *NdeI/Eco*RI band was gel purified and subcloned into pET28a(+). The resulting plasmid was cut with *Hin*dIII and *Xho*I and the *Hin*dIII/*Xho*I band was inserted. This construct was transformed in *E. coli* strains BL21(DE3), JM109(DE3), or HMS174(DE3) competent cells for expression. Cells were grown to an OD_{600nm} of 0.4 to 0.6 in LB containing 50 μg/ml kanamycin (Sigma, Poole) before being induced for 4 h with 0.5-1.5 mM IPTG at 25 or 37°C.

For expression in *E. coli* of the truncated pro-mature cathepsin B of *E. tenella* with a N-terminal His-tag, the cloning was performed between the *Nde*I and the *Xho*I sites. The procedure was the same as above. The primers used were pETCatB31for and pETCatB31rev. The protein was expressed in the BL21(DE3) cells, by induction 4 h with 1 mM IPTG at 37°C.

For expression in *E.coli* of the pro-mature cathepsin B of *E. tenella* with a N-terminal His-tag and carrying the mutation for the Factor Xa and the TEV protease cleavage sites in the protein, proteins were expressed in the BL21(DE3) cells, by induction 4 h with 1 mM IPTG at 37°C.

For expression in *E. coli* of the native pyroglutamyl peptidase I of *L. major* with a N-terminal His-tag, the cloning was performed between the *Nde*I and the *Sal*I sites. The procedure was the same as before. The primers used were LmPPfor and LmPPrev. The protein was expressed in the BL21(DE3) cells, by induction 5 h with 1 mM IPTG at 20°C.

For expression in *E. coli* of the active site mutated pyroglutamyl peptidase I of *L. major* with a N-terminal His-tag, proteins were expressed in the BL21(DE3) cells, by induction 5 h with 1 mM IPTG at 20°C.

For expression in *E. coli* of the native and truncated glutaminyl cyclase (QC) of *L. major* with a N-terminal His-tag, the cloning was performed between the *Nde*I and the *Hin*dIII sites. The procedure was the same as before. The primers used for cloning of the full length QC were QCfor and QCrev. The primers used for cloning of the truncated QC were QCshortfor and QCrev The protein was expressed in the BL21(DE3) cells, by induction 4 h with 1 mM IPTG at 15, 25, or 37°C.

2.3.1.2. In *Pichia pastoris*

2.3.1.2.1. Constitutive system

The pGAPZαA vector (Invitrogen, Paisley) was chosen for constitutive expression. Constructs were designed for constitutive expression of the pro-mature cathepsin B with a C-terminal His-tag (using the primers pGAPCatBfor and pGAPCatBrev), and the pro-mature cathepsin B without a His-tag (using the primers pGAPCatBfor and pGAPCatBrev2), with subsequent release into the medium of the culture. The DNA fragments were cloned between the *Xho*I and *Sac*II sites. The final constructs and control empty vector were linearised with *Avr*II prior to transfection into the host *Pichia pastoris* cells X-33. Cells were spread on plates selective for Zeocin resistance (100 μg/ml) and incubated 3 to 4 days at 30°C until colonies developed, as described in the *Pichia* expression kit manual (Invitrogen, Paisley). Insertions of the fragments were verified by PCR on genomic DNA extracted from positive clones as described in the *Pichia* expression kit manual, using the couples of PCR primers α-factor and 3'AOX1.

One positive colony for each of the constructs was selected and tested for expression of the target protein as described in the manufacturer's manual (final volume of 50 ml of YPD medium). One ml aliquots were taken a regular time intervals, cells were separated from the supernatant of culture by centrifugation and samples of supernatant were used for analysis using gelatin SDS-PAGE, the azocasein assay, and either directly on Coomassie blue-stained SDS-PAGE gels and western-blots, or after tricholoacetic acid (TCA) precipitation on Coomassie blue-stained SDS-PAGE gels and western-blots. The cells were resuspended in 1 ml of 50 mM Tris pH 8.0 and analysed on Commassie blue-stained SDS-PAGE gels. Once the optimum time of growth had been determined, the cells were directly grown to that time point and harvested by centrifugation at 3000 *g* for 5 min, and the supernatant and cells were stored at -20°C until further treatment and analysis.

2.3.1.2.2. Inducible system

A construct was designed for inducible expression of the pro-mature enzyme cathepsin B-like enzyme of *E. tenella*, with subsequent release into the culture medium. The pPIC9 vector (Invitrogen, Paisley) was used, and the DNA fragment was cloned between the *Sna*BI and *Avr*II sites. The primers used for PCR were pPIC9CatBfor and pPIC9CatBrev. The final construct pPIC9catB and control empty vector were linearised with *Sac*I prior to transfection in the host *Pichia pastoris* cells KM71. As the KM71 strain contains a mutation in the endogenous *AOX1* gene (encoding the alcohol oxidase enzyme), only His$^+$ Mut^S phenotypes will be generated by successful recombination (His$^+$ refers to histidine prototrophy, and Mut^S refers to "Methanol utilization slow" caused by the loss of the alcohol oxidase activity encoded by the *AOX1* gene). Cells were spread on plates selective for histidine prototrophy (His$^+$) and incubated 3 to 4 days at 30°C until colonies developed, as described in the *Pichia* expression kit manual (Invitrogen, Paisley). Insertions of the fragments were verified by PCR on genomic DNA extracted from positive clones as described in the *Pichia* expression kit manual, using the couples of PCR primers α-factor and 3'AOX1, or 5'AOX1 and 3'AOX1.

Positive colonies for each of the transfection were selected and tested for expression of the cathepsin B as described in the manufacturer's manual. One positive colony for each of the transfection was selected, and used to inoculate 100 ml of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate pH 6.0, 1.34% (w/v) YNB, 4 x 10^{-5}% biotin (w/v), 1% (v/v) glycerol). The culture was grown at 30°C shaking at 300 rpm, in a 1 l flask, until the culture reached an OD_{600} of 5-10 (24-48 h). The cells were then harvested by centrifugation at 3000 *g* for 5 min and resuspended in 20 ml of BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate pH 6, 1.34% (w/v) YNB, 4 x 10^{-5}% biotin (w/v), 0.5% (v/v) methanol), placed in a 250 ml flask and the culture was grown again at 30°C shaking at 300 rpm. The cells were induced by addition of methanol to 0.5% (v/v, final concentration) every 24 h, 0.5 ml aliquots were taken a regular time intervals and cells were separated from the supernatant of culture by centrifugation and samples of supernatant were analysed using gelatin SDS-PAGE, the azocasein assay, and either directly on Coomassie blue-stained SDS-PAGE gels, or after TCA precipitation on Coomassie blue-stained SDS-PAGE gels and western-blots. Once the optimum time of growth had been determined, cells were directly grown to that time point and harvested by centrifugation at 3000 *g* for 5 min, and the supernatant and cells were stored at -20°C until further treatment and analysis.

2.3.2. Protein defolding, refolding and activation procedures

A summary of the conditions tested for unfolding, refolding, purification and activation of various preparations of recombinant cathepsin B of *E. tenella* is presented in Table 2.2.

The Ni-agarose-purified soluble pro-mature cathepsin B expressed in *E. coli* BL21(DE3) or in HMS174(DE3) was treated as described in Sanderson *et al.*, 2000, for unfolding and re-folding attempts of the enzyme.

The Ni-agarose-purified soluble pro-mature cathepsin B expressed in *E. coli* BL21(DE3) was treated as described in Hellberg *et al.*, 2002, with a few modifications. The sample was diluted 1:10 (v/v) in 8 M urea or 6 M guanidium, 100 mM NaH_2PO_4, 10 mM Tris/HCl pH 6.3 to unfold the protein. A fraction of this sample was then diluted dropwise 1:100 (v/v) in a refolding buffer composed of 50 mM Tris/HCl pH 6.8, 300 mM NaCl, 5% glycerol (v/v), 0.1 mM EDTA, 0.5 mM oxidised glutathione, 3 mM reduced glutathione and the mixture was incubated overnight at 4°C. Another fraction of the sample was directly dialysed against 100 times its volume of 50 mM Tris pH 6.8 overnight at 4°C. The diluted sample in refolding buffer was concentrated with a Centricon device through YM-10 membrane (Amicon, Watford) from 1 l to 3 ml to a final concentration of 100 µg/ml and dialysed over night against 500 ml of 50 mM Tris/HCl, pH 6.8 at 4°C. The 2 samples treated separately with or without refolding by dilution were diluted 1:1 (v/v) in 100 mM Tris/HCl, pH 8.8, 0.08% SDS, 20 mM DTT at 37°C. Aliquots were taken at various time from 0 min to 1 h and subsequently analysed.

The inclusion bodies of pro-mature cathepsin B expressed in *E. coli* HMS174(DE3) were treated as described in Sanderson *et al.*, 2000, and Sijwali *et al.*, 2001.

Samples of Nickel-agarose purified truncated cathepsin B were diluted 1:1 (v/v) in 0.1 M sodium acetate buffer, pH 5.0 or 0.1 M Tris/HCl buffer, pH 8.0 supplemented with 20 mM DTT and incubated at 37°C for 20 min, with or without the inhibitor E64 (final concentration of 300 µM).

A sample of purified truncated cathepsin B was diluted 1:4 (v/v) to a final concentration of about 100 µg/ml in 0.1 M Tris/HCl pH 8.0, 8 M urea (total volume of 500 µl) and incubated 1 h at 37°C, to unfold the protein without rupturing the disulfide bonds. The sample was then dialysed for 2 h at 4°C against 500 ml of 50 mM Tris/HCl, pH 8.0 and stored at -20°C for later analysis.

Table 2.2. **Summary of the conditions tested for unfolding, refolding, purification and activation of various preparations of recombinant cathepsin B of *E. tenella*.** BL21 and HMS174 are *E. coli* strains used for expression.

Sample / *Step of protein preparation*	**Expressed in BL21/soluble**	**Expressed in HMS174 at 25°C/ Soluble fraction**	**Expressed in HMS174 at 37°C/ Solubilisation of inclusion bodies**
Unfolding/ solubilisation of inclusion bodies	6 M guanidium 8 M urea	X	Sanderson *et al.*, 2000 Sijwali *et al.*, 2001 +/- 10 mM DTT
Refolding	Dilution, pH 8.0, EDTA, DTT, GSSG/GSH, 4°C, 12 h	X	pH 8.0
Purification of refolded protein	Nickel-agarose	X	Anion-exchange
Activation (combinations of the listed components)	37°C pH 4.0 to 7.0 5 mM EDTA 10 mM DTT 0.9 M NaCl During 1 to 24 h	37°C pH 4.0 to 7.0 5 mM EDTA 10 mM DTT 0.9 M NaCl During 1 to 6 h	37°C pH 4.0 to 7.0 protein concentration from 50 to 500 μg/ml 5 mM EDTA 10 mM DTT 0.9 M NaCl During 1 to 7 h

2.3.3. Protein processing

2.3.3.1. Non-specific

The idea behind using non-specific peptidases in an attempt to cleave the pro-domain from the mature domain of the cathepsin B of *E. tenella* was that domains within a protein are tightly folded units (Richardson, 1981) and a correctly folded protein will have its loops and hinges most accessible to proteolytic cleavage. Using mild conditions, i.e. low enzyme/substrate (protein that requires cleavage) ratios and short incubation times might make it possible to separate domains by limiting the number of bonds cleaved and hopefully limiting processing to accessible bonds located in these loops between domains.

Three commercially available peptidases were chosen: trypsin-TPCK from bovine pancreas (ref. T1426, 12700 U/mg), α-chymotrypsin-TLCK from bovine pancreas (ref. C3142, 57 U/mg) and subtilisin A from *Bacillus licheniformis* (ref. P5380, 10.6 U/mg) (all Sigma, Poole) for their broad substrate specificity and ease of use. A stock solution of each peptidase at a concentration of 1 mg/ml was prepared in 50 mM Tris/HCl, pH 8.0. The Ni-agarose-purified soluble pro-mature cathepsin B expressed in *E. coli* BL21(DE3) was used as the substrate. To start with, enzyme/substrate ratios were fixed at 1:50 and 1:500 (w/w). Reactions were in 50 mM Tris/HCl, pH 8.0 and samples were incubated at 37°C for 2 min to 1 h. The reactions were terminated by dilution in an equal volume of 2x SDS-PAGE loading buffer (0.1 M Tris/HCl pH 6.8, 15% (v/v) glycerol, 3% (w/v) SDS, 0.8% (v/v) β-mercaptoethanol, 0.0001% (w/v) Pyronin Y). Then ratios were adjusted if necessary (where there was too extensive cleavage) down to 1:5000 (w/w).

2.3.3.2. Specific

A method to achieve specific cleavage at a particular location is to use a specific peptidase that will only cut bonds in this location. As no obvious specific cutting site could be detected within the region thought to make the link between pro- and mature domains of the cathepsin B of *E. tenella*, it was decided to engineer the specific cleavage sites for the Factor Xa and the TEV protease at that junction.

Biotinylated Factor Xa from the Factor Xa Cleavage Capture Kit (Novagen, Nottingham), and recombinant TEV protease (Invitrogen, Paisley) were used.

Digestions with Factor Xa used 10 to 50 μg of nickel-agarose purified cathepsin B of *E. tenella* produced in *E. coli* and carrying the Factor Xa cleavage site, 5 μl of 10X cleavage buffer, 1, 0.2, or 0.01 U of Factor Xa in a final volume of 50 μl at 20 °C for various lengths of incubation. Non-specific cleavage on the purified pro-mature cathepsin B of *E. tenella* produced in *E. coli* not carrying the mutations for the Factor Xa cleavage site was also tested.

Twenty μg of nickel-agarose purified cathepsin B of *E. tenella* produced in *E. coli* and carrying the TEV cleavage site was incubated with 7.5 μl 20X cleavage buffer, 1.5 μl 0.1 M DTT, 10 U TEV in a final volume of 150 μl at 20°C for various lengths of incubation. Non-specific cleavage on the purified pro-mature cathepsin B of *E. tenella* produced in *E. coli* not carrying the mutations for the TEV cleavage site was also tested.

2.3.4. Determination of protein concentrations

Protein concentrations were either determined spectrophotometrically at 280 nm (provided the extinction coefficient ε of the protein was known), or using the BioRad protein assay, based on the Bradford method (Bradford, 1976) and using bovine serum albumin (BSA) as the protein standard.

2.3.5. Protein precipitation

To reduce quickly the volume of protein solution to analyse, one of the methods used was trichloroacetic acid (TCA) precipitation. This was used in particular prior to SDS-PAGE analysis. The disadvantage is that the protein will be irreversibly denatured. TCA was added to the protein sample at 10% (v/v) final concentration, and the mixture was incubated overnight at -20°C. The sample was then centrifuged at 13000 *g* for 20 min at 4°C, the pellet was washed in 1 ml 100% ice-cold acetone, and resuspended usually in 10-30 μl 2x SDS-PAGE loading buffer (0.1 M Tris/HCl pH 6.8, 15% (v/v) glycerol, 3% (w/v) SDS, 0.8% (v/v) β-mercaptoethanol, 0.0001% (w/v) Pyronin Y).

2.3.6. Protein concentration

Devices used for protein concentration were YM-10 Centricon Centrifugal devices (Amicon, Watford), following the manufacturer's instructions.

2.3.7. Protein purification

2.3.7.1. Nickel-agarose chromatography

Soluble, or solubilised, recombinant N-terminal His-tagged cathepsin B-like proteins of *E. tenella* produced in *E. coli*, as well as soluble recombinant N-terminal His-tagged PPI and PPI active site mutants of *L. major* produced in *E. coli*, were purified by gravity on a 3 ml nickel-agarose column (Qiagen, Crawley) following the manufacturer's instructions. The column was pre-equilibrated with 5 volumes of 10 mM Tris/HCl, pH 8.0, the sample was applied, and the column was washed with 5 volumes of the same buffer, followed by a washing with 5 volumes of 10 mM Tris/HCl, pH 8.0, with 30 mM imidazole. Protein was eluted in 3 times 1 ml of 10 mM Tris/HCl, pH 8.0, with 500 mM imidazole. Eluted fractions were dialysed against various buffers (see 2.3.2. for the cathepsin B-like proteins of *E. tenella*), or against 100 volumes of 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM Tris(2-carboxyethyl)phosphorine hydrochloride (TCEP, Sigma, Poole) (for the PPI, and PPI active site mutants of *L. major*) overnight at 4°C.

2.3.7.2. Use of BioCAD

The same proteins were also purified using the BioCAD 700 E workstation (PE Biosystems, Foster City, CA, US), with a 1.7 ml POROS MC 4.6 mm column. The POROS beads were charged with nickel ions. The purification work was partially done by myself, and by Alan Scott (University of Glasgow). The column was equilibrated with 50 mM Tris/HCl, pH 8.0, and flow rate of 5-10 ml/min. Sample was applied, flow-through was re-loaded, and the column was washed with 20 ml of 50 mM Tris/HCl, pH 8.0, and 15 ml of 50 mM Tris/HCl, pH 8.0, 30 mM imidazole. The protein was eluted with a gradient of 50-500 mM imidazole in the same Tris buffer, over 15 ml. Protein was detected by absorbance at 280 nm. Fractions containing the protein were pooled and treated as described 2.3.7.1.

2.3.7.3. Ammonium sulfate saturation

This method was used for the purification of the recombinant cathepsin B of *E. tenella* produced in *P. pastoris*. Different aliquots of supernatant of culture containing the released recombinant cathepsin B were incubated with various percentages of a saturating solution of ammonium sulphate for 2 h at 4°C with mild agitation to fractionate the proteins present in the culture supernatant according to their solubility at various concentrations of ammonium sulphate. The samples were then centrifuged for 15 min at 13000 *g* at 4°C and the pellet was resuspended in the same volume of 50 mM sodium acetate, pH 5.0. The samples of resuspended pellets and supernatant were tested for enzyme activity and the protein profile was assessed by SDS-PAGE.

2.3.8. Protein native molecular mass determination

To determine the molecular mass of the recombinant pyroglutamyl peptidase I of *L. major* with a N-terminal His-tag, expressed in *E. coli*, and purified using a nickel-agarose column, gel-filtration analysis was performed. The column used was a high-resolution Sepharose 12 (1 cm diameter, 30 cm long), which was equilibrated in 50 mM Tris and 150 mM NaCl, pH 7.4, and run at 0.5 ml/min. Calibration was with the following proteins: ribonuclease A (13.7 kDa), chymotrypsinogen (25 kDa), ovalbumin (43 kDa), BSA (67 kDa), alcohol dehydrogenase (150 kDa) and β-amylase (200 kDa) (all Sigma, Poole). The pure soluble protein was provided to Dr. Helen Denton (University of Glasgow), who carried out the experiment.

2.3.9. Protein deglycosylation

Supernatant of *P. pastoris* culture containing the released recombinant cathepsin B *of E. tenella* was treated with 40% followed by 70% saturation of ammonium sulphate and aliquots (equivalent to about 20 μg of enzyme) were deglycosylated using 100 U of PGNase F (500000 U/ml, New England Biolabs, Hitchin) for 2.5 h at 37°C. The reactions were terminated by freezing at -20°C. Deglycosylation was assessed by SDS-PAGE and western-blot analysis.

2.3.10. SDS-PAGE

Proteins analysed by SDS-PAGE included recombinant proteins, and *E.coli*, *P. pastoris* and *L. major* lysates. The samples were mixed with equal volumes of 2X loading buffer and boiled for 5 min. Proteins were separated by SDS-PAGE according their molecular masses. The method was as described by Laemmli (1970). Gels used 12% (w/v) polyacrylamide (BioRad, Hemel Hempstead) and were run using a Mini-Protean II slab system (BioRad, Hemel Hempstead) according the manufacturer's instructions. Gels were stained with 0.25% (w/v) Coomassie Brilliant Blue R250 and destained with 10% (v/v) acetic acid, 12.5% (v/v) methanol, or used for electro-blotting for specific protein analysis.

2.3.11. Antibody production

Full length, N-terminal His-tagged, nickel-agarose-purified recombinant cathepsin B-like enzyme of *E. tenella* produced in *E. coli* was used to raise antiserum in a rabbit (done by Diagnostic Scotland, using standard methods). Full length, N-terminal His-tagged nickel-agarose purified recombinant pyroglutamyl peptidase I (PPI) of *L. major* produced in *E. coli*, and full length, N-terminal His-tagged nickel-agarose purified recombinant glutaminyl cyclase (QC) of *L. major* produced in *E. coli* were used to raise antiserum in a rat (done by Biological Services, University of Glasgow), by inoculation of 50 µg of protein diluted to a ratio of 1:1 (v/v) in complete Freund's adjuvant (Sigma, Poole) in a total volume of 100 µl, followed by two other inoculations of 50 µg of protein diluted to a ratio of 1:1 (v/v) in incomplete Freund's adjuvant (Sigma, Poole), with one month time separating each inoculation.

2.3.12. Western-blot analysis

Samples were in 1X loading buffer and proteins were separated on 12% (v/v) SDS-PAGE gels. Separated proteins were transferred to Hybond-C nitrocellulose membrane (Amersham, Chalfont St. Giles) by electroblotting in transfer buffer (20 mM Tris/HCl, 15 mM glycine, 20% (v/v) methanol) using a BioRad (Hemel Hempstead) mini transblot cell, at 4°C for 1 h at 100 V. Ponceau S staining allowed protein ladder visualization and labeling on the membrane. The membrane was then blocked for 1 h at room temperature in 20 mM Tris/HCl pH 7.6, 13.7 mM NaCl, 0.2% (v/v) Tween-20, and 5% (w/v) low fat dried milk. It was then incubated overnight with primary antibodies at 4°C in 20 mM Tris/HCl pH 7.6, 13.7 mM NaCl, 0.1% (v/v) Tween-20. Anti-cathepsin B of *E. tenella* antisera were used at a dilution 1:8000 (v/v), anti-PPI of *L. major* antiserum was used at 1:500 (v/v), and anti-His-tag antibody was used at 1:2000 (v/v). Blots were washed 4 times for 30 min with 20 mM Tris/HCl pH 7.6, 13.7 mM NaCl, 1% (w/v) low fat dried milk, and then incubated with secondary antibody for 2 h at room temperature in 200 mM Tris/HCl pH 7.6, 137 mM NaCl, 1% (w/v) low fat dried milk. Secondary antibodies were anti-rabbit IgG-conjugated horseradish peroxidase (HRP) (Promega, Southampton) at 1:5000 (v/v), anti-rat IgG-conjugated HRP (Pierce, Cramlington) at 1:1500 (v/v), or anti-mouse IgG-conjugated HRP (Pierce, Cramlington) at 1:2000 (v/v), depending on the primary antibody used. The membrane was washed 3 times for 30 min with 20 mM Tris/HCl pH 7.6, 13.7 mM NaCl, 1% (w/v) low fat dried milk. Bound antibodies were detected by using the Supersignal Enhanced Chemiluminescence (ECL) reagents (Pierce, Cramlington) according the manufacturer's instructions. Fluorescence was detected by autoradiography with reflexion autoradiography film (NEN Life Sciences, Boston, MA, US).

2.3.13. Indirect immunofluorescence

2.3.13.1. Analysis of the subcellular localisation of the cathepsin B-like enzyme in *E. tenella*

The anti-cathepsin B of *E. tenella* antiserum raised in rabbit was sent to Professor David J. P. Ferguson (Oxford University, John Radcliffe Hospital, Oxford), who performed all of the cathepsin B localisation studies on various *E. tenella* stages of the life cycle.

2.3.13.2. Analysis of the subcellular localisation of the pyroglutamyl peptidase I enzyme in *L. major*

All steps were at room temperature. Stationary phase parasites were collected by centrifugation for 10 min at 1300 *g*, washed in PBS, and resuspended at 2 x 10^7 cells/ml in PBS. An equal volume of 6% (w/v) paraformaldehyde in PBS was added and the mixture was incubated on ice for 1 h. Five volumes of PBS were added and the cells were washed and harvested as above. Cells were resuspended at 2 x 10^7 cells/ml in PBS. Slides were pre-treated with 200 μl of 0.01% (w/v) poly-L-lysine solution (Sigma, Poole), incubated for 5 min and then allowed to dry for 2 h. Then 200 μl of parasite suspension was applied to the slide and incubated 15 min. The cells were permeabilised in 0.1% (v/v) Triton X-100 in PBS for 10 min (in a glass chamber), slides were washed 3 times in PBS for 5 min, and blocked for 1 h with 20% (v/v) HIFCS in PBS. The primary antibody at a dilution 1:25, 1:100 or 1:500 (v/v) (anti-pyroglutamyl peptidase of *L. major* raised in rat) in PBS, with 20% (v/v) HIFCS was incubated on the slides for 1 h. The slides were washed 3 times for 5 min with PBS, and the secondary antibody at a dilution of 1:500 (v/v) (Alexa Fluor® 568 goat anti-rat IgG [Molecular Probes, Paisley]) in PBS with 20% HIFCS was incubated for 1 h at room temperature. The secondary antibody solution also contained the 4',6-diamidino-2-phenylindole DAPI staining (5 μg/ml, final concentration). The slides were washed 3 times for 5 min with PBS and mounted with anti-quenching agent (MOWIOL-DABCO in 50% (v/v) PBS). The edges of the coverslips were sealed with varnish and the slides were stored at 4°C in the dark until analysis. Fluorescence signals were visualized by UV fluorescence microscopy on a Zeiss Axioplan fluorescence microscope with a Hamamstsu Digital Camera using the OpenLab software (Improvision, University of Warwick).

2.3.14. Enzyme activity measurements

2.3.14.1. Substrate-gel electrophoresis

Reducing gelatin SDS-PAGE was as described in Robertson *et al.*, 1990. Native gelatin SDS-PAGE was as described in Pandey *et al.*, 2003. All gelatin gels contained gelatin at 0.2% (w/v) final concentration. After electrophoresis, the gels were incubated for 1 h in 2.5% (v/v) Triton X-100, and then in 0.1 M sodium acetate, pH 5.0, supplemented with 10 mM DTT at 37°C for 6-16 h before staining with Coomassie blue.

2.3.14.2. Azocasein assay

The proteolytic activity of various enzyme fractions towards azocasein (Sigma, Poole) was tested. Fifty μl of a 1% (w/v) azocasein solution in 1X PBS was mixed with 400 μl of 0.1 M Tris/HCl pH 8.0 or 0.1 M sodium acetate pH 5.0 with or without 10 mM DTT and/or 1 mM (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl}carbamoyl)oxirane-2-carboxylic acid (E64), and 50 μl of sample to test (about 5-10 μg of recombinant protein) and incubated at 37°C overnight. Controls without sample or with 25 μg of α-chymotrypsin (Sigma, Poole, 57 U/mg) were also included. After incubation, 500 μl of ice cold 1% TCA were added to the samples, incubated for 10 min on ice and centrifuged at 13000 *g* for 5 min at room temperature. The absorbance of the supernatant was measured at 400 nm.

2.3.14.3. Spectrophotometric assays

In all experiments, 1 unit of activity represents 1 μmol of substrate hydrolysed per min. All measurements were in triplicate.

The extinction coefficient at 410 nm for the substrate was measured using a standard curve of free chromogenic substituent in the buffer used for the assays. The extinction coefficient used for the paranitroanilide (pNA) substrates (all Bachem, St. Helens) was 9.5×10^3 l /mol/cm in all buffer systems. The light path was 1 cm.

2.3.14.3.1. Recombinant cathepsin B of *E. tenella*

The assays using the chromogenic substrates benzyloxycarbonyl-Phe-Arg-pNA (Z-Phe-Arg-pNA) and Z-Arg-Arg-pNA consisted of 50 mM sodium acetate pH 5.0, 2 mM EDTA, 10 mM DTT, with 333 μM of chromogenic substrate in a total volume of 600 μl. The buffer was incubated at 37°C with the enzyme sample for 10 min and the reaction was started by addition of the substrate. The reaction was followed continuously for 10 min at 410 nm.

2.3.14.3.2. Recombinant pyroglutamyl peptidase I of *L. major*

The assays using the chromogenic substrates pGlu-pNA (Bachem, St. Helens) consisted of different buffers, with 2 mM of chromogenic substrate in a total volume of 1 ml. The buffer was incubated at 32°C with the enzyme sample for 5 min and the reaction was started by addition of the substrate. The reaction was followed continuously for 10 min at 410 nm.

The dependence on DTT was measured in 25 mM potassium phosphate buffer pH 8.0. The reaction was started by addition of 2 mM of substrate and measured continuously for 10 min.

The pH dependence was determined using 2 different methods. The first comprised assaying the activity in a mix of 4 buffers (25 mM acetic acid, 25 mM 2-(N-morpholino) ethanesulphonic acid (MES), 75 mM Tris, 25 mM glycine) adjusted to the appropriate pH, with 10 mM DTT. The reaction was started by addition of 2 mM pGlu-pNA. The second method used a set of buffers, which was composed of "Good" buffers (Good *et al.*, 1966), as they are biologically and chemically non-reactive and their pK_a show a minimum dependence on the temperature and the ionic strength. These were MES/NaOH (25 mM, pH 5.4 and 6.5), piperazine-NN'-bis-2-ethanesulphonic acid (PIPES)/NaOH (25 mM, pH 6.3 and 7), 3-(N-morpholino) propanesulphonic acid (MOPS)/ NaOH (25 mM, pH 6.5 and 7.5), HEPES/NaOH (25 mM, pH 7.0 and 8.0), NN-bis (2-hydroxyethyl) glycine (BICINE)/NaOH (25 mM, pH 7.8 and 8.8), 2-(cyclohexylamino) ethanesulphonic acid (CHES)/NaOH (25 mM, pH 8.5, 9.0, 10.0 and 10.5), 3-(cyclohexylamino)-1-propanesulphonic acid (CAPSO)/NaOH (25 mM, pH 9.0, 9.5 and 10.0). Assays also included 10 mM DTT. The reaction was started by addition of 2 mM of pGlu-pNA.

The effect of EDTA, and magnesium ions was measured in 50 mM HEPES pH 8.0, 10 mM DTT using the substrate pGlu-pNA. The reaction was started by addition of 2 mM of substrate and measured continuously for 10 min.

Enzyme stability was tested in 50 mM HEPES pH 8.0, 2 mM Tris(2-carboxyethyl)phosphorine hydrochloride (TCEP), 1 mM EDTA with the substrate pGlu-pNA, at various time points. Enzyme stability was tested at –20, 4, and 37°C.

The measurement of the K_m for hydrolysis of pGlu-pNA was performed using 10 μM to 2 mM of substrate, in 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM TCEP, at 32°C.

For testing of inhibitors' effects, the inhibitor (or ion to test) was added to the buffer-enzyme mix and incubated for 10 min at 32°C. Assays were in 50 mM HEPES pH 8.0, 2 mM TCEP. The reaction was started with 200 μM of pGlu-pNA, and was then as described for the peptidase assay.

2.3.14.4. Fluorometric assays

In all experiments, 1 unit of activity represents 1 μmol of substrate hydrolysed per min. All measurements were in triplicate.

The extinction coefficients for the different products of cleavage were measured using a standard curve of free fluorescent substituent. All substrates were from Bachem, St. Helens. The extinction coefficient for β-naphtylamine (β-NA) was found to be 2.6 x 10^9 l/mol/cm in all buffer systems (excitation 320/ emission 410 nm). The extinction coefficient for 7-amino-4-methylcoumarin (AMC) was found to be 5.8 x 10^9 l/mol/cm in all buffer systems. The light path was 1 cm (excitation 380/ emission 465 nm).

2.3.14.4.1. Recombinant cathepsin B of *E. tenella*

Peptidase activity was assessed using the fluorogenic substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC at a concentration of 200 μM unless otherwise stated in 50 mM sodium acetate, pH 5.0, 5 mM EDTA, and 10 mM DTT. Reactions were in a total volume of 1 ml, at 37°C and started by addition of the substrate after a pre-incubation of the buffer-enzyme mix for 10 min at 37°C. The appearance of AMC was measured using excitation at 380 nm and emission at 465 nm. Activities were calculated from the rate over the first 10 min.

The dependence on DTT and on salt was measured using 10 μl of crude *Pichia* supernatant in 50 mM sodium acetate, pH 5.0, using the substrate Z-Phe-Arg-AMC. The reaction was started by addition of 50 μM of substrate and measured continuously for 5 min.

The effect of EDTA was measured using 10 μl of crude *Pichia* supernatant in 50 mM sodium acetate pH 5.0, 10 mM DTT with the substrate Z-Phe-Arg-AMC. The reaction was started by addition of 30 μM of substrate and measured continuously for 5 min (excitation 380 nm/emission 465 nm).

The pH dependence was determined using 2 different methods. The first comprised assaying the activity in a mix of 4 buffers (25 mM acetic acid, 25 mM MES, 75 mM Tris, 25 mM glycine) adjusted to the appropriate pH, with 10 μl of crude *Pichia* supernatant and 10 mM DTT. The reaction was started by addition of 30 μM of Z-Phe-Arg-AMC. The second method involved using 3 different buffers separately, depending on their pH range. These buffers were 0.1 M acetic acid, 0.1 M sodium acetate and 0.1 M potassium phosphate. Each assay used 2 μl of crude *Pichia* supernatant. All buffers were supplemented with 10 mM DTT. The reaction was started by addition of 30 μM of Z-Phe-Arg-AMC.

K_m measurements towards Z-Phe-Arg-AMC and Z-Arg-Arg-AMC used 0.5 μg of enzyme per assay and were in 50 mM sodium acetate pH 5.0, supplemented with 10 mM DTT and 5 mM EDTA. The enzyme sample was incubated for 10 min at 37°C in the assay buffer before start of the reaction by addition of the substrate (10 μM to 1.5 mM). The assays were continuous for 10 min at 37°C.

For testing of inhibitors' effects, the inhibitor was added to the buffer-enzyme mix and incubated for 10 min at 37°C. The reaction was then as described for K_m measurements.

2.3.14.4.2. Recombinant pyroglutamyl peptidase I of *L. major*

All assays were at 32°C in a total volume of 1 ml. The buffer was incubated at 32°C with the enzyme sample for 5 min and the reaction was started by addition of the substrate. Excitation and emission wavelengths were, respectively, 320 and 410 nm for the β-NA fluorogenic substrates.

The measurement of the K_m for hydrolysis of pGlu-βNA and H-Glu-βNA was performed using 0.1 μM to 1 mM of substrate, in 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM TCEP. All other assays against pGlu-βNA used 200 μM of this substrate.

Assays against the H-Gln-βNA substrate were in 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM TCEP, with 200 μM of this substrate.

2.3.14.4.3. Native pyroglutamyl peptidase I of *L. major*

L. major promastigote cell lysis is described in 2.1.3. Typically 5-10 x 10^7 *L. major* stationary phase promastigotes equivalent were tested per assay. The buffer consisted of 1 ml of 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM TCEP, or 1 ml of 50 mM Tris, pH 8.0. The buffer was pre-incubated at 32°C for 5 min with

200 μM of pGlu-βNA and the assay was started by addition of the cell extract to analyse. The measurements were continuous for 10 min. Excitation and emission wavelengths were, respectively, 320 and 410 nm.

2.4. *L. major* infectivity

2.4.1. *In vitro* infectivity in macrophages

Peritoneal macrophages were extracted from peritoneal lavage of CD-1 mice (work done by Susan Baillie, University of Glasgow). Macrophages were resuspended in RPMI medium (GibcoBRL, Paisley) with 10% (v/v) HIFCS at a concentration of 5 x 10^5 cells/ml, 200 μl of this suspension was placed in each well of cavity slides (VWR, Lutterworth), incubated at 37°C for 24 h in 5% CO_2, 95% air. Stationary phase *L. major* promastigotes or purified metacyclics of *L. major* were diluted in RPMI with 10% (v/v) HIFCS, and 200 μl were added to the macrophage suspension at a ratio of 2:1 or 0.5:1 respectively, following the method described in Frame *et al.*, 2000. After 24 h in 5% CO_2, 95% air at 37°C, the *L. major* promastigotes were washed off and 400 μl of RPMI with 10% (v/v) HIFCS was added. After 3 days of further incubation at 37°C the medium was removed and replaced by fresh medium and slides were further incubated for 3 days at 37°C. The cells were then fixed with methanol and stained with Giemsa stain, and the parasite loads were determined microscopically by counting 200 macrophages. Each cell line of *L. major* being analysed was infected into 3 wells of the cavity slide.

2.4.2. Infectivity in BALB/c mice

Groups of 6 mice were inoculated in the footpad with 5 x 10^5 stationary-phase *L. major* promastigotes, or 10^5 purified metacyclic promastigotes of *L. major* resuspended in 20 μl of PBS, pH 7.4. The thickness of the infected footpads was measured regularly over a 5-12 weeks period.

2.5. Bioinformatic analyses

Vector NTI (v. 6.0) was used to analyse DNA and protein sequences (searching for open reading frames, restriction sites, design of oligonucleotides for PCR, construction of cloning and expression plasmids). Sequence alignments were studied with Align X (Clustal X) and ContigExpress was used to align overlapping sequences. For phylogenetic analyses, sequences were aligned using Align X (Vector NTI program), cut to the same amino acid length (166), and insertions were removed manually. The edited sequences were realigned using Align X. The unrooted phylogenetic tree was generated using the neighbour-joining method in the Molecular Evolutionary Genetics Analysis (MEGA) program, version 2.1 (Kumar *et al.*, 2001), with 1000 bootstrap replicates.

The following web sources were used for database mining and analysis of DNA and protein sequences:

TIGR databases: www.tigr.org

The Sanger Institute: www.sanger.ac.uk

GeneDB: www.genedb.org

NCBI: www.ncbi.nlm.nih.gov

MEROPS: www.merops.sanger.ac.uk

Protein domain predictions were with:

SignalP (Nielsen *et al.*, 1996): www.cbs.dtu.dk/services/SignalP

Target P v 1.0: www.cbs.dtu.dk/services/TargetP

TMpred: www.ch.embnet.org

2.6. Statistical analyses

Values were expressed either as mean ± standard deviation (SD) when the number of repetitions was more than 2, or mean ± standard error (SE) when the number of repetitions was 2. Levels of significance were calculated by unpaired *t* tests using the GraphPad Prism program (San Diego, CA, US). Differences were considered significant at $P<0.05$.

Chapter 3

Expression and characterisation of the recombinant cathepsin B of *E. tenella*

3.1. Introduction

Invasion of host cells by *Eimeria*, and other apicomplexan parasites, involves the regulated release of specialised secretory organelles; namely micronemes, rhoptries and dense granules, located at the anterior end of the parasite (Dubremetz, 1998) and forming the apical complex. Maturation, trafficking and secretion of many of the secretory organelles' proteins rely upon proteolytic processing (Soldati *et al.*, 1998; Carruthers *et al.*, 2000; Reiss *et al.*, 2001). After several cycles of asexual multiplication (schizogony or merogony), gametogony begins; macrogametocytes (female gamete) and microgametocytes (male) are formed, and fuse to form oocysts, that are released. The oocyst wall ensures the parasite survival in the external environment before finding the next host. The oocyst wall formation therefore constitutes an essential process for disease transmission. The wall is sequentially formed by the release of three macrogamete organelles structurally similar to the dense granules of the infectious stages (Ferguson *et al.*, 2000); the veil forming bodies, the wall forming bodies type 1 (associated to the Golgi apparatus and located in the cytoplasm), and the wall forming bodies type 2 (associated to the endoplasmic reticulum and found in vacuolar spaces) (Ferguson *et al.*, 2003). Peptidases have been shown to be involved in the oocyst wall formation in *Eimeria* (Belli *et al.*, 2003a; Belli *et al.*, 2003b), by maturation of precursor proteins found in the wall forming bodies.

Some cysteine peptidases from apicomplexan parasites have been identified as important factors for invasion of host cells (Kim, 2004). These enzymes include the cathepsin B-like toxopain-1 in *Toxoplasma gondii* (Que *et al.*, 2002), and falcipain-1 in *Plasmodium falciparum* (Greenbaum *et al.*, 2002). Furthermore, parasite cysteine peptidases have attracted particular attention over the recent years because of their importance in parasite survival, interaction with the host cells and pathogenicity. Some of them seem to be promising targets for new selective inhibitors and anti-parasite agents (Selzer *et al.*, 1997; Rosenthal, 1999). The well studied cathepsin L-like enzyme, CPB2.8, of *Leishmania mexicana* provides a good example (Mottram *et al.*, 1996; Coombs and Mottram, 1997).

Cathepsin B enzymes, which belong to the clan CA, family C1 of cysteine peptidases comprise a signal peptide, a pro-domain and a mature domain. The pro-domain, which might possess in some cases folding information (Pandey *et al.*, 2004), has been shown to be a strong inhibitor of the catalytic domain (Sijwali *et al.*, 2002), and so maintains the peptidase in an inactive form during trafficking to the lysosome. Activation of the enzyme is a regulated process involving the proteolytic cleavage (endogenous or exogenous) of the pro-peptide, facilitated by decreasing the pH (Turk *et al.*, 2000), which is favoured in the acidic pH environment of the lysosome.

The ever-increasing number of gene sequences and EST data published in the public domain provides a great resource for analyses and potential identification of targets for therapeutics. Up to now, the anticoccidial drugs dominating the market for the treatment of avian *Eimeria* have been polyether ionophores, which perturb ion gradients in the parasite (Dutton *et al.*, 1995). But because of problems of drug resistance and occurrence of mutant populations, toxicity and limited efficacy of current drugs, identification and evaluation of new drugs and new targets is constantly and urgently needed.

The aim of this part of my work was to obtain active recombinant cathepsin B of *E. tenella* for biochemical characterisation, and drug screening. The difficulty of maintaining *Eimeria* species *in-vitro* somewhat limited the extent of the possible functional investigation in the parasite. Nonetheless, cysteine peptidase inhibitors have been shown to block sporozoite invasion (Samantha Brown, unpublished data), and to get an insight on what might be the role of the enzyme, immunolocalisation studies on invasive and sexual stages of *Eimeria tenella* were performed in collaboration with Prof David J. Ferguson, Oxford.

3.2. Results

3.2.1. Cloning of the cathepsin B gene

Database searches on the genomic database of *E. tenella* using the cathepsin B of *Toxoplasma gondii* (accession number AAL60053) as a query allowed detection of a 1535 bp open reading frame (Contig5436, www.geneDB.org). Amplification by PCR of the complete putative sequence coding for the cathepsin-B like protein (MEROPS clan CA, family C1) on both gDNA and cDNA gave a single product of the same size, about 1.6 kb, which corresponds to the predicted size for the amplification product using the specific primers designed (Fig. 3.1). The amplified fragments were sequenced and both fragments presented the same nucleotide sequences, confirming the absence of introns in the gene. The results of 5' RACE permitted detection of the presence of a mRNA comprising a 40 bp 5' untranslated region (5' UTR), confirming the start methionine (Fig. 3.2). Furthermore, the start methionine conforms to a Kozak eukaryotic consensus (Kozak, 1999) and the consensus assigned to protozoan parasites (Yamauchi, 1991) favouring the presence of a purine at position -3 and +4 of the start ATG. The complete ORF encodes a pre-pro-enzyme of 512 amino acids (M_r = 55,937).

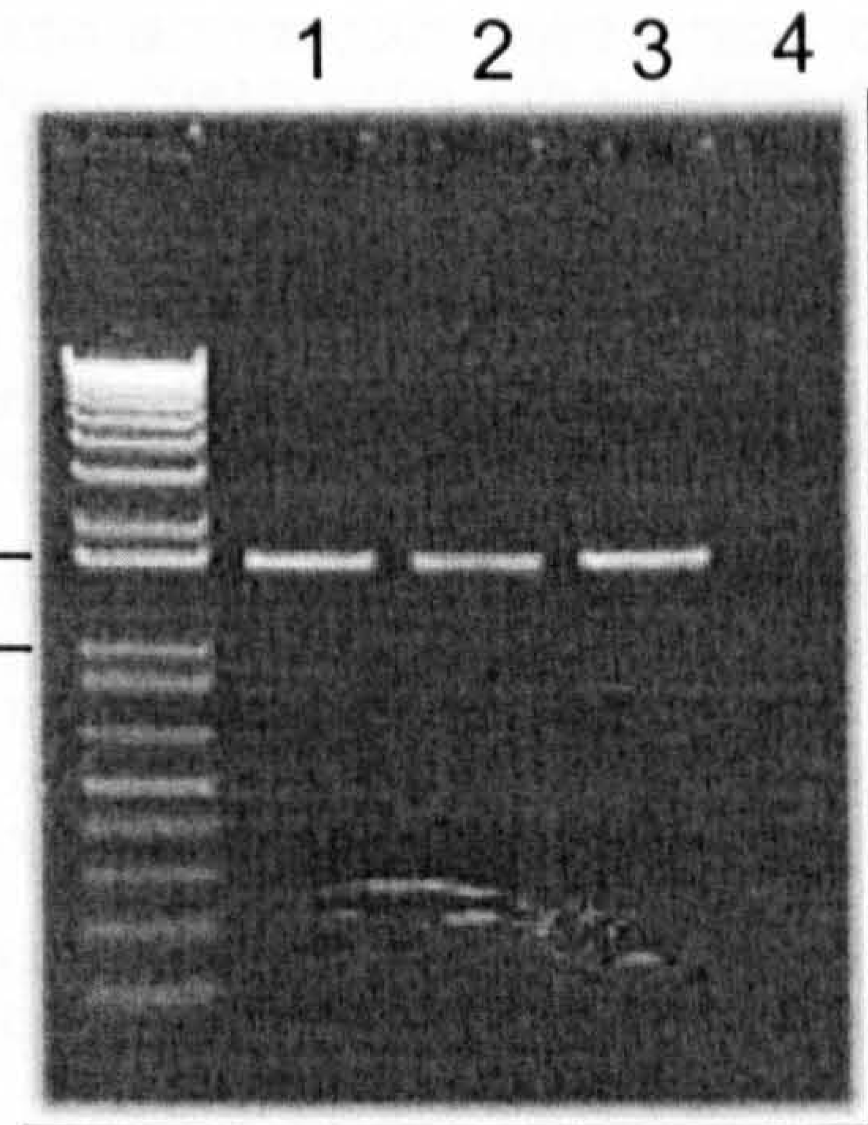


FIG. 3.1. **PCR amplification of the cathepsin B-like gene of *E. tenella*.** Electrophoresis on 1% agarose gel of 10 µl of PCR products from the amplification on genomic DNA (lane 1) and cDNA (lanes 2 and 3) of the putative full sequence coding for the cathepsin B-like enzyme *of E. tenella*. Negative control PCR, representative of the amplification on genomic DNA present in the cDNA preparation after reverse transcription, is shown on lane 4. The marker used is the 1kb Plus. The size of fragments is indicated on the left of the figure in kb.

```
                                                       M  R  I  G   K  L  W
1       TATTCGATTT CCACTTAGCT ACTACGCTTT GCATTCTGAG TTTGTGAAAA TGAGGATAGG GAAGCTGTGG
        ATAAGCTAAA GGTGAATCGA TGATGCGAAA CGTAAGACTC AAACACTTTT ACTCCTATCC CTTCGACACC
        A  T  L  A   I  S  S   P  L  I   F  S  C  A   M  P  S   D  D  L   G  G  A  K ·
71      GCCACCTTGG CCATTTCCAG CCCCCTGATC TTCAGTTGTG CGATGCCCTC CGATGATTTG GGCGGCGCTA
        CGGTGGAACC GGTAAAGGTC GGGGGACTAG AAGTCAACAC GCTACGGGAG GCTACTAAAC CCGCCGCGAT
        ·  E  D  L   Q  S  V   G  N  A  I   D  A  I   F  D  R   I  L  R  D   V  E  K ·
141     AGGAAGATCT GCAGTCTGTG GGCAATGCCA TCGATGCCAT ATTCGACCGG ATTCTGCGAG ATGTAGAGAA
        TCCTTCTAGA CGTCAGACAC CCGTTACGGT AGCTACGGTA TAAGCTGGCC TAAGACGCTC TACATCTCTT
        · D  I  G   H  G  K  A   L  Q  A   C  G  P   S  G  S  H   T  Q  G   V  E  N
211     GGATATTGGA CATGGCAAAG CCCTCCAAGC TTGCGGTCCC AGCGGTTCCC ACACGCAAGG TGTGGAGAAC
        CCTATAACCT GTACCGTTTC GGGAGGTTCG AACGCCAGGG TCGCCAAGGG TGTGCGTTCC ACACCTCTTG
         C  R  R  V   L  K  G   E  P  D   E  S  H  E   L  L  P   Y  L  L   K  D  Q  E ·
281     TGCAGGCGTG TGCTGAAAGG AGAGCCTGAC GAATCTCATG AGTTGCTGCC GTACCTGCTT AAGGACCAGG
        ACGTCCGCAC ACGACTTTCC TCTCGGACTG CTTAGAGTAC TCAACGACGG CATGGACGAA TTCCTGGTCC
        ·  A  M  R   L  W  R   D  H  F  R   R  H  V   G  E  L   V  S  S  A   S  E  H ·
351     AGGCCATGAG GTTGTGGCGG GATCACTTCC GTCGGCACGT GGGCGAGCTG GTATCCAGTG CATCGGAACA
        TCCGGTACTC CAACACCGCC CTAGTGAAGG CAGCCGTGCA CCCGCTCGAC CATAGGTCAC GTAGCCTTGT
        · S  G  A   D  T  P  L   Q  G  P   V  L  K   S  L  A  E   S  E  F   W  G  S
421     TAGCGGGGCA GATACCCCCC TTCAGGGGCC GGTATTGAAA AGCCTCGCGG AAAGCGAATT CTGGGGAAGC
        ATCGCCCCGT CTATGGGGGG AAGTCCCCGG CCATAACTTT TCGGAGCGCC TTTCGCTTAA GACCCCTTCG
        R  P  A  V   S  N  G   A  L  Q   H  L  R  V   K  M  Q   R  L  K   L  Q  A  A ·
491     CGGCCAGCTG TTAGCAACGG AGCTTTGCAG CACTTGAGGG TGAAGATGCA GCGGCTGAAG CTGCAAGCAG
        GCCGGTCGAC AATCGTTGCC TCGAAACGTC GTGAACTCCC ACTTCTACGT CGCCGACTTC GACGTTCGTC
        ·  E  Q  G   L  D  P   E  Q  A  V   T  W  E   A  E  V   S  P  R  F   K  Y  H ·
561     CTGAGCAGGG TCTAGACCCT GAACAGGCGG TGACGTGGGA AGCAGAAGTG TCTCCTCGTT TCAAATACCA
        GACTCGTCCC AGATCTGGGA CTTGTCCGCC ACTGCACCCT TCGTCTTCAC AGAGGAGCAA AGTTTATGGT
        · S  I  K   D  A  K  R   H  M  G   T  Y  L   S  F  Y  S   D  P  D   K  P  E
631     CAGCATCAAA GACGCGAAAA GGCATATGGG TACTTACCTT TCTTTCTATT CAGATCCAGA CAAGCCAGAG
        GTCGTAGTTT CTGCGCTTTT CCGTATACCC ATGAATGGAA AGAAAGATAA GTCTAGGTCT GTTCGGTCTC
         V  P  L  G   E  P  L   P  V  K   V  F  A  E   T  Q  Q   V  L  E   T  D  K  F ·
701     GTGCCGCTTG GGGAGCCACT GCCTGTGAAG GTGTTTGCGG AGACGCAGCA GGTTCTGGAA ACTGACAAGT
        CACGGCGAAC CCCTCGGTGA CGGACACTTC CACAAACGCC TCTGCGTCGT CCAAGACCTT TGACTGTTCA
        ·  D  A  R   E  A  F   P  Q  C  A   E  V  I   G  H  V   R  D  Q  G   D  C  G ·
771     TTGACGCGAG GGAGGCCTTT CCTCAATGCG CAGAGGTGAT TGGCCATGTC CGAGACCAAG GCGACTGCGG
        AACTGCGCTC CCTCCGGAAA GGAGTTACGC GTCTCCACTA ACCGGTACAG GCTCTGGTTC CGCTGACGCC
        · S  C  W   A  F  A  S   T  E  A   L  N  D   R  F  C  I   K  S  G   G  R  H
841     CTCGTGCTGG GCCTTTGCTA GCACAGAGGC CCTGAACGAC CGTTTCTGCA TCAAGTCAGG CGGCCGCCAT
        GAGCACGACC CGGAAACGAT CGTGTCTCCG GGACTTGCTG GCAAAGACGT AGTTCAGTCC GCCGGCGGTA
         R  E  A  L   S  P  Q   H  T  T   S  C  C  D   L  L  H   C  L  S   F  G  C  S ·
911     CGGGAGGCTC TGTCGCCGCA GCATACGACT TCATGCTGCG ACCTACTGCA CTGTTTGTCT TTTGGCTGCT
        GCCCTCCGAG ACAGCGGCGT CGTATGCTGA AGTACGACGC TGGATGACGT GACAAACAGA AAACCGACGA
        ·  G  G  Q   P  R  M   A  W  R  W   F  S  N   D  G  V   V  T  G  G   D  Y  N ·
981     CCGGGGGCCA GCCGAGGATG GCTTGGCGGT GGTTTTCGAA CGATGGGGTG GTAACCGGTG GTGACTACAA
        GGCCCCCGGT CGGCTCCTAC CGAACCGCCA CCAAAAGCTT GCTACCCCAC CATTGGCCAC CACTGATGTT
        · E  L  H   T  G  K  S   C  W  P   Y  E  I   P  F  C  R   H  H  S   E  G  P
1051    CGAGCTGCAC ACGGGTAAAA GCTGCTGGCC TTACGAGATT CCCTTCTGCC GGCACCACTC GGAAGGGCCC
        GCTCGACGTG TGCCCATTTT CGACGACCGG AATGCTCTAA GGGAAGACGG CCGTGGTGAG CCTTCCCGGG
         Y  P  K  C   E  G  P   L  P  K   A  P  K  C   R  K  D   C  E  E   A  E  Y  T ·
1121    TACCCAAAGT GTGAGGGGCC CTTGCCCAAG GCTCCGAAGT GCCGCAAGGA TTGTGAGGAA GCAGAGTACA
        ATGGGTTTCA CACTCCCCGG GAACGGGTTC CGAGGCTTCA CGGCGTTCCT AACACTCCTT CGTCTCATGT
        ·  S  K  V   K  P  F   K  D  D  L   H  F  A   T  S  A   Y  S  V  E   G  R  D ·
1191    CCAGCAAAGT GAAGCCATTC AAGGATGATC TACATTTCGC CACGAGTGCA TACAGTGTGG AAGGCCGGGA
        GGTCGTTTCA CTTCGGTAAG TTCCTACTAG ATGTAAAGCG GTGCTCACGT ATGTCACACC TTCCGGCCCT
        · Q  I  K   R  E  L  M   E  N  G   T  L  T   G  A  F  L   V  Y  E   D  F  L
1261    CCAAATTAAA CGGGAACTTA TGGAAAATGG CACTCTCACT GGCGCTTTCC TCGTGTATGA GGACTTCCTG
        GGTTTAATTT GCCCTTGAAT ACCTTTTACC GTGAGAGTGA CCGCGAAAGG AGCACATACT CCTGAAGGAC
         L  Y  K  E   G  V  Y   H  H  V   T  G  M  P   M  G  G   H  A  V   K  V  I  G ·
1331    CTGTACAAGG AGGGGGTGTA TCACCACGTG ACAGGCATGC CCATGGGTGG TCACGCAGTG AAGGTGATTG
        GACATGTTCC TCCCCCACAT AGTGGTGCAC TGTCCGTACG GGTACCCACC AGTGCGTCAC TTCCACTAAC
        ·  F  G  N   E  D  G   R  D  Y  W   L  A  V   N  S  W   N  E  Y  W   G  D  K ·
1401    GCTTCGGGAA CGAAGACGGG AGGGACTACT GGCTAGCTGT GAACAGCTGG AACGAATACT GGGGAGATAA
        CGAAGCCCTT GCTTCTGCCC TCCCTGATGA CCGATCGACA CTTGTCGACC TTGCTTATGA CCCCTCTATT
        · G  T  F   K  I  E  M   G  E  A   G  I  D   K  E  F  C   G  G  E   P  K  V
1471    AGGCACTTTC AAAATTGAAA TGGGAGAGGC AGGAATAGAT AAAGAATTCT GTGGAGGGGA GCCGAAAGTG
        TCCGTGAAAG TTTTAACTTT ACCCTCTCCG TCCTTATCTA TTTCTTAAGA CACCTCCCCT CGGCTTTCAC
         P  N  D  K   N  A  S   L  L  P   S  A  Q  D   L
1541    CCCAATGACA AAAACGCCTC ATTGCTGCCG TCAGCGCAGG ACCTATGAAA CTAAAGGCGA TTCAAGGGC
        GGGTTACTGT TTTTGCGGAG TAACGACGGC AGTCGCGTCC TGGATACTTT GATTTCCGCT AAGTTCCCG
```

FIG. 3.2. **Complete sequence of the contig containing an open reading frame encoding a cathepsin B-like enzyme of *E. tenella*.** The sequence of the 5' UTR obtained from 5' RACE experiment is on grey background. The translation of the putative ORF is shown in bold.

The amino acid sequence was analysed with the SignalP algorithm (Nielsen *et al.*, 1997) and a pre-peptide of 21 hydrophobic amino acids could be predicted. By comparison with cathepsins B from other organisms (Fig. 3.3), a pro-domain of 213 amino acids could be predicted as well. As a comparison, the cathepsin B-like enzyme of *T. gondii* (Que *et al.*, 2002) has a pre-peptide of 34 amino acids, a pro-domain of 239 amino acids and the mature enzyme is composed of 296 residues (Fig. 3.4). The cathepsin B-like enzyme of *E. tenella* has a long pro-domain very similar to the one from *T. gondii.* But both enzymes lack the ERFNIN motif, usually found in the pro-domain of cathepsins L- and H-like enzymes. The C-terminal end of the enzyme contains a potential site for asparagine-linked glycosylation (N-X-S/T) at position 502. The complete pre-pro-enzyme showed 42% similarity with the *T. gondii* cathepsin B and the mature enzyme of 491 amino acids showed good similarity to the human cathepsin B (45%), the cathepsin B of *L. major* (39%), *T. gondii* (57%), and *S. mansoni* (44%). The position of the catalytic cysteine, histidine and asparagine in the cathepsin B-like enzyme of *E. tenella* are respectively 263, 445, and 465 (*E. tenella* cathepsin B numbering). The pro-enzyme contains 16 cysteine residues, which can possibly be involved in the formation of 8 disulfide bridges. The positions for 10 of the cysteine residues in the different enzymes of the different organisms compared were conserved. The occluding loop diagnostic of cathepsin B enzymes can also be identified at His^{352} and His^{353}. This loop might confer to the enzyme an exopeptidase activity in addition to the endopeptidase activity. By comparison to the human cathepsin B, a C-terminal extension of 8 amino acids can be detected. This short extension is similar to the one found for *T. gondii* (26 amino acids) but differs from the very long extensions found for some cathepsin L-like enzymes.

```
              1                                                                    70
Et      (1) -------------MRIGKLWATLAISSPLIFSCAMPSDDLGGAKEDLQSVGN---AIDAIFDRILR--DV
Tg      (1) MEGRKSFRVLGTPLPFAALAAILLLGCMYTRAAGTPDDSLFPLSEDTSVDPRESFSAEDVLNAFVSPESV
Sm      (1) ----------------------------------------------------------------------
Lm      (1) ----------------------------------------------------------------------
Hum     (1) ----------------------------------------------------------------------
            71                                                                  140
Et     (53) EKDIGHGKALQACGPSGSHTQGVENCRRVLK----------GEPDESHELLPYLLKDQEAMRLWRDHFRR
Tg     (71) ESLFDSIVAEQVVATSGNLTESAPRDRDSAPERRSDGGTLRGHPEDAGELLRLLLADSEDMELWKANFFR
Sm      (1) ----------------------------------------------------------------------
Lm      (1) ----------------------------------------------------------------------
Hum     (1) ----------------------------------------------------------------------
            141                                                                 210
Et    (113) HVG----------------ELVSSASEHSGADTPLQGPVLKSLAESEFWGSRPAVSNGALQHLRVKMQRL
Tg    (141) HLTHSMRHIVRDSVLVSEKAFPSEQAGEAGRDPGDLKKALEETGEDVFWESRPASSNAAVALIKKKMEKQ
Sm      (1) ---------------------------------MNQYSCYLLQLYIIILLSYGTLNEIDARRHKRMYQPL
Lm      (1) ---------------------------------MALRAKSALCLVAVFALLLATTVSGLYAKPSDFPLLG
Hum     (1) --------------------------------------------MWQLWASLCCLLVLANARSRPSFHPL
            211                                                                 280
Et    (167) KLQAAEQGLDPEQAVTWEAEVSPRFKYHSIKDAKRHMGTYLSFYSDPDKPEVPLGEPLPVKVFAETQQVL
Tg    (211) AGKTGEG----AG-HTWEPEVSLRFRYLSLKDAKKLMGTFLVN-TKVEGFPTPKGMPLPAKEFENATEPV
Sm     (38) SMELINFINY---EANTTWKAAPTTRFRTVSDIRRMLGALPDP----------NGEQLETLCTGYISDEL
Lm     (38) KSFVAEVNSKAKGQWTASANNGYLVTGKSLGEVRKLMGVTDMS---------TEAVPPRNFSVEELQQDL
Hum    (27) SDELVNYVN----KRNTTWQAGHNFYNVDMSYLKRLCGTFLG------------GPKPPQRVMFTEDLKL
            281                         *                                       350
Et    (237) ETDKFDAREAFPQCAEVIGHVRDQGDCGSCWAFASTEALNDRFCIKSGGRHREALSPQHTTSCCDLLHCL
Tg    (275) P-AHFDARTAFPACKDVVGHVRDQGDCGSCWAFASTEAFNDRLCIRSQGKRLMPLSAQHTTSCCNAIHCA
Sm     (95) P-KSFDARVEWPHCPSIS-EIRDQSSCGSCWAFGAVEAMSDRICIKSKGKHKPFLSAENLVSCC---SSC
Lm     (99) P-EFFDAAEHWPMCLTIS-EIRDQSNCGSCWAIAAVEAISDRYCTFG-GVPDRRMSTSNLLSCC---FIC
Hum    (81) P-ASFDAREQWPQCPTIK-EIRDQGSCGSCWAFGAVEAISDRICIHTNAHVSVEVSAEDLLTCCG--SMC
            351                                                                 420
Et    (307) SFGCSGGQPRMAWRWFSNDGVVTGGDYNELHTGKSCWPYEIPFCRHHSEGP-YPKCEGPLP--KAPKCRK
Tg    (344) SFGCNGGQPGMAWRWFERKGVVTGGDFDALGKGTTCWPYEVPFCAHHAKAP-FPDCDATLVPRKTPKCRK
Sm    (160) GMGCNGGFPHSAWLYWKNQGIVTGDLYNTTNG---CQPYEFPPCEHHVIGP-LPSCDGDVE---TPSCKT
Lm    (163) GLGCHGGIPTVAWLWWVWVGIAT-------ED---CQPYPFDPCSHHGNSEKYPPCPSTIYD--TPKCNT
Hum   (147) GDGCNGGYPAEAWNFWTRKGLVSGGLYESHVG---CRPYSIPPCEHHVNGS-RPPCTGEGD---TPKCSK
            421                                                                 490
Et    (374) DCEEAEYTSKVKPFKDDLHFATSAYSVEG-RDQIKRELMENGTLTGAFLVYEDFLLYKEGVYHHVTGMPM
Tg    (413) DCEEQAYADNVHPFDQDTHKATSAYSLRS-RDDVKRDMMTHGPVSGAFMVYEDFLSYKSGVYKHVSGLPV
Sm    (223) NCQPGYN----IPYEKDKWYGEKVYRIHSNPEAIMLELMRNGPVEVDFEVYADFPNYKSGVYQHVSGALL
Lm    (221) TCERNEM-------DLVKYKGSTSYSVKG-EKELMIELMTNGPLELTMQVYSDFVGYKSGVYKHVLGDFL
Hum   (210) ICEPGYS----PTYKQDKHYGYNSYSVSNSEKDIMAEIYKNGPVEGAFSVYSDFLLYKSGVYQHVTGEMM
           491*              *                                                  560
Et    (443) GGHAVKVIGFGNEDGRDYWLAVNSWNEYWGDKGTFKIEMG--EAGIDKEFCGGEPKVPNDKNASLLPSAQ
Tg    (482) GGHAIKIIGWGTENGEEYWHAVNSWNTYWGDGGQFKIAMG--QCGIDGEMVAGEAAWQETEGVVNGEEEL
Sm    (289) GGHAVRLLGWGEENNVPYWLIANSWNSDWGDKGYFKIVRGKNECGIESDVNAGIPKIKN-----------
Lm    (283) GGHAVKLVGWGTQDGVPYWKVANSWNTDWGDKGYFLIQRGNNECKIESGGVAGIPAQE------------
Hum   (276) GGHAIRILGWGVENGTPYWLVANSWNTDWGDNGFFKILRGQDHCGIESEVVAGIPRTDQYWEKI------
            561                580
Et    (511) DL------------------
Tg    (550) PGQRAAGARAGAHAEEEREM
Sm    (348) --------------------
Lm    (341) --------------------
Hum   (340) --------------------
```

FIG. 3.3. **Alignment of the cathepsin B of *Eimeria tenella* with cathepsin B enzymes from different organisms.** Numbering is based on the cathepsin B of *Toxoplasma gondii*. Dashes indicate gaps in the alignment. The predicted signal peptides are underlined. Active site residues are starred. Conserved residues are on grey background. The starts for the mature enzymes are in bold. Et, *Eimeria tenella* ; Tg, *Toxoplasma gondii* (accession number AAL60053, Que *et al.*, 2002); Sm, *Schistosoma mansoni* (CAC85211, Klinkert *et al.*, 1994); Lm, *Leishmania major* (AAB48119, Sakanari *et al.*, 1997); Hum, human cathepsin B (NP 680093, Chan *et al.*, 1986).

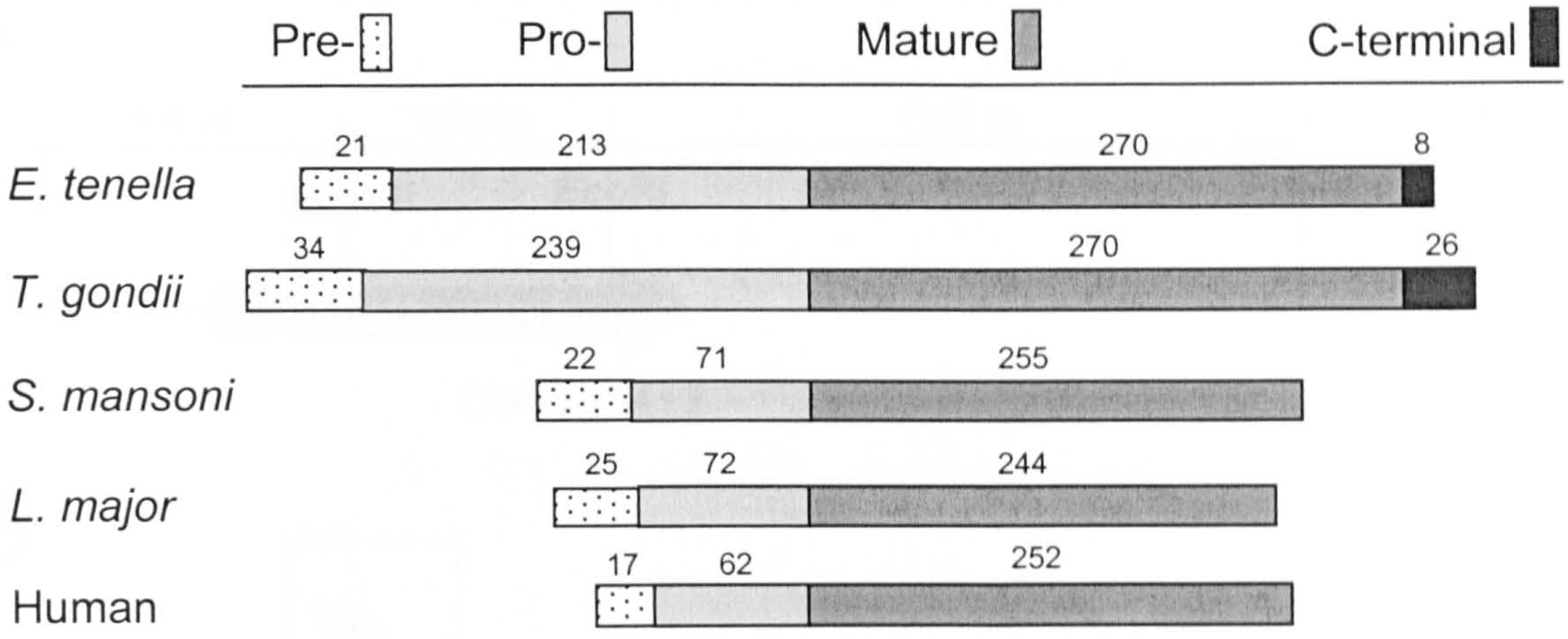


FIG. 3.4. **Schematic comparative representation of the different cathepsin B sequences.** The sequences used are from the alignment in figure 3.3. The pre-, pro-, mature and C-terminal domains are represented as blocks and the size in amino acids for each domain is indicated on top of each block.

3.2.2. The cathepsin B gene may be single copy in the *E. tenella* genome

By Southern blotting of gDNA and hybridization with a probe specific to the cathepsin B gene, a single DNA fragment at about 1 kb was detected (Fig. 3.5). Another DNA fragment at about 1.4 kb should be detected too, but the amount of DNA run on the gel was low, and as the intensity of the band depends on the length of hybridization, this band may not have been labeled enough for detection. The result obtained is indicative that the gene encoding a cathepsin B-like enzyme in *E. tenella* occurs in one copy in the genome. This result is supported by the sequence data from the Sanger database.

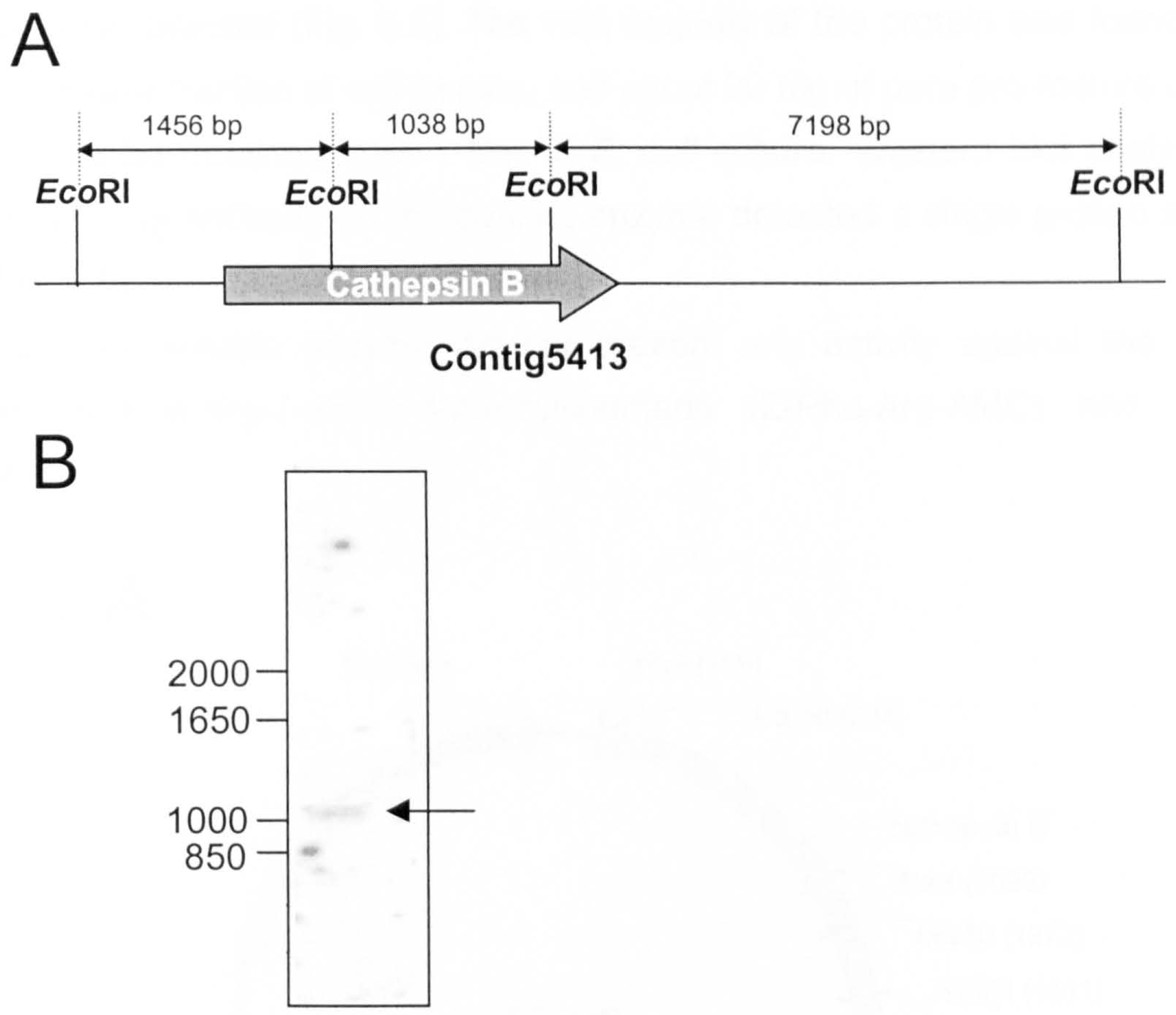


FIG. 3.5. **Southern blot analysis of *E. tenella* genomic DNA using the cathepsin B sequence as a probe.** (A) Schematic representation of the cathepsin B gene on Contig5413 (*Eimeria* assembly, Wellcome Trust Sanger Institute). The ORF is shown as an arrow. *Eco*RI sites are represented as well as the predicted sizes of the resulting digestion fragments. (B) Southern blot analysis. Two μg of genomic DNA were digested with *Eco*RI, separated on 1% agarose gel, blotted on nylon membrane and hybridised with ^{32}P-labelled cathepsin B full gene. The blot was exposed for 6 days. The positions of the molecular size markers (bp) are shown on the left of the figure. The arrow indicates the band at about 1 kb.

3.2.3. Expression of recombinant full-length cathepsin B in *E. coli*

3.2.3.1. Expression in *E. coli* BL21(DE3)

The sequence encoding the pro-mature cathepsin B was amplified from cDNA and cloned in the pET28a(+) vector for expression in *E. coli* with a N-terminal His-tag. The plasmid obtained and predicted protein are schematised in Fig. 3.6. Expression in *E.coli* BL21(DE3) resulted in detection of an abundant protein at 56 kDa corresponding to the expected size for the pro-mature enzyme, and no protein corresponding to the size of the mature enzyme (around 30 kDa) could be detected (Fig. 3.7). The vast majority of the protein was found to be in the soluble fraction of cell lysates, and about 20 mg of pure pro-mature cathepsin B could be obtained from 1 liter of *E. coli* culture. Western blot analysis using anti-His-tag antibody on the purified enzyme detected a single protein at 57 kDa (Fig. 3.8).

The pure soluble enzyme did not present any activity against the substrates benzoyl-Phe-Arg-7-amino-4-methylcoumarin (Z-Phe-Arg-AMC) and Z-Arg-Arg-AMC.

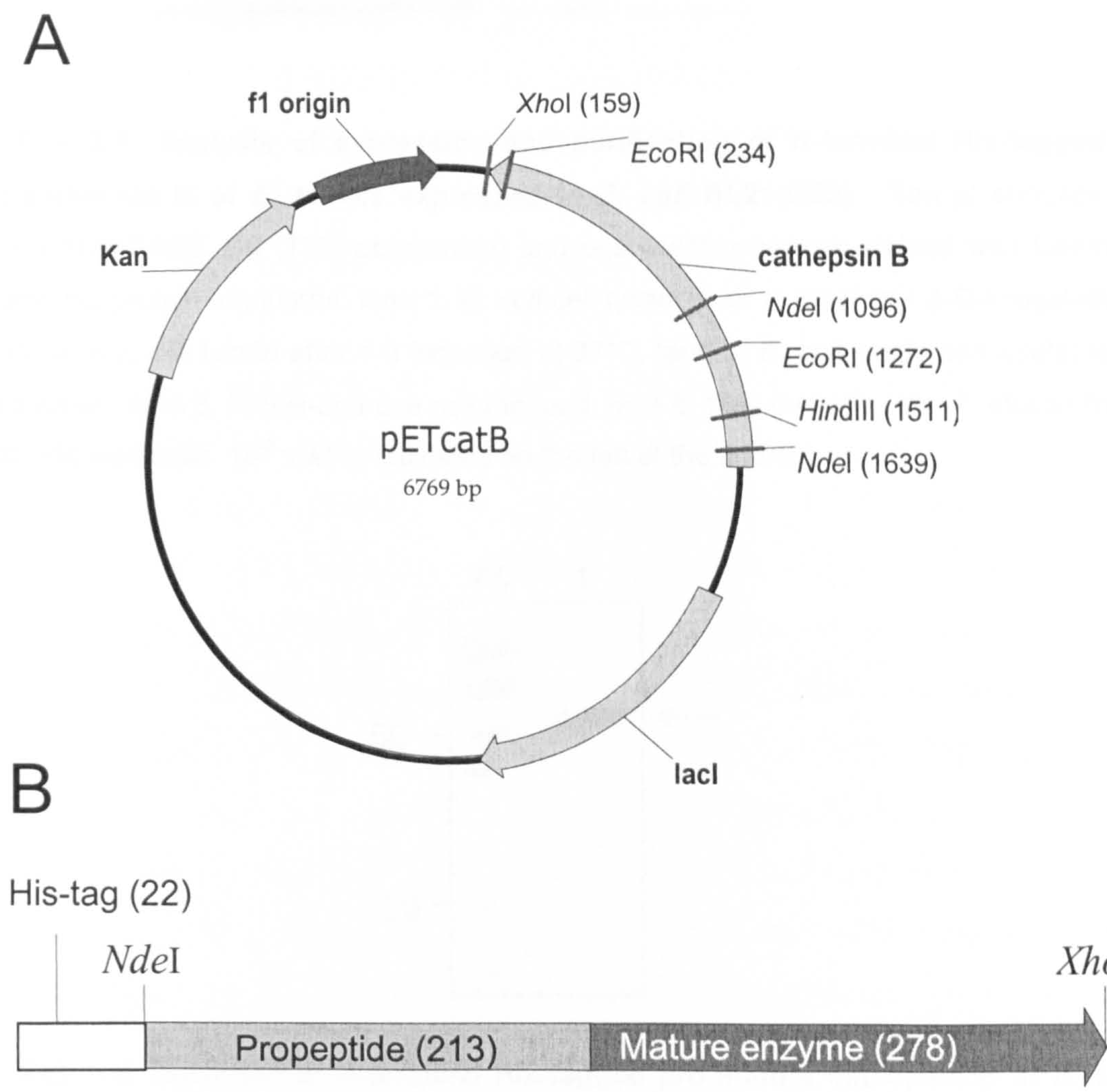


FIG. 3.6. **Construct for expression of recombinant pro-cathepsin B of *E. tenella* in *E. coli*.** (A) Schematic representation of the plasmid resulting from the cloning of the pro-mature sequence of the cathepsin B of *E. tenella* in the pET28a(+) vector (Novagen) for expression of pro-mature cathepsin B with a N-terminal His-tag. Kan, kanamycin resistance gene; f1 origin, origin of replication of the phage f1; lacI, gene encoding the *lac* repressor protein. (B) Schematic representation of the protein expressed from the pETcatB plasmid. The size in amino acids of the different fragments is indicated between brackets.

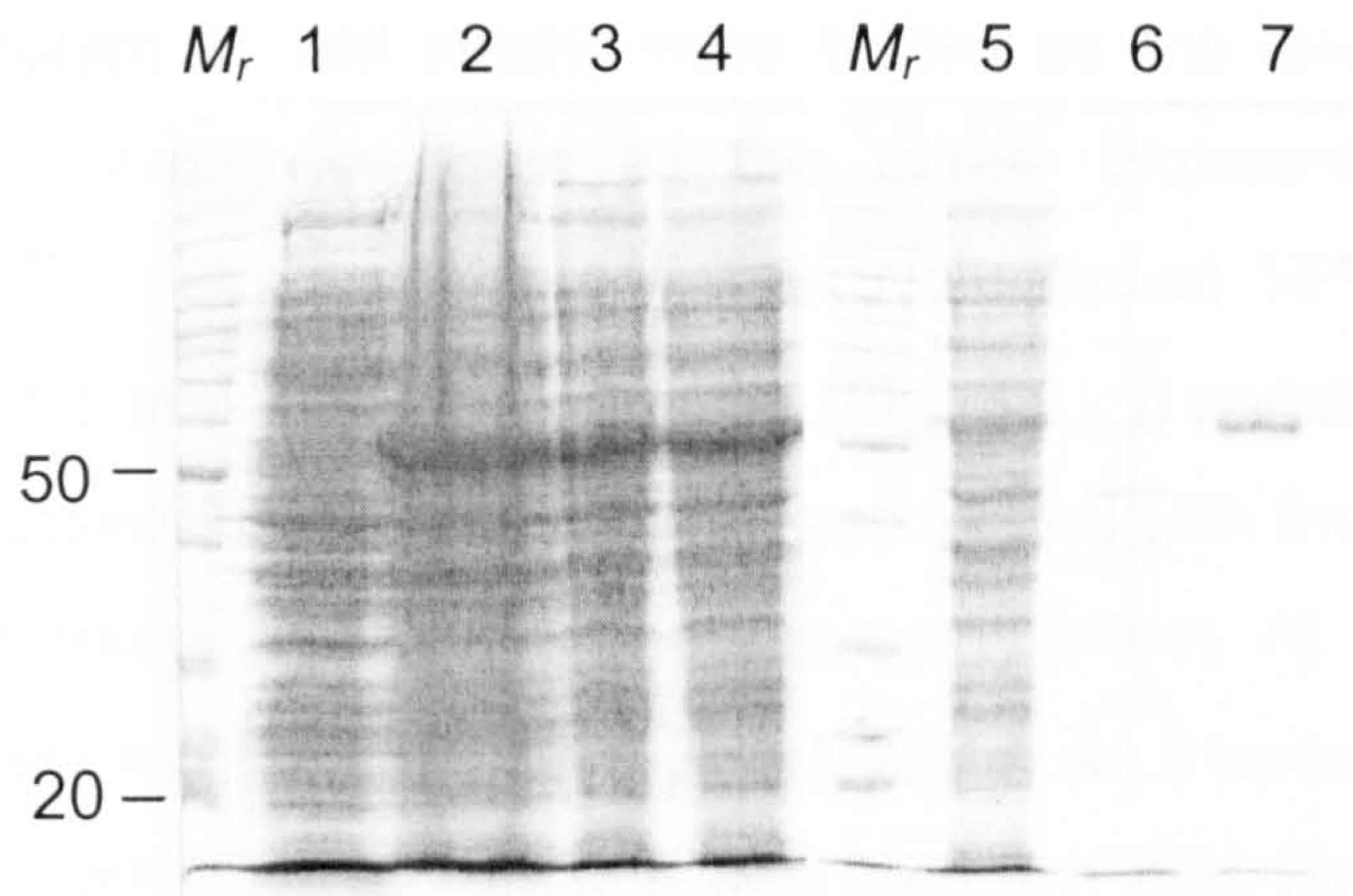


FIG. 3.7. **Analysis of expression and purification of N-terminal His-tagged pro-mature cathepsin B of *E. tenella* expressed in *E. coli* BL21(DE3).** Ten μl samples were loaded on SDS-PAGE gel (12% acrylamide) and post-electrophoresis stained with Coomassie blue. Lane M_r, protein standards; lane 1, *E. coli* cell lysate prior to isopropyl β-D-thiogalactoside induction; lane 2, cell lysate after 4 h induction at 37°C; lane 3, *E. coli* whole cell lysate; lane 4, soluble fraction; lane 5, Nickel-agarose column load; lane 6, flow through; lane 7, elution fraction with 500 mM imidazole. 10^{-3} x M_r is indicated on the left of the figure.

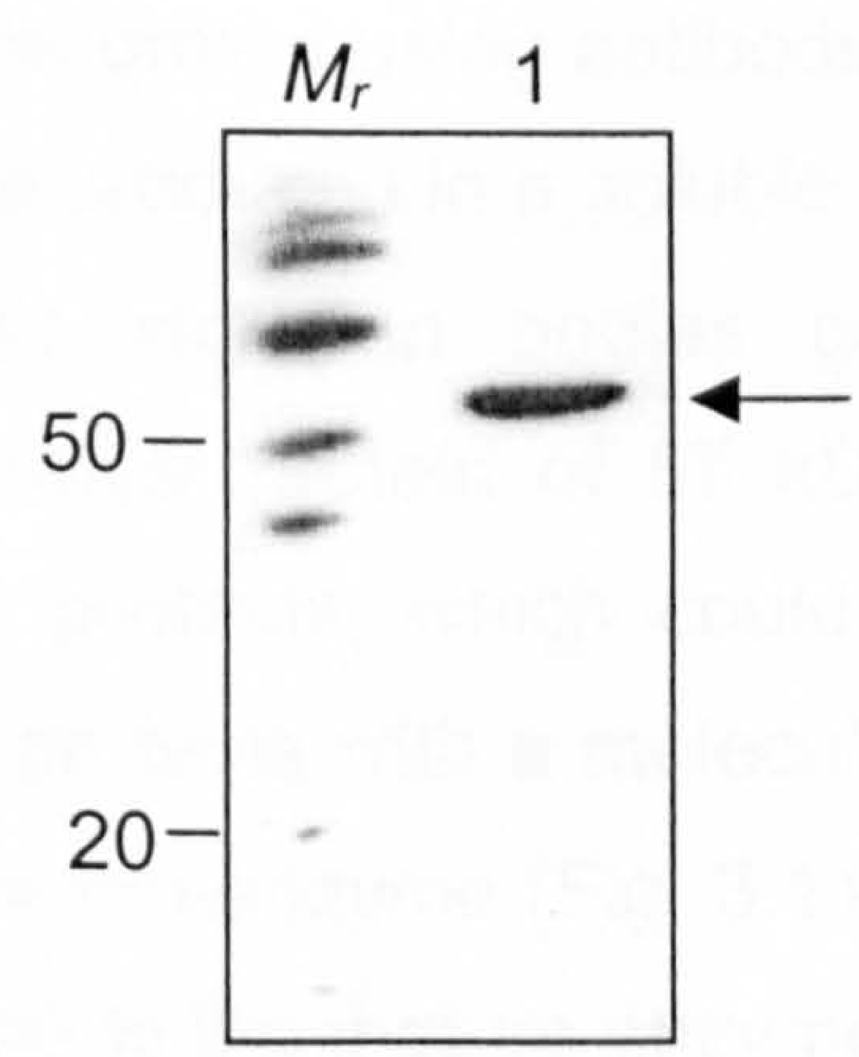


FIG. 3.8. **Detection of N-terminal His-tagged pro-mature cathepsin B of *E. tenella* expressed in *E. coli* BL21(DE3) using anti His-tag antibody.** Membrane exposure of 2 min, anti-His-tag antibody dilution of 1:2000 (v/v). Lane M_r, protein standards; lane 1, 10 μl of recombinant protein expressed in BL21(DE3) and purified on nickel-agarose column. 10^{-3} x M_r is indicated on the left of the figure. Arrow indicates the His-tagged pro-mature enzyme.

3.2.3.2. Expression in *E. coli* JM109(DE3) and HMS174(DE3)

Different *E. coli* strains were tested as the levels of soluble expression may vary from one strain to the other. Expression in JM109(DE3) and HMS174(DE3) strains of *E. coli* was evaluated at 15°C, 25°C and 37°C, using 0.5, 1, or 1.5 mM IPTG for induction (Fig. 3.9). It revealed that at 37°C the over-expressed protein was mainly found associated with the inclusion bodies and that the concentration in IPTG did not have any effect. At 15°C, the expression was considerably reduced and hardly detectable. As it seemed that the concentration in IPTG did not have an effect, it was fixed at 1 mM. Expression in both strains at 25°C with 1 mM IPTG resulted in over-expression of a protein of 57 kDa, occurring approximately equally in the soluble fraction of whole *E. coli* cell lysates and inclusion bodies (Fig. 3.10). As the expression pattern in the two strains was very similar, further analysis were performed on the HMS174(DE3) strain only (arbitrary choice).

Inclusion bodies produced in E. coli HMS174(DE3) were solubilised (see Table 2.2, Chapter 2), and the purity of the enzyme could be increased using nickel-agarose column, though the purification was not complete (Fig. 3.11). Western blot analysis was performed using antibodies raised in rabbit against the purified recombinant enzyme produced in a soluble manner in *E. coli* BL21(DE3). Analysis on the solubilised inclusion bodies generated via expression in HMS174(DE3) detected a major protein of 57 kDa corresponding to the pro-mature enzyme, plus other proteins, which could correspond to *E. coli* cross reacting proteins or, for the proteins with a molecular mass less than 57 kDa, to products of processing of the pro-enzyme (Fig. 3.11). Nevertheless, no protein at 30 kDa that would correspond to the mature enzyme was detected.

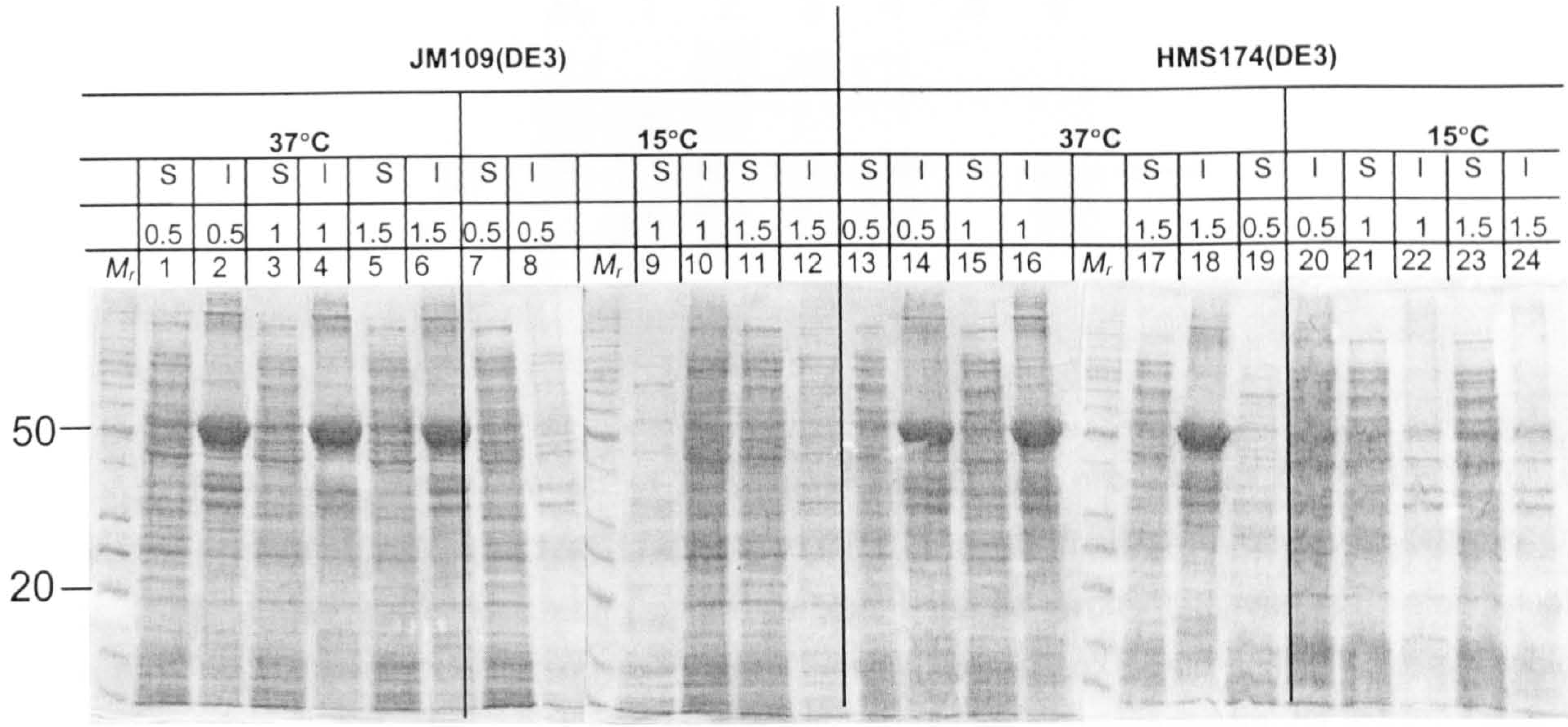


FIG. 3.9. **Analysis of expression of the N-terminal His-tagged pro-mature cathepsin B of *E. tenella* in *E. coli* JM109(DE3) and HMS174(DE3) at 37°C and 15°C**. Ten μl samples were loaded on SDS-PAGE gel (12% acrylamide)and post-electrophoresis stained with Coomassie blue. Lanes M_r, protein standards; lanes 1 to 12, JM109(DE3); lanes 13 to 24, HMS174(DE3). Above the lane numbers are indicated the concentrations in mM of IPTG used for induction, as well as the temperature of induction. Inductions were for 4 h. S: soluble fraction, I: insoluble fraction. 10^{-3} x M_r is indicated on the left of the figure.

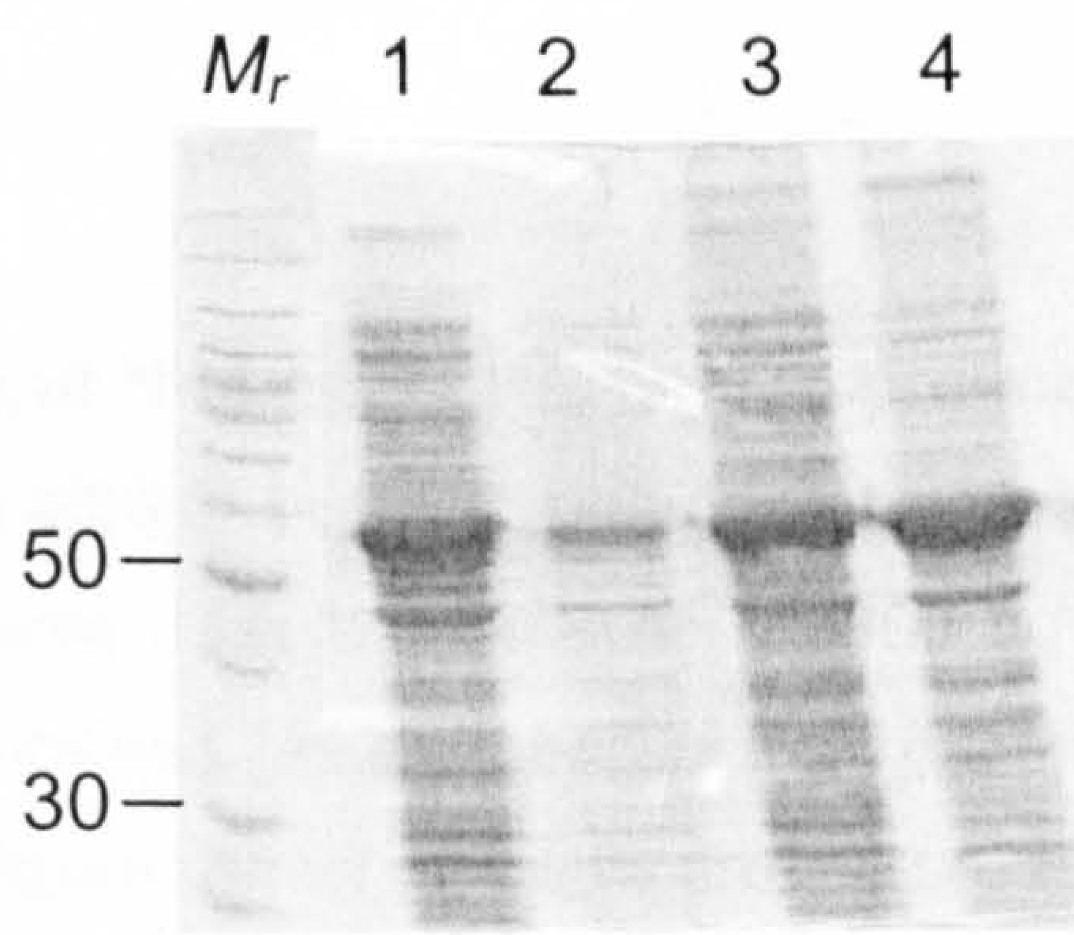


FIG. 3.10. **Analysis of expression of the N-terminal His-tagged pro-mature cathepsin B of *E. tenella* in *E. coli* JM109(DE3) and HMS174(DE3) at 25°C.** Induction was with 1 mM IPTG for 4 h. Ten μl samples were loaded on SDS-PAGE gel (12% acrylamide) and post-electrophoresis stained with Coomassie blue. Lane M_r, protein standards; lane 1, JM109(DE3) soluble fraction; lane 2, JM109(DE3) inclusion bodies; lane 3, HMS174(DE3) soluble fraction; lane 4, HMS174(DE3) inclusion bodies. 10^{-3} x M_r is indicated on the left of the figure.

FIG. 3.11. **Solubilisation of N-terminal His-tagged pro-mature cathepsin B of *E. tenella* produced in *E. coli* HMS174(DE3) at 37°C and purification.** Analysis on SDS-PAGE gel (12% acrylamide) stained with Coomassie blue. Lane M_r, protein standards; lane 1, soluble fraction of *E. coli* HMS174(DE3) lysate; lane 2, insoluble fraction; lane 3 and 4, inclusion bodies washes; lane 5, solubilised inclusion bodies before loading on anion-exchange column; lane 6, eluted fraction. 10^{-3} x M_r is indicated on the left of the figure. 10 μl samples were loaded. Arrow indicates the pro-mature enzyme.

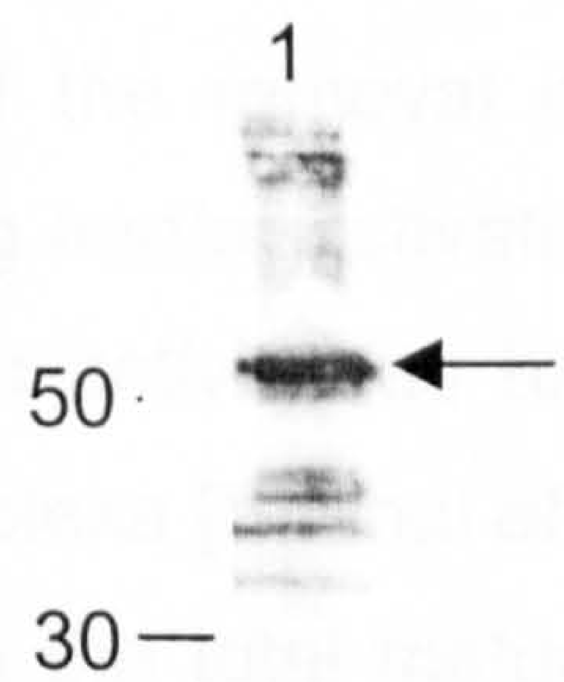


FIG. 3.12. **Detection of N-terminal His-tagged pro-mature cathepsin B of *E. tenella* produced in *E. coli* HMS174(DE3) at 37°C using anti cathepsin B of *E. tenella* antibody raised in rabbit.** Antibody was used at a dilution 1:5000 (v/v). Membrane was exposed 15 seconds. Lane 1, 10 μl of solubilised inclusion bodies from expression in HMS174(DE3). 10^{-3} x M_r is indicated on the left of the figure. Arrow indicates pro-mature enzyme.

3.2.4. Attempts to activate the pro-mature cathepsin B produced in *E. coli*

Various sample of enzyme produced in a soluble manner in *E. coli* BL21(DE3) or HMS174(DE3), or as inclusion bodies in HMS174(DE3) were treated using methods and combinations of methods described in the literature for activation of cysteine peptidases. The summary of which samples were treated is presented in Table 2.2, chapter 2. The aim in all cases was to try to improve folding and disulfide bridge formation and so induce activation (removal of the pro-domain) and obtain mature active enzyme. The samples treated were analysed on Coomassie blue-stained SDS-PAGE, western-blots, gelatin gels and assayed for activity using the azocasein assay and against the "typical" cathepsin B substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC. Non-induced *E. coli* cell lysates treated in the same manner and no enzyme controls were included in the measurements of activity.

In all cases, no maturation to the mature size enzyme (about 30 kDa) could be detected, and no activity against azocasein or gelatin could be detected (results not shown). The maximum activity observed among all the samples tested against the substrate Z-Phe-Arg-AMC was 30 mU/mg, but this may have been due to the pro-mature enzyme itself.

3.2.5. Expression of recombinant truncated cathepsin B in *E. coli*

The idea was to test if the removal of most of the pro-domain would facilitate the folding, processing and/or activation process of the enzyme in order to get a significant increase in activity (as has been the case for some other cysteine peptidases of Apicomplexa [Shenai *et al.*, 2000; Sijwali *et al.*, 2001]).

The sequence encoding the total mature domain of the cathepsin B and 31 amino acids of the pro-domain was amplified from cDNA and cloned in the pET28a(+) vector for expression in *E. coli* BL21(DE3) with a N-terminal His-tag. The plasmid obtained and predicted protein are schematised in Fig. 3.13. Cells transformed with the construct were induced for either 4 h at 37°C with 1 or 0.5 mM IPTG, or 4 h at 20°C with 1 or 0.5 mM IPTG. An over-expressed protein of 36 kDa, corresponding to the expected size for the truncated cathepsin B, could be detected in all cases, but it was found in the insoluble fraction of *E. coli* lysates if expressed at 37°C (results not shown). The highest rate of expression in a soluble manner was obtained by induction at 20°C with 0.5 mM IPTG, and the soluble protein produced was purified using Ni-agarose affinity chromatography (Fig. 3.14).

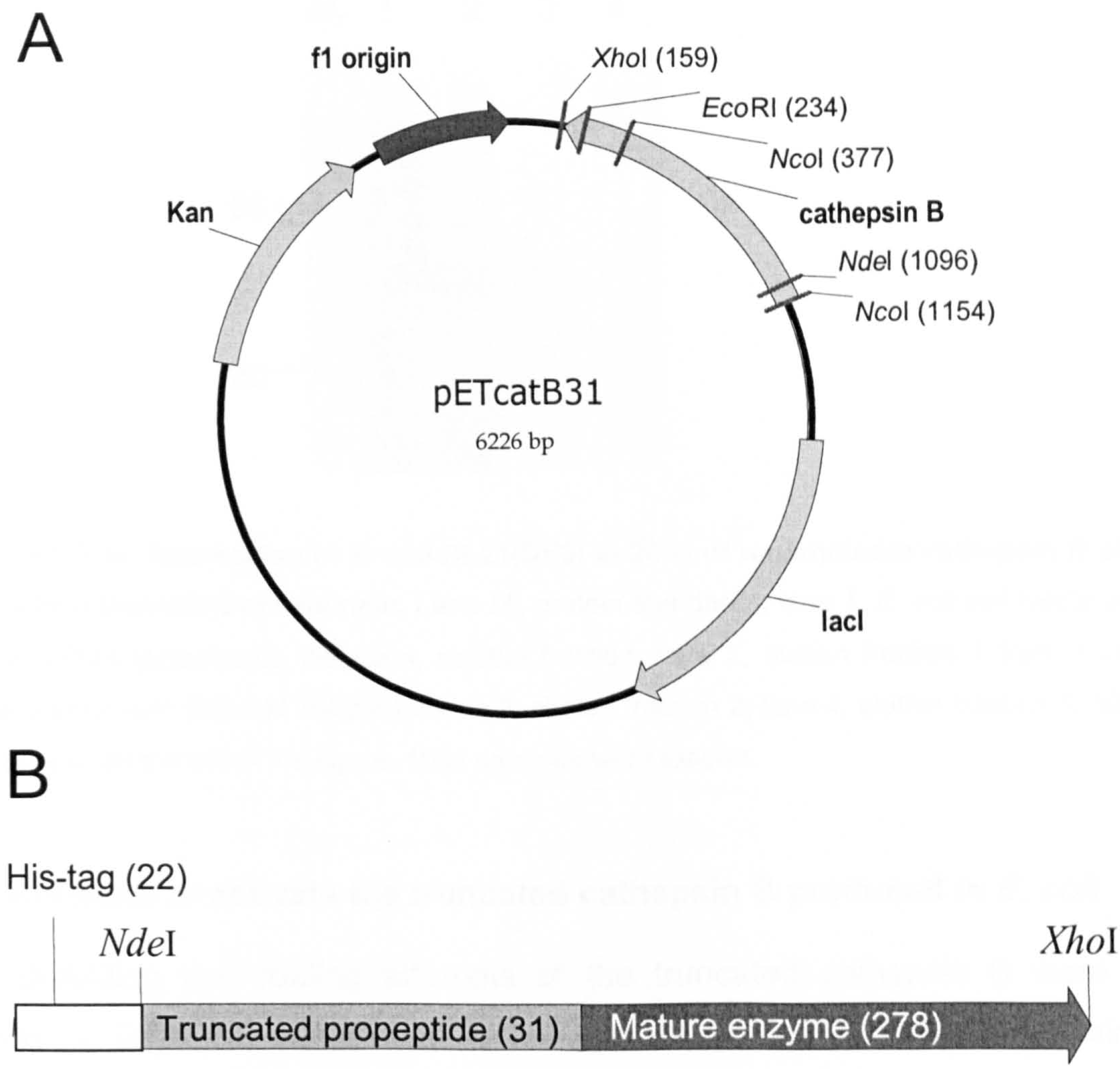


FIG. 3.13. **Construct for expression in *E. coli* BL21(DE3) of recombinant cathepsin B of *E. tenella* with 31 amino acids of pro-domain.** (A) Schematic representation of the plasmid resulting from the cloning of the sequence of the cathepsin B of *E. tenella* with 31 amino acids of the pro-domain in the pET28a(+) vector (Novagen) for expression of mature cathepsin B with a N-terminal His-tag and 31 amino acids of pro-domain. Kan, kanamycin resistance gene; f1 origin, origin of replication of the phage f1; lacI, gene encoding the *lac* repressor protein. (B) Schematic representation of the protein expressed from the pETcatB plasmid. The size in amino acids of the different fragments is indicated between brackets.

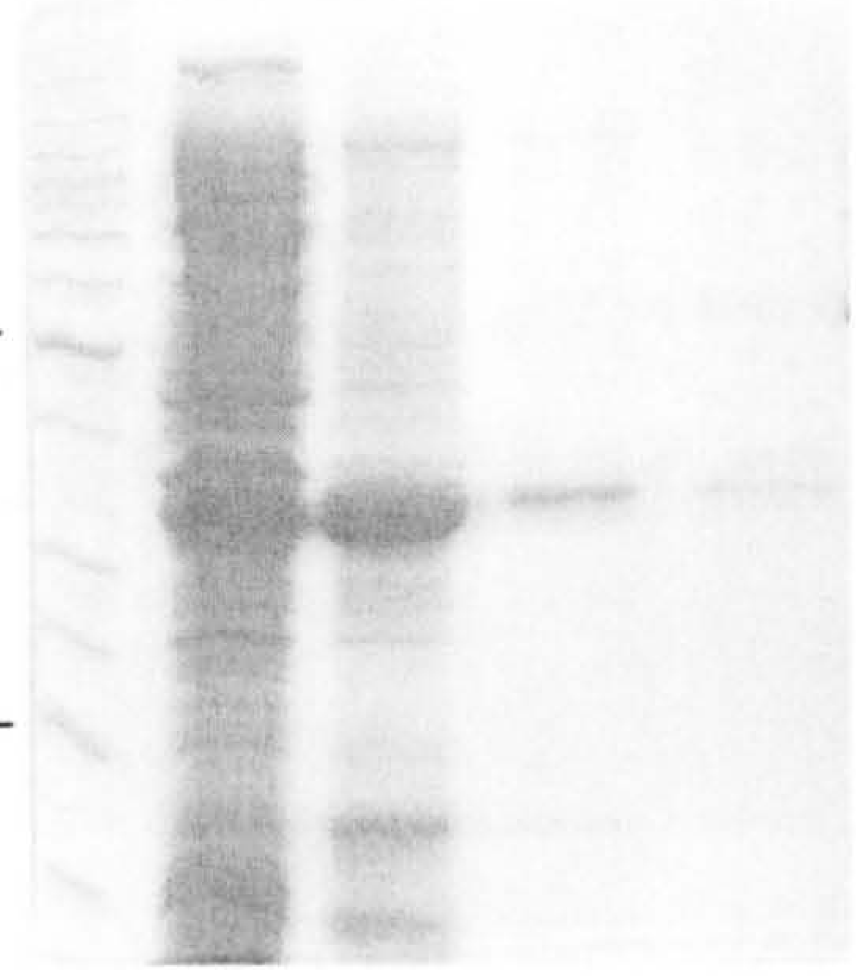


FIG. 3.14. **Expression in *E. coli* BL21(DE3) at 20°C of recombinant cathepsin B of *E. tenella* with a truncated pro-domain.** Lane M_r, protein standards; lane 1, *E. coli* cell lysate after isopropyl β-D-thiogalactoside induction, soluble fraction; lane 2, elution fraction 1 from Nickel-agarose column with 500 mM imidazole; lane 3, elution fraction 2; lane 4, elution fraction 3. 10^{-3} x M_r is indicated on the left of the figure. 10 µl samples were loaded.

3.2.6. Attempts to activate the truncated cathepsin B produced in *E. coli*

Unfolding and folding attempts of the truncated cathepsin B were as described in Table 2.2. Samples of Nickel-agarose purified truncated cathepsin B diluted 1:1 (v/v) in 0.1 M sodium acetate buffer pH 5.0 or 0.1 M Tris/HCl buffer pH 8.0 supplemented with 20 mM DTT and incubated at 37°C for 20 min showed a shift in size of about 2 kDa (Fig. 3.15). Precipitation was observed in the samples incubated at pH 5.0, which resulted in lower amount of protein being visible on the gel. The shift was not inhibited when the samples were incubated in the presence of 300 µM of E64 (Fig. 3.15).

Analysis on western blot using anti-cathepsin B antibodies detected a single protein of 34 kDa, 2 kDa short compared to the truncated cathepsin B sample not incubated in the described conditions (Fig. 3.16). Using anti-His-tag antibodies, the protein at 34 kDa could still be detected (Fig. 3.17), suggesting some processing of the protein, but not at the N-terminus. The samples incubated at pH 5.0 or pH 8.0 were loaded onto gelatin gels and post-electrophoresis and washing the gels were incubated at pH 5.0 and pH 8.0 with DTT, but no activity could be detected. The same samples were assayed at pH 5.0 or 8.0 with 10 mM DTT against Z-Phe-Arg-pNA, but no activity was detected.

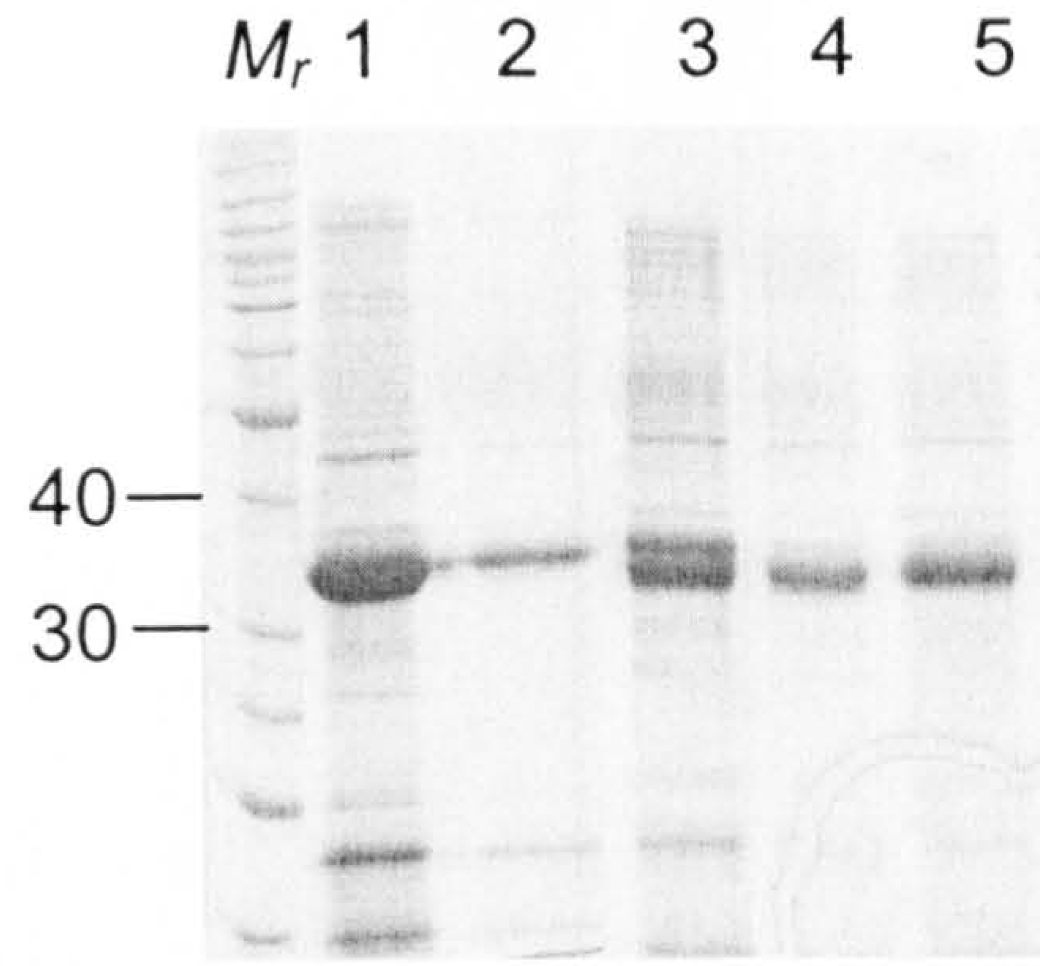


FIG. 3.15. **Analysis of the processing of the truncated cathepsin B of *E. tenella* expressed in *E. coli*.** SDS-PAGE gel (12% acrylamide) stained with Coomassie blue. Lane M_r, protein standards; lane 1, 10 μg of purified protein; lane 2, 3 μg of the purified protein; lane 3, 10 μg of purified protein left 18 h at 4°C after purification; lane 4, 10 μg of purified truncated cathepsin B incubated in 0.1 M Tris/HCl pH 8.0, 10 mM DTT at 37°C for 20 min; lane 5, 10 μg of purified truncated cathepsin B incubated in 0.1 M Tris/HCl pH 8.0, 10 mM DTT at 37°C for 20 min with 300 μM E64 . 10^{-3} x M_r is indicated on the left of the figure.

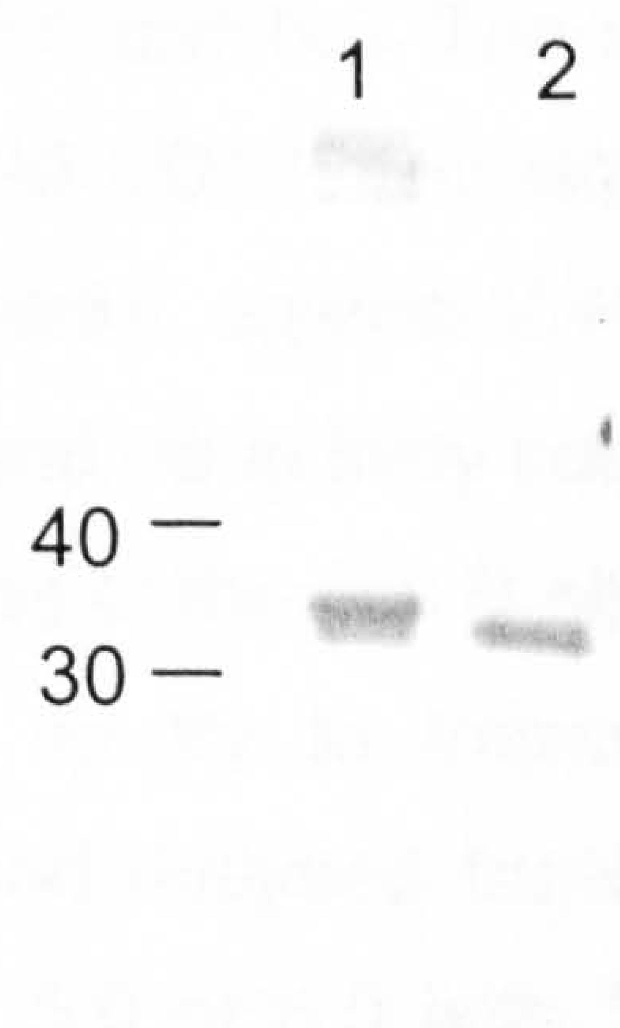


FIG. 3.16. **Detection of the truncated cathepsin B of *E. tenella* expressed in *E. coli* with anti-cathepsin B antibodies.** Antibody dilution of 1:5000 (v/v). Membrane was exposed 30 seconds. Lane 1, 10 μg of purified protein; lane 2, 10 μg of purified truncated cathepsin B incubated in 0.1 M Tris/HCl pH 8.0, 10 mM DTT at 37°C for 20 min. 10^{-3} x M_r is indicated on the left of the figure.

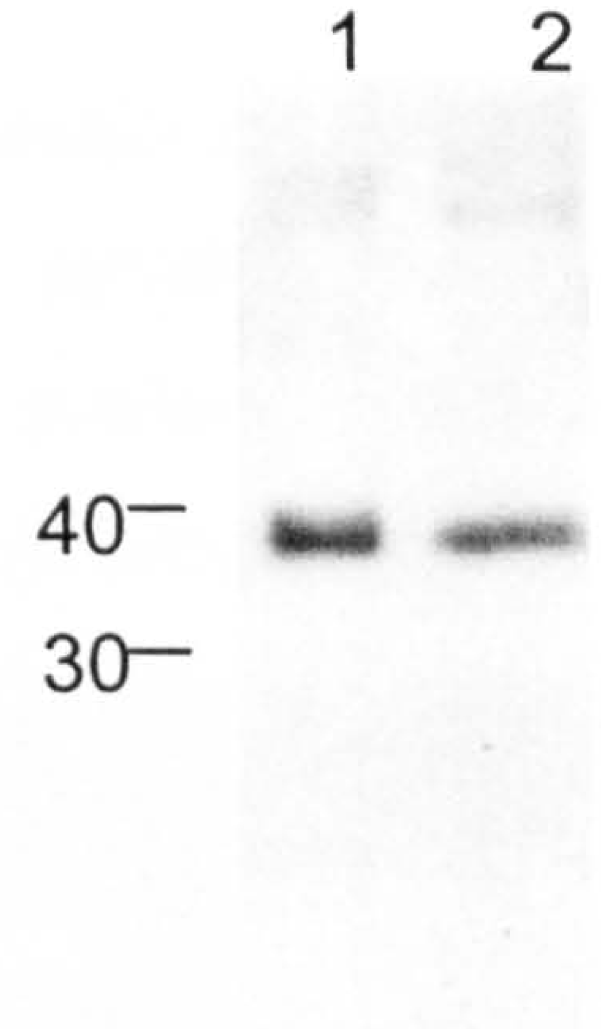


FIG. 3.17. **Detection of the truncated cathepsin B of *E. tenella* expressed in *E. coli* with anti-His-tag antibodies.** Antibody was used at a dilution of 1:2000 (v/v). Membrane was exposed 30 seconds. Lane 1, 10 μg of purified protein; lane 2, 10 μg of purified truncated cathepsin B incubated in 0.1 M Tris/HCl pH 8.0, 10 mM DTT at 37°C for 20 min. $10^{-3} \times M_r$ is indicated on the left of the figure.

Truncated cathepsin B purified on Ni-agarose was directly assayed at pH 4.0, 5.0, 6.0 or 7.0 in the presence of 5 mM EDTA, 10 mM DTT and 0.9 M NaCl against Z-Phe-Arg-AMC, but no activity could be detected. The protein fell out of solution rapidly at pH 4.0, 5.0 and 6.0. The same sample was then assayed at pH 5 or 7 with or without 5 mM EDTA and with or without 10 mM DTT, to test the influence of these components, against Z-Phe-Arg-AMC or Z-Arg-Arg-AMC. The protein fell out of solution and no activity could be detected.

The soluble truncated cathepsin B obtained was also treated to unfold and slowly refold the protein, to try to improve the folding and disulfide bridge formation. The unfolded and dialysed truncated cathepsin B (Sanderson *et al.*, 2000) was assayed at pH 5.0 or 8.0 with 10 mM DTT against Z-Phe-Arg-AMC, but no activity could be detected.

3.2.7. Non-specific processing of the soluble recombinant pro-mature cathepsin B expressed in *E. coli* BL21(DE3)

The idea of this approach was that a correctly folded protein will have primarily its loops and hinges accessible to proteolytic cleavage in mild conditions. Using low enzyme/substrate ratio and short incubation time, digestions with enzymes of wide substrate specificity might allow the separation of different domains; i.e. the separation of the pro-domain from the mature domain of the enzyme. Based on the structure of the human cathepsin B (Fig. 3.18), the domain at the junction between the pro- and the mature domains seems accessible to the solvent. Provided the folding might be similar in the cathepsin B of *E. tenella*, cleavage might be possible at that junction.

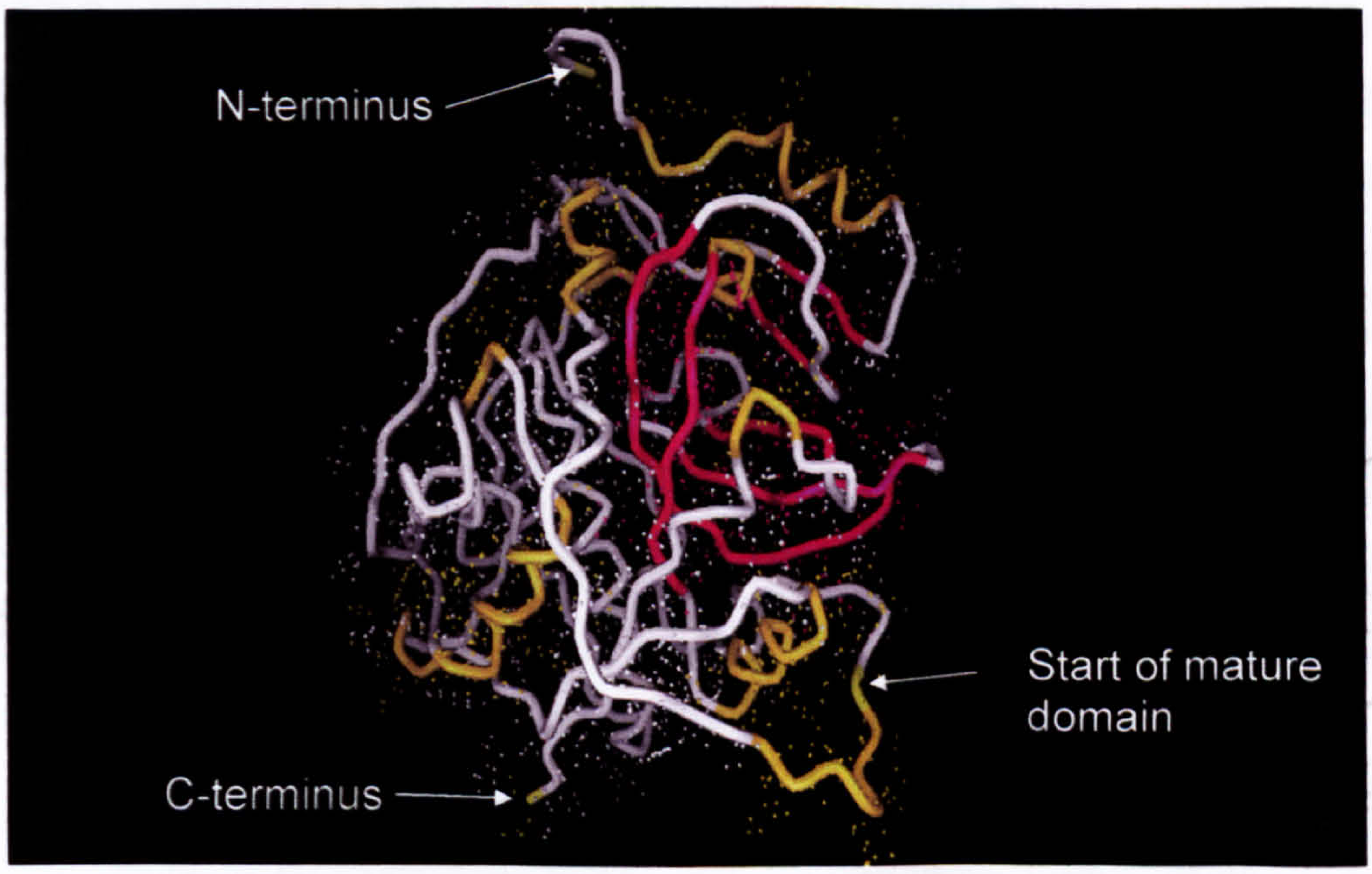


FIG. 3.18. **Crystal structure of human procathepsin B at 3.3 Angstrom resolution (Turk *et al.*, 1996).** α-helixes are in orange, β-sheets are in pink and strands are in grey. Residues at the N-, and C-termini and at the start of the mature domain are in yellow and marked by arrows.

An analysis using SDS-PAGE of a trypsin digest of the partially purified (on Ni-agarose) pro-mature cathepsin B using enzyme/substrate ratios of 1:500 or 1:5000 is shown on Fig. 3.19A. The pro-mature cathepsin B seems to be reasonably quickly degraded into fragments. In particular fragments at 31, 29, 20 and 16 kDa seemed to increase in amount over time. The sample corresponding to 30 min incubation with a 1:5000 ratio of trypsin seemed to yield a good representation of these fragments and was therefore chosen for further analyses.

A chymotrypsin digest analysis at ratios of 1:500 and 1:5000 is shown on Fig. 3.19B. The ratio of 1:5000 seemed to be again adequate for analysis as several fragments (at 34, 28, 18, 16 kDa, for example) seemed to increase in amount over time in a sensible time for analysis.

In the case of the use of subtilisin, the ratio had to be adjusted to 1:50000, as at a ratio of 1:5000, practically all cathepsin B was degraded within 2 min of incubation (Fig. 3.19C). Analysis of a digest at ratio 1:50000 is presented on Fig. 3.19D. Within 30 to 60 min of incubation, 2 major fragments at 42 and 28 kDa seemed to clearly increase in amount.

Samples of cathepsin B incubated either with trypsin 1:5000 for 30 min, chymotrypsin 1:5000 for 60 min, or subtilisin 1:50000 for 45 min were analysed using gelatin SDS-PAGE, but no activity could be detected when gels were incubated at pH 5.0 or 8.0 with 20 mM DTT. The same samples were tested for activity against Z-Phe-Arg-pNA at pH 5.0, 7.0 or 8.0 +/- 5 mM EDTA, +/- 10 mM DTT. No activity could be detected except in the samples with trypsin at pH 7.0 or 8.0. This activity was not E64 inhibited (1 mM) but could be inhibited with TLCK (0.1 mg/ml), an inhibitor of serine peptidases, at pH 7.0 but not 8.0. Knowing TLCK is not stable at pH above 7.5, it can be concluded that the activity detected was due to trypsin.

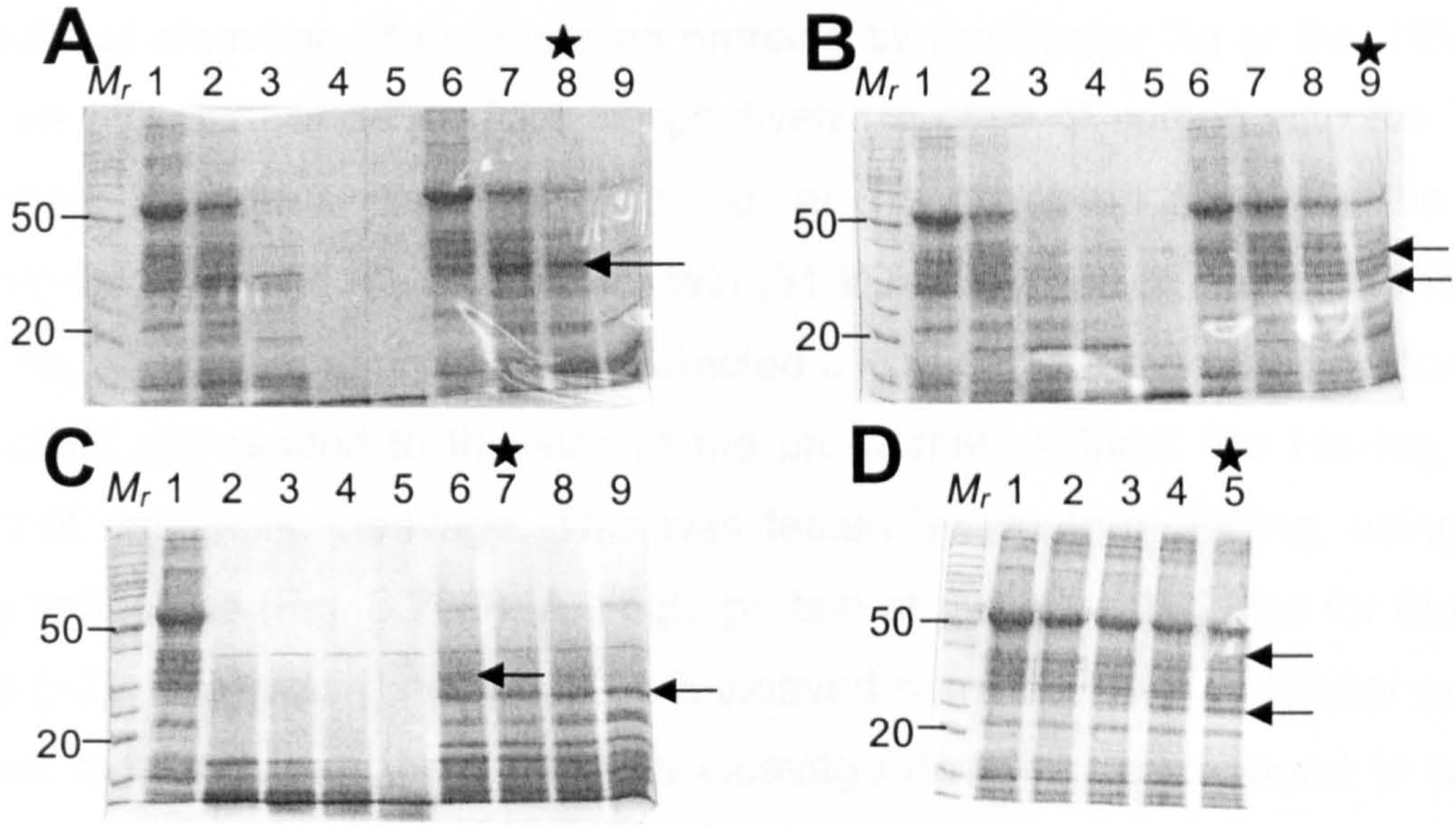


FIG. 3.19. **Analysis of the non-specific cleavage of the recombinant N-terminal His-tagged cathepsin B of *E. tenella* produced in *E. coli*.** 40 μg of cathepsin B were used per test. Cleavages were at room temperature and analysed using SDS-PAGE (12% acrylamide) and staining with Coomassie blue. Lane M_r, protein standards. Arrows mark bands appearing with time. 10^{-3} x M_r is indicated on the left of the figure. (A) Cleavage with trypsin. Trypsin ratio was of 1:500 (w/w) (lanes 1 to 5) or 1: 5000 (w/w) (lanes 6 to 9). Lane 1, incubation 0 min; lanes 2 and 6, 2 min; lanes 3 and 7, 15 min; lanes 4 and 8, 30 min; lanes 6 and 9,1 h. (B) Cleavage with chymotrypsin. Same labelling as in (A). (C) Cleavage with subtilisin. Same labelling as in (A). (D) Cleavage with subtilisin, ratio 1:50000. Lane 1, incubation of 0 min; lane 2, 2 min; lane 3, 15 min; lane 4, 30 min; lane 5, 1 h. Samples marked with a star were further analysed by gelatin SDS-PAGE and tested for activity.

3.2.8. Specific processing of the recombinant pro-mature cathepsin B

The idea of this approach was to introduce various very specific cleavage sites for commercially available enzymes at the limit between the predicted pro- and mature domains, predictions based on the domains of the *T. gondii* cathepsin B (Que *et al.*, 2002). This would be done by site-directed mutagenesis on the existing complete cathepsin B expression construct in pET28a(+). The hypothesis was that it might allow for easy removal of the pro-domain and activation of the enzyme, provided the soluble enzyme was properly folded.

Thrombin and enterokinase did cleave the non-mutated cathepsin B, whereas the Factor Xa and the TEV protease did not, and mutagenesis to introduce the specific cleavages sites for these 2 peptidases was performed. The resulting proteins were expressed in a soluble manner in *E. coli* BL21(DE3) at 25°C and purified on nickel-agarose column. A scheme of the method used and the results of cleavage of the mutated proteins by the Factor Xa or the TEV are shown on Fig. 3.20 and Fig. 3.21, respectively. In case of cutting with the TEV, two protein fragments could be detected, at the predicted sizes for the pro-domain (~26 kDa) and the mature domain (31 kDa). In case of cleavage with the Factor Xa, an additional protein was detected 2 kDa smaller than the pro-domain, which could correspond to the size of the pro-domain without the His-tag, or a product of unspecific cleavage. This was tested by western-blotting, using anti His-tag antibodies (Fig. 3.20D). A single protein at the expected size for the pro-domain (~26 kDa) was detected in both cleaved samples. The additional protein fragment resulting from the Factor Xa cleavage does not correspond to a His-tagged protein.

The cleaved samples were analysed using gelatin SDS-PAGE with incubation at pH 5.0 or 8.0, but no activity could be detected. Based on the possibility that the enzyme might get inactivated irreversibly during the electrophoresis, the azocasein assay was used. The activity was tested at pH 5.0 or 8.0 with 24 h incubation at 37°C. From the measurements taken, no real positive cleavage of the azocasein occurred compared to the negative control (unprocessed recombinant cathepsin B of *E. tenella*) (results not shown). The same samples were also tested for activity continuously for 30 min against Z-Phe-Arg-AMC but no cleavage of the substrate could be detected.

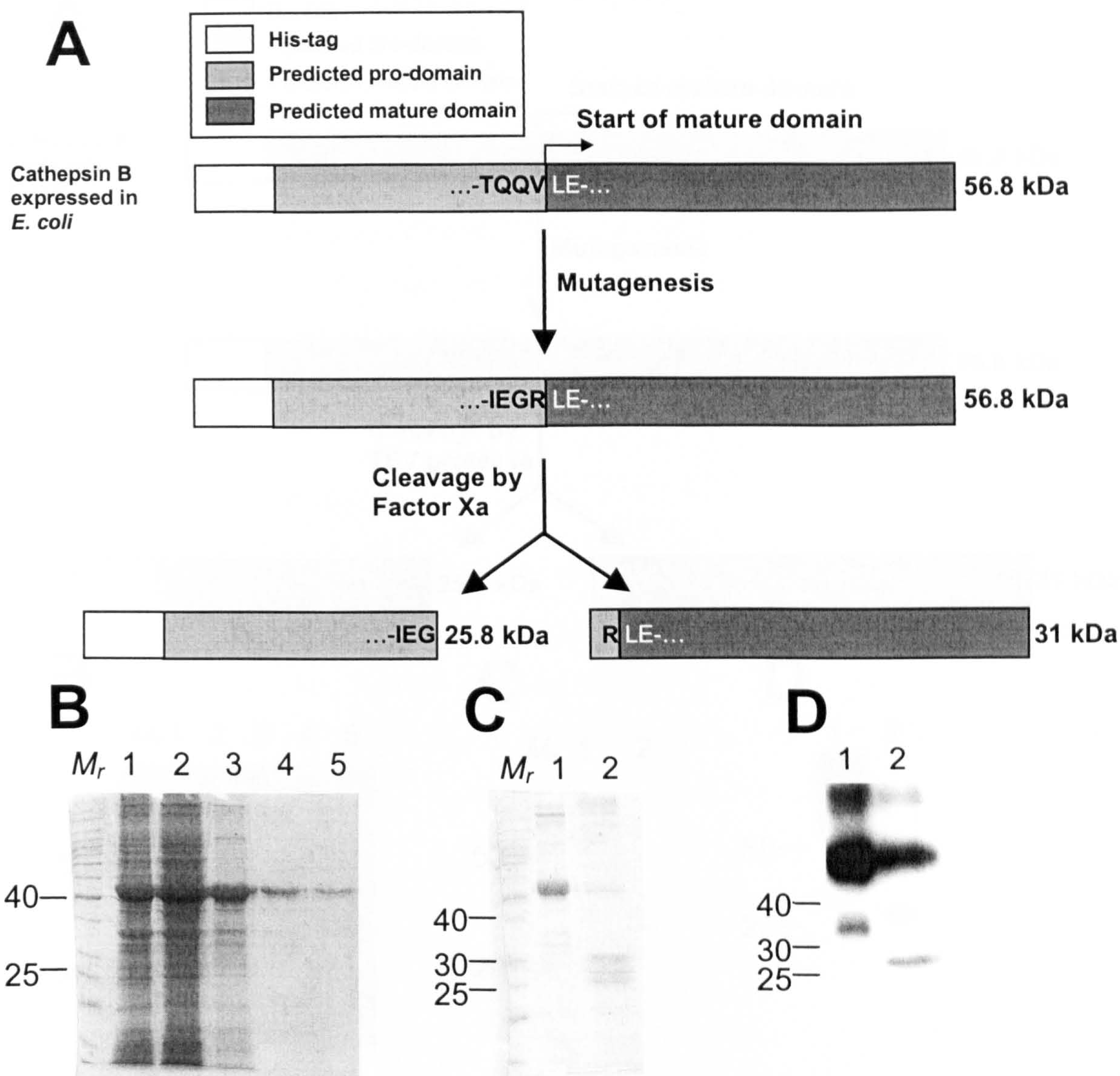


FIG. 3.20. **Mutagenesis for specific cleavage by the Factor Xa at the predicted junction between the pro- and the mature domains of the recombinant cathepsin B of *E. tenella* expressed in *E. coli*.** (A) Schematic representation of the method used. Amino acids are represented by the single letter code. Sizes of fragments are not to scale. (B) Expression of recombinant N-terminal His-tagged protein carrying the mutation for the Factor Xa cleavage site. Lane M_r, protein standards; lane 1, *E. coli* cell lysate after IPTG induction 4 h at 25°C, insoluble fraction; lane 2, *E. coli* cell lysate, soluble fraction; lane 3, elution fraction 1 from Nickel-agarose column; lane 4, elution fraction 2; lane 5, elution fraction 3. 10^{-3} x M_r is indicated on the left of the figure. 10 µl samples were loaded. (C) Analysis of cleavage (overnight) of recombinant cathepsin B (40 µg) with Factor Xa (1 Unit) at room temperature using SDS-PAGE (12% acrylamide) and staining with Coomassie blue. Lane 1, non-mutated cathepsin; lane 2, mutated cathepsin B. (D) Detection of His-tagged proteins on gel on figure C. Anti His-tag antibody dilution of 1:2000 (v/v). Membrane exposure of 1 min. Lane 1, non-mutated cathepsin; lane 2, mutated cathepsin B.

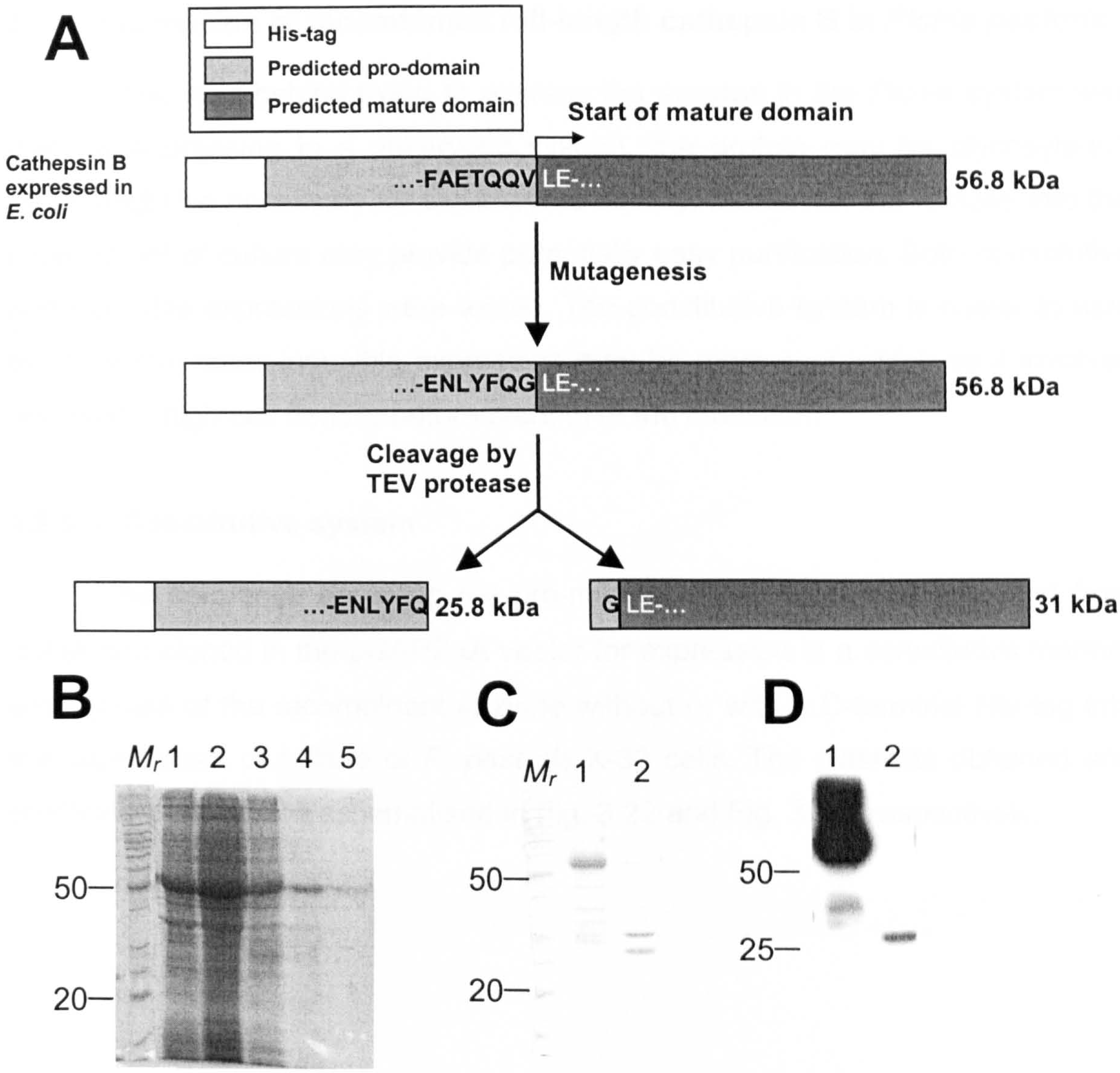


FIG. 3.21. **Mutagenesis for specific cleavage by the TEV (Tabacco etch virus) NIa protease at the predicted junction between the pro- and the mature domains of the recombinant cathepsin B of *E. tenella* expressed in *E. coli*.** (A) Schematic representation of the method used. Amino acids are represented by the single letter code. Sizes of fragments are not to scale. (B) Expression of recombinant N-terminal His-tagged protein carrying the mutation for the TEV protease cleavage site. Lane M_r, protein standards; lane 1, *E. coli* cell lysate after IPTG induction 4 h at 25°C, insoluble fraction; lane 2, *E. coli* cell lysate, soluble fraction; lane 3, elution fraction 1 from Nickel-agarose column; lane 4, elution fraction 2; lane 5, elution fraction 3. $10^{-3} \times M_r$ is indicated on the left of the figure. 10 µl samples were loaded. (C) Analysis of cleavage (overnight) of recombinant cathepsin B (40 µg) with TEV protease (10 Units) at room temperature using SDS-PAGE (12% acrylamide) and staining with Coomassie blue. Lane 1, non-mutated cathepsin B; lane 2, mutated cathepsin B. (D) Detection of His-tagged proteins on gel on figure C. Anti His-tag antibody dilution of 1:2000 (v/v). Membrane was exposed 1 min. Lane 1, non-mutated cathepsin B; lane 2, mutated cathepsin B.

3.2.9. Expression of recombinant full-length cathepsin B in *Pichia pastoris*

The idea behind trying to express the enzyme in the *Pichia* system was that, by expressing in a eukaryotic system, the protein may be glycosylated, which might be necessary for the enzyme activity. Moreover, the release into the supernatant of culture may provide potentially easy purification. Both constitutive and inducible expressions were tested. The constitutive system is easier to use, but in some cases inducible expression may be more successful, as it involves reaching a high cell density before starting of the induction.

3.2.9.1. Constitutive system

The sequence encoding the pro-mature cathepsin B was amplified from cDNA and cloned in the pGAPZαA vector for expression in a constitutive manner and release of the recombinant enzyme without or with a C-terminal His-tag into the supernatant of culture of *P. pastoris* X-33 cells. The plasmids obtained and predicted proteins are schematised in Fig. 3.22 and Fig. 3.23, respectively.

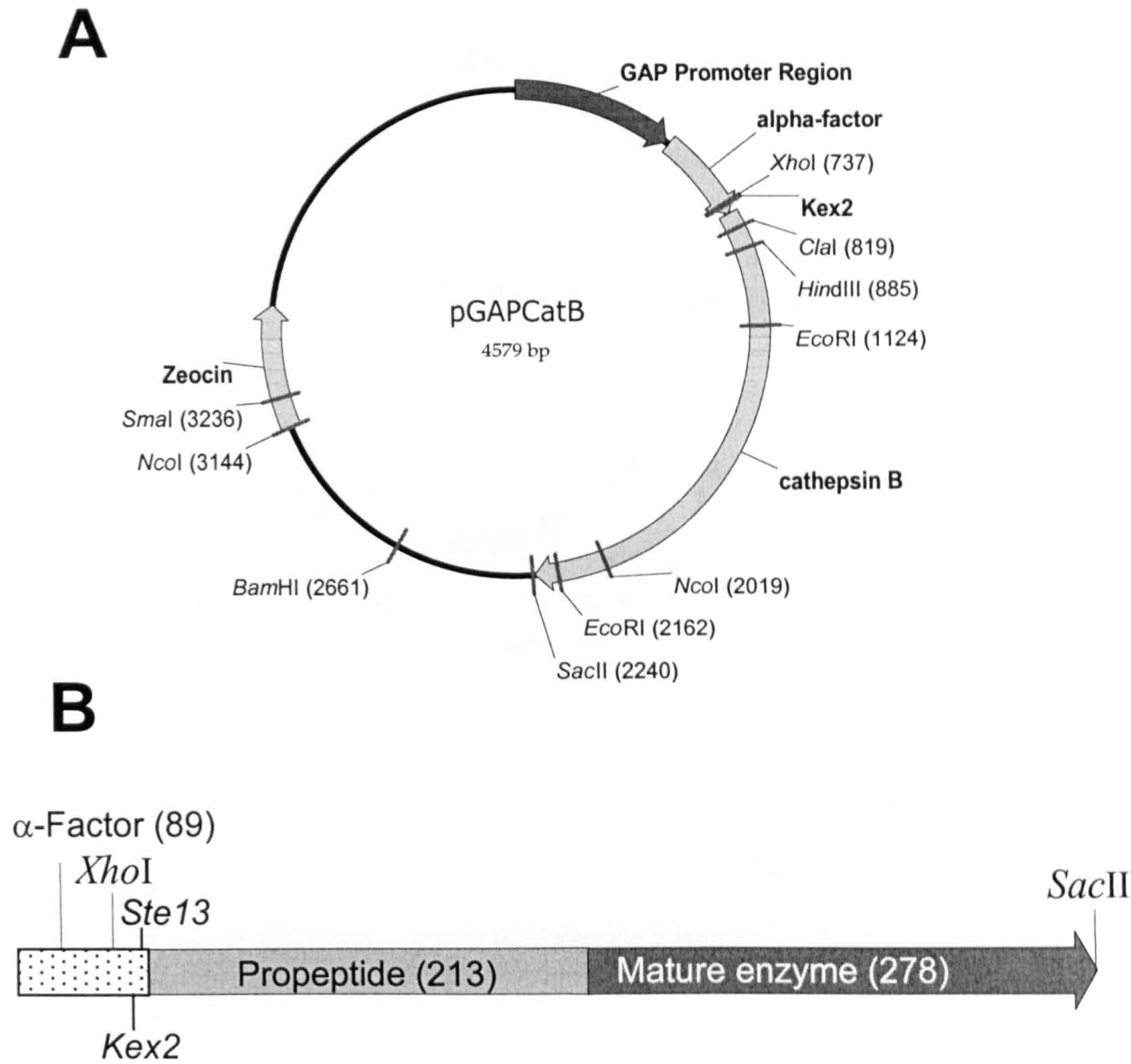


FIG. 3.22. **Construct for expression of recombinant pro-cathepsin B of *E. tenella* in *Pichia pastoris*, constitutive system.** (A) Schematic representation of the plasmid resulting from the cloning of the pro-mature sequence of the cathepsin B of *E. tenella* in the pGAPZα vector for constitutive expression of pro-mature cathepsin B in the supernatant of culture. (B) Schematic representation of the protein expressed from the pGAPcatB plasmid. The size in amino acids of the different fragments is indicated between brackets.

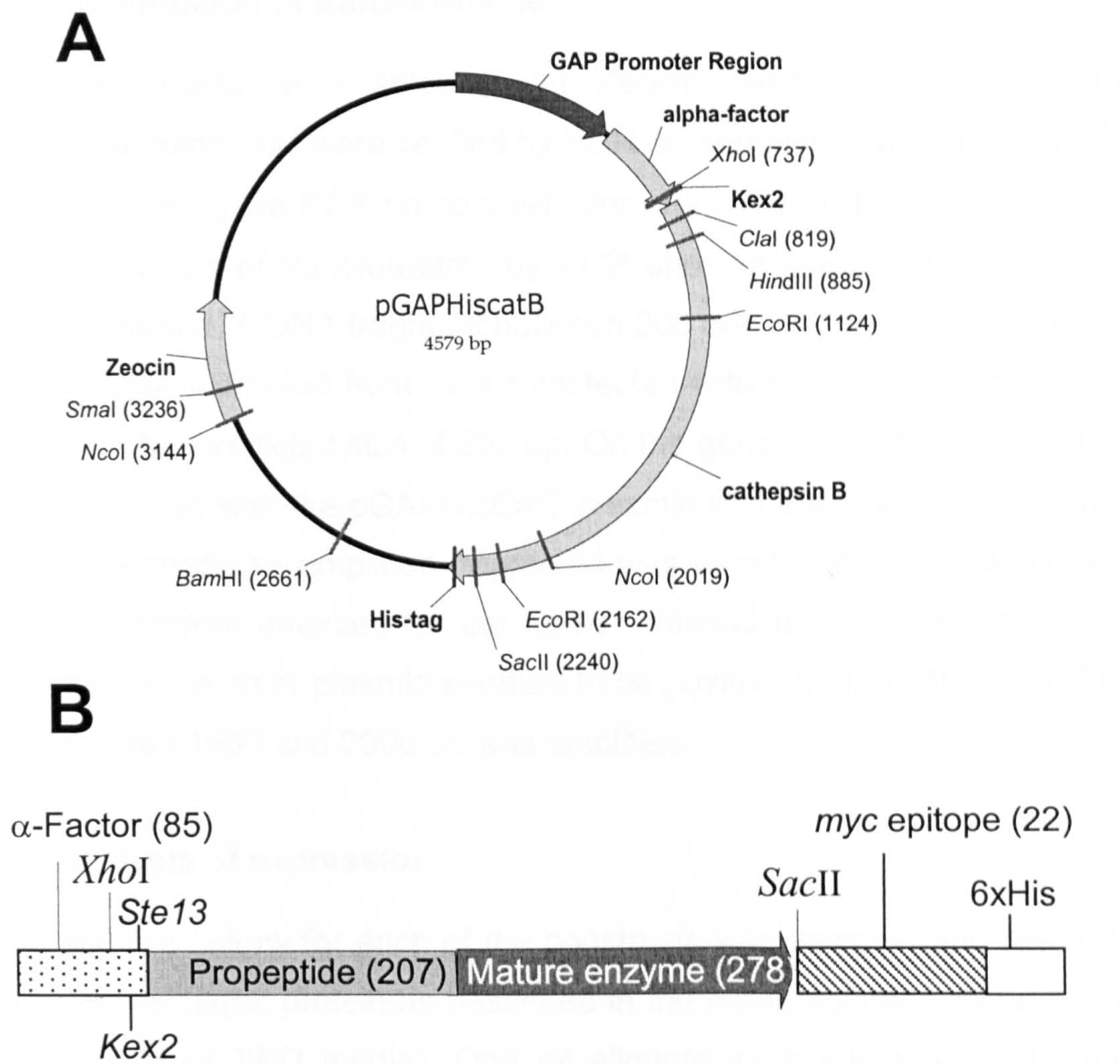


FIG. 3.23. **Construct for expression of recombinant pro-cathepsin B of *E. tenella* with a C-terminal His-tag in *Pichia pastoris*, constitutive system.** (A) Schematic representation of the plasmid resulting from the cloning of the pro-mature sequence of the cathepsin B of *E. tenella* in the pGAPZα vector for constitutive expression of pro-mature cathepsin B in the supernatant of culture. (B) Schematic representation of the protein expressed from the pGAPcatB plasmid. The size in amino acids of the different fragments is indicated between brackets.

3.2.9.1.1. Confirmation of transformants

Transformants were selected on Zeocin plates (100 μg/ml), and insertions of the fragments were verified by PCR on genomic DNA extracted from positive clones, using the PCR primers α-factor and 3'AOX1. Results are shown in Fig 3.24. Analysis of transformants by PCR showed integration for all three plasmids transfected. A DNA fragment between 200 and 300 bp was amplified on the genomic DNA extracted from cells transfected with the empty vector, which corresponds to the predicted size of 295 bp. On the genomic DNA extracted from the cells transfected with the pGAPHisCatB plasmid and the pGAPCatB plasmid, the size of the predicted amplified fragments was about 1800 bp. Two positive clones were therefore detected for the former plasmid (Fig. 3.24A) and all the clones tested for the latter plasmid seemed to be positive (Fig. 3.24B), as a DNA fragment between 1650 and 2000 bp was amplified.

3.2.9.1.2. Analysis of expression

One positive colony for each of the constructs was selected and tested for expression of the target protein as described in the manufacturer's manual (final volume of 50 ml of YPD media). One ml aliquots were taken a regular time intervals, cells were separated from the supernatant of culture by centrifugation and samples of supernatant and cell lysates were used for analysis using gelatin SDS-PAGE, the azocasein assay, and either with Coomassie blue-stained SDS-PAGE or western blot analysis.

Analysis of the supernatant of culture of the *Pichia* cells having integrated the empty vector or the pGAPHisCatB or pGAPCatB plasmids showed in all cases the presence of a faint single protein between 40 and 50 kDa after 24 to 48 h of culture (results not shown). This protein is unlikely to be related to the protein of interest, as it is expressed in the negative control cells that had integrated the empty vector. No protein at the expected size for the pro-mature cathepsin B (55.4 kDa) could be detected in crude supernatant or in the cell extracts by SDS-PAGE (results not shown). By western-blot analysis, a protein corresponding to the expected size for the pro-mature cathepsin B at about 55 kDa could be detected in the crude supernatant of culture and the cell extracts of the *Pichia* cells transfected with the pGAPCatB and pGAPHisCatB plasmids but not in the control cells transfected with the empty vector (Fig. 3.25). Signs of processing could also be detected as smaller bands could be visualised, especially in the supernatant of culture of the cells transfected with the pGAPCatB plasmid. The proteins trapped inside the cells were of less interest, as purification would be more difficult.

By TCA precipitation of a small volume of supernatant of culture of *Pichia* cells transfected with the pGAPHisCatB and the pGAPCatB plasmids, the protein between 40 and 50 kDa was more easily detected by SDS-PAGE (Fig. 3.26). However, it does not seem to be related to the cathepsin B, as it was not recognised with anti-cathepsin B antibodies (Fig. 3.25). An additional protein at about 55 kDa, corresponding to the expected size for the pro-mature cathepsin B could also faintly be detected after 48 to 72 h of culture (Fig. 3.26), and the expression seemed stable for at least 8 days (results not shown). By western-blot analysis of the TCA-precipitated supernatant of culture of *Pichia* cells transfected with the pGAPHisCatB plasmid and the pGAPCatB plasmid, the results obtained by SDS-PAGE were confirmed. That is, a 55 kDa band corresponding to the predicted size for the pro-mature cathepsin B could be detected after 48 h of culture (Fig. 3.27). Moreover, signs of processing of the protein could be observed in the supernatant of cells transfected with the pGAPCatB plasmid (Fig. 3.27).

Aliquots of supernatant of the cultures described above were analysed using gelatin SDS-PAGE with incubation at pH 5.0 and pH 8.0 for 24 h at 37°C, but no difference in activity could be detected between the cathespin B-expressing cells and the negative control cells. The same pattern of proteins was detected in all cases (proteins were in high molecular weight range, above 50 kDa). 50 µl of supernatant of culture were tested as well for activity using the azocasein assay, but no activity could specifically be detected compared to the control cells supernatant. The activity detected on azocasein was in all cases DTT inhibited and not E64 inhibited (the activity was probably due to the release of a peptidase into the supernatant of culture not related to the cathepsin B) (Table 3.1). This peptidase activity did degrade the cathepsin B.

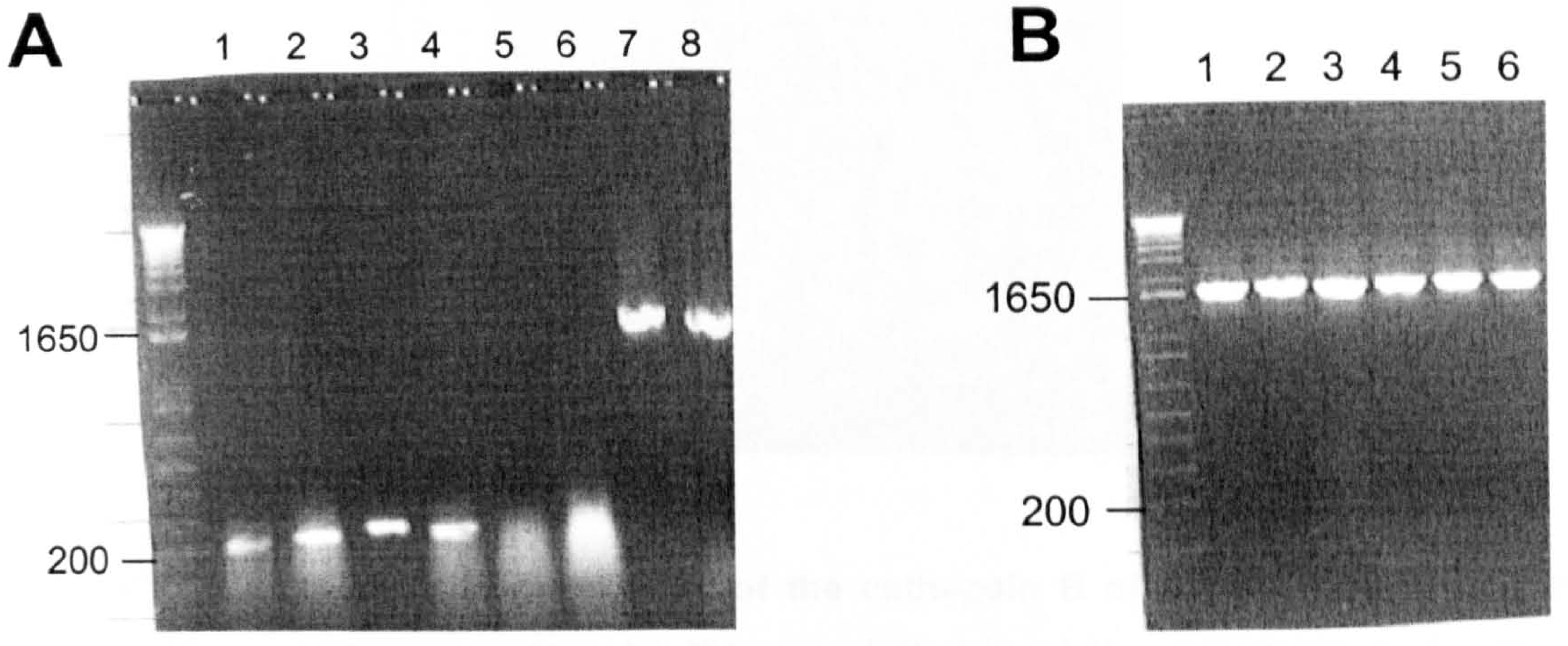


FIG. 3.24. **Verification of insertion of the pGAPHiscatB and pGAPcatB constructs in the *Pichia* X-33 genome.** PCR analysis of *Pichia* transformants, using1% agarose gel and ethidium bromide staining. (A) Cells transfected with the pGAPZαA empty vector (lanes 1 to 4), or with the pGAPHiscatB plasmid (lanes 5 to 8). (B) Cells transfected with the pGAPcatB plasmid. Size in bp is indicated on the left of the figure. M_r corresponds to molecular mass in kDa.

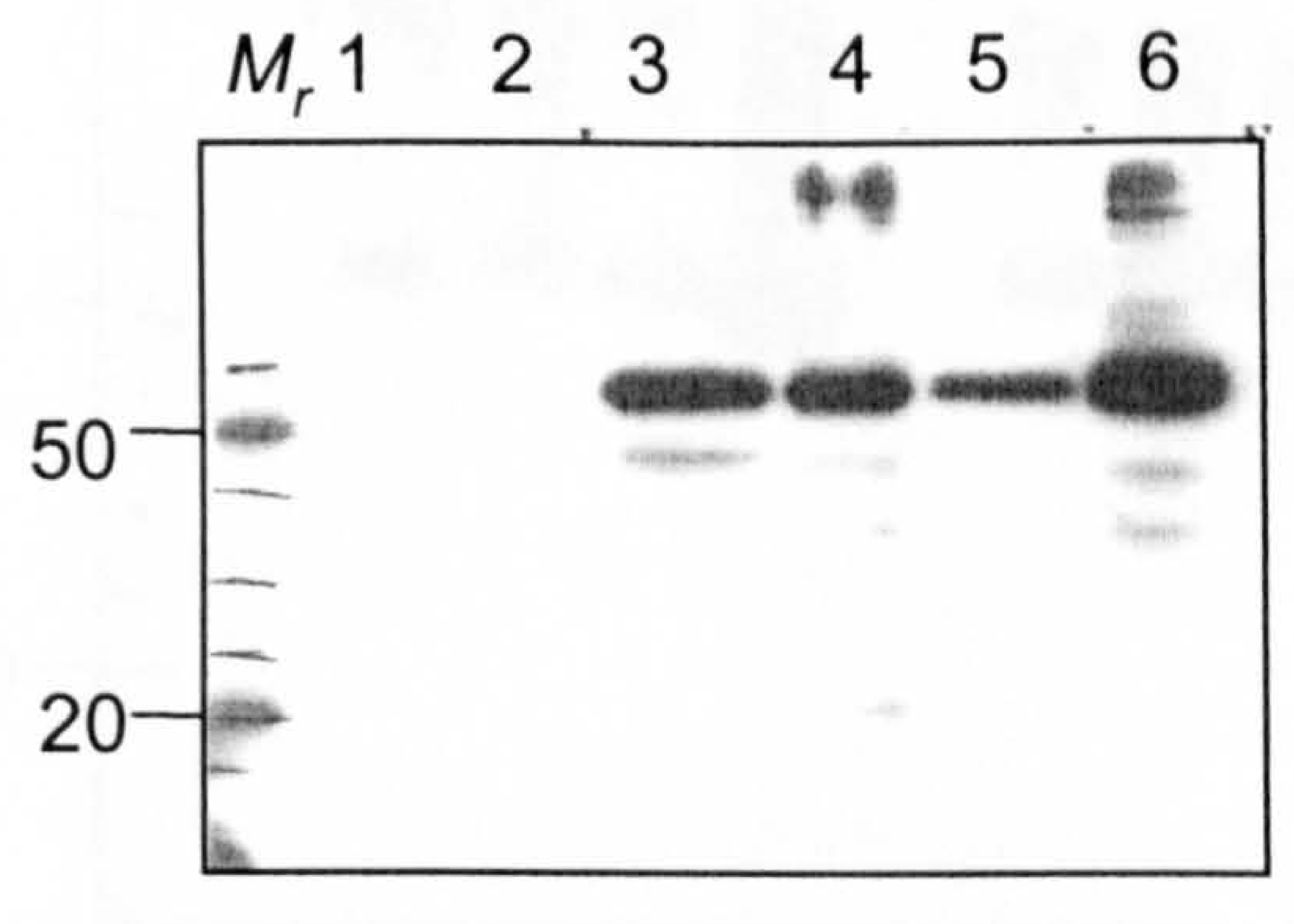


FIG. 3.25. **Analysis of expression of the cathepsin B of *Eimeria tenella* in *Pichia pastoris* X-33, constitutive system.** Analysis by western blot (1:8000 dilution of last bleed anti cathepsin B antiserum; 1:5000 dilution of anti-rabbit antibody; 2 min exposure) of the supernatant of culture (lanes 1, 3 and 5) and the *Pichia* cell extracts (lanes 2, 4 and 6) transfected with the pGAPZαA vector (lanes 1 and 2), the pGAPcatB plasmid (lanes 3 and 4) and the pGAPHiscatB plasmid (lanes 5 and 6) at t= 72 h. 10 μl of suspension were loaded. M_r corresponds to molecular mass in kDa.

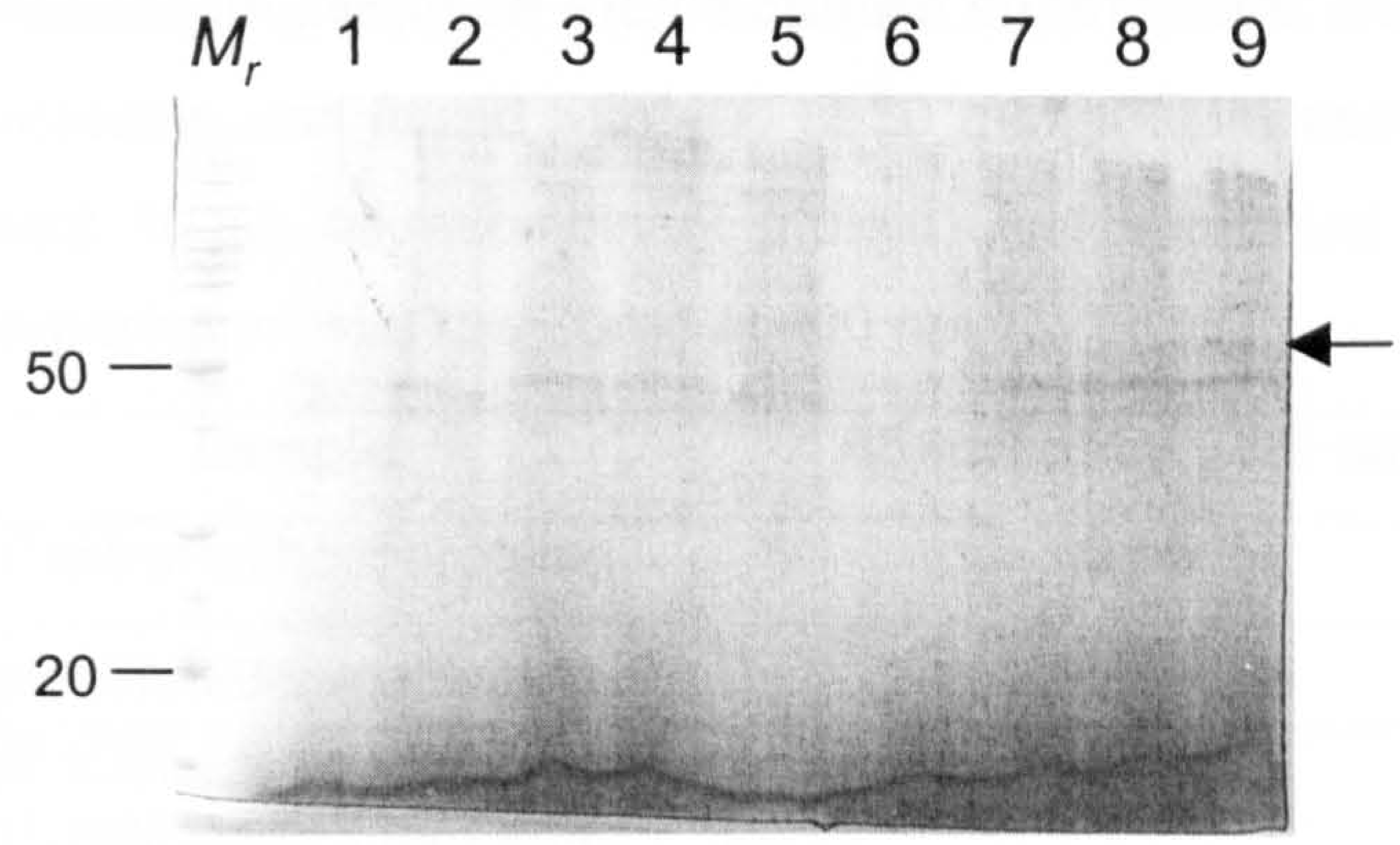


FIG. 3.26. **Analysis of expression of the cathepsin B of *Eimeria tenella* in *Pichia pastoris* X-33, constitutive system, by TCA precipitation of the supernatant of cultures.** Analysis on post-electrophoresis Coomassie blue-stained 12% (w/v) SDS-PAGE gel of the equivalent of 0.2 ml of TCA-precipitated supernatant of culture of *Pichia* cells transfected with the pGAPHisCatB plasmid (lanes 1 to 5) and the pGAPCatB plasmid (lanes 6 to 9). Lane 1, 24 h; lane 2, 48 h; lane 3, 72 h; lane 4, 96 h; lane 5, 168 h; lane 6, 24 h; lane 7, 48 h; lane 8, 72 h; lane 9, 96 h. The arrow indicates the 55 kDa protein. M_r corresponds to molecular mass in kDa.

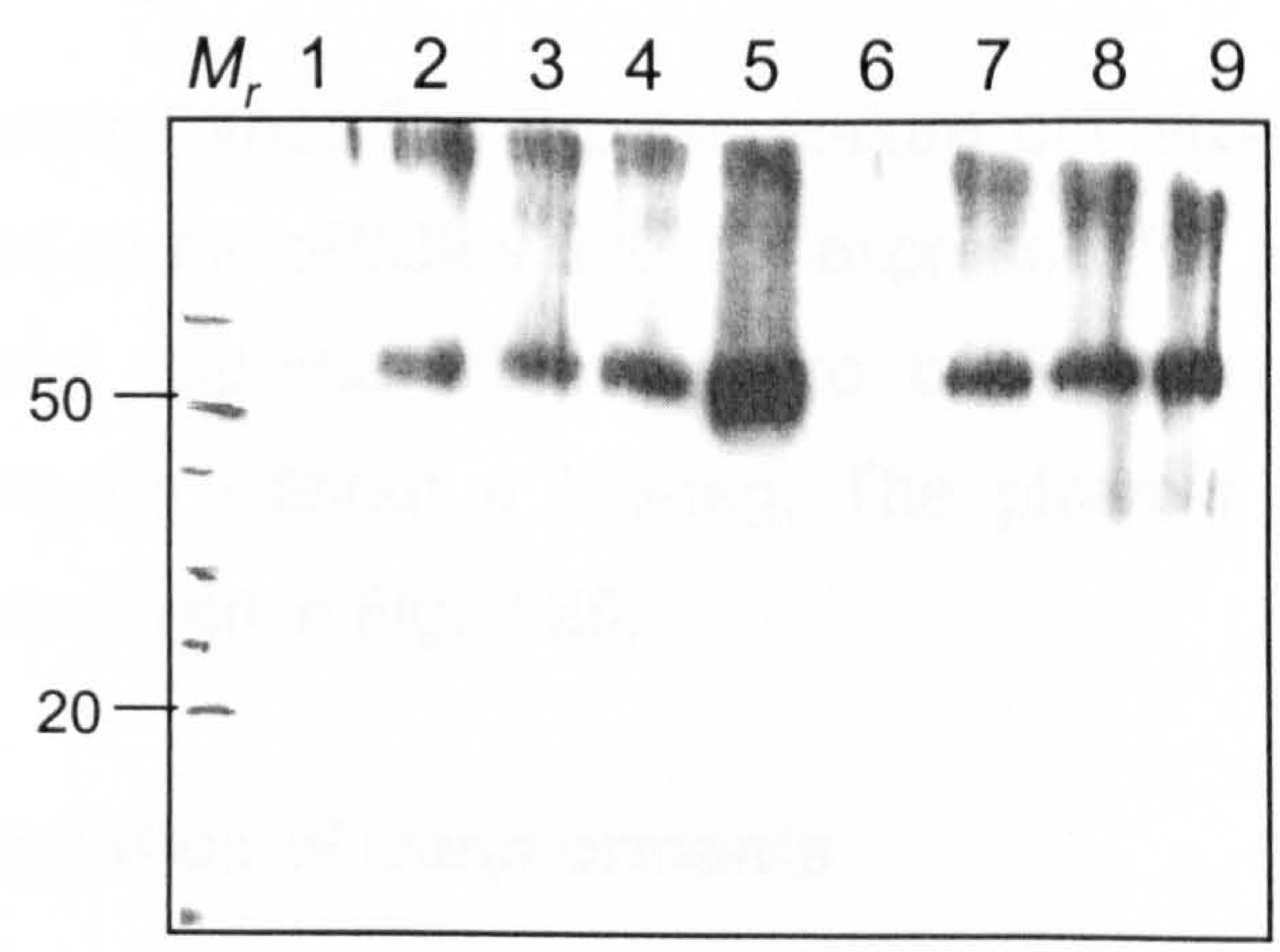


FIG. 3.27. **Detection of the cathepsin B of *Eimeria tenella* expressed and released by *Pichia pastoris* X-33, constitutive system, in TCA-precipitated supernatant of cultures.** Analysis by western blot (1:8000 dilution of last bleed anti cathepsin B antiserum; 1:5000 dilution of anti-rabbit antibody; 2.5 min exposure) of the equivalent of 0.2 ml of TCA-precipitated supernatant of culture of *Pichia* cells transfected with the pGAPHisCatB plasmid (lanes 1 to 5) and the pGAPCatB plasmid (lanes 6 to 9). Lane 1, 24 h; lane 2, 48 h; lane 3, 72 h; lane 4, 96 h; lane 5, 168 h; lane 6, 24 h; lane 7, 48 h; lane 8, 72 h; lane 9, 96 h. M_r corresponds to molecular mass in kDa.

Table 3.1. **Azocasein assay on *Pichia* culture supernatant after 192 h of culture.** Fifty µl of a 1% (w/v) azocasein was mixed with 400 µl 50 mM sodium acetate pH 5.0 and 50 µl of sample to test (about 10 µg of recombinant protein) and incubated at 37°C overnight. The absorbance of the supernatant was measured at 400 nm.

Sample	Absorbance (400 nm)
Control without protein	0.09
pPIC9catB supernatant	0.948
pPIC9catB supernatant 10 mM DTT, 1 mM E64	0.464
pPIC9catB supernatant 1 mM E64	0.975
pPIC9catB supernatant 10 mM DTT	0.471
pPIC9 supernatant	0.861
pPIC9 supernatant 1 mM E64	0.921
pPIC9 supernatant 10 mM DTT	0.459

3.2.9.2. Inducible system

The sequence encoding the pro-mature cathepsin B was amplified from cDNA and cloned in the pPIC9 vector for expression in an inducible manner and secretion into the supernatant of culture of *P. pastoris* KM71 cells of the recombinant enzyme without a His-tag. The plasmid obtained and predicted protein are schematised in Fig. 3.28.

3.2.9.2.1. Confirmation of transformants

Insertions of the fragments were verified by PCR on genomic DNA extracted from positive clones, using the pairs of PCR primers α-factor and 3'AOX1, or 5'AOX1 and 3'AOX1 (see Table 2.1, chapter 2). Results of PCRs on genomic DNA extracted from KM71 His^+ Mut^S are shown on Fig. 3.29. By PCR using the α-factor/3'AOX1 primers, a DNA fragment of 1661 bp is predicted to be amplified on the genomic DNA from KM71 cells transfected with the pPIC9catB plasmid, and a DNA fragment of 195 bp is predicted to be amplified on the genomic DNA from KM71 cells transfected with the pPIC9 empty vector. Using the 5'AOX1/3'AOX1 primers, a DNA fragment at 3.6 kb corresponding to the amplification from the endogenous mutated AOX1 gene should be detected in all

His^+ Mut^s cells. The amplification of the inserted fragment from the pPIC9catB construct would yield a product at 1968 bp and the amplification of the inserted fragment from the pPIC9 empty vector would yield a product at 492 bp. Therefore, from the results on Fig. 3.29, the clone selected for each transfection seems to correspond to a positive clone.

3.2.9.2.2. Analysis of expression

One positive colony for each of the transfections was selected and tested for expression of the target protein. Cells were induced with 0.5 % methanol (v/v, final concentration) every 24 h and 0.5 ml aliquots were taken a regular time intervals. Cells were separated from the supernatant of culture by centrifugation and samples of supernatant were used for analysis using gelatin SDS-PAGE, the azocasein assay, and, either directly or after TCA-precipitation, Coomassie blue-stained SDS-PAGE and western blot analysis. The control strain transfected with the empty vector was induced and its supernatant of culture was also analysed after TCA-precipitation.

Time course analysis of expression showed that the recombinant protein was already present about 24 h after the start of induction, that the amount of protein increased slightly until 96 h after induction and then remained stable until at least 192 h after induction (data not shown). The cells were routinely harvested 192 h after the start of induction. Analysis of expression showed a major protein at about 30 kDa on Coomassie blue-stained SDS-PAGE gels in the supernatant of cells transfected with the pPIC9catB plasmid, but not in the supernatant of the cells transfected with the empty pPIC9 vector (Fig. 3.30). This corresponds to the size expected for the mature cathepsin B-like enzyme. Using specific anti-pro-cathepsin B antiserum, a major protein at about 30 kDa (Fig. 3.30, lane 6), the size expected for the mature enzyme, was found in the supernatant of culture of the KM71 *Pichia* cells transfected with the pPIC9CatB plasmid for inducible expression. Based on Coomassie blue-stained SDS-PAGE analysis, the expression was estimated to be about 100 mg of protein per litre of culture. In the TCA-precipitated supernatant of culture at the time of harvest, 2 major proteins were detected using specific anti-pro-cathepsin B antiserum (Fig. 3.30, lane 7). One band was the protein at 30 kDa, the other was slightly larger (about 32 kDa).

Fainter proteins at about 54 kDa and 22 kDa could also be detected, which could correspond, respectively, to the pro-mature enzyme and the pro-domain.

By western blot analysis of the cell extracts of *Pichia* cells transfected with the pPIC9catB plasmid and the empty vector, only traces of the 54 kDa protein could be detected in the cell extracts of the cells transfected with the pPIC9catB plasmid but not the empty vector, suggesting that most of the protein expressed was released into the medium (results not shown).

Supernatants of cultures of both the cathepsin B-expressing and control cells were tested for activity against the Z-Phe-Arg-AMC substrate and there was significantly greater activity (3000-fold) in the supernatant of culture of cells expressing a processed protein recognised by anti-cathepsin B antiserum, compared to the same volume of supernatant of culture of cells transfected with the empty vector. The culture supernatant containing the active enzyme was analysed immediately or stored at -20°C for later analysis.

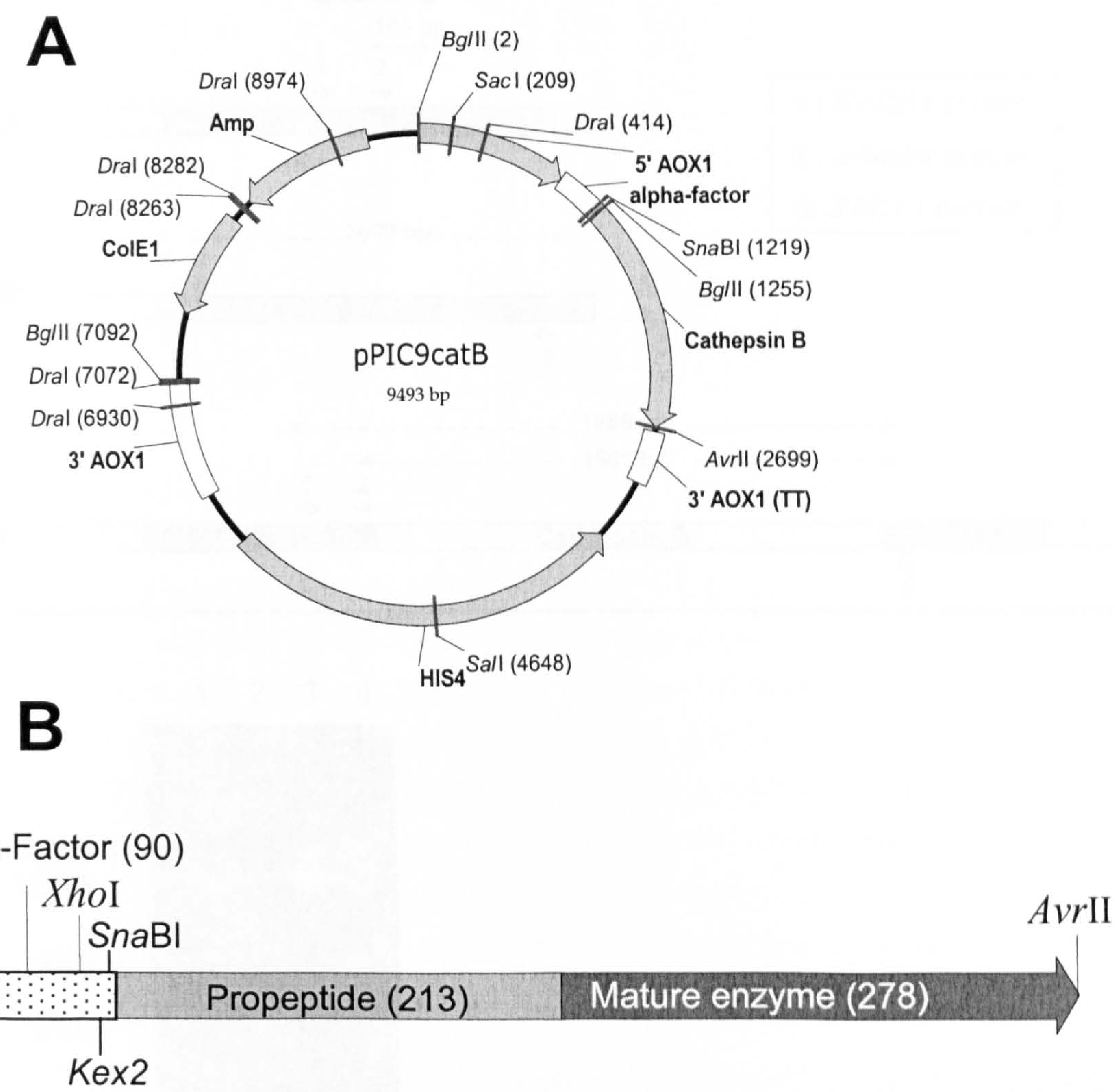
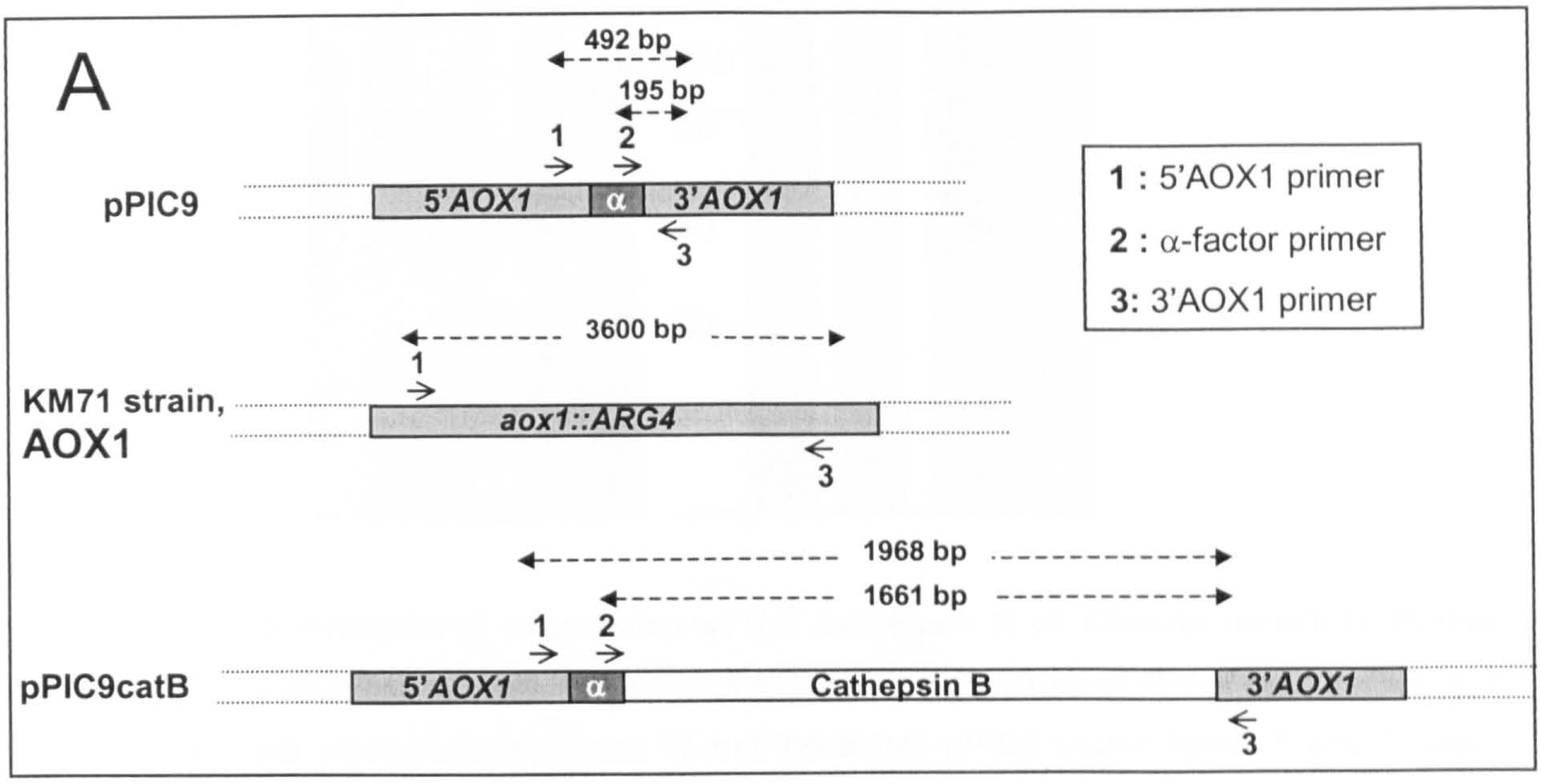


FIG. 3.28. **Construct for expression of recombinant pro-cathepsin B of *E. tenella* in *Pichia pastoris*, inducible system.** (A) Schematic representation of the plasmid resulting from the cloning of the pro-mature sequence of the cathepsin B of *E. tenella* in the pPIC9 vector for inducible expression of pro-mature cathepsin B in the supernatant of culture. Amp, ampicillin resistance gene; ColE1, origin of replication; *AOX1*, alcohol dehydrogenase gene; *HIS4*; histidine prototrophy gene. (B) Schematic representation of the protein expressed from the pPIC9catB plasmid. The size in amino acids of the different fragments is indicated between brackets.



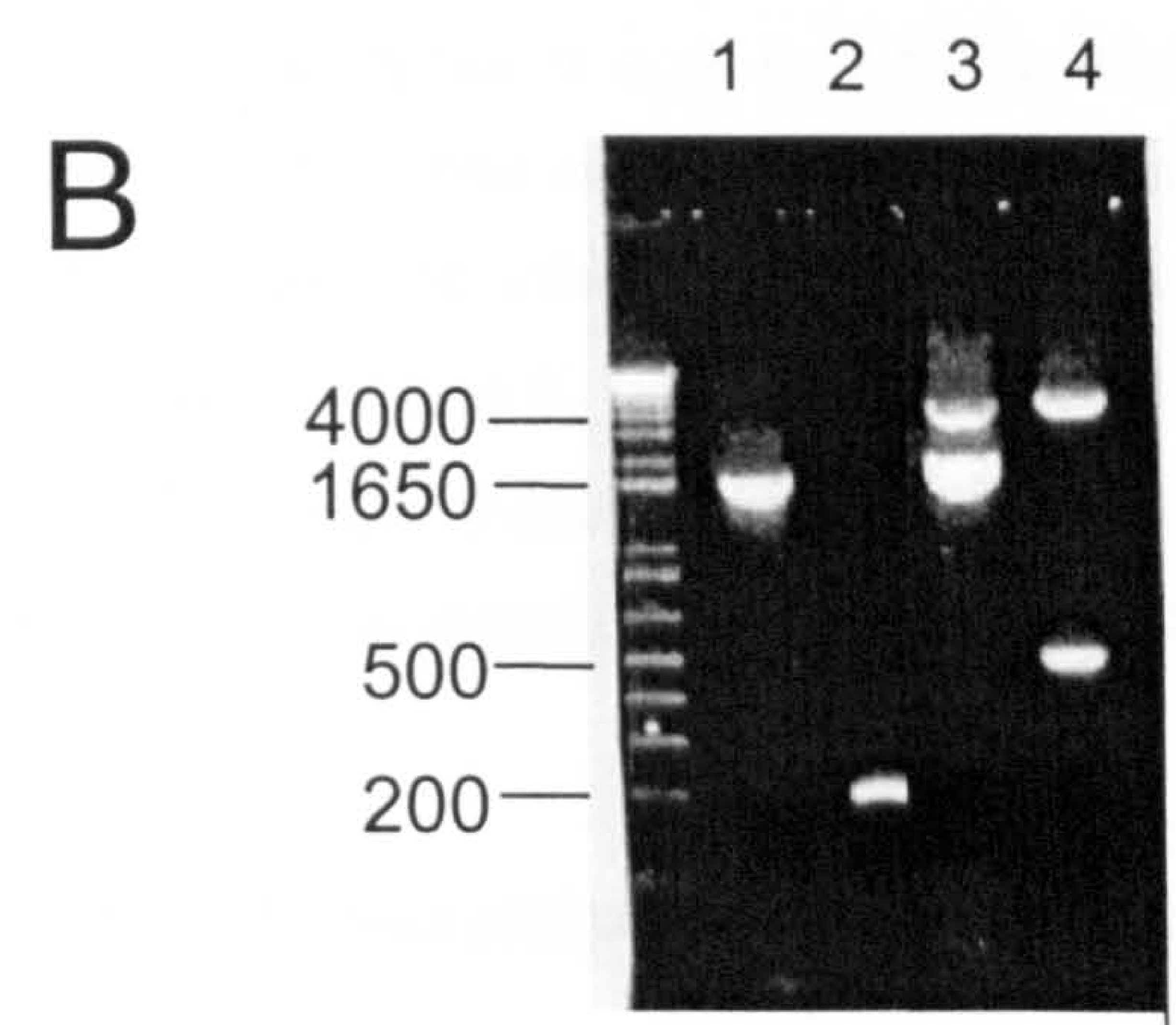


FIG. 3.29. **Verification of insertion of the pPIC9catB construct in the *Pichia* KM71 genome.** (A) Schematic representation of the position of the different primers used for PCR and their position. α, α-factor sequence; *AOX1*, alcohol dehydrogenase gene; *ARG4*, gene for arginine prototrophy. (B) PCR analysis of *Pichia* transformants using 1% agarose gel and ethidium bromide staining. Ten μl of reactions were loaded. PCRs used the primers α-factor/3'AOX1 (lanes 1 and 2), and 5'AOX1/3'AOX1 (lanes 3 and 4). Genomic DNA from KM71 cells transfected with the pPIC9catB plasmid was analysed on lanes 1 and 3. Genomic DNA from KM71 cells transfected with the pPIC9 empty vector was analysed on lanes 2 and 4. Size in bp is indicated on the left of the figure.

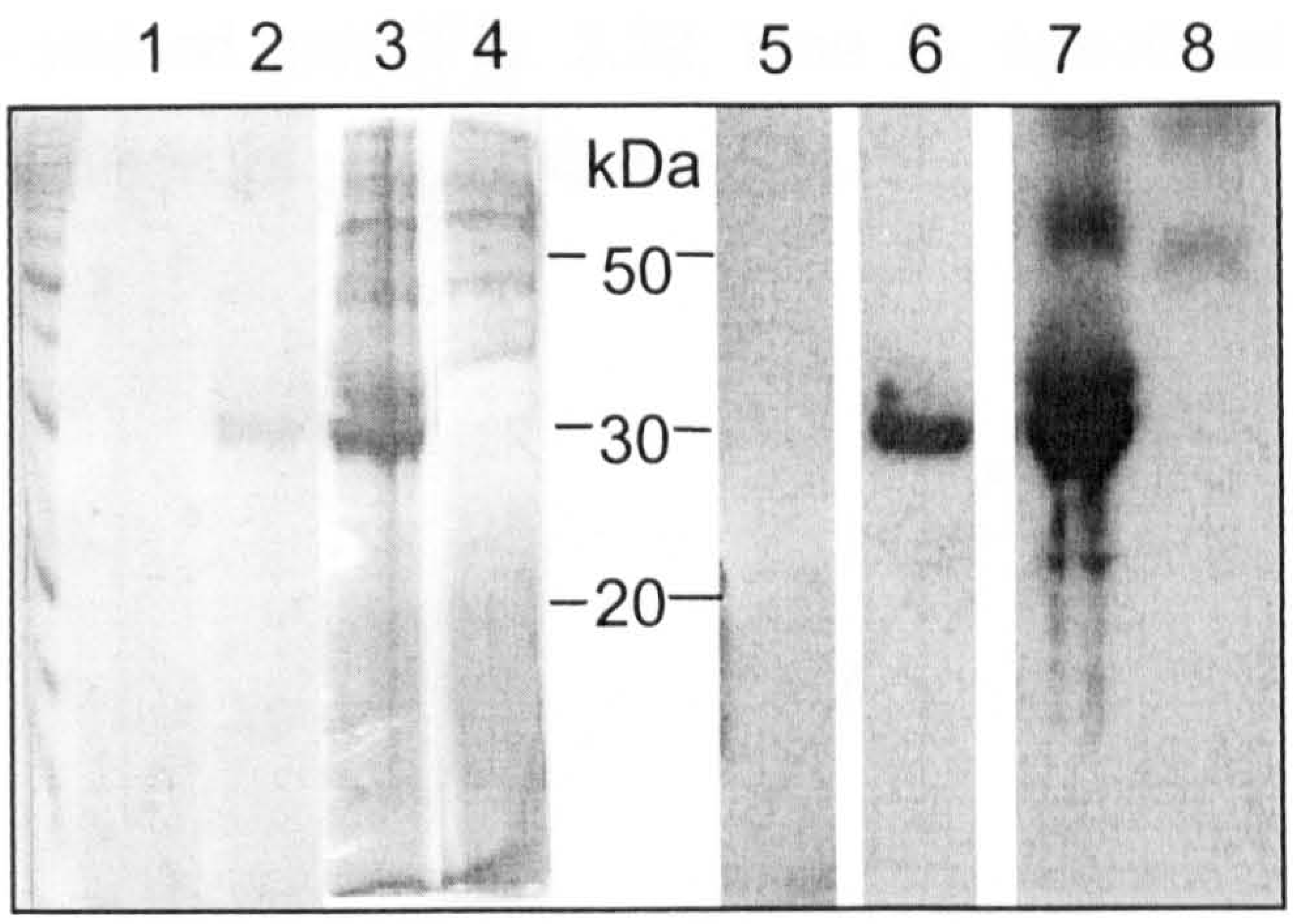


FIG. 3.30. **Analysis of expression of the cathepsin B of *Eimeria tenella* in *Pichia pastoris* KM71, inducible system.** Fifteen μl of supernatant of culture of *Pichia* cells transfected with the pPIC9catB plasmid (lanes 2 and 6) and the empty pPIC9 vector (lanes 1 and 5) were analysed using Coomassie blue-stained 12% (w/v) SDS/PAGE (lanes 1 and 2) and by western blot analysis (lanes 3 and 4; 1:8000 dilution of last bleed anti-cathepsin B antiserum; 1:5000 dilution of anti-rabbit antibody; 5 min exposure). The equivalent to 0.5 ml of supernatant of culture of *Pichia* cells transfected with the pPIC9catB plasmid (lanes 3 and 7) and the empty pPIC9 vector (lanes 4 and 8) after TCA-precipitation were analysed using Coomassie blue-stained 12% (w/v) SDS/PAGE (lanes 3 and 4) and by western blot analysis (lanes 7 and 8; 1:8000 dilution of last bleed anti-cathepsin B antibody; 1:5000 dilution of anti-rabbit antibody; 2 min exposure).

3.2.10. Purification of recombinant full-length cathepsin B expressed and secreted by *Pichia pastoris*

The enzyme was purified from the supernatant of culture by ammonium sulfate precipitation. The enzyme was found to precipitate mainly between 40 and 70 % ammonium sulfate saturation (Fig. 3.31). The protein appeared to be slightly purer using 70% rather than 80% saturation precipitation and the smear apparent in the low molecular mass range of the gels with crude extract disappeared (Fig. 3.32, lane 1). In addition, the yellow-brown colour of the culture supernatant (due to the abundant polysaccharides present) also was less in the pellet resuspensions compared with the crude extract. Thus, ammonium sulfate precipitation improved the purity of the enzyme significantly. A 40 % saturation precipitation followed by a 70 % saturation precipitation on the supernatant of precipitation resulted in the presence of 2 proteins at about 30 and 32 kDa on

Coomassie blue-stained gels (Fig. 3.32, lane 2), specifically recognised by anti-pro-cathepsin B antiserum (Fig. 3.32, lane 4).

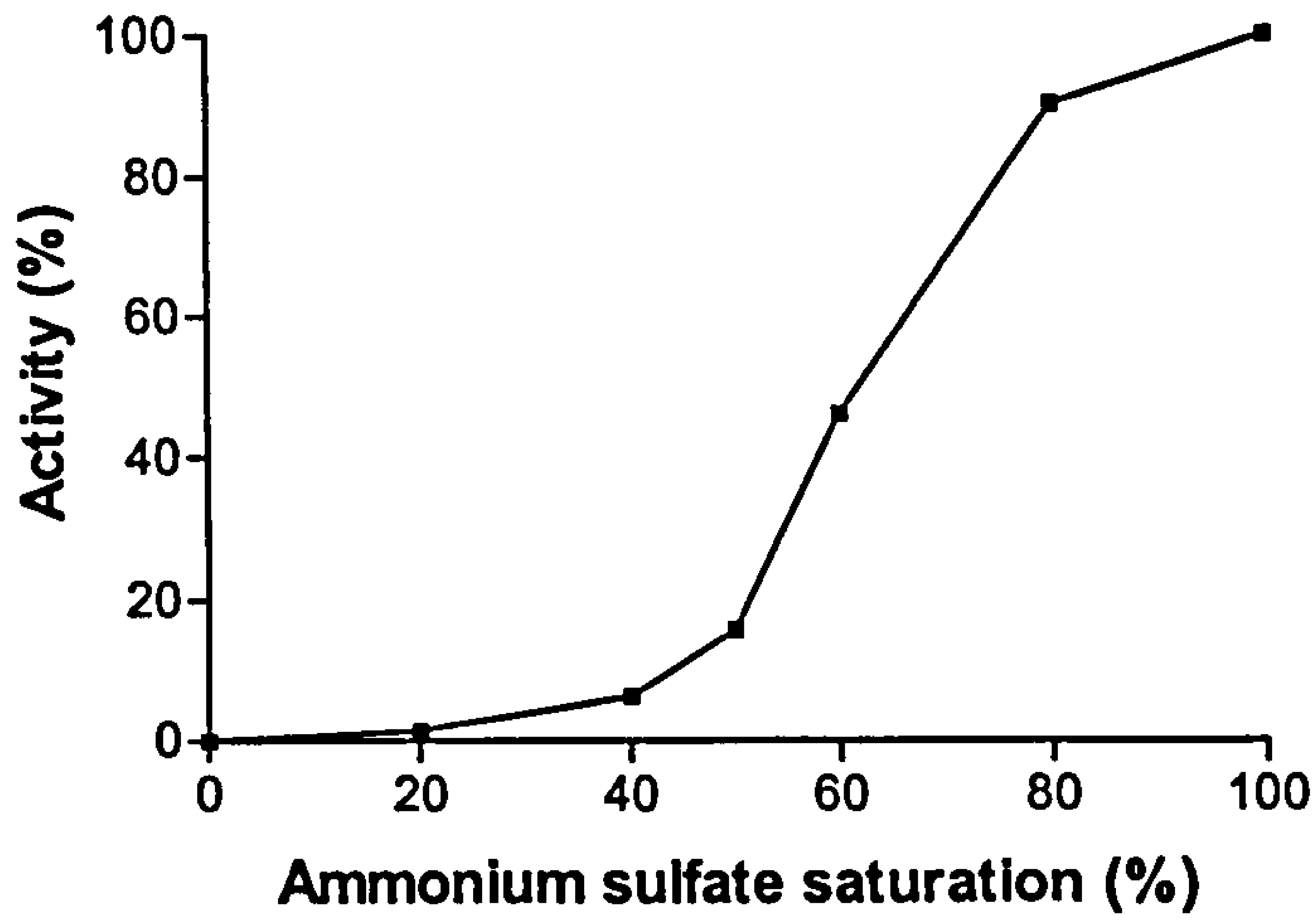


FIG. 3.31. **Ammonium sulfate precipitation of the cathepsin B of *Eimeria tenella* produced by *Pichia pastoris* KM71, inducible system.** Aliquots of supernatant of culture of *Pichia* cells expressing the cathepsin B were incubated for 2 h at 4°C with gentle agitation in the presence of increasing ammonium sulfate saturation concentrations. The precipitated proteins were pelleted for 15 min at 13000 *g*, 4°C and resuspended in the initial volume of 50 mM sodium acetate, pH 5.0. The activity was measured and compared to the non-treated sample. The activity present in the pellet was expressed as a percentage of the initial activity.

3.2.11. Deglycosylation of recombinant full-length cathepsin B expressed in *Pichia pastoris*

After treatment with the PGNase F for deglycosylation, a single protein at 30 kDa corresponding to the mature cathepsin B could be detected on Coomassie blue-stained SDS-PAGE gel (Fig. 3.32, lane 3), specifically recognised by anti-pro-cathepsin B antibodies (Fig. 3.32, lane 5).

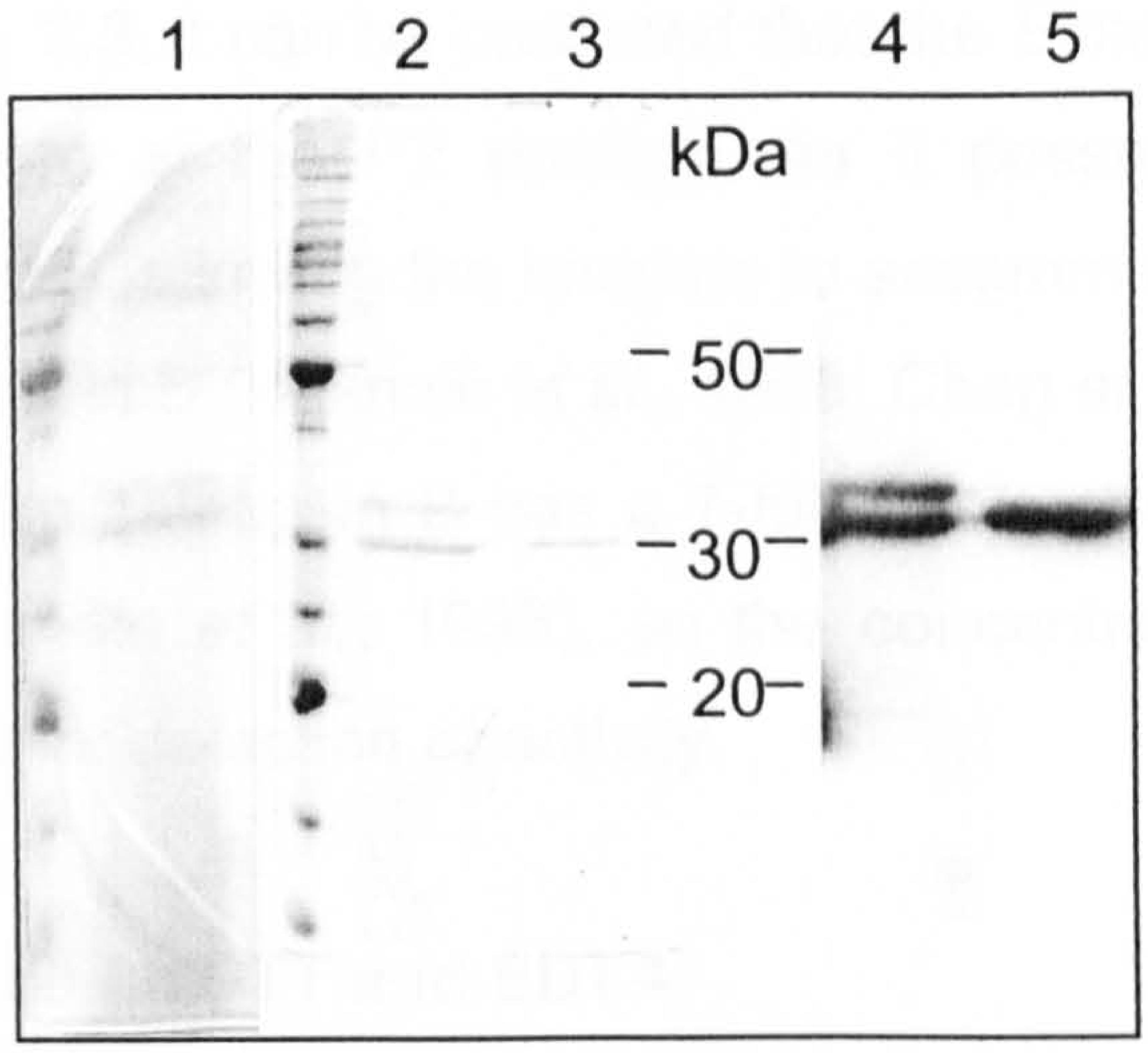


FIG. 3.32. **Deglycosylation of the cathepsin B of *Eimeria tenella* produced and released by *Pichia pastoris* KM71, inducible system.** The supernatant of culture of cathepsin B-expressing *Pichia* cells (lane 1) was treated with 40% ammonium sulfate for 2 h at 4°C with gentle agitation; the supernatant of this treatment was then incubated in the presence of 70% ammonium sulfate for 2 h at 4°C with gentle agitation, and the precipitated fraction after this step was resuspended and analysed on a Coomassie blue-stained 12% (w/v) SDS-PAGE gel (lane 2) and by western blot analysis (lane 4, 1:8000 dilution of last bleed anti-cathepsin B antiserum; 1:5000 dilution of anti-rabbit antibody; 3 min exposure). The purified enzyme was treated with PGNAse F (0.5 U, 1 h at 37°C) and analysed on a SDS-PAGE gel (lane 3) and by western-blot (lane 5) as described for the non-deglycosylated sample. Fifteen μl of each sample were loaded on the gels.

3.2.12. Characterisation of the active recombinant cathepsin B expressed in *Pichia pastoris*

An activity of about 0.6 U/ml towards the substrate Z-Phe-Arg-pNA (333 μM) in 50 mM sodium acetate pH 5.0, 2 mM EDTA, 10 mM DTT, could be detected in the sample of crude culture supernatant (concentration of 100 μg protein/ml), whereas the crude supernatant of culture of *Pichia* cells transfected with the empty pPIC9 vector showed a maximum of activity of 0.2 mU/ml against the same substrate. The K_m for Z-Phe-Arg-pNA could not be measured as the substrate was falling out of solution at higher concentrations. No activity towards 100 μM of Z-Arg-Arg-pNA could be detected under the conditions used. Higher concentrations of substrate provoked its precipitation in the assay buffer. From the alignment on Fig. 3.3, it can be predicted that the *Eimeria* cathepsin B should be able to cleave Arg at the P2 position, as it possesses Glu-325 (human cathepsin B numbering), allowing the enzyme to accommodate Arg at the P2 site like the human cathepsin B (Hasnain *et al.*, 1993; Chan *et al.*, 1999). But it is also known that the human cathepsin B has a 7-fold preference for Phe over Arg at the P2 position (Hasnain *et al.*, 1993), so the concentration of Z-Arg-Arg-pNA may not be sufficient for detection of activity.

3.2.12.1. Dependence on DTT and EDTA

The enzyme activity was strictly dependent on the presence of DTT, 10 mM being the minimum required for maximum activity. EDTA seemed to have a beneficial effect on the activity, with 5 mM being the minimum required for observation of maximum activity (Fig. 3.33).

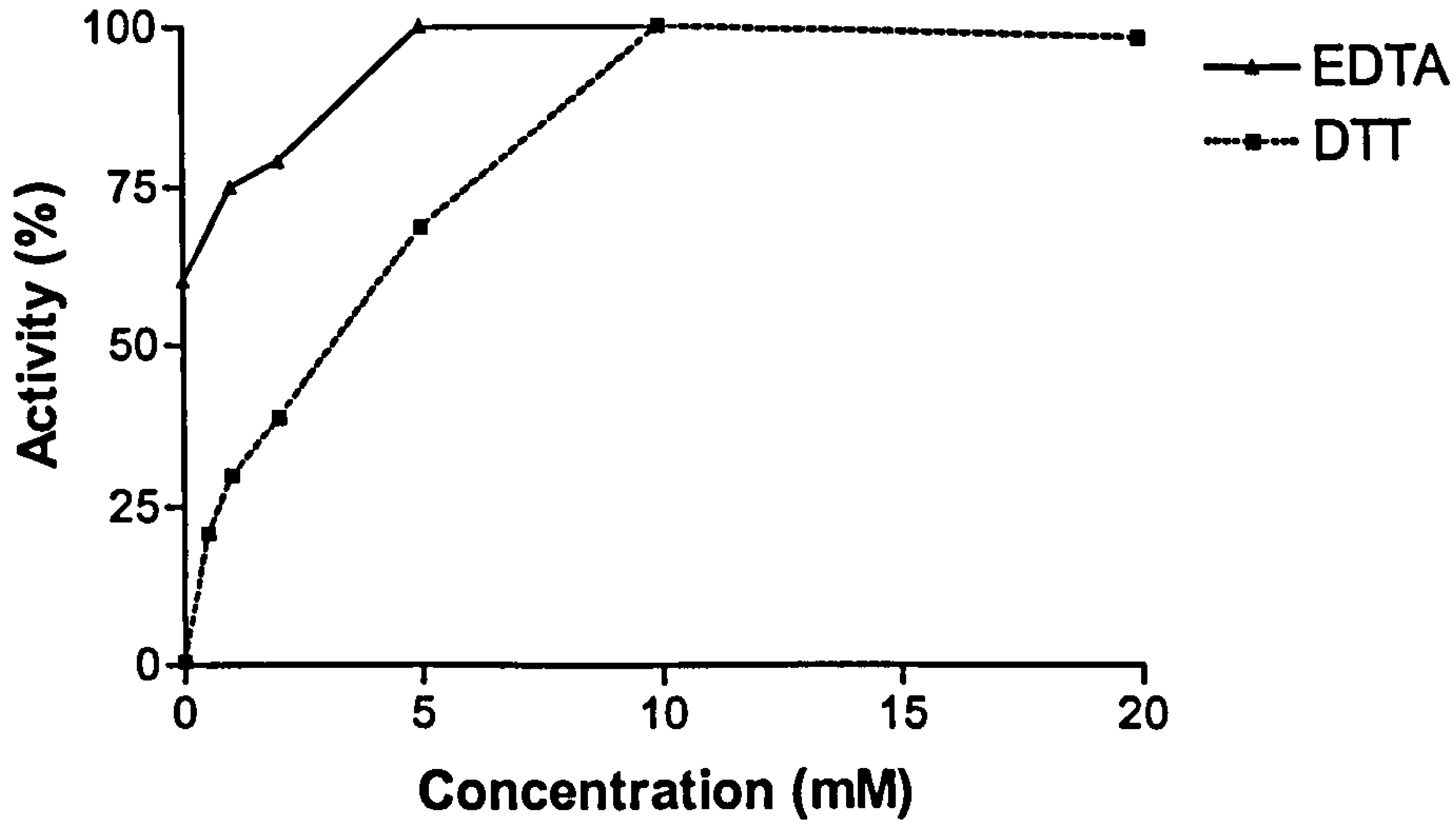


FIG. 3.33. **Dependence of the activity of the recombinant cathepsin B of *Eimeria tenella* on DTT and EDTA.** Enzyme activity was tested for dependence on DTT and EDTA, and measured in 50 mM sodium acetate pH 5.0, using the substrate Z-Phe-Arg-AMC, in a total volume of 0.6 ml. 30 μl of crude *Pichia* supernatant were used for each assay. The reaction was started by addition of 50 μM of substrate and measured continuously for 5 min at 37°C (Excitation: 380 nm/Emission: 465 nm). Dependence of the activity on DTT was measured in the absence of EDTA, and dependence of the activity on EDTA was measured in the presence of 10 mM DTT. Activity towards Z-Phe-Arg-AMC in the control supernatants (from *Pichia* transfected with the empty plasmid) was negligible under these conditions.

3.2.12.2. Dependence on salt

Sodium chloride, which may in some cases stabilise enzyme activity, did not seem to have any beneficial effect at concentrations up to 200 mM; on the contrary, above 200 mM it seemed to be slightly detrimental, reducing the activity by 20% (Fig. 3.34). Sodium chloride was omitted from all subsequent assays.

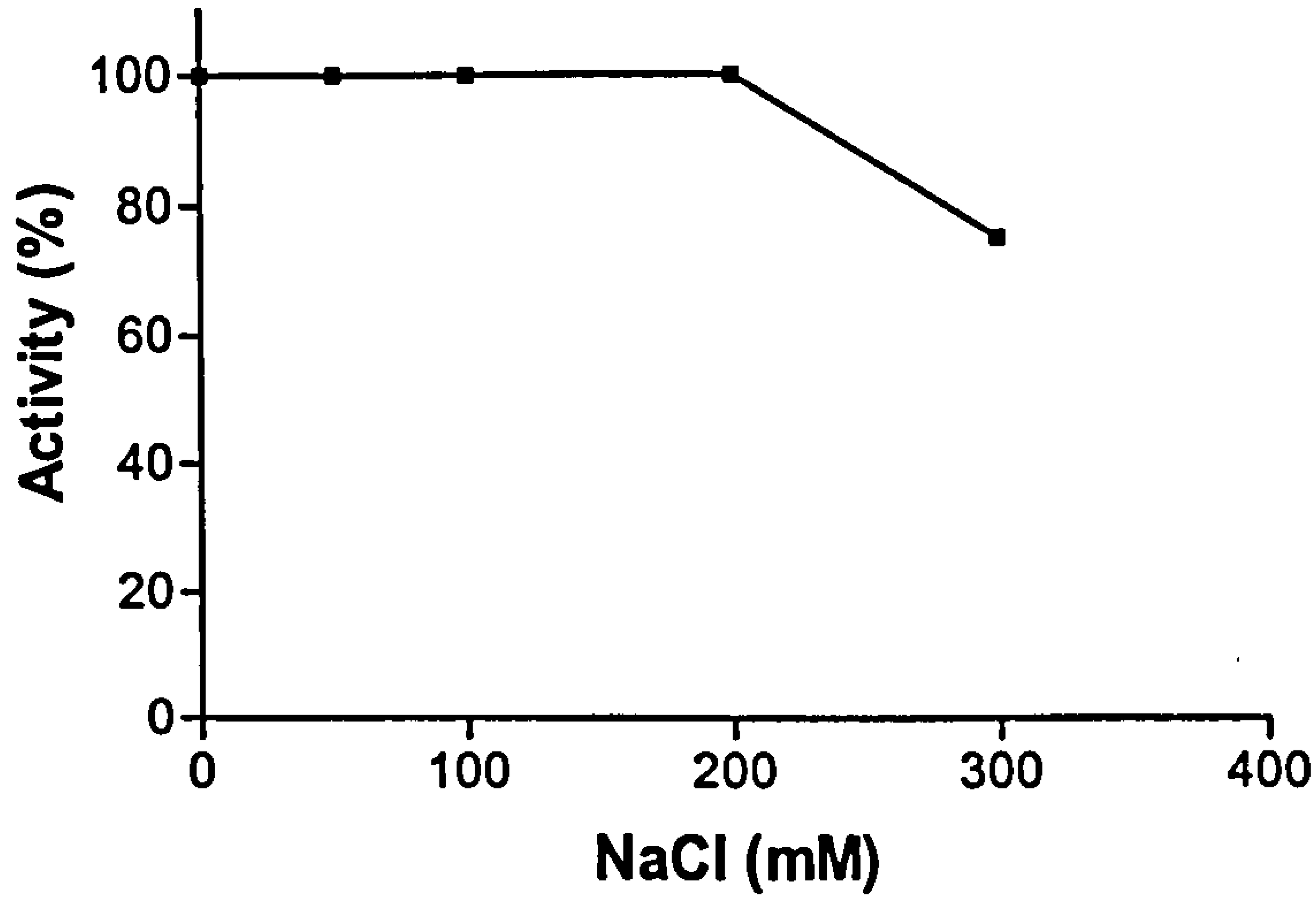


FIG. 3.34. **Dependence of the activity of the recombinant cathepsin B of *Eimeria tenella* on salt.** Enzyme activity was tested for dependence on NaCl, and measured in 50 mM sodium acetate pH 5.0, using the substrate Z-Phe-Arg-AMC, in a total volume of 0.6 ml. 30 µl of crude *Pichia* supernatant were used for each assay. The reaction was started by addition of 50 µM of substrate and measured continuously for 5 min at 37°C (Excitation: 380 nm/Emission: 465 nm).

3.2.12.3. pH optimum

In terms of pH dependence, the activity was close to zero when the pH was under 4.0 or above 9.0 and seemed to present 2 maxima, at pH 5.0 and pH 8.0, in the mix of buffers (Fig. 3.35A). The reason why the activity seems to drop at pH 5.5 and 6.0 is unclear at present, but might be due to changes in ionic strength in the buffer. Using individual buffers, the activity dropped drastically when the pH was under 4.0 or above 9.0. The activity seemed to present 2 peaks in these conditions, one between pH 4.0 and 5.5, one at pH 7.5. The activity decreased rapidly or slightly (depending on the buffer) between pH 5.5 and 6.5 (Fig. 3.35B).

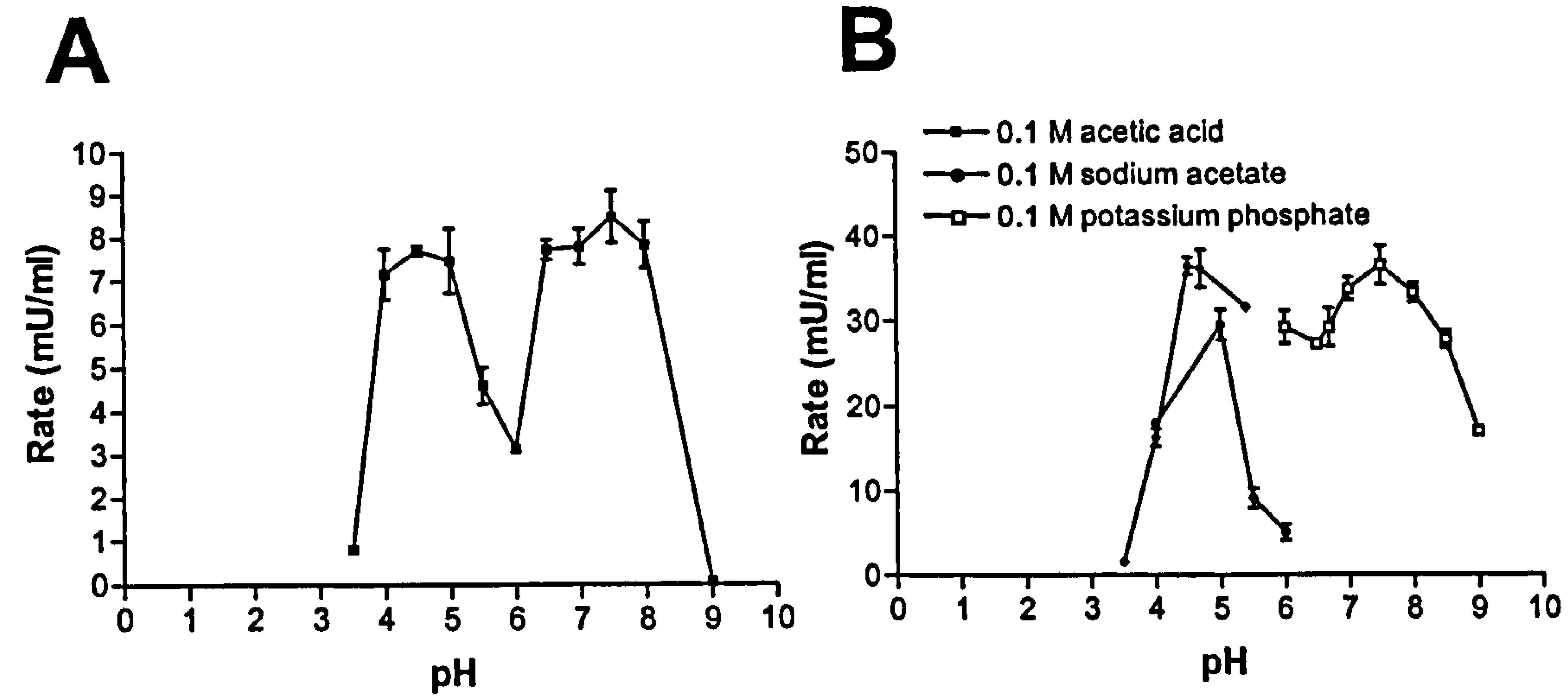


FIG. 3.35. **Dependence of the activity of the recombinant cathepsin B of *Eimeria tenella* on pH.** (A) The pH dependence was determined using a mix of 4 buffers (25 mM acetic acid, 25 mM MES, 75 mM Tris, 25 mM glycine) adjusted to the appropriate pH, with 2 µl of crude *Pichia* supernatant and 10 mM DTT. The reaction was started by addition of 30 µM of Z-Phe-Arg-AMC. (B) Three different buffers were used separately, depending on their pH range. These buffers were 0.1 M acetic acid, 0.1 M sodium acetate and 0.1 M potassium phosphate. All buffers were supplemented with 10 mM DTT. The reaction was started by addition of 30 µM of Z-Phe-Arg-AMC. Reactions used 10 µl of crude *Pichia* supernatant and were measured continuously for 5 min at 37°C (Excitation: 380 nm/Emission: 465 nm). Results are the means ± SD from 3 independent measurements.

3.2.12.4. K_ms against Z-Phe-Arg-AMC and Z-Arg-Arg-AMC

Purified enzyme was active against the substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC; activity was 20-fold higher against Z-Phe-Arg-AMC than against Z-Arg-Arg-AMC (Fig. 3.36A). The strong preference for a phenylalanine residue at the P2 site compared to arginine is a characteristic for this class of enzymes. The K_m and specific activities were respectively 168 ± 33 µM and 6 U/mg protein against Z-Phe-Arg-AMC; and 1.40 ± 0.25 mM and 0.3 U/mg protein against Z-Arg-Arg-AMC (Fig. 3.36B). The purified enzyme had very comparable characteristics to the purified and deglycosylated enzyme, with a variation of K_ms lower than 8% and variation of activity lower than 10%. The use of culture supernatant from control cultures (*Pichia* with the empty vector) showed that it contained no detectable activity towards Z-Phe-Arg-AMC (< 0.2 mU/ml).

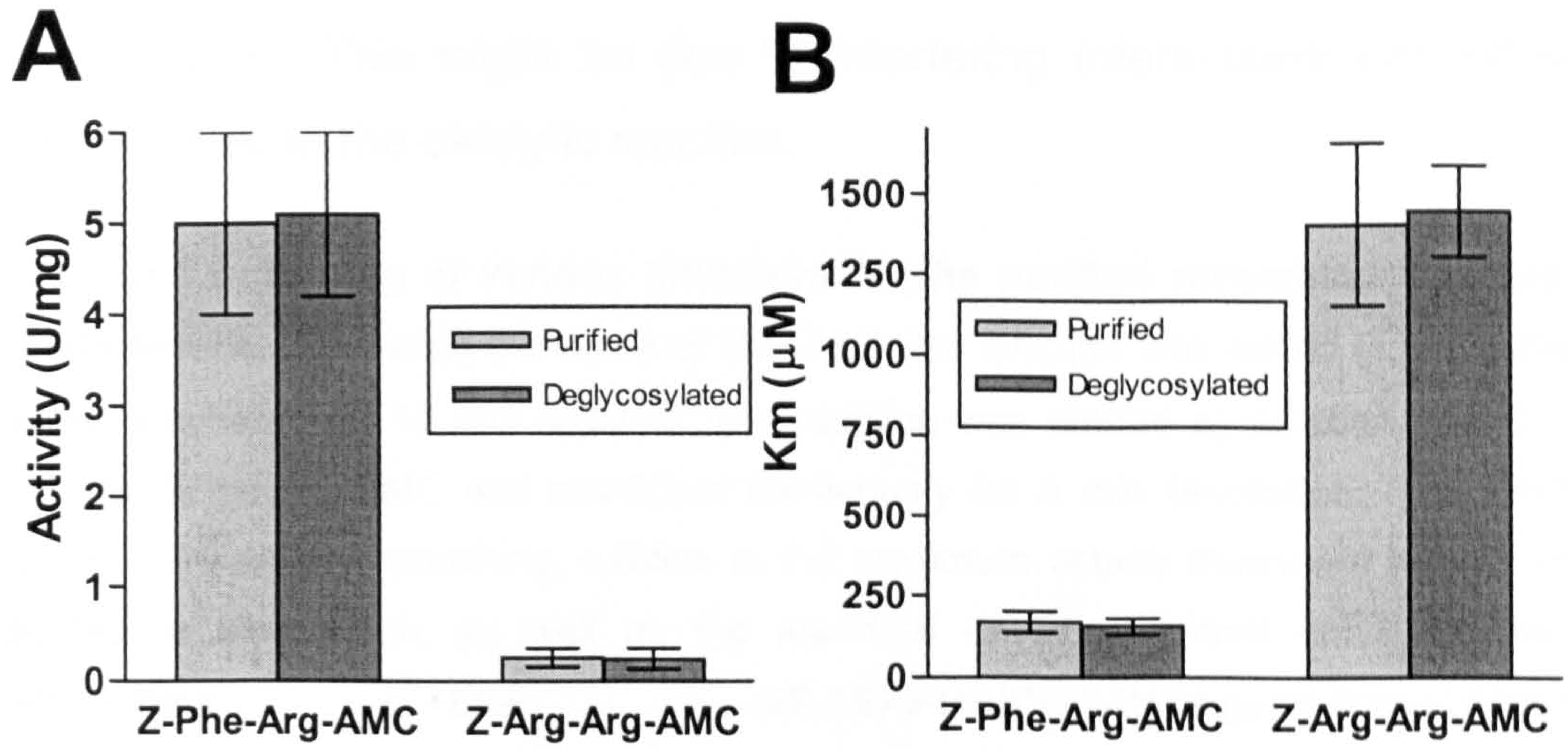


FIG. 3.36. **Analysis of the activity of the non-deglycosylated and deglycosylated cathepsin B of *Eimeria tenella* produced in *Pichia pastoris* against the synthetic fluorogenic substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC.** Bars indicate the standard deviation around the mean, based on 3 repetitions of each measurement. 0.5 µg of enzyme was used per assay. (A) Representation of the activity (U/mg protein) of the different enzyme fractions against the 2 substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC. Enzyme samples were tested for activity in a final volume of 1 ml. The enzyme sample was incubated for 10 min at 37°C in the assay buffer before start of the reaction by addition of the substrate (200 µM, unless for K_m determination were 10 µM to 1.5 mM were used). Excitation and emission wavelengths were respectively 380 and 465 nm. The assays were continuous for 10 min at 37°C. (B) Representation of the K_m (µM) of the different enzyme fractions against the 2 substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC. Results are the means ± SD from 3 independent measurements.

3.2.12.5. Inhibitors

Various peptidase inhibitors were tested on both the non-deglycosylated and deglycosylated enzymes (Tables 3.1A and B). As expected, the cysteine peptidases inhibitors (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl}carbamoyl)oxirane-2-carboxylic acid (E64), N-methyl-piperazine-Phe-homoPhe-vinylsulfone-phenyl (K11777), Z-Phe-Arg-diazomethylketone (ZFRDMK), leupeptin and N-tosyl-L-lysine-chloromethyl ketone (TLCK) strongly inhibited the enzymes. The sulfhydryl-blocking agent, iodoacetamide, showed strong inhibition too, while N-ethyl maleimide (NEM) did not. The metallopeptidase inhibitor 1,10-phenanthroline, the serine peptidase inhibitors phenylmethylsulphonyl fluoride (PMSF) and benzamidine, and the aspartyl peptidase inhibitor pepstatin A did not show any affect unless used at very high concentrations. This might be due to interfering interactions with other amino acids involved in the catalytic reaction.

Table 3.2. **Testing of various inhibitors on the purified recombinant cathepsin B of *Eimeria tenella*.** For testing the effect of inhibitors, the inhibitor was added to the buffer-enzyme mix and incubated for 10 min at 37°C. The reaction was started by addition of 200 μM of the substrate Z-Phe-Arg-AMC and monitored continously for 5 min (excitation: 380 nm/Emission: 465 nm). The activity remaining, relative to the maximum activity measured without inhibitor, is indicated in percentage, as well as the standard deviation based on 3 repeats of each measurement. E64: (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl}carbamoyl)oxirane-2-carboxylic acid; K11777: N-methyl-piperazine-Phe-homoPhe-vinylsulfone-phenyl; ZFRDMK: Z-Phe-Arg-diazomethylketone; TLCK: N-tosyl-L-lysine-chloromethyl ketone; PMSF: phenylmethylsulphonyl fluoride; NEM: N-ethyl maleimide. (A)The purified enzyme was used (0.5 μg per assay). (B) The purified deglycosylated enzyme was used (0.5 μg per assay).

A

Inhibitor	Concentration (μM)	Activity (%)
E64	100	0
	10	2 ± 0.2
	1	7 ± 2.4
K11777	10	0
	1	1.3 ± 0.3
ZFRDMK	100	0
	10	1.8 ± 0.3
Leupeptin	100	0.4 ± 0.1
	10	3.3 ± 0.2
TLCK	100	1.3 ± 0.2
	10	4.3 ± 0.6
Iodoacetamide	100	0.7 ± 0.1
	10	14 ± 2.5
Bestatin	100	3 ± 0.2
	10	59 ± 5
1,10-phenanthroline	1000	50 ± 8
	100	97 ± 5
PMSF	1000	66.8 ± 4
	100	99 ± 7
Benzamidine	1000	72.1 ± 6
	100	98 ± 6
Pepstatin A	10	99 ± 1.5
	1	100 ± 5
NEM	100	86 ± 3
	10	91 ± 2
2-pyrrolidinone	100	95 ± 3
	10	100 ± 0.5

B

Inhibitor	Concentration (μM)	Activity (%)
E64	100	0
	10	0
	1	4 ± 1.5
K11777	10	0.2 ± 0.1
	1	1.6 ± 0.3
ZFRDMK	100	0
	10	2.1 ± 0.1
Leupeptin	100	0
	10	2.1 ± 0.8
TLCK	100	0.6 ± 0.4
	10	5.6 ± 0.9

3.2.10.6. Gelatin and azocasein are not substrates for the enzyme

Samples of purified and deglycosylated, or not, cathepsin B produced and released by *P. pastoris* were analysed on gelatin SDS/PAGE. No activity could be detected in the conditions tested on either reducing or native gels.

Samples of purified and deglycosylated, or not, cathepsin B produced and released by *P. pastoris* were analysed using the azocasein assay. The activity detected on azocasein was in all cases close to the control without enzyme (absorbance of ~0.1, see Table 3.1).

3.2.13. The cathepsin B enzyme is expressed in *E. tenella* oocysts and sporozoites

The analysis by western-blot of total protein content of sporulated oocysts and sporozoites with specific anti-cathepsin B antibodies detected a protein at ~30 kDa, which may correspond to the mature cathepsin B. Faint proteins at just above 50 kDa and about 20 kDa may correspond to the pro-mature protein and the pro-domain, respectively (Fig. 3.37).

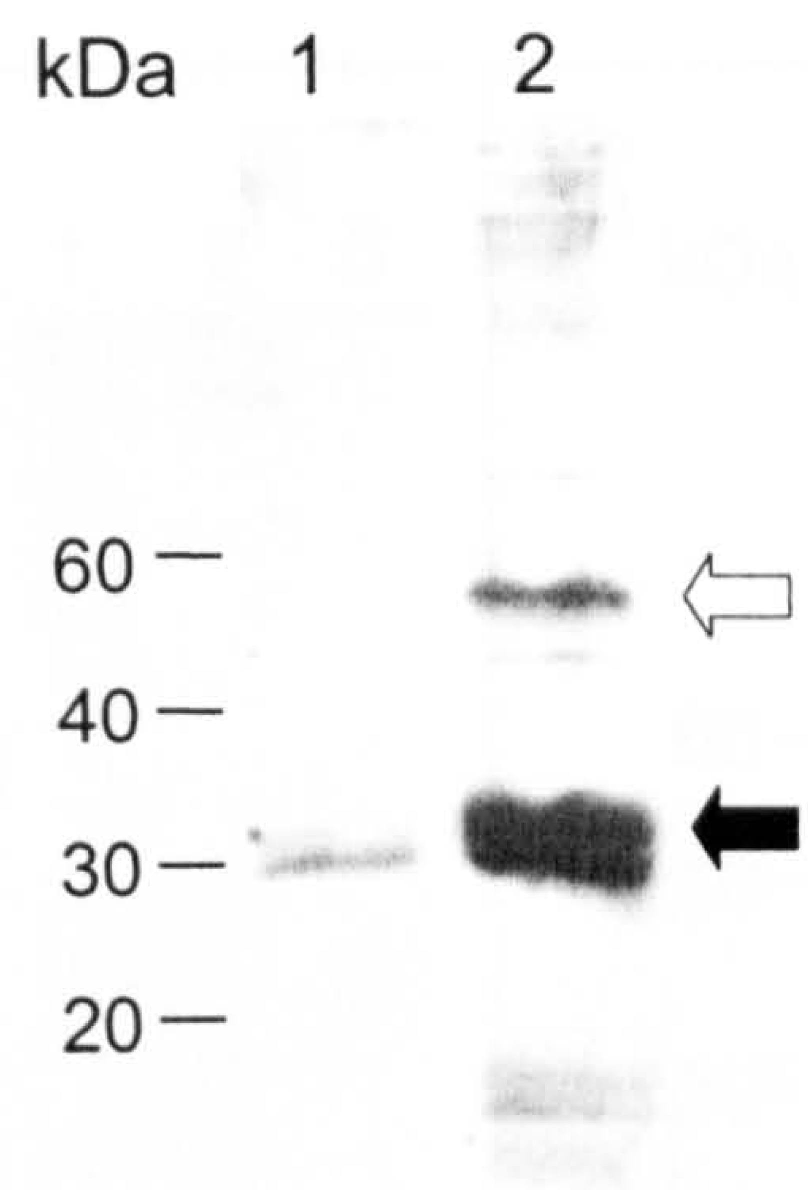


FIG. 3.37. **Immuno-detection of the cathepsin B of *Eimeria tenella* on sporozoite and oocyst extracts.** Detection using the anti-cathepsin B of *E. tenella* antibodies raised in rabbit (1:1000), followed by HRP-conjugated anti-rabbit antibodies (1:2000). Membrane exposure of 2.5 min. Lane 1, total protein from ~10^5 oocysts equivalent; lane 2, total protein from ~10^6 sporozoites. Molecular mass standards are indicated on the left of the figure. The full arrow indicates the expected size for the mature cathepsin B enzyme (31 kDa), and the hollow arrow indicates the size expected for the pro-cathepsin B (57 kDa).

3.2.14. Localisation of the cathepsin B in the asexual stages of the life cycle

In sporozoites, it was shown that the anti-cathepsin B antibodies were detecting proteins at the sizes of the pro-mature and mature enzymes (Fig. 3.37). The antibody was also tested against purified organelles from *E. tenella* sporozoites, on fractions thought to correspond to micronemes and rhoptries (samples extracted and purified on sucrose gradient by Dr Fiona Tomley's group, Institute of Animal Health, Compton). In the lane corresponding to sporozoites, 2 proteins were detected at respectively about 55 kDa and 30 kDa, which correspond to the expected sizes for the pro-mature and mature cathepsin B, respectively 57 and 31 kDa (Fig. 3.38). The same 2 proteins were also detected in the fraction thought to correspond to micronemes, as well as an additional protein at about 100 kDa of unknown origin. The same bands could not be detected in the fraction thought to correspond to rhoptries (Fig. 3.38B, lane 3). Two proteins at respectively 45 and 70 kDa were detected though it is unknown what these proteins may be (Fig. 3.38, lane 3).

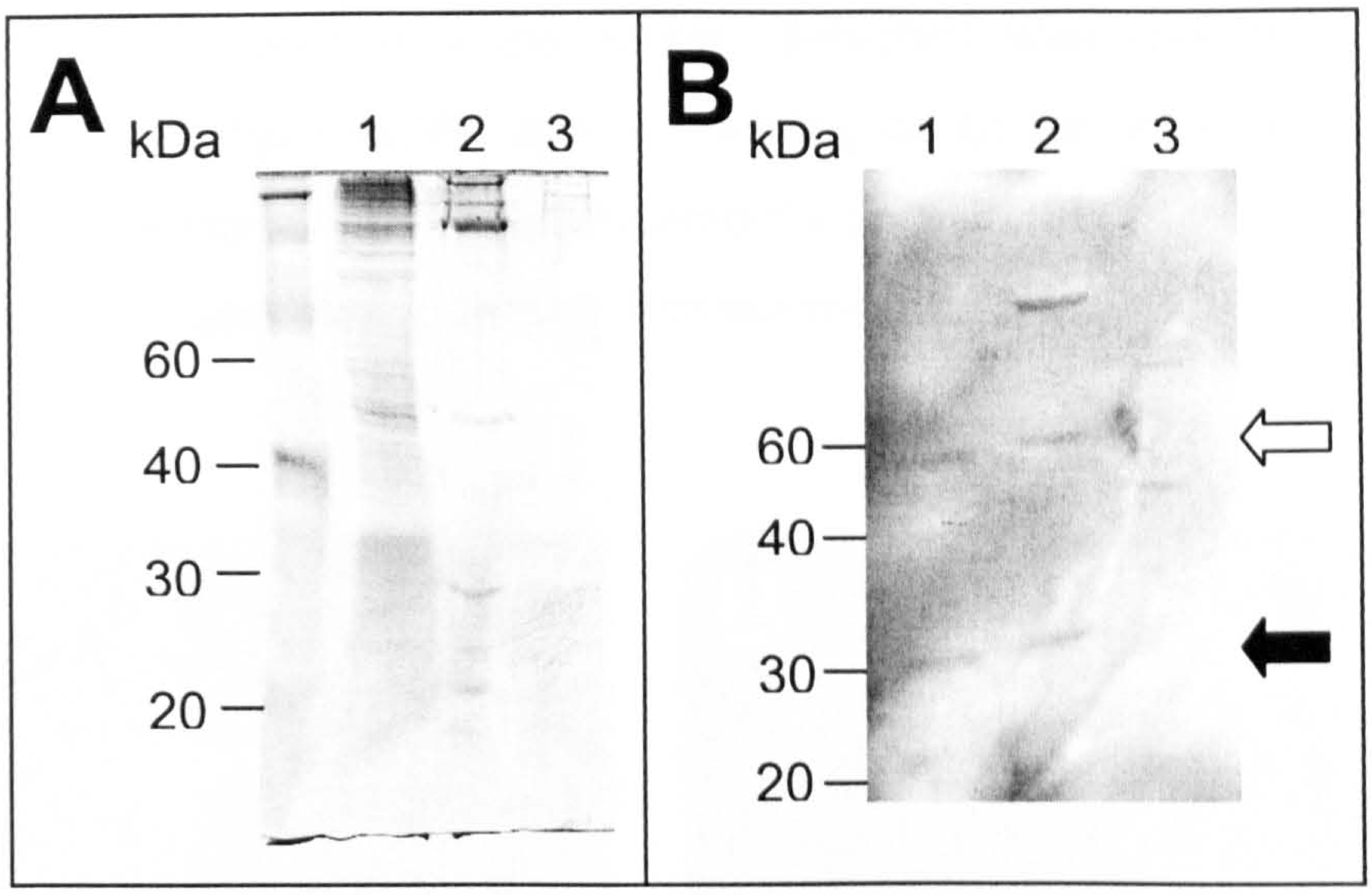


FIG. 3.38. **SDS-PAGE and western-blot analysis of *Eimeria tenella* sporozoite, microneme and rhoptry preparations.** Lane 1, total protein from ~10^6 sporozoites equivalent; lane 2, total microneme proteins from ~4 x 10^6 sporozoites equivalent; lane 3, total rhoptry proteins from ~4 x 10^6 sporozoites equivalent. Molecular weight standards are indicated on the left of the figure. (A) Coomassie blue-stained gel. (B) Detection using the anti-cathepsin B of *E. tenella* antiserum raised in rabbit (1:750), followed by HRP-conjugated anti-rabbit antibodies (1:5000). Membrane exposure of 30 sec. The full arrow indicates the expected size for the mature cathepsin B enzyme (31 kDa), and the hollow arrow indicates the size expected for the pro-cathepsin B (57 kDa).

The immunolocalisation study of the cathepsin B in the asexual stages of *E. tenella* was done by Prof David Ferguson (Oxford). The antibody against the cathepsin B was used to stain sections of the caecum of chickens infected with *E. tenella*, or on purified sporozoites. In purified sporozoites, the staining was localised to the anterior end of the parasite using the anti-cathepsin B antibodies, and the staining pattern was very similar to the one obtained with the anti-MIC2 antibodies (Fig. 3.39A). MIC2 is an adhesive *E. tenella* microneme protein (Bromley *et al.*, 2003). The apparent microneme staining using anti-cathepsin B antibodies is consistent with the results obtained by western blot analysis on sporozoites organelle preparations (Fig. 3.38). During schizogony, a discrete zone at the periphery of the schizont was strongly stained (Fig. 3.40). The nature of the structure stained was unclear. Very little signal was detected from the merozoites. However, it was possible to identify a single small spherical structure positively stained within the merozoite cytoplasm (Fig. 3.40). But the signal was very faint, and it is unclear if the signal detected was real and what these structures might be. No specific apical staining could be detected compared to the anti-MIC2 staining in the mature schizonts (Fig. 3.39B), and the cathepsin B might therefore be cytoplasmic in fully formed merozoites.

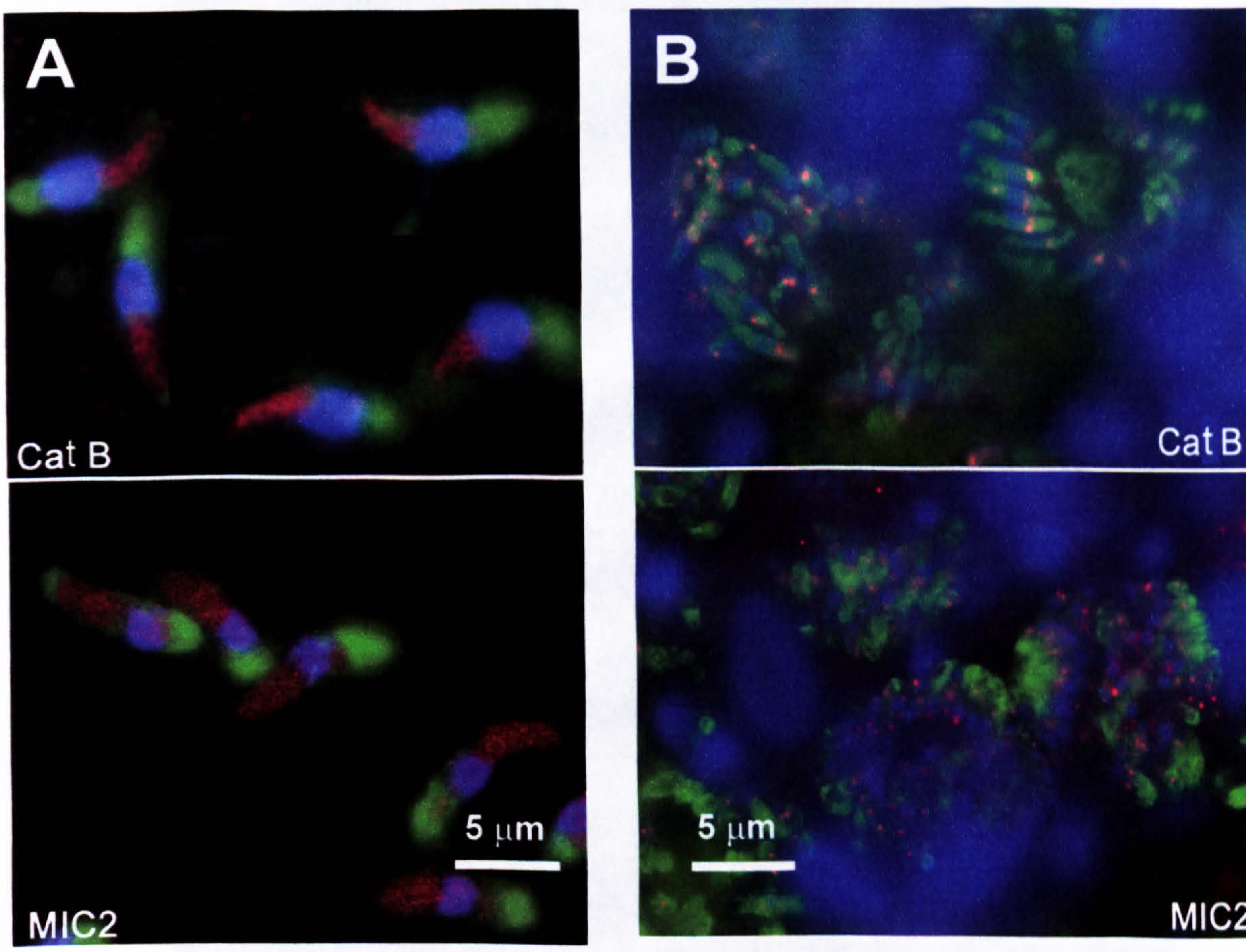


FIG. 3.39. **Immunolocalisation of the cathepsin B in sporozoites (A) and merozoites (B) of *E. tenella*.** Blue colour corresponds to DAPI staining of DNA. Red colour corresponds to auto-fluorescence of the posterior refractile body in A, and staining of a perinuclear structure in B. CatB: anti-cathepsin B antibodies. MIC2: anti-MIC2 antibodies. Sections of the caecum of chickens infected for 112 h were used in B.

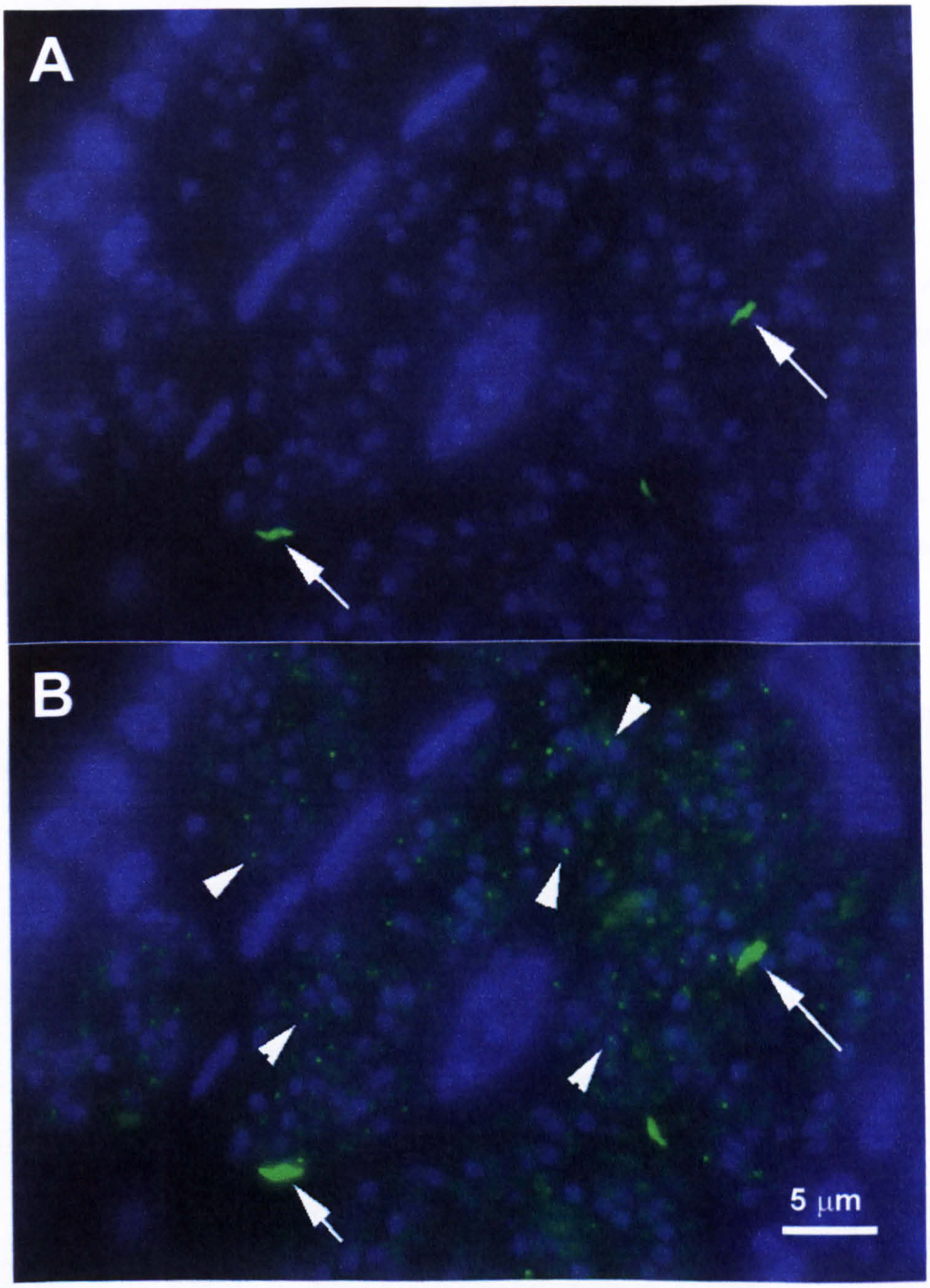


FIG. 3.40. **Immunolocalisation of the cathepsin B in mature schizonts of *E. tenella*.** Sections of the caecum of chickens infected for 96 h were used. (A) and (B) are the same photographs, the signal being increased in photograph (B). Long arrows indicate staining on the periphery of the early schizont, short arrows indicate staining at the tip of mature merozoites.

3.2.15. Localisation of the cathepsin B in the sexual stages of the life cycle of *E. tenella*

This work was done by Prof David Ferguson (Oxford). In the sexual stages, no staining could be found in the microgametes (male), and there was a light staining of the cytoplasm of early macrogametocytes (Fig. 3.41). However, in the mature macrogametocyte, granules at the periphery of the cells were stained, and more specifically, the periphery of these granules was stained (Fig 3.41 and 3.42). The size and location of the granules would be consistent with being wall-forming bodies (Fig. 3.43), probably of type 1, due to their location and appearance in the mature macrogametes (Vieira *et al.*, 1997; Mouafo *et al.*, 2002). In early oocysts within epithelial cells and in the gut's lumen, there was a strong staining of the cytoplasm, as well as some staining of cytoplasmic structures probably related to wall forming bodies (Fig. 3.44).

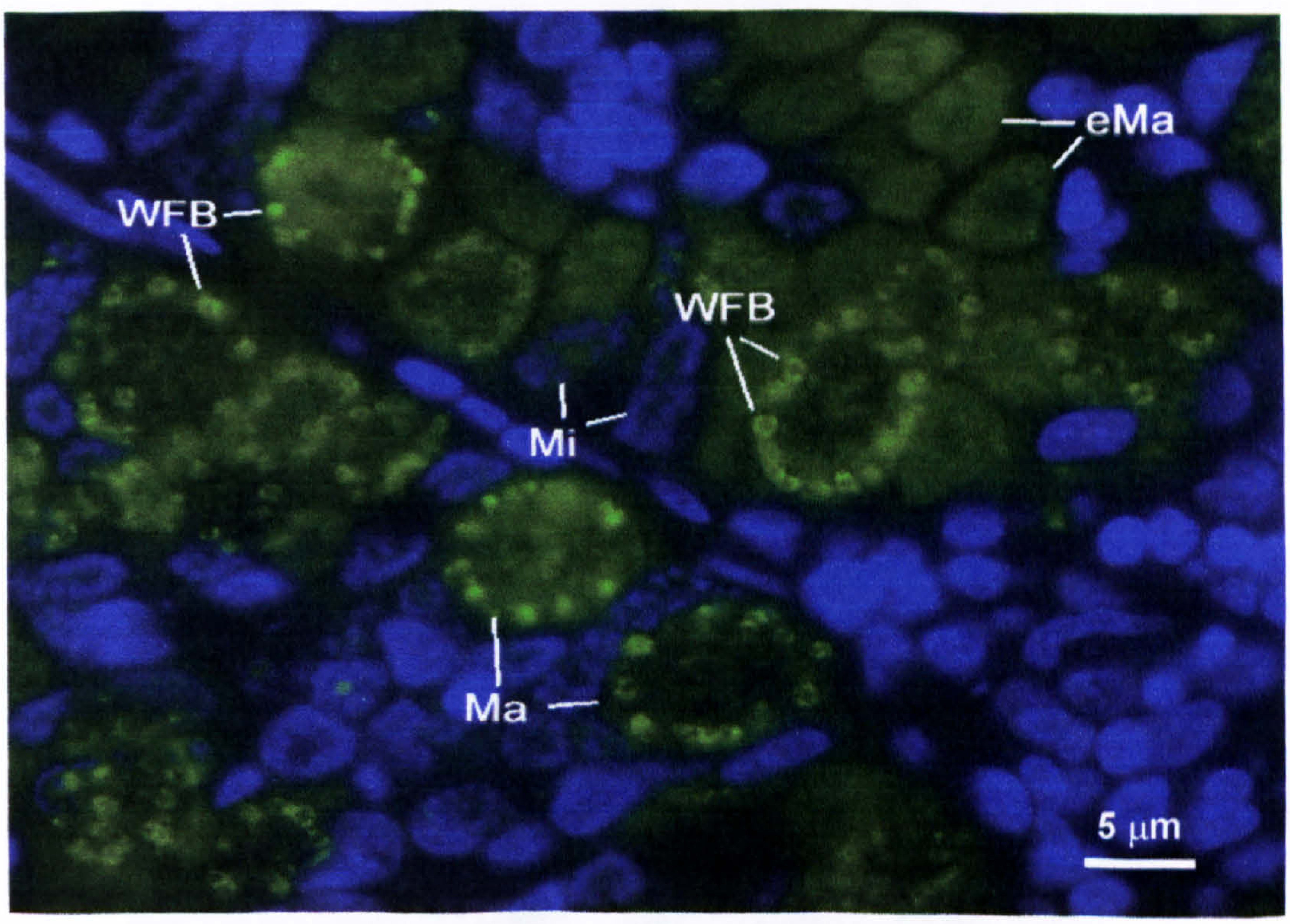


FIG. 3.41. **Immunolocalisation of the cathepsin B in the sexual stages of *E. tenella*.** Sections of the caecum of chickens infected for 96 h were used. Mi: microgamete; eMa: early macrogamete; Ma: macrogamete; WFB: wall forming bodies.

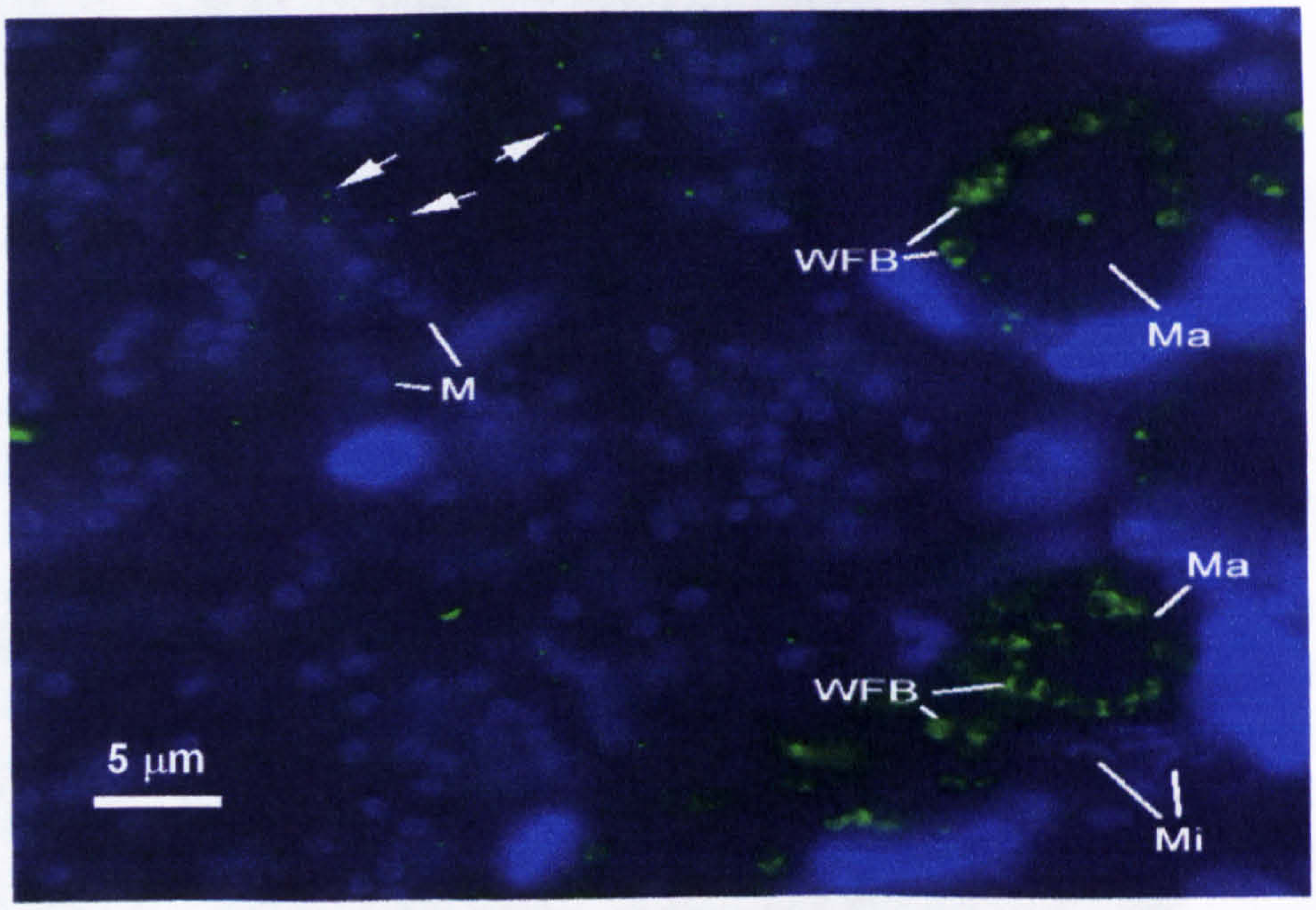


FIG. 3.42. **Immunolocalisation of the cathepsin B in the sexual stages of *E. tenella*.** Sections of the caecum of chickens infected for 96 h were used. M: merozoites; Mi: microgamete; Ma: macrogamete; WFB: wall forming bodies. Arrows indicate the staining at the tip of the merozoites.

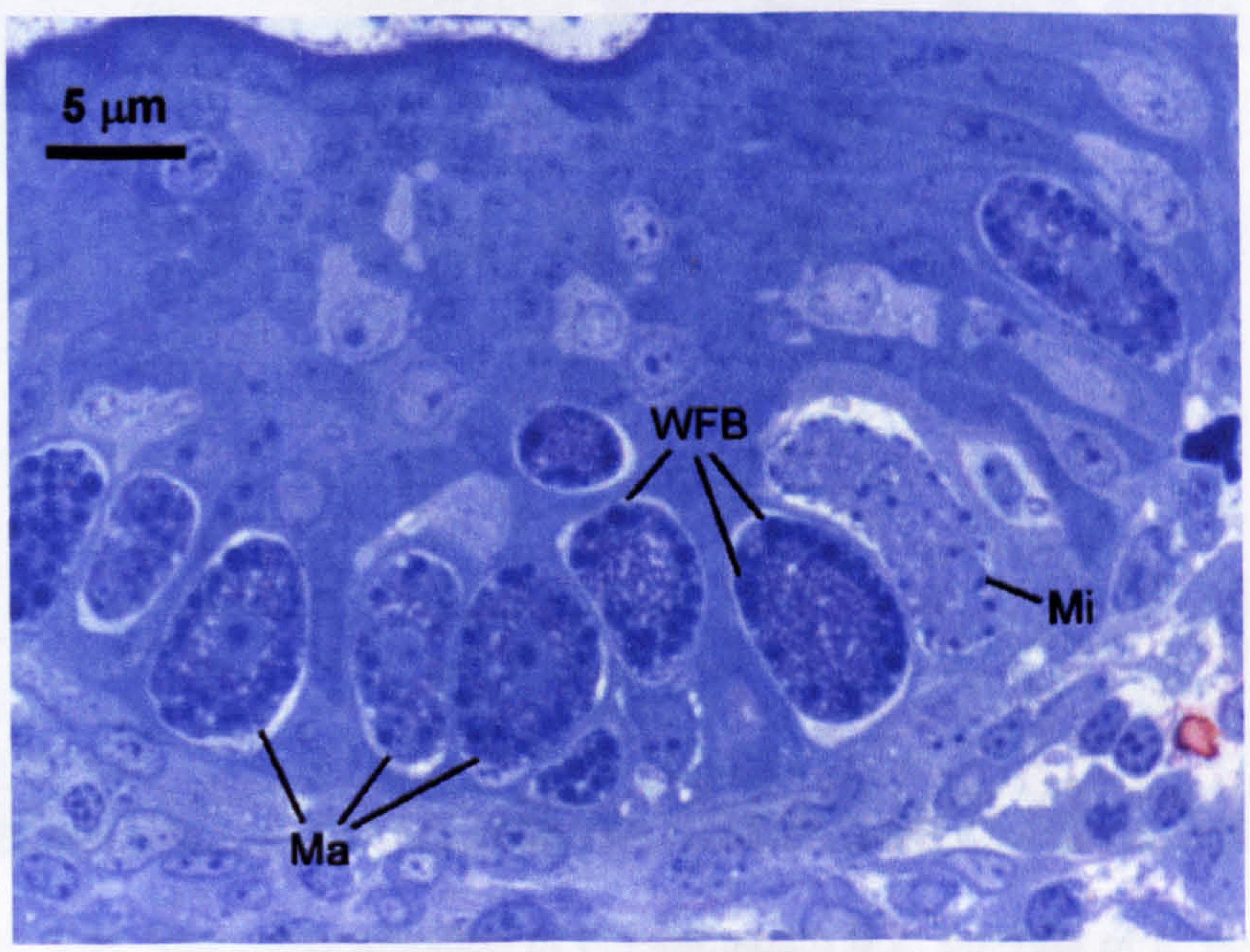


FIG. 3.43. **Phase microscopy of sexual stages of *E. tenella*.** Sections of the caecum of chickens infected for 96 h stained with azure A were used. Mi: microgamete; Ma: macrogamete; WFB: wall forming bodies.

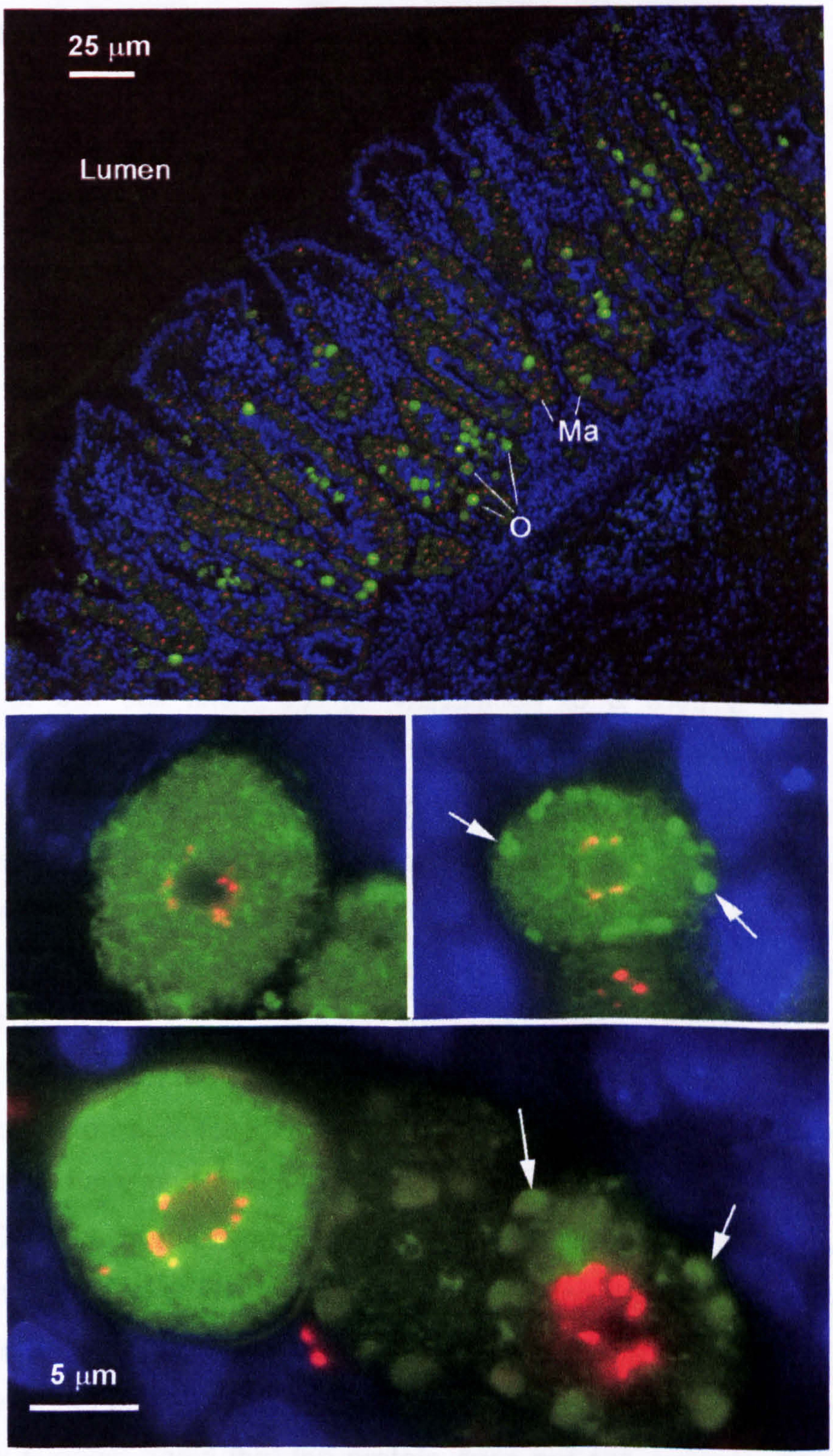


FIG. 3.44. **Immunolocalisation of the cathepsin B in oocysts of *E. tenella*.** Sections of the caecum of chickens infected for 136 h were used. Blue staining corresponds to DAPI staining of DNA. Red staining corresponds to staining of a perinuclear structure. Ma: macrogamete; O: oocyst. Arrows indicate structures that may be wall forming bodies.

3.3. Discussion

Peptidases have been shown to take part in essential events of the life cycle in *Eimeria*, such as the formation of the oocyst wall (Belli *et. al*, 2003a; Belli *et. al*, 2003b) and host cell invasion (Fullet and McDougald, 1990). By database mining using the cathepsin B sequence of *T. gondii* (Que *et al.*, 2002), a single complete sequence coding for a cathepsin B-like enzyme was identified in the *Eimeria tenella* EST database. The sequence was amplified on gDNA and cDNA and this showed that no introns are present. The gene was shown to be single copy. The predicted enzyme contained the cysteine, asparagine, histidine catalytic triad and showed good similarity with the cathepsin B of *T. gondii* (42%) and other cathepsin Bs from different organisms (Fig. 3.3).

Soluble recombinant pro-cathepsin B could be successfully expressed in various *E. coli* cell lines, but no maturation to a mature size protein with significant activity against the substrate Z-Phe-Arg-AMC, which has been shown to be a substrate for cathepsin B enzymes (Caffrey *et al.*, 2002), could be observed, despite the numerous conditions tested for unfolding, refolding and activation. This may suggest that the enzyme did not self-activate, or that the enzyme did not properly fold, and conditions for proper folding and self-activation were not obtained.

Expression of the cathepsin B-like enzyme without the full pro-domain may have been favourable for folding and/or activation. In this case, however it does not seem to have had a beneficial effect. No activity towards gelatin or Z-Phe-Arg-AMC could be detected in any of the conditions tested. Incubation at 37°C with 10 mM DTT seemed to yield some processing but the western blot analysis suggested that it did not happen at the N-terminal end, which would be where a cleavage would be expected in case of removal of the pro-region and activation. This may suggest again that the protein does not self-activate. The fact that the protein fell out of solution in the assays at pH below 6 tends to indicate a sub-optimal folding. As cathepsins usually fold around the pro-domain (Wiederanders, 2000; Cappetta *et al.*, 2002), this would not be unexpected. There has been proof though, that some cathepsins can fold and be active when expressed without their pro-domain (Sijwali *et al.*, 2001, Pandey *et al.*, 2004), hence the reason for testing this possibility.

It has recently been shown that the cathepsin B of *Schistosoma mansoni* that the activation process is not autocatalytic, rather it can be processed and activated using the endogenous asparaginyl endopeptidase (Sajid *et al.*, 2003). This is consistent with the conclusions we drew from our attempts to activate the *E. tenella* cathepsin B-like enzymes expressed in *E.coli*. Whether or not there is an asparaginyl endopeptidase in *E. tenella* that activates the cathepsin B-like enzyme is not known as yet (although no asparaginyl endopeptidase gene in the genome database of *Eimeria* could be identified). A variety of commercially available peptidases were tested for their ability to activate the soluble cathepsin B from *E. tenella* produced in *E. coli*. Although processing could be detected, there was no indication that any of the cleaved fragments corresponded to the mature cathepsin B, and no activity could be detected in any of the methods tested. Non-specific peptidases would in theory preferentially cut in the most accessible domains of a protein, that is to say, in accessible hinges and loops. By looking at the three dimensional structure of the human cathepsin B, one could expect that the *Eimeria* peptidase, provided it folds into a comparable structure, would not be cut in multiple places, as it looks like a rather compact globular protein. Nevertheless this method did not generate any active protein.

As no active peptidase could be generated by non-specific cleavage of the pro-mature protein, the idea came to introduce a cleavage site for specific peptidases commercially available, by mutagenesis at the predicted junction between mature and pro-domains. Specific cleavage using Factor Xa and TEV protease generated a protein the size of the mature cathepsin B. The mature fragment produced did not show activity in the conditions tested, even when assayed at a pH that would promote the dissociation of the pro-domain that might stay bound and act as an inhibitor (Carmona *et al.*, 1996). The fact that still no active enzyme could be produced might be due to the fact that some proteins expressed in *E. coli* do not get fully their native conformation even if expressed in a soluble manner (Jean *et al.*, 2000). It also might be due to the fact that the expression in *E. coli* lacks the typical eukaryotic phenomenon of glycosylation and the peptidase might need them to be active, as it is the case for the human glutamate carboxypeptidase II (Barinka *et al.*, 2004). This is why three *Pichia* systems were tested, in order to get soluble, released and active *Eimeria* cathepsin B.

Pro-mature cathepsin B could successfully be produced in the constitutive *Pichia* expression system, and the protein seemed to be successfully released into the culture medium. Furthermore, the suppression of the C-terminal His-tag seemed to favour processing of the pro-mature protein. No signs of processing could be detected in the supernatant of culture of the cells transfected with the plasmid encoding for the pro-mature cathepsin B C-terminal His-tagged, though some apparent processing was observed in the cell extracts. It is unknown why no complete processing and no significant activity against Z-Phe-Arg-AMC compared to the control cells could be detected in these cases.

Active mature recombinant cathepsin B of *Eimeria tenella* was successfully produced in the *Pichia pastoris* inducible system (Fig. 3.30). The enzyme was released in the supernatant of culture at a concentration of about 100 µg/ml, and was readily purified by sequential ammonium sulfate precipitation (Fig. 3.31). The characteristics of the enzyme conform to what is expected for a cathepsin B; it is strictly dependant on reducing agents (Fig. 3.33), it cleaves Phe-Arg and Arg-Arg substrates with a preference for Phe-Arg (Fig. 3.36), it is inhibited strongly by general cysteine peptidase inhibitors like E64 and iodoacetamide, as well as by the specific inhibitor K11777 (Table 3.1). Interestingly, it is active over a wide spectrum of pHs (Fig. 3.35), whereas most cathepsin Bs are mostly active at acidic pH. This might reflect a broad distribution of the enzyme in various locations in the cell.

The reason why the soluble enzyme produced in *E. coli* was not active remains unclear. Glycosylations, which do not occur in bacterial expression systems, did not seem to play an important role in the enzyme activity. It is still unknown if the cathepsin B of *Eimeria tenella* is capable of *cis*-maturation by autocatalytically removing its pro-domain, as it seems to be the case for the cathepsin B of *T. gondii* (Que *et al.*, 2002) and has been reported for other cysteine peptidases of the same family (Turk *et al.*, 2001), or if it requires an heterologous enzyme for processing, which may be provided in this case by the expressing host. If, indeed, a host enzyme is required for maturation, this may explain why different sorts of maturation patterns were observed in different

Pichia host cell lines. Recombinant expression in the *Pichia* inducible system of a version of the cathepsin B carrying mutations at the active site residues would allow determining if there is indeed *cis*- or *trans*-activation.

Immuno-blotting using anti-cathepsin B of *Eimeria tenella* polyclonal antibodies allowed detection of the enzyme in sporozoite and oocyst extracts (Fig. 3.37). The enzyme seems therefore to be expressed at these stages of the life cycle. The pro- and mature enzyme could also be detected in the fraction thought to correspond to microneme proteins, but not rhoptry proteins (Fig. 3.38). The absence of specific anti-microneme protein or anti-rhoptry protein antibodies of *E. tenella* does not allow to give a definite proof of the specificity of the anti-cathepsin B antibody binding, and other contaminating proteins could be detected in these fractions. Also, the organelle fractions are not completely pure, although preparations appear clean by observation by electron microscopy (Bromley *et al.*, 2003).

Immunolocalisation studies of the cathepsin B, on various *E. tenella* stages of the life cycle, were done by Prof David Ferguson (Oxford). The enzyme showed an anterior end localisation in sporozoites, similar to the pattern found for MIC2, a microneme protein of *E. tenella*. In merozoites, the enzyme seemed to localise to a single dot at one tip, but the signal was very faint and might be not real. The signal was diffuse in the cytoplasm and no apical localisation could be detected like with the anti-MIC2 antibodies. So the cathepsin B may localise to micronemes in sporozoites, but may not in merozoites. This might reflect a stage specific expression of the cathepsin B. But a cytoplasmic localisation in sporozoites might still be possible, as it cannot be excluded that microneme preparations might be contaminated by cytoplasmic material. Nevertheless, there was reasonable evidence that the cathepsin B of *E. tenella* did not localise to rhoptries. This is in contrast to the results obtained for toxopain of *T. gondii* (Que *et al.*, 2000), which was reported to localise to rhoptries, but could reflect a different role for the enzyme. Cysteine peptidases from *E. tenella* appear to have a role in host cell invasion, as invasion can be inhibited by cysteine peptidase inhibitors (Samantha Brown, PhD thesis); the sequencing of the genome of *E. tenella* will provide a good tool for investigation of the parasite's cysteine peptidase repertoire, though the presence of introns in most genes and lack of definite intron prediction system makes protein prediction still problematic. In sporozoites, the cathepsin B seemed to be found in microneme preparations. In the sexual stages of the parasite, the enzyme is located in mature macrogametes and early oocysts to granules, the size and location of which would be consistent with being wall-forming bodies. With formation of the oocyst wall, the localisation in the oocysts spread to the entire cytoplasm. Peptidases had previously been reported to be involved in the oocyst wall formation in *Eimeria* (Belli *et al.*, 2003a), by maturation of precursor proteins found in the wall forming bodies, but this is the first a time a cysteine peptidase has actually been shown to localise to such formations. Similar roles for cysteine peptidases have been suggested in *Giardia*. Three cathepsin B-like cysteine proteases, CP1, CP2 and CP3, were purified from *G. lamblia* (Ward *et al.*, 1997). CP2 might be involved in excystation of trophozoites as excystation is inhibited in the presence of CP2 specific inhibitor (Ward *et al.*, 1997), though this has not been definitely established. Another stage regulated cathepsin B-like cysteine protease known as ESCP has been reported to process a cyst wall protein and may therefore be involved in the parasite's encystation (Touz *et al.*, 2002). The oocyst wall formation being an essential process for parasite's survival in the external environment and propagation of the disease, these may be exciting findings. The enzyme is now available for drug screening, and provided specific inhibitors can be obtained, the cathepsin B of *E. tenella* might present a good new drug target.

Chapter 4

Expression and characterisation of recombinant pyroglutamyl peptidase I of *L. major*

4.1. Introduction

Pyroglutamyl peptidase (also referred to as pyrrolidonecarboxylate peptidase, pyroglutamate aminopeptidase, pyrrolidonecarboxylyl peptidase, 5-oxoprolyl peptidase, pyroglutamyl aminopeptidase and pyroglutamyl peptide hydrolyse) belongs to the class of omega peptidases (EC 3.4.19.-) and hydrolyses N-terminal L-pyroglutamate residues (L-pGlu) from peptides and proteins containing this modification (Browne and O'Cuinn, 1983; Cummins and O'Connor, 1996). Pyroglutamyl peptidase activity was first detected in *Pseudomonas fluorescens* (Doolittle and Armentrout, 1968) and has been since described in tissues from mammals (Dando *et al.*, 2003; Cummins and O'Connor, 1996), bacteria (Ito *et al.*, 2001; Ogasahara *et al.*, 2001), birds (Tsuru *et al.*, 1982), fish, reptiles (Prasad *et al.*, 1982) and plants (Kembhavi *et al.*, 1993). They are divided into two classes, the cysteine peptidase pyroglutamyl peptidase I (EC 3.4.19.3., MEROPS C15 [180]), and the metallopeptidase pyroglutamyl peptidase II (EC 3.4.19.6., MEROPS M1 [181]). Pyroglutamyl peptidase I is a soluble, intracellular enzyme (Cummins and O'Connor, 1998), whereas pyroglutamyl peptidase II is membrane-bound (O'Connor and O'Cuinn, 1984).

4.1.1. Pyroglutamyl metabolism in human

4.1.1.1. Pyroglutamyl peptidase I

Pyroglutamyl peptidase I (PPI) occurs as a monomeric enzyme of relatively low molecular mass (~ 25 kDa) and with a pH optimum from 6.5 to 8.5, that is located in the cytosol in mammals (Mantle *et al.*, 1991; Cummins and O'Connor, 1996; Dando *et al.*, 2003). PPI activity exhibits a strict requirement for reducing environment (Dando *et al.*, 2003). It has broad substrate specificity in that it can hydrolyse a variety of natural or synthetic peptides containing L-pGlu at the N-terminal position (Cummins and O'Connor, 1996; Dando *et al.*, 2003). The rate of hydrolysis of the various peptides depends on the amino acid residue following the pGlu in the sequence and it has been shown that mammalian PPI does not cleave pGlu-Pro bonds (Browne and O'Cuinn, 1983). Although PPI has been shown to have broad substrate specificity, it also has a strong specificity for pGlu residues and will not cleave modified pGlu residues (Capecchi and Loudon, 1985). A schematic representation of the catalytic reaction is shown on Fig. 4.1.

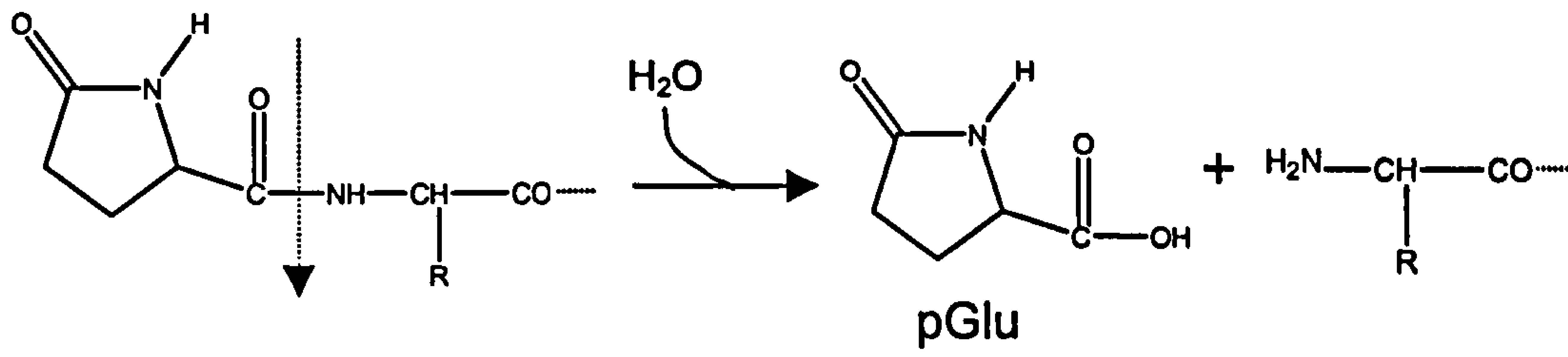


FIG. 4.1. Schematic representation of the reaction catalysed by PPI. The dotted arrow represents the cleavage site by PPI. The dashes at the C-terminus represent the continuation of the amino acid chain. The reaction releases free pyroglutamate (pGlu) and the amino acid chain. R: variable side group of the amino acid.

The catalytic triad of the PPI was found to be composed of Glu-81, Cys-144 and His-168 (*Bacillus amylotumefaciens* numbering, Yoshimoto *et al.*, 1993). These residues are conserved in the human and the mouse enzymes (Dando *et al.*, 2003).

Due to the nature of the active site and the requirement of the catalytic cysteine, sulfhydryl-blocking agents (iodoacetate, iodocetamide) are potent inhibitors (Dando *et al.*, 2003). Metallic ions such as Cu^{2+}, Zn^{2+} and Ni^{2+} were also shown to be strong inhibitors, reversible by addition of a chelating agent in excess or by dialysis (Dando *et al.*, 2003). PPI was not inhibited by serine peptidase inhibitors, such as phenylmethylsulphonyl fluoride (PMSF) or cysteine peptidase (family C1) inhibitors, such as (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl}carbamoyl)oxirane-2-carboxylic acid (E64). Among active site specific inhibitors, chloromethyl ketone derivatives of pGlu are irreversible, potent inhibitors (Fujiwara *et al.*, 1982). Other specific inhibitors include aldehyde analogues of pGlu, natural compounds isolated from *Streptomyces*, and the pyroglutamyl analogue 2-pyrrolidone, also used to stabilise PPI activity during storage (Cummins and O'Connor, 1998).

The biological activity of some peptides is regulated via modifications such as cyclisation of N-terminal glutamine to L-pGlu, acetylation of the N-terminal residue or amidation of the carboxy terminal residue. The pyroglutamyl group confers a relative stability towards aminopeptidase degradation to these peptides. In fact, 12.3% of proteins with signal peptides are initiated with Gln once the signal peptide is removed (Liao *et al.*, 2003). Also, removal of the N-terminal translation initiator methionine is often crucial for the function and stability of proteins (Hirel *et al.*, 1989; Liao *et al.*, 2003), and pGlu-modified peptides can then occur if the second amino acid is a Gln, though this is not systematic (Predel *et al.*, 1999). Many of the naturally occurring peptides containing the pGlu modification are substrates for PPI (see Table 4.1). Among them are neuropeptides. Neuropeptides are up to 40 amino acid-long peptides released to allow inter-cellular communication, and regulate processes such as differentiation, growth and metabolism in mammals (Cummins and O'Connor, 1998). In addition to giving resistance to aminopeptidase cleavage, the pGlu modification has also been shown to be essential to the bioactivity of some peptides. For instance, the chemotactic protein-2 activity in human is dependent on its pGlu modification, as receptor recognition involves the N-terminal part of the molecule (Van Coillie *et al.*, 1998). In human, free pGlu has been shown to improve learning and to help prevent age-associated memory loss (Grioli *et al.*, 1990).

As the L-pGlu N-terminal modification can be determinant in biological activity of peptides (Perlman *et al.*, 1994), PPI has been attributed a role of neuropeptide regulation and metabolism in mammalian systems, though the lack of its release from the cells still causes doubts regarding its function (Salers *et al.*, 1991). PPI may participate in the protein metabolism (Lauffart and Mantle, 1988) and increased levels of PPI have been observed in the spinal cord of patients with motor neurone disease (Mantle *et al.*, 1995). However, only very few mammalian PPIs have been fully characterised so far and a definite physiological role still has to be attributed.

Table 4.1. **Some natural bioactive compounds cleaved by mammalian PPI.** TRH, thyrotropin-releasing hormone; LHRH, luteinizing hormone-releasing hormone; ACE, angiotensin converting enzyme. [1], Fujiwara *et al.*, 1979; [2], Cummin and O'Connor, 1996; [3], Dando *et al.*, 2003.

Protein/peptide	N-terminal cleavage site	Organism	Reference
TRH	pGlu-His	Human	[1,2]
LHRH	pGlu-His	Human	[1,2,3]
Bombesin	pGlu-Gln	Frog	[2,3]
Bradykinin potentiator B	pGlu-Gly	Human	[1,3]
ACE inhibitor	pGlu-Trp	Human	[3]
Leukopyrokinin	pGlu-Thr	Human	[3]
Neurotensin	pGlu-Leu	Human	[1,2,3]
Litorin	pGlu-Gln	Frog	[1]

4.1.1.2. Pyroglutamyl peptidase II

A distinct PP activity membrane-bound and inhibited by reducing and chelating agents was demonstrated in several mammals (Friedman *et al.*, 1986; Gallagher *et al.*, 1997). Compared to PPI, PPII has a relatively high molecular mass (~ 230 kDa) and is not inhibited by PPI inhibitors (Friedman *et al.*, 1985). PPII purified from rat and porcine brain showed a dimeric architecture characteristic of membrane-bound peptidases (Bauer, 1994). The activity of this metallopeptidase is mainly located in the central nervous system (CNS) in mammals, where it is found mainly associated with neurons (Bauer *et al.*, 1990; Cruz *et al.*, 1991b). A characteristic of PPII is that it is highly specific for thyrotropin-releasing hormone (TRH) and TRH-like peptides (Elmore *et al.*, 1990), which stimulate the release of thyroid-stimulating hormones. The increased release of thyroid hormones increases the metabolic rate, glucose absorption, blood pressure and heart rate, decreases the cholesterol level in plasma, stimulates growth hormone production, and regulates the CNS development (Cummins and O'Connor, 1998). PPII is therefore thought to be an important regulatory element that influences the activity of TRH at target sites (Schomburg and Bauer, 1995). PPII expression and activity is indeed regulated by TRH (Joseph-Bravo *et al.*, 1998). The only site-specific inhibitor of PPII known to date is CPHNA (*N*-[1(*R,S*)-carboxy-2-phenylethyl]-*N*-imidazole benzyl-histidyl β-naphthylamide) (Charli *et al.*, 1989).

A third PP activity was isolated from serum and its narrow specificity for TRH-like peptides lead to the name "serum thyroliberinase" (Bauer and Nowak, 1979). However, given its high resemblance to the PPII activity, it is possible, but not yet known, that it is a product of the same gene. There has been recent evidence that the thyroliberinase activity results from the cleavage of the PPII liver enzyme (Schmitmeier *et al.*, 2002).

4.1.1.3. Glutaminyl cyclase

Glutaminyl cyclase (QC), or glutaminyl-peptide cyclotransferase (EC 2.3.2.5., MEROPS clan MH, family M28), is the enzyme involved in cyclisation of N-terminal glutamine, which is unmasked by proteolytic processing of precursor proteins by, for example, prohormone convertases (Seidah *et al.*, 1998). Peptides modified by a QC are constitutively secreted in humans, and tissue-specific forms of the enzyme may exist (Sykes *et al.*, 1999). The enzyme localises to the secretory pathway (Pohl *et al.*, 1991). The cyclisation of glutamine can occur spontaneously, especially under the catalytic influence of phosphate ions (Chikuma *et al.*, 2004), but previous studies have demonstrated that the formation of pGlu residues in cells must be enzymatically catalysed because the non-enzymatic conversion is slow under physiological conditions (Consalvo *et al.*, 1988). QCs have also been identified in various mammals (Busby *et al.*, 1987) and plants (Oberg *et al.*, 1998). It was previously proposed the mammalian QC has some features in common with zinc aminopeptidases, and human pituitary glutaminyl cyclase and bacterial zinc aminopeptidase share a common fold and active site residues. In contrast to the aminopeptidase, however, QC does not appear to require zinc for enzymatic activity (Booth *et al.*, 2004). The human QC possesses 6 key active site residues: histidines 140 and 330, glutamates 201 and 202, and aspartates 159 and 248 (Bateman *et al.*, 2001), although the 2 aspartates have no catalytic role. Little is known about the biological role of QC at present. Researchers have suggested that QC is responsible for modification of storage proteins during seed germination in plants (Gololobov *et al.*, 1996), as well as in vivo modification of bioactive peptides in mammals (Hinke *et al.*, 2000). These findings support the view that pGlu modification might be important for protection of the N-terminus of bioactive peptides against exopeptidases.

Moreover, this enzymatically catalysed N-terminal formation of the pyroglutamic acid residue could be important in developing the proper receptor-binding conformation of such peptides (Ziegler *et al.*, 1998).

4.1.2. Pyroglutamyl peptidase activity in bacteria

Most bacterial PPs examined so far are soluble, have broad substrate specificity, require a reducing environment and occur as multimers (Fujiwara *et al.*, 1979; Awade *et al.*, 1992a; Awade *et al.*, 1992b). In terms of substrate specificity, catalytic residues and inhibitors, they resemble the mammalian PPI enzymes. Structures for the PPs of *Thermococcus litoralis* and *Bacillus amylotumefaciens* are known (Protein Data Bank [PDB] identification numbers 1A2Z and 1AUG, respectively). The known PPs bacterial genes are single copy, and are characterised by inverted repeats of the gene found upstream and downstream of the ORFs, suggesting transcriptional control (Cummins and O'Connor, 1998). The expression of the PP gene of *Pseudomonas fluorescens* is regulated by the intracellular levels of pyroglutamic acid and iron (Le Saux and Robert-Baudouy, 1997). Like for the mammalian PPIs, a definite function still has to be attributed. It has been suggested they may play a role in intracellular protein metabolism (Lazdunski, 1989), though the presence of PP activity is not ubiquitous among the bacterial kingdom.

The only indication of the presence of a QC in bacteria was presented by Cook and Russell in 1991. They suggested the glutamine cyclase reaction of *Streptococcus bovis* was a mechanism of deriving energy from non-oxidative and non-reductive deamination.

4.1.3. Bioactive peptides containing pGlu

Aside from neuropeptides, various bioactive peptides possess a N-terminal L-pGlu modification and can be synthesised by various organisms. As examples, a variety of anti-microbial peptides have been isolated from mammals (Tang *et al.*, 2002; Boman, 2003), insects (Rabel *et al.*, 2004), shrimp (Destoumieux *et al.*, 1997), spider (Silva *et al.*, 2000), and scorpions (Cociancich *et al.*, 1993). The widespread occurrence of these compounds suggests they may play a role in innate immunity against microorganisms and other pathogens

(Selsted and Ouellette, 1995; Ganz and Lehrer, 1999). Gomesin, an 18 amino acid N-terminal pGlu-modified peptide isolated from a spider (Silva *et al.*, 2000) showed strong bacterial growth inhibition, affected development of filamentous fungi and yeast, as well as the viability of the parasite *Leishmania amazonensis*. The mechanism of inhibition of growth is unknown. However, this observation might support a need for bacteria and other pathogens, such as parasites, to possess a PP activity, for degradation of such bioactive compounds. There is some evidence that the innate immune response of humans involves various lytic factors, including antibacterial peptides. These peptides are classified in different families, depending on their structure. Humans possess 2 classes of defensins, α- and β-defensins, which are cationic 15-45 amino acid long peptides with a pattern of disulfide bridges essential to the activity (Torres and Kuchel, 2004). Defensins are produced in the phagolysosomes of neutrophils and macrophages as well as in epithelial cells (Selsted and Ouellette, 1995) and play a role in innate immunity acting as anti-microbial agents and chemokines (Hoover *et al*, 2001; Schibli *et al.*, 2002). Some of the β-defensins have a pGlu modification at the N-terminus (Selsted *et al.*, 1993). Another family of anti-microbial peptides in humans, the cathelicidins, are produced in epithelial cells as pre-pro-proteins. The pro-protein possesses a N-terminal pGlu (Andersson *et al.*, 2002), and upon maturation, the mature C-terminal domain acts as a potent anti-microbial agent (Isogai *et al.*, 2003). In insects, various homologues of the defensins exist, namely the cecropins (Cornet *et al.*, 1995). The only reported proofs of action of human anti-microbial peptides on parasitic protozoa are that resistance in humans against *Trypanosoma brucei brucei* can occur as the parasite is killed by lytic factors (anti-microbial peptides) found in the serum (Lugli *et al.*, 2004), and that the enteric parasite *Giardia lamblia* can be killed by cationic neutrophil peptides produced by the intestine epithelium (Aley *et al.*, 1994). In *Drosophila*, an anti-microbial peptide with a N-terminal pGlu has been identified (Rabel *et al.*, 2004). So there is some evidence that pGlu-modified peptides with anti-microbial activity exist in human and insects, the two hosts of *Leishmania*.

4.1.4. Pyroglutamyl peptidase in parasites

The sequence for a PP with resemblance to mammalian PPI was identified in the genomes of *L. major*, *L. infantum*, *T. brucei* and *T. cruzi*. Interestingly, no sequence with identity to the human PPI could be identified in the genomes of *P. falciparum*, *P. yoelii*, *T. gondii*, *E. tenella*, *E. histolytica*, *S. mansoni*, or *T. vaginalis*. It is still unclear if the presence of a PPI may be specific to kinetoplastids, as the sequencing and annotation of most of the genomes cited are uncompleted. The sequence for a possible glutaminyl cyclase, the enzyme responsible for the cyclisation of Gln residues to pGlu, and sequences homologous to the mammalian PPII enzyme were also identified in the genomes of *L. major*, *L. infantum*, *T. brucei* and *T. cruzi*. No studies have previously been reported on PP activity in parasites and it is unknown what role it may play. But one hypothesis we considered is that it might be involved in bioactive peptide inactivation in the mammalian host or in the insect vector. As these parasite PPs seem to resemble mammalian PPIs, it may be predicted that they may have a broad substrate specificity and be able to cleave various regulatory and anti-microbial factors.

This study focused on the PPI of *L. major*. The aim was to express recombinant active PPI for biochemical characterisation in order to get an insight on which of the eukaryotic or prokaryotic enzymes it is the most closely related to, in terms of oligomerisation, substrate specificity, kinetic features and inhibitors. A second aim was to try to obtain recombinant glutaminyl cyclase, for characterisation and antibody production, in order to investigate the occurrence and possible role of the enzyme in the parasite.

4.2. Results

4.2.1. Cloning of the pyroglutamyl peptidase I gene of *L. major*

Database searches on the genomic database of *L. major* using the human PPI as a query (accession number NP 060182, www.ncbi.nlm.nih.org/entrez/) detected a 834 bp open reading frame (Fig. 4.1) translating into a 277 amino acid protein (LmjF34.2000, www.genedb.org) with the PPI features, although it was annotated as a hypothetical protein of unknown function. No other sequence with PPI features could be detected in the *L. major* database (now totally sequenced), which indicates the PPI is a single copy gene in *L. major*. The amino acid sequence was analysed with the SignalP algorithm (Nielsen *et al.*, 1997), and TMpred (Hofmann and Stoffel, 1992), and no signal peptide or trans-membrane domain were predicted.

5' RACE on total cDNA was performed using a 5' primer with a sequence complementary to the splice leader (SL primer) and a 3' primer with a sequence complementary to a sequence downstream of the predicted start ATG. The PCR product obtained was used as a template for a semi-nested PCR using the SL and the PPrace2 primers (see Table 2.1). The results of the 5' RACE showed the presence of a mRNA comprising a 49 bp 5' untranslated region (UTR), and confirming the start methionine attributed in LmjF34.2000 (Fig. 4.2).

```
  1  GTCCTCAGCA CTGAAAGGGA ATCATCCTCC ACCTTCCCTC CCCAACGACC ACCGCACACG TGCACCGCAC
     CAGGAGTCGT GACTTTCCCT TAGTAGGAGG TGGAAGGGAG GGGTTGCTGG TGGCGTGTGC ACGTGGCGTG
 71  GATCAATGTC TCACTATTGC AAGCGGCCTT CTACGAGCAC ACAGGGCTCC TCGCAGTTCC GCACCGCTTA
     CTAGTTACAG AGTGATAACG TTCGCCGGAA GATGCTCGTG TGTCCCGAGG AGCGTCAAGG CGTGGCGAAT
       M  L  H   S  K  A   G  I  V  V   F  I  T   G  Y  G   P  F  A  T   V  K  V ·
141  AAATGCTGCA CTCCAAGGCG GGCATCGTCG TGTTCATCAC GGGCTACGGG CCCTTCGCCA CTGTGAAGGT
     TTTACGACGT GAGGTTCCGC CCGTAGCAGC ACAAGTAGTG CCCGATGCCC GGGAAGCGGT GACACTTCCA
     · N  P  S   S  D  I  A   L  R  V   A  E  G   L  K  R  H   P  D  V   A  E  V
211  GAACCCCAGC TCCGACATTG CCCTTCGTGT CGCGGAGGGC CTGAAGCGCC ACCCCGACGT CGCGGAGGTG
     CTTGGGGTCG AGGCTGTAAC GGGAAGCACA GCGCCTCCCG GACTTCGCGG TGGGGCTGCA GCGCCTCCAC
      R  Y  T  E   L  D  V   S  V  T   S  V  A  A   Y  F  E   K  V  E   R  D  T  A ·
281  CGCTACACGG AGCTGGATGT CAGCGTGACC AGCGTCGCAG CGTACTTTGA GAAGGTGGAA AGGGACACTG
     GCGATGTGCC TCGACCTACA GTCGCACTGG TCGCAGCGTC GCATGAAACT CTTCCACCTT TCCCTGTGAC
     ·  D  I  I   A  E  H   G  A  G  Q   V  K  I   L  L  C   H  L  G  V   H  N  D ·
351  CCGACATCAT CGCCGAACAC GGCGCCGGCC AGGTGAAGAT CCTTCTCTGC CACCTAGGTG TTCACAATGA
     GGCTGTAGTA GCGGCTTGTG CCGCGGCCGG TCCACTTCTA GGAAGAGACG GTGGATCCAC AAGTGTTACT
     · T  T  G   L  I  C  V   E  V  Q   G  C  N   E  L  F  S   S  V  P   D  V  D
421  CACAACGGGG CTCATCTGCG TTGAGGTGCA GGGCTGCAAC GAGCTCTTTT CTAGCGTCCC TGACGTGGAC
     GTGTTGCCCC GAGTAGACGC AACTCCACGT CCCGACGTTG CTCGAGAAAA GATCGCAGGG ACTGCACCTG
      G  K  V  L   N  H  E   P  I  V   P  E  D  G   A  I  E   V  F  H   E  S  W  F ·
491  GGTAAGGTGC TCAACCACGA GCCCATTGTA CCGGAGGACG GCGCCATTGA AGTATTTCAC GAGAGCTGGT
     CCATTCCACG AGTTGGTGCT CGGGTAACAT GGCCTCCTGC CGCGGTAACT TCATAAAGTG CTCTCGACCA
     ·  G  K  E   G  S  P   Q  L  E  K   L  E  R   L  I  Q   Q  V  N  D   T  V  A ·
561  TCGGCAAGGA GGGCTCGCCG CAGCTCGAGA AGCTGGAGCG CTTGATTCAA CAGGTAAACG ACACCGTTGC
     AGCCGTTCCT CCCGAGCGGC GTCGAGCTCT TCGACCTCGC GAACTAAGTT GTCCATTTGC TGTGGCAACG
     · E  S  W   H  H  W  V   T  G  A   V  T  N   N  E  V  T   S  A  D   T  D  R
631  CGAGTCATGG CATCACTGGG TGACGGGTGC CGTGACCAAC AATGAGGTGA CATCGGCGGA CACAGACAGA
     GCTCAGTACC GTAGTGACCC ACTGCCCACG GCACTGGTTG TTACTCCACT GTAGCCGCCT GTGTCTGTCT
      K  D  M  A   M  P  T   F  Q  A   P  S  S  A   I  S  R   N  A  G   R  Y  L  C ·
701  AAGGACATGG CGATGCCAAC GTTTCAGGCG CCGAGCTCTG CAATTTCGCG CAACGCAGGG CGCTACCTGT
     TTCCTGTACC GCTACGGTTG CAAAGTCCGC GGCTCGAGAC GTTAAAGCGC GTTGCGTCCC GCGATGGACA
     ·  N  C  A   L  Y  H   A  L  R  L   Q  E  K   N  P  G   V  V  Y  G   I  F  V ·
771  GCAACTGCGC CCTCTACCAC GCACTTCGAC TTCAGGAAAA GAACCCGGGT GTCGTCTATG GCATCTTCGT
     CGTTGACGCG GGAGATGGTG CGTGAAGCTG AAGTCCTTTT CTTGGGCCCA CAGCAGATAC CGTAGAAGCA
     · H  V  V   D  P  I  R   G  K  T   E  I  E   G  G  P  I   V  A  Y   N  P  P
841  TCACGTTGTC GATCCGATTC GCGGCAAGAC GGAGATCGAG GGCGGTCCCA TTGTGGCTTA CAACCCACCG
     AGTGCAACAG CTAGGCTAAG CGCCGTTCTG CCTCTAGCTC CCGCCAGGGT AACACCGAAT GTTGGGTGGC
      T  I  V  Q   S  V  Q   V  Q  C   L  M  H  G   L  L  S   L  M  T   M
911  ACGATCGTAC AGTCGGTGCA AGTGCAGTGT CTCATGCACG GCCTGCTGTC TTTGATGACC ATGTGACGAA
     TGCTAGCATG TCAGCCACGT TCACGTCACA GAGTACGTGC CGGACGACAG AAACTACTGG TACACTGCTT
981  GACATCGTTC GCAGCGA
     CTGTAGCAAG CGTCGCT
```

FIG. 4.2. **Complete sequence encoding a PPI enzyme in *L. major*.** The sequence of the 5' UTR obtained from the 5' RACE experiment is shown with a grey background. The translation of the putative ORF is shown in bold (single letter amino acid code). Dots indicate codons that are split over the line break.

The 834 bp ORF translating into a 277 amino acid protein with PPI features was located on chromosome 34 of *L. major*. The protein was aligned with known PPIs from mammals, plant and bacteria (Fig. 4.3) and the corresponding identity table is presented in Table 4.2. The *L. major* protein showed little similarity with the various PPIs, but the catalytic residues Glu-81, Cys-144 and His-168 were conserved in all the sequences (*Thermococcus litoralis* numbering), placing the *L. major* enzyme in the C15 family of cysteine peptidase (Barrett and Rawlings, 2001). Only 13 residues were fully conserved in the 5 sequences. The *L. major* enzyme also has a large insertion (position 145-199) of unknown function.

```
                1                                                50
  AthPPI    (1) MGSEGPTGVTIHITGFKKFHGVAENPTEKMANNLKEYLAKNCVSKDVNLG
  BamPPI    (1) ------MEKKVLLTGFDPFGGETVNPSWEAVKRLNGAAEGP---------
 TlitPPI    (1) -------MKKVLITGFEPFGGDSKNPTEQIAKYFDRKQIGN---------
   HsPPI    (1) ---MEQPRKAVVVTGFGPFGEHTVNASWIAVQELEKLGLGDS--------
   LmPPI    (1) -MLHSKAGIVVFITGYGPFATVKVNPSSDIALRVAEGLKRHPDVAEVR--
                51                                              100
  AthPPI   (51) SCTVLETAGQGALASLYQLLQSAVNTKESESLTGKTIWVHFGVNSGAT-K
  BamPPI   (36) -------ASIVSEQVPTVFYKSLAVLREAIKKHQPDIIICVGQAGGRM-Q
 TlitPPI   (35) -------AMVYGRVLPVSVKRATIELKRYLEEIKPEIVINLGLAPTYS-N
   HsPPI   (40) -------VDLHVYEIPVEYQTVQRLIPALWEKHSPQLVVHVGVSGMAT-T
   LmPPI   (48) YTELDVSVTSVAAYFEKVERDTADIIAEHGAGQVKILLCHLGVHNDTTGL
                101                                             150
  AthPPI  (100) FAIEQQAVNEATFRCPDELGWKPQNLPIVPSDGPISTVRKTNLPVEE---
  BamPPI   (78) ITPERVAINLNEARIPDNEGNQPVGEDIS-QGGPAA--YWTGLPIKR---
 TlitPPI   (77) ITVERIAVNIIDARIPDNDGYQPIDEKIE-EDAPLA--YMATLPVRA---
   HsPPI   (82) VTLEKCGHNKGYKGLDNCR-FCPGSQCCV-EDGPES--IDSIIDMDA---
   LmPPI   (98) ICVEVQGCNELFSSVPDVDGKVLNHEPIVPEDGAIEVFHESWFGKEGSPQ
                151                                             200
  AthPPI  (147) ----------------------ITKALEKN------------------GF
  BamPPI  (122) ----------------------IVEEIKKE------------------GI
 TlitPPI  (121) ----------------------ITKTLRDN------------------GI
   HsPPI  (125) ----------------------VCKRVTTLGL----------------DV
   LmPPI  (148) LEKLERLIQQVNDTVAESWHHWVTGAVTNNEVTSADTDRKDMAMPTFQAP
                201                                             250
  AthPPI  (157) EVITSDDAGRFVCNYVYYHSLRFAEQNKT--RSLFVHVPLFVAVDEET--
  BamPPI  (132) PAAVSYTAGTFVCNHLFYGLMDEISRHHPHIRGGFIHIPYIPEQTL---Q
 TlitPPI  (131) PATISYSAGTYLCNYVMFKTLHFSKIEGYPLKAGFIHVPYTPDQVVNKFF
   HsPPI  (137) SVTISQDAGRYLCDFTYYTSLYQS-----HGRSAFVHVPPLGKPYNADQL
   LmPPI  (198) SSAISRNAGRYLCNCALYHALRLQEKNPGVVYGIFVHVVDPIRGKTEIEG
                251                                     292
  AthPPI  (203) ----------------QMRFTVSLLEVLASICK---------
  BamPPI  (179) KS---APSLSLDHITKALKIAAVTAAVHEDDIETGGGELH--
 TlitPPI  (181) LLGKNTPSMCLEAEIKAIELAVKVSLDYLEKDRDDIKIPL--
   HsPPI  (182) G--------------RALRAIIEEMLDLLEQSEGKINYCHKH
   LmPPI  (248) GP------------IVAYNPPTIVQSVQVQCLMHGLLSLMTM
```

FIG. 4.3. **Alignment of the pyroglutamyl peptidase I (PPI) of *L. major* with various known PPIs from different organisms.** Dashes indicate gaps in the alignment. The catalytic triad residues are starred. Identical residues are on black background. Conservative residues are on grey background Ath, *Arabidopsis thaliana*; (GenBank Accession No.: NP 173758); Bam, *Bacillus amyloliquefaciens* (JX0244); Tlit, *Thermococcus litoralis* (CAA74299); Hs, *Homo sapiens* (NP 060182), Lm, *L. major*; PPI, pyroglutamyl peptidase I.

Table 4.2. **Similarity table of the *L. major* PPI with various PPIs.** Numbers represent percentage of identity. ArabTh : *Arabidopsis thaliana*; Bamyl : *Bacillus amyloliquefaciens*; Tlit : *Thermococcus litoralis*; Hum : *Homo sapiens*; Lm : *L. major.*

	AthPPI	BamPPI	TlitPPI	HsPPI	LmPPI
AthPPI	100	32	33	33	23
BamPPI		100	49	38	15
TlitPPI			100	37	15
HsPPI				100	20
LmPPI					100

Using the human PPI as a query (accession number NP 060182, www.ncbi.nlm.nih.org/entrez/) also allowed detection of PPIs in the genomic databases of other parasites. For instance, a PPI-like sequence was detected in the genome of *L. infantum*, *Trypanosoma brucei* and *T. cruzi*. These sequences were aligned as before (Fig. 4.4) and a somewhat higher similarity was detected among them (Table 4.3). *T. brucei* and *T. cruzi* PPs are quite similar to each other, as are the *L. infantum* and *L. major* PPs. Interestingly, the *L. major* PP was only just under 30% homologous to the PPs of the 2 *Trypanosoma* species, which is the percentage of similarity between the human and the bacterial enzymes (Dando *et al.*, 2003).

To study the structure that the *L. major* PP might have, it was sent to SwissModel for structure prediction, using the PPs from *T. litoralis* and *B. amylotumefaciens* (PDB 1A2Z and 1AUG, respectively) as templates. Only the residues 5 to 41 and 193 to 240 could be assigned a structural prediction by superposition to residues 1 to 36 of the *B. amylotumefaciens* PP or residues 128 to 171 of *T. litoralis*, respectively. The rest of the sequence did not show sufficient similarity for three-dimensional structure prediction.

```
               1                                                 50
LiPPI    (1)   MLHSKAGIVVFITGYGPFATVKVNPSSDIALRVVENLKRHPDVAEMRYTE
LmPPI    (1)   MLHSKAGIVVFITGYGPFATVKVNPSSDIALRVAEGLKRHPDVAEVRYTE
TbPPI    (1)   --MKPTKPLLYITGYGPFLEVTENPSATIAQSVAEQVRQS-GEADVHHET
TcPPI    (1)   --MRRSNLQLYMTGFGPFKSITVNPSAMIGESVSKRMMED-SELQVHYEK
               51                                                99
LiPPI    (51)  LDVSVTTVAAYFEKVEREIADIIAEHGAGKVKILLCHLGVHN-DTTGLIR
LmPPI    (51)  LDVSVTSVAAYFEKVERDTADIIAEHGAGQVKILLCHLGVHN-DTTGLIC
TbPPI    (48)  LDVNLEAVSKYFNRLNESVTAHLEATHPEN-RVLLVNVGLHSREKEKVLR
TcPPI    (48)  LEVGLEAVSAYFDGLEGALAQHMEDNPDG--CVLLLHIGLHSREEEGKMR
                                                                149
LiPPI    (100) VEVQGYNELFASVPDVDGKVLNHEPIVPEDGTIEVFHESWFGKAGSPQLE
LmPPI    (100) VEVQGCNELFSSVPDVDGKVLNHEPIVPEDGAIEVFHESWFGKEGSPQLE
TbPPI    (97)  LEVRAFNELEGNP-------------IDDELPLSTCKDSAFVKG--CKLE
TcPPI    (96)  LEVCGYNELEGSP-------------IDESIPMDVCLESMLVRDGCDESK
               150                                              199
LiPPI    (150) KLERLIQQVNDTIAESWHHSVTGAVTKNEVTPSDADRKDMAMPAFQAPSW
LmPPI    (150) KLERLIQQVNDTVAESWHHWVTGAVTNNEVTSADTDRKDMAMPTFQAPSS
TbPPI    (132) TTTALIEELNAIERN-------------------------GSDHHEKPRW
TcPPI    (133) ATAELIEELNAAISK---------RGR-------------REGDRQQPHW
               200                                               249
LiPPI    (200) AISRNAGRYLCNCALYHALRLQEKNPGIVYGVFIHVVDPIRGKTEIEGGP
LmPPI    (200) AISRNAGRYLCNCALYHALRLQEKNPGVVYGIFVHVVDPIRGKTEIEGGP
TbPPI    (157) IISYDAGRYYCNYALYRGVKMQEALNSRVFAVFLHIVASTVVC-------
TcPPI    (161) IISRDAGRYYCNYTLYRSLKLQKKWRGRVFAVFVHVANPFTCN-------
               250                              283
LiPPI    (250) IVAYNPTTMVQSVQVQRFMHGLLSLMT------
LmPPI    (250) IVAYNPPTIVQSVQVQCLMHGLLSLMTM-----
TbPPI    (200) -------MEEQVAQVRMLVSHLLKHMEAVE---
TcPPI    (204) ----NPSLEEQTAQVLSLVSGLVRIVRNKHIV-
```

FIG. 4.4. **Alignment of the pyroglutamyl peptidase I (PPI) of *L. major* with various predicted PPIs in different trypanosomatids.** Dashes indicate gaps in the alignment. The catalytic triad residues are starred. Identical residues are on black background; conservative residues are on grey background. Li, *Leishmania infantum* (GeneDB CDS: LinJ34.1440); Tb, *Trypanosoma brucei* (GeneDB CDS: Tb04.1H19.1350); Tc, *Trypanosoma cruzi* (GeneDB CDS: Tc00.1047053506635.60); Lm, *L. major* (GeneDB CDS: LmjF34.2000); PPI, pyroglutamyl peptidase I.

Table 4.3. **Similarity table of the *L. major* PPI with various parasites' PPIs.** Numbers represent percentage of identity. Li : *L. infantum*; Lm : *L. major*; Tb : *T. brucei*; Tc : *T. cruzi*.

	LiPPI	LmPPI	TbPPI	TcPPI
LiPPI	100	78	26	27
LmPPI		100	27	28
TbPPI			100	56
TcPPI				100

4.2.2. Expression of recombinant *L. major* PPI in *E. coli*

The full open reading frame (ORF) for the PPI of *L. major* was cloned in the pET28a(+) vector (Novagen) for expression of PPI with a N-terminal His-tag. The resulting pETLmPP plasmid is represented on Fig. 4.5. Cells transformed with the pETLmPP plasmid were induced either for 4 h at 37°C with 1 mM IPTG, or for 4 h at 20°C with 1 mM IPTG. An over-expressed protein of about 32 kDa, corresponding to the expected size for N-terminal His-tagged protein, could be detected in all cases, but it was found in the insoluble fraction of *E. coli* lysates if expressed at 37°C. The highest rate of expression in a soluble manner was obtained by induction at 20°C with 1 mM IPTG, and the soluble protein produced was purified on Ni-agarose column (Fig. 4.6). The BioCAD purification on Ni-agarose column resulted in a protein above 95% pure, although a few contaminating *E. coli* proteins remained. From a 1 liter induced culture, about 4 mg of enzyme were routinely purified. As a control, non-transformed BL21(DE3) *E. coli* cells were treated in the same manner and both the insoluble and soluble fractions were analysed using Coomassie-blue stained SDS-PAGE. No over-expressed protein could be observed in any of the fractions (results not shown). The soluble fraction was kept as a control for the analysis of the enzyme activity.

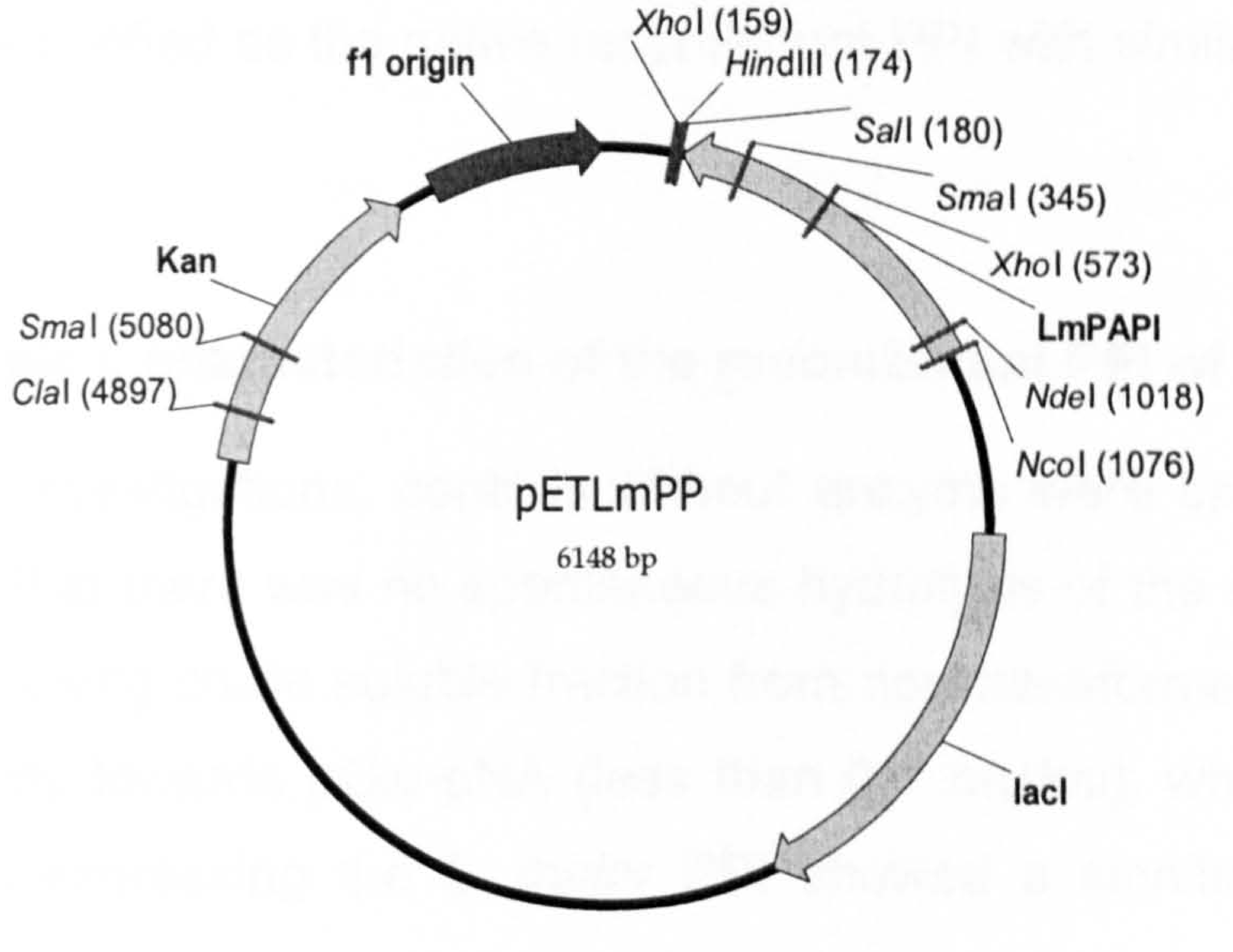


FIG. 4.5. **Construct for expression of recombinant pyroglutamyl peptidase I (PPI) of *L. major* in *E. coli*.** Schematic representation of the plasmid resulting from the cloning of the sequence of the PPI of *L. major* in the pET28a(+) vector (Novagen) for expression of PPI with a N-terminal His-tag. Kan, kanamycin resistance gene; f1 origin, origin of replication of the phage f1; lacI, gene encoding the *lac* repressor protein.

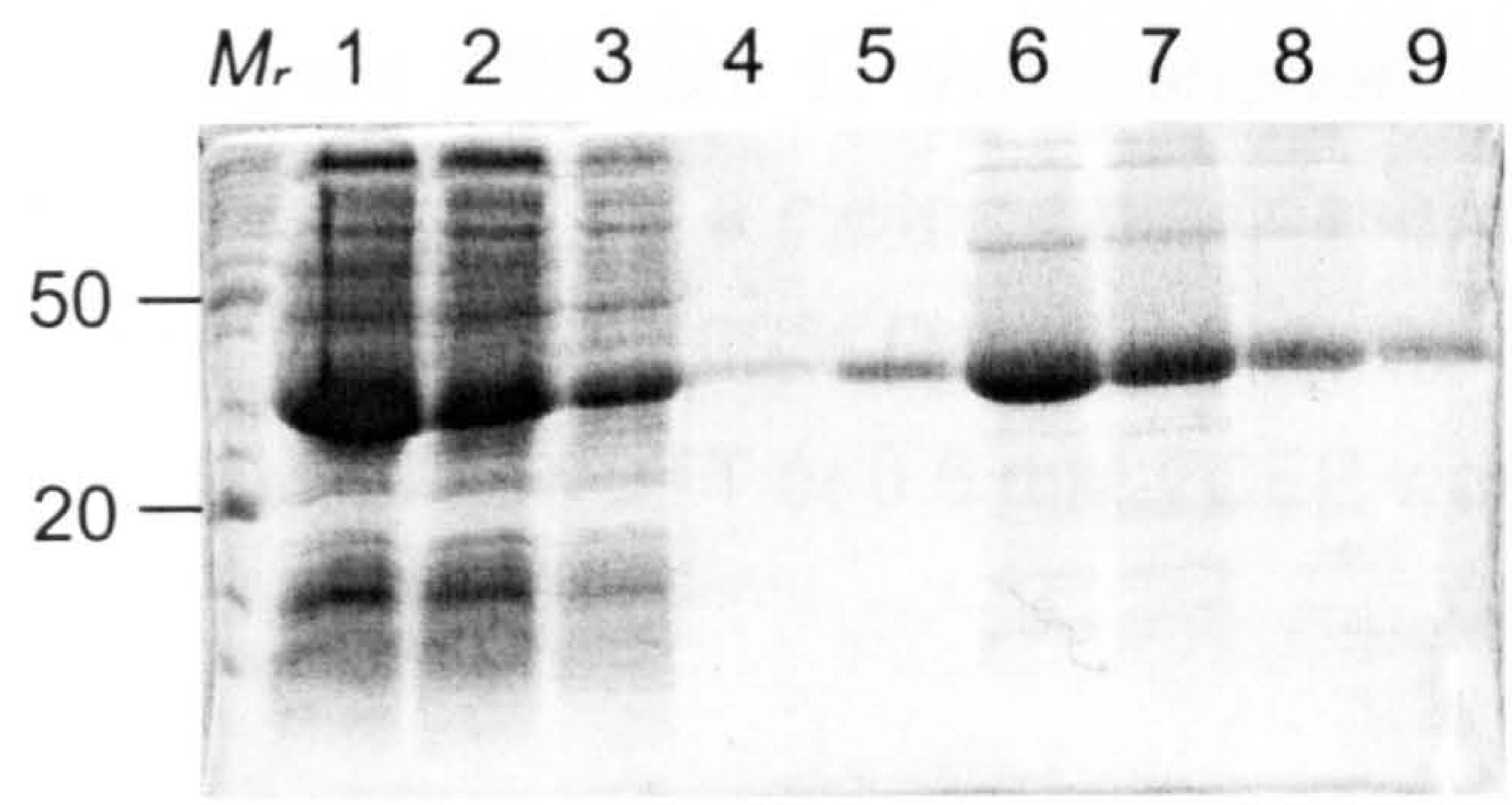


FIG. 4.6. **Analysis of expression and purification of the *L. major* PPI expressed in *E. coli* BL21(DE3).** A SDS-PAGE gel (12% acrylamide) stained with Coomassie blue. Lane M_r, protein standards; lane 1, *E. coli* cell lysate after IPTG induction, crude lysate; lane 2, load for Ni-agarose column; lane 3, flow-through; lane 4, 30 mM imidazole wash; lanes 5 to 9, elution fractions from Ni-agarose column with 500 mM imidazole. 10^{-3} x M_r is indicated on the left of the figure. 10 μl samples were loaded.

Active site mutants of the PPI of *L. major* were generated by site-directed mutagenesis. Based on the work of Le Saux and co-workers (Le Saux *et al.*, 1996), cysteine-210 was mutated to alanine, histidine-234 to serine, glutamic acid-101 to glutamine and glutamic acid-107 to glutamine (*L. major* PPI numbering). The mutated *L. major* PPIs were all partially soluble when induced at 20°C, and were purified as the native recombinant PPI with similar yields (results not shown).

4.2.3. Biochemical characteristics of the recombinant PPI of *L. major*

For all investigations, controls without enzyme were carried out, and it was confirmed that there was no spontaneous hydrolysis of the substrates used. Control assays using crude soluble fraction from non-transformed *E. coli* did not show any activity towards pGlu-pNA (less than 0.1 mU/ml), whereas the same extract of cells expressing the *L. major* PPI showed a significant activity (~1 U/ml).

4.2.3.1. Dependence on reducing conditions, EDTA and magnesium ion

The recombinant PPI from *L. major* showed a strong dependence on reducing agents, as expected for a cysteine peptidase and as reported for other PPIs (Cummins and O'Connor, 1998; Dando *et al.*, 2003). The presence in the assay buffer of at least 10 mM DTT or 0.5 mM TCEP was required to observe full activity (Fig. 4.7).

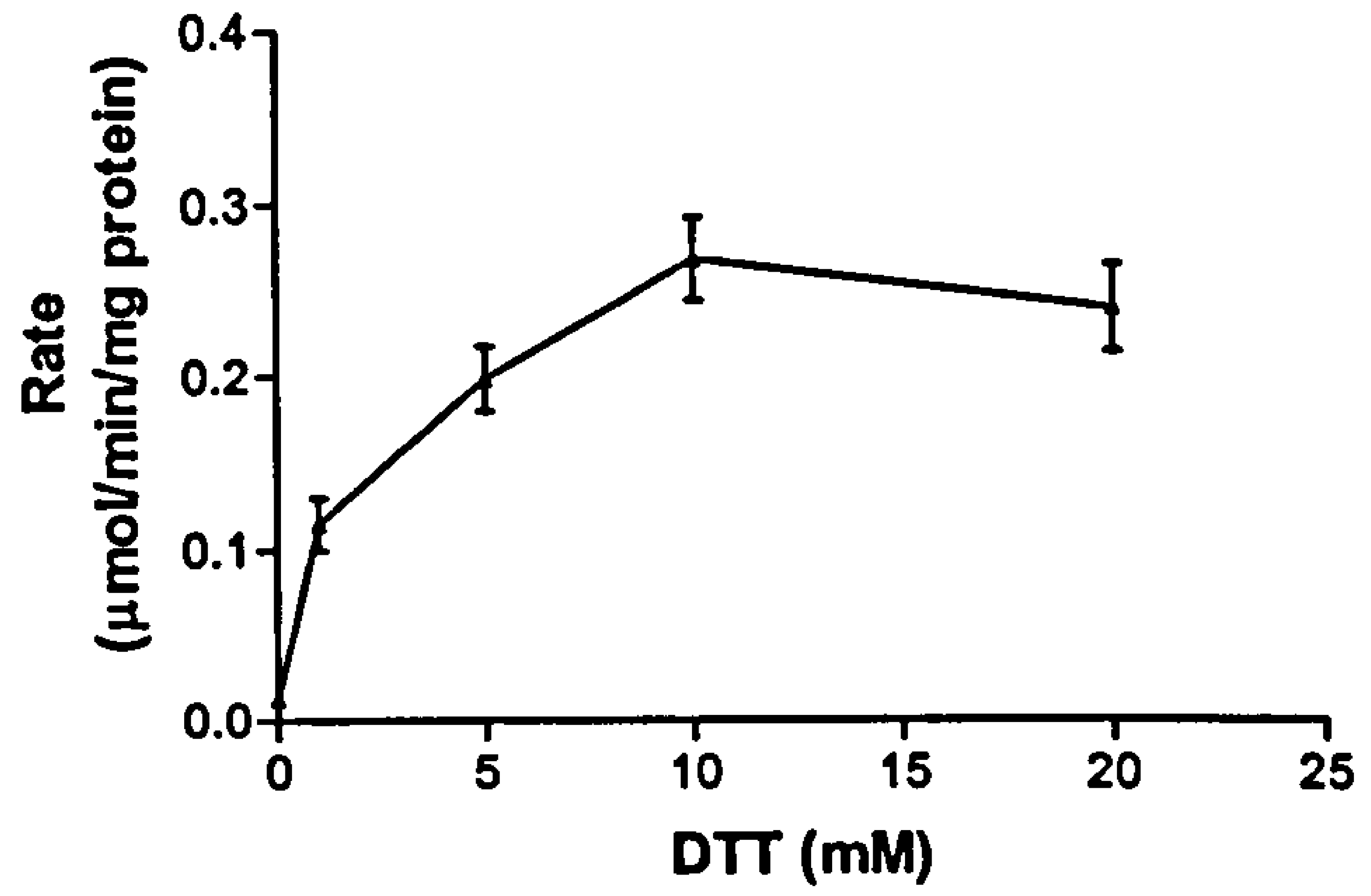


FIG. 4.7. **Dependence of *L. major* PPI activity on DTT.** Assays used 10.6 µg enzyme, 2 mM pGlu-pNA and were in 25 mM potassium phosphate buffer, pH 8.0. Results are the means ± SD from 3 independent measurements.

As for the human PPI, no stabilising effect was observed in the presence of EDTA (Dando *et al.*, 2003) for the *L. major* recombinant enzyme (Fig. 4.8), but 1 mM EDTA was routinely included in the assay buffer to chelate ions like Ni^{2+}, which may be present in traces due to Nickel-agarose purification and have been shown to be inhibitory on PPI enzymes (Dando *et al.*, 2003, section 4.2.3.6.).

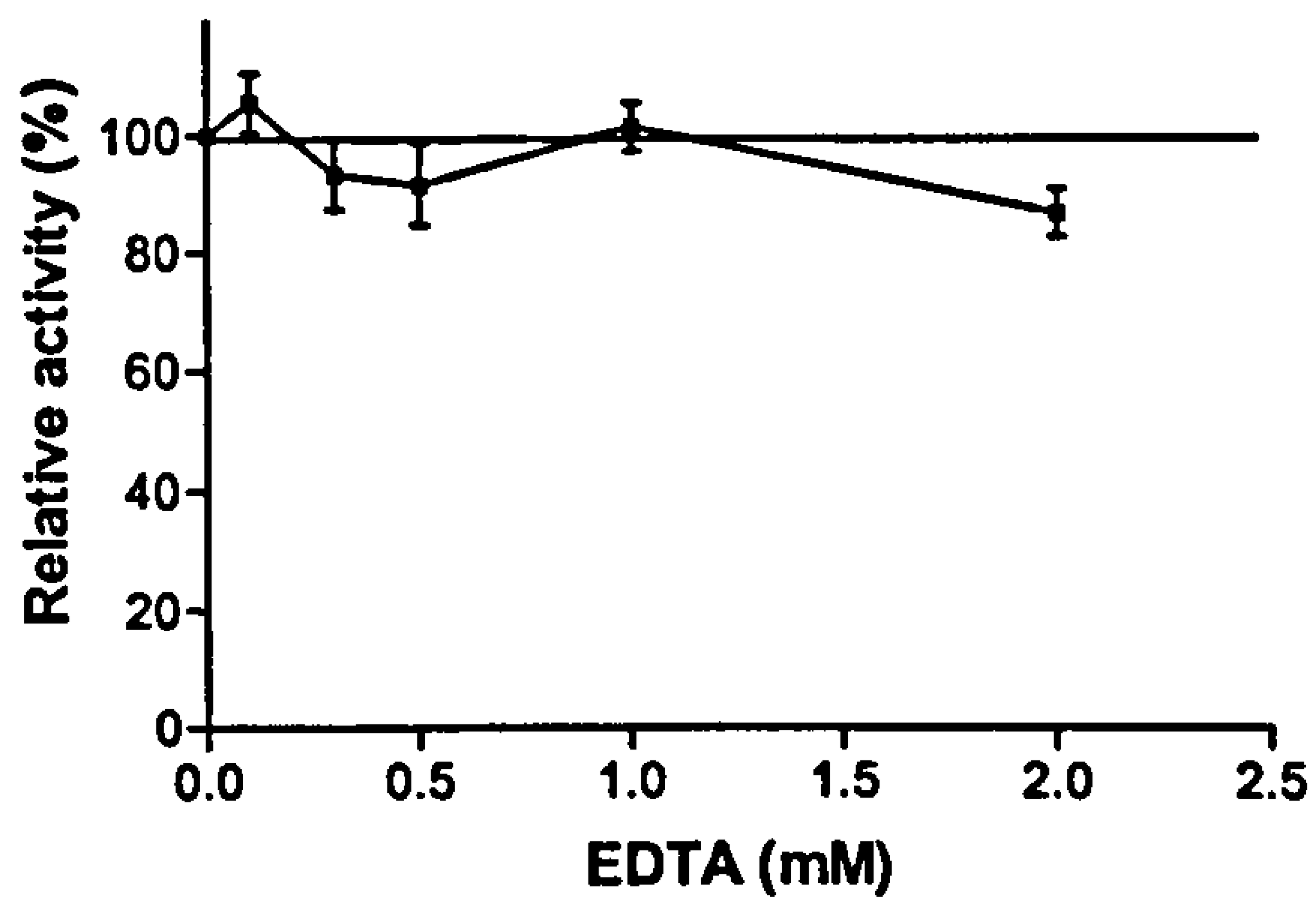


FIG. 4.8. **Effect of EDTA on *L. major* PPI activity.** Experiments used 11.4 µg enzyme, 2 mM pGlu-pNA and were in 50 mM HEPES buffer, pH 8.0, with 10 mM DTT. Results are the means ± SD from 3 independent measurements.

The effect of Mg^{2+} ions, which would be chelated by the EDTA included in the buffer and might be beneficial to enzyme activity, was investigated. No beneficial effect was observed (Fig. 4.9), and therefore Mg^{2+} was not added to the assays.

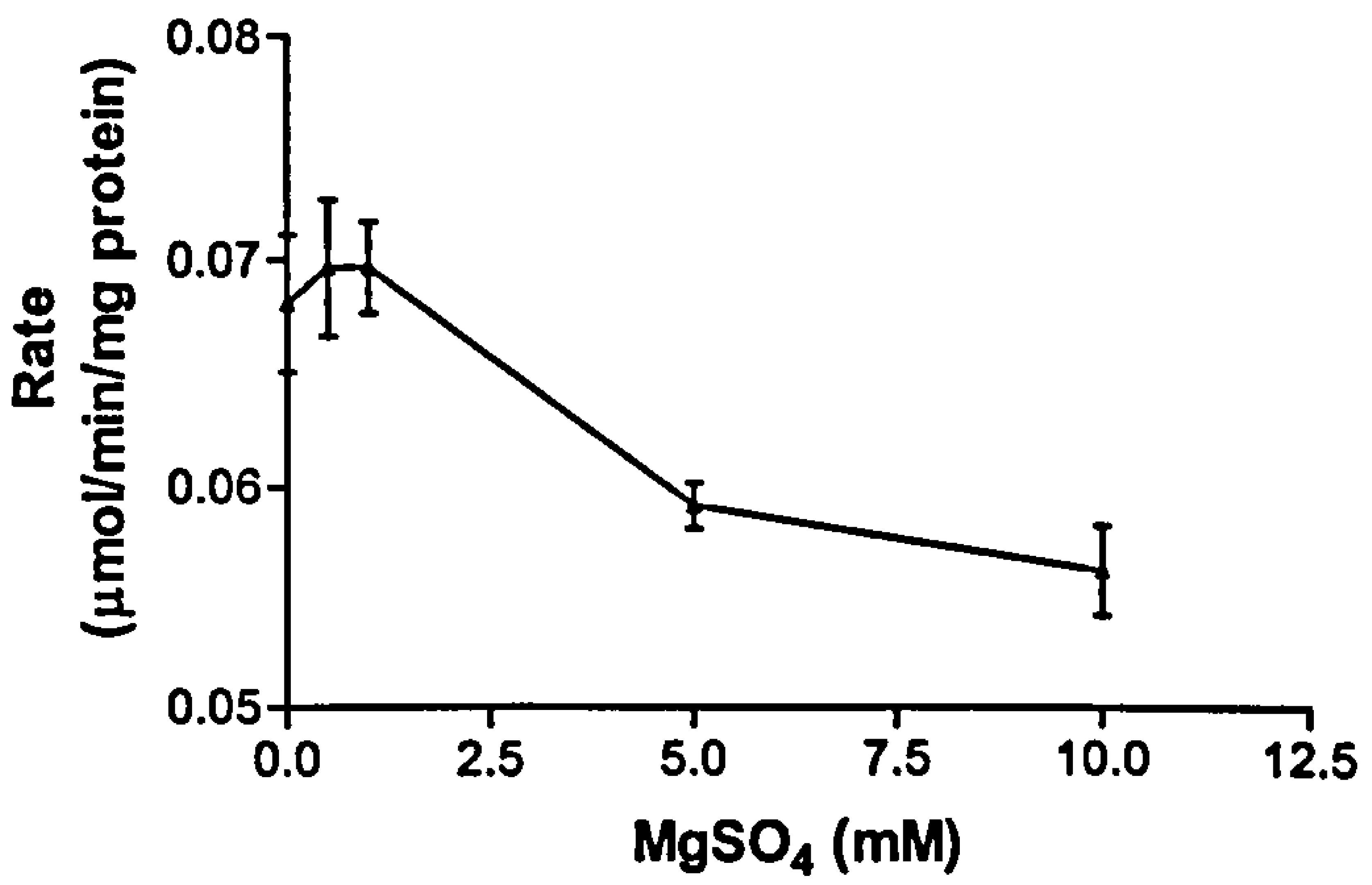


FIG. 4.9. **Effect of Mg^{2+} on *L. major* PPI activity.** Experiments used 11.4 µg enzyme, 2 mM pGlu-pNA and were in 50 mM HEPES buffer, pH 8.0, with 10 mM DTT. Results are the means ± SD from 3 independent measurements.

4.2.3.2. pH optimum

Activity towards the synthetic substrate pGlu-pNA was observed over a wide spectrum of pHs, with maximum activity being between pH 8.8 and 10.5 (Fig. 4.10). This is slightly higher than has been reported for other PPIs (Awade *et al.*, 1992a, Dando *et al.*, 2003), but confirms an alkali optimum for the enzyme. The activity decreased rapidly below pH 8.0. Though the observed pH optimum was above 8, assays were performed at pH 8.0, for biological significance.

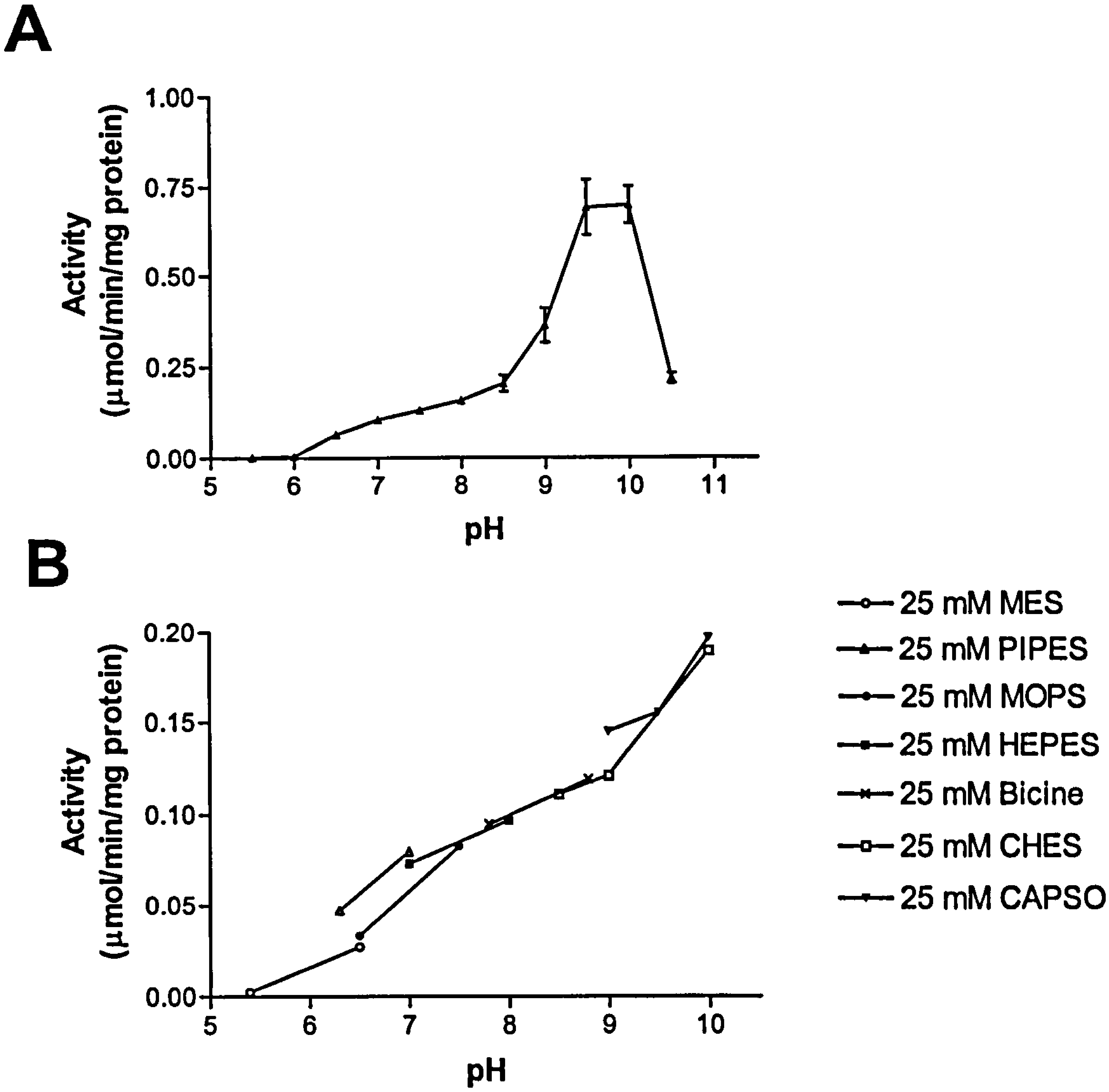


FIG. 4.10. **Effect of pH on *L. major* PPI activity.** (A) Experiments used 11.4 µg enzyme and were in the mix of buffers (25 mM acetic acid, 25 mM MES, 75 mM Tris-base, 25 mM Glycine), supplemented with 1 mM EDTA, 10 mM DTT, 2 mM pGlu-pNA. (B) Experiments used 11.4 µg enzyme and were in the buffers indicated in the legend, supplemented with 1 mM EDTA, 10 mM DTT, 2 mM pGlu-pNA. Results are the means ± SD from 3 independent measurements.

4.2.3.3. Storage and stability

The recombinant *L. major* PPI was found to be stable for at least 3 months at -20°C in 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM TCEP, and 3 h at 37°C but lost almost all activity in 6 days storage at 4°C in the same suspension buffer (Fig. 4.11). The enzyme was thus routinely stored at -20°C, and fresh aliquots were defrosted immediately prior to analysis.

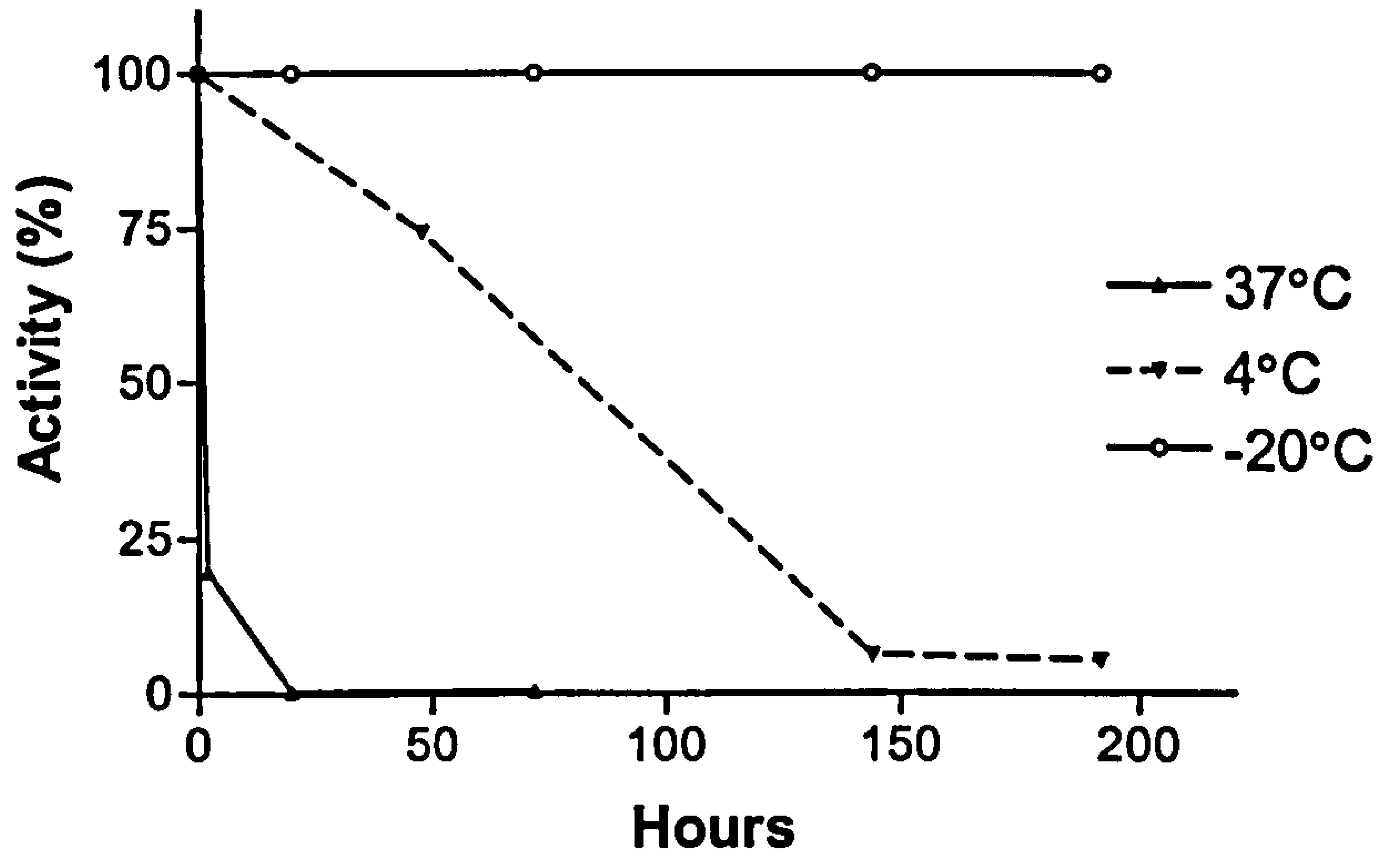


FIG. 4.11. **Stability of *L. major* PPI, stored in 50 mM HEPES pH 8.0, 2 mM TCEP, 1 mM EDTA.** Experiments used 10.5 µg enzyme, 2 mM pGlu-pNA and were in 50 mM HEPES pH 8.0, supplemented with 2 mM TCEP and 1 mM EDTA.

4.2.3.4. *L. major* PPI is a monomer

Gel filtration analysis showed that the active recombinant enzyme has a molecular mass of about 36 kDa (Fig. 4.12). The predicted molecular mass for a monomer of the enzyme is 32.5 kDa. The fraction containing the protein showed significant activity towards pGlu-βNA (1 U/mg protein). Mammalian PPIs are also monomers (Dando *et al.*, 2003).

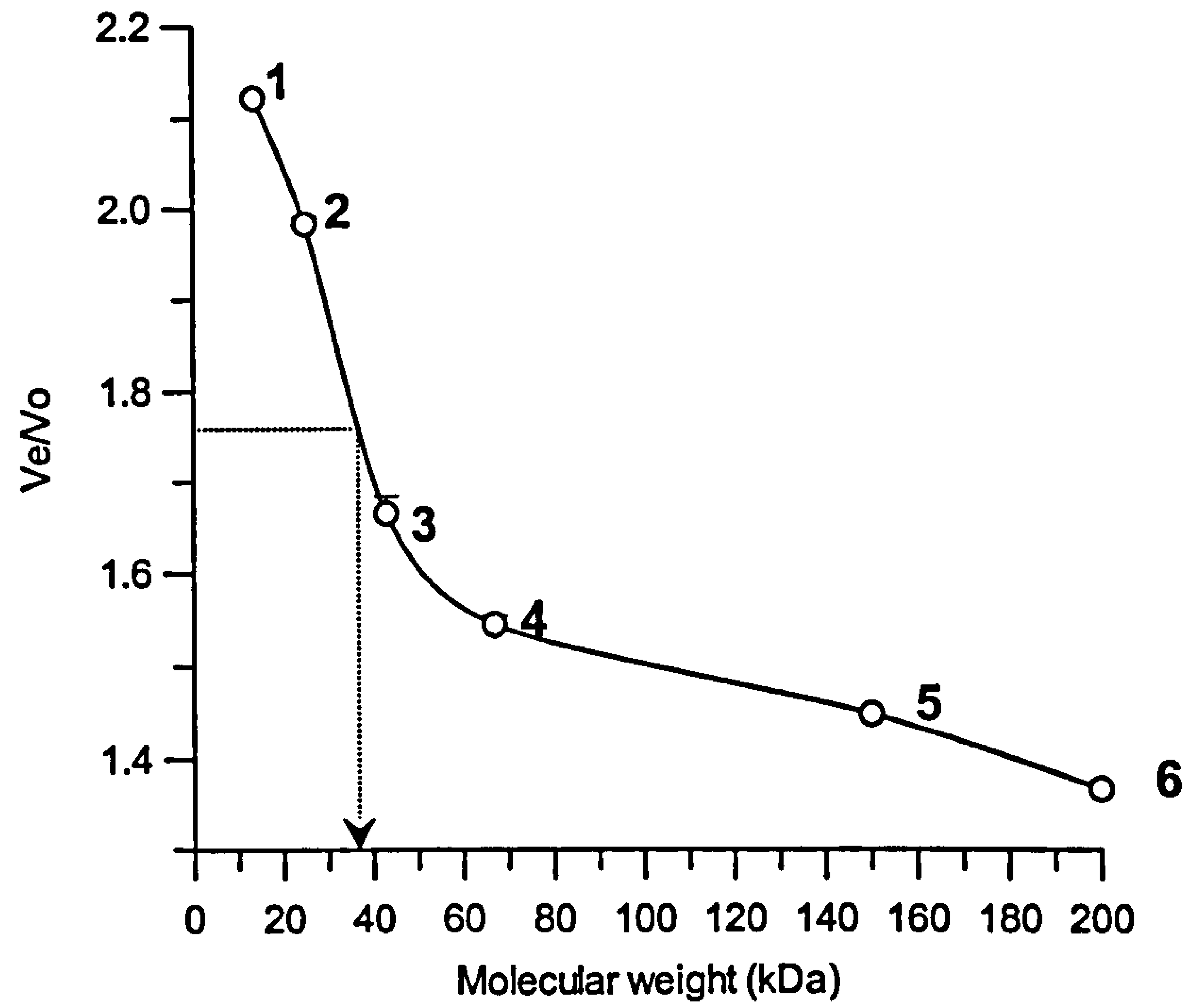


FIG. 4.12. **Determination of the molecular mass of the recombinant *L. major* PPI by gel-filtration analysis.** The enzyme was produced in *E. coli* with a N-terminal His-tag. The purified PPI was applied to a Sepharose 12 10 x 300mm gel-filtration column, which had been calibrated with the following proteins: 1, ribonuclease A; 2, chymotrypsinogen; 3, ovalbumin; 4, BSA; 5, alcohol dehydrogenase; 6, β-amylase. The broken line indicates the ratio V_e (elution volume) to V_o (void volume) of recombinant, active *L. major* PPI and the arrow head indicates the inferred molecular mass (36 kDa).

4.2.3.5. Activity towards pGlu-, H-Glu- and H-Gln- substrates

Purified enzyme was active towards the substrates pGlu-pNA, pGlu-βNA, H-Glu-βNA, with a strong preference for the pGlu substrates (20-fold greater activity) (Fig. 4.13). The enzyme did not present any activity towards H-Gly-βNA. Other PPIs have been reported not to cleave Glu at the N-terminus (Dando *et al.*, 2003). The activity towards H-Glu-βNA being low, it cannot be excluded that this activity might be due to cleavage of pGlu-βNA present at a low concentration in the H-Glu-βNA substrate, and resulting from spontaneous cyclisation of the H-Glu-βNA substrate. The K_ms and specific activities were, respectively, 4.5 ± 0.5 μM and 1.02 ± 200 mU/mg protein towards pGlu-βNA; 33.5 ± 2.5 μM and 904 ± 310 mU/mg against pGlu-pNA; 85 ± 18 μM and 50 ± 12 mU/mg protein towards H-Glu-βNA. This falls within the range of expected K_m values reported for mammalian PPIs towards pGlu-pNA (20 to 80 μM) (Mantle *et al.*, 1995; Cummins and O'Connor, 1998; Dando *et al.*, 2003), and not the range for the bacterial enzymes (180 to 250 μM) (Le Saux *et al.*, 1996; Ito *et al.*, 2001).

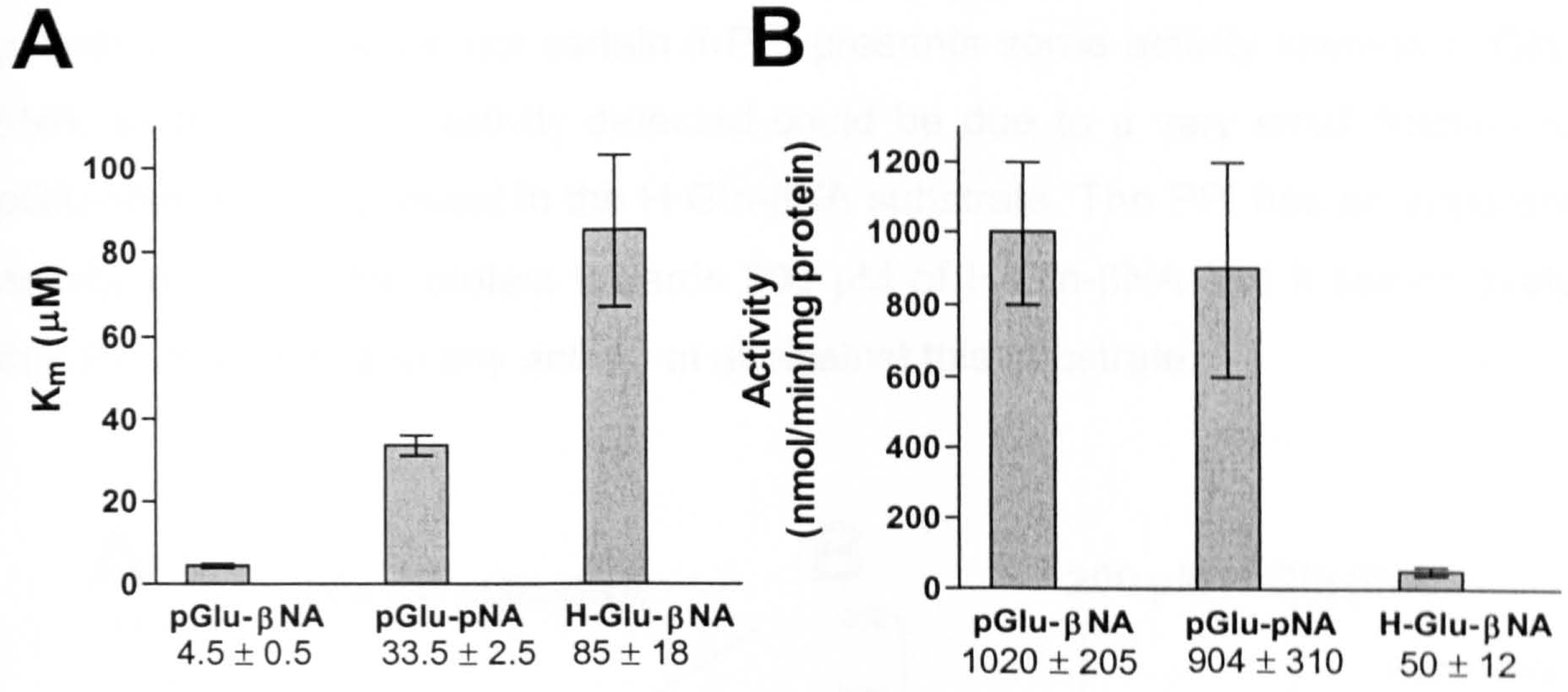


FIG. 4.13. **Analysis of the activity of the PPI of *L. major* produced in *E. coli* against the synthetic substrates pGlu-βNA, pGlu-pNA, and H-Glu-βNA.** (A) K_ms (μM) of the 3 substrates. (B) Activities (U/mg protein) of the enzyme against the 3 substrates. Enzyme samples were tested for activity in a final volume of 1 ml. The enzyme sample was incubated for 5 min at 32°C in 50 mM HEPES pH 8.0 buffer, supplemented with 2 mM TCEP and 1 mM EDTA, before start of the reaction by addition of the substrate (200 μM, unless for K_m determination 0.1 μM to 1 mM were used). Excitation and emission wavelengths were respectively 320 and 410 nm for the βNA fluorogenic substrates. Reading wavelength was 410 nm for the pNA chromogenic substrate. The assays were continuous for 10 min at 32°C. Results are the means ± SD from 3 independent measurements.

Activity towards H-Gln-βNA was also investigated, but the results were more difficult to analyse. There was an initial low rate for about 5 min, which quickly accelerated to reach the rate that occurred towards the pGlu-βNA substrate (Fig. 4.14). The initial rate only was observed when there was no pre-incubation of the substrate in the assay buffer at 32°C. If pre-incubation was used, the rate detected was immediately the same as maximum rate towards pGlu-βNA. It was concluded that a small proportion of the H-Gln-βNA substrate was quickly converting to pGlu-βNA. The K_m towards pGlu-βNA is about 4 μM, thus a spontaneous cyclisation of only about 5% of the 200 μM of H-Gln-βNA used in the assay would result in the maximum rate towards the pGlu-βNA substrate. It had previously been shown that up to 6% of H-Gln-βNA could convert spontaneously to pGlu-βNA when incubated at 37°C for 2 h (Chikuma *et al.*, 2004). Lower concentrations of H-Gln-βNA were also investigated but the initial rate then was at the limit of detection and precise measurements were not possible. It is therefore not certain if PPI presents some activity towards H-Gln-βNA, as the very low activity detected could be due to a very small fraction of pGlu-βNA already present in the H-Gln-βNA substrate. The PPI has an apparent activity of 10 mU/mg protein towards 200 μM of H-Gln-βNA and it seems likely that PPI may not have any activity at all against this substrate.

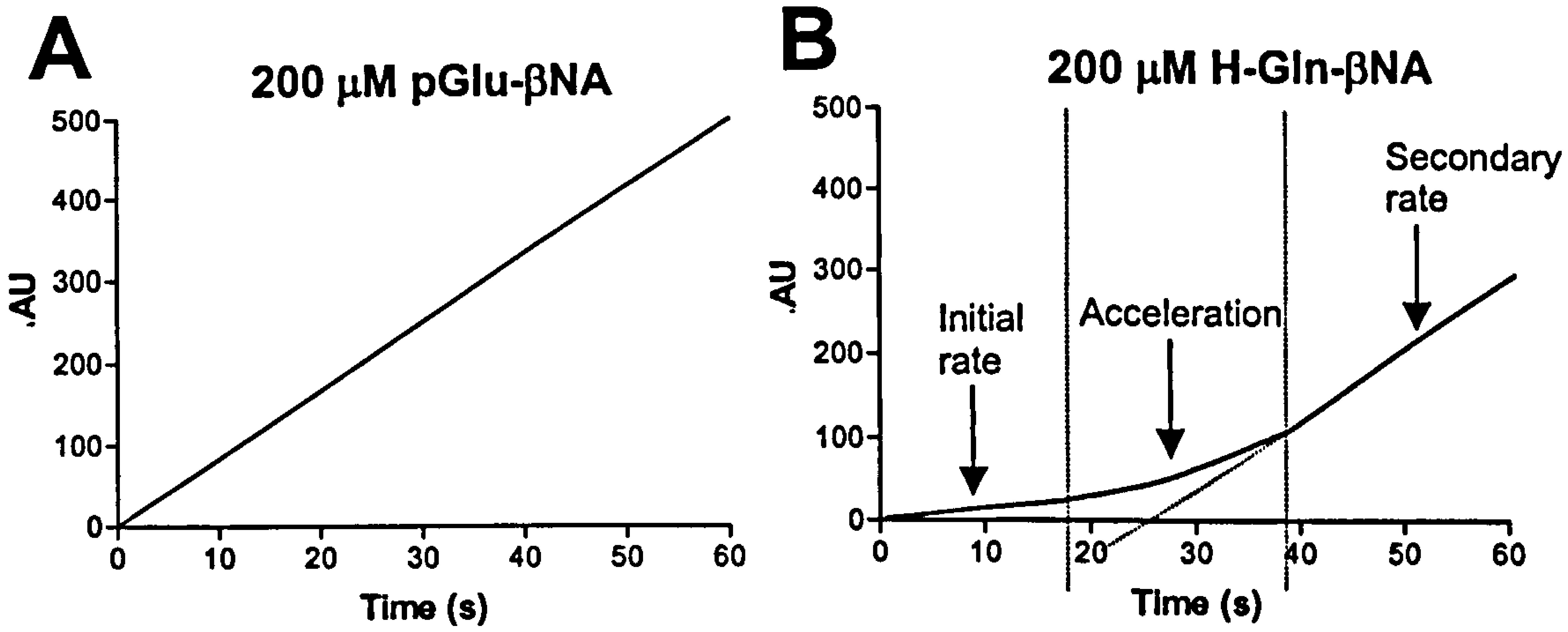


FIG. 4.14. **Typical traces obtained with 0.2 mU of recombinant PPI of *L. major* produced in *E. coli*, and the substrates 200 μM pGlu-βNA (A), and 200 μM H-Gln-βNA (B).** Assays were in 50 mM HEPES pH 8.0, 2 mM TCEP, 1 mM EDTA, at 32°C. In both A and B, the buffer was pre-incubated for 5 min at 32°C, and the reaction was started by simultaneous addition of PPI and substrate. Measurements were continuous (excitation 320 nm, emission 410 nm). AU, arbitrary fluorescence units; s, seconds.

4.2.3.6. Inhibitors

The human PPI had been reported to be inhibited by transition metal ions (Dando *et al.*, 2003). Various ions were tested on the recombinant PPI of *L. major* (Table 4.4). Nickel, copper and zinc ions showed strong inhibition of the enzyme; nickel almost completely inhibited the enzyme when at 1 mM. As expected, calcium, magnesium, sodium and lithium ions did not have any effect.

Table 4.4. **The effects of various ions on the activity of the purified recombinant PPI of *L. major*.** Assays were in 50 mM HEPES pH 8.0, 2 mM TCEP. The ions were added to the buffer-enzyme mix and incubated for 10 min at 32°C. The reaction was then started by addition of 200 μM pGlu-pNA and monitored continuously for 5 min at 410 nm. The activity remaining is indicated as a percentage of the activity of the non-treated enzyme in the same conditions. Values correspond to mean ± SD from 3 measurements.

Ion	Concentration (mM)	Activity (%)
Ni^{2+}	1	1.0 ± 0.3
Ni^{2+}	0.1	52 ± 1.0
Mg^{2+}	1	94 ± 2.1
Na^{+}	1	100 ± 2.5
Ca^{2+}	1	88 ± 1.6
Cu^{2+}	1	8.7 ± 0.3
Li^{+}	1	92 ± 0.7
Zn^{2+}	1	19.8 ± 0.4

Other inhibitors were tested, but only sulfhydryl blocking agents such as iodoacetamide, iodoacetic acid and N-ethyl maleimide (NEM) showed strong inhibition of the enzyme (Table 4.5). Though the PPI is a cysteine peptidase, the common cysteine peptidase inhibitor (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl}carbamoyl)oxirane-2-carboxylic acid (E64) did not inhibit the enzyme's activity. This reflects the very different catalytic mechanism of PPI compared with clan CA, family C1 peptidases, such as papain. This is consistent to what was found for the human enzyme (Dando *et al.*, 2003). Surprisingly, the serine peptidase inhibitor phenylmethylsulphonyl fluoride (PMSF) showed some inhibition, but only at the mM range, and the pGlu analogue 2-pyrrolidinone, reported to be an inhibitor for the bacterial enzymes (Cummins and O'Connor, 1998), inhibited only when at 1 mM.

Table 4.5. **The effects of various inhibitors on the activity of the purified recombinant PPI of *L. major*.** Assays were in 50 mM HEPES pH 8.0, 2 mM TCEP. Inhibitors were added to the buffer-enzyme mix and incubated for 10 min at 32°C. The reaction was then started by addition of 200 µM pGlu-pNA and monitored continuously for 5 min at 410 nm. The activity remaining is indicated as a percentage of the activity of the non-treated enzyme in the same conditions, with the standard deviation based on 3 repeats of each measurement. PMSF: phenylmethylsulphonyl fluoride; NEM: N-ethyl maleimide.

Inhibitor	Concentration (µM)	Activity (%)
E64	100	79 ± 1.2
	10	91.8 ± 2.1
Pepstatin A	10	93 ± 1.6
	1	94.5 ± 1.4
PMSF	1000	46.6 ± 0.8
	100	101 ± 1.0
Iodoacetamide	100	1.7 ± 0.3
	10	7.4 ± 0.8
Iodoacetic acid	100	0
	10	1.2 ± 0.4
1,10-phenanthroline	1000	98.2 ± 0.7
	100	98.8 ± 1.9
Benzamidine	1000	96 ± 2.2
	100	100 ± 1.2
Bestatin	10	101 ± 1.7
NEM	100	1.9 ± 0.4
	10	3.1 ± 0.3
2-pyrrolidinone	1000	66 ± 2.5
	100	86.7 ± 3.0
	10	99 ± 2.8

4.2.4. Biochemical analysis of active site mutants of the recombinant PPI of *L. major*

As expected for a pyroglutamyl peptidase I (Dando *et al.*, 2003), the active site cysteine and histidine predicted for the PPI of *L. major*, at respectively positions 210 and 234 (PPI of *L. major* numbering), from the alignment with PPI from different organisms (Fig. 4.3 and 4.4) were found to be essential for the enzyme's activity. Mutants lacking the active site cysteine (C210A) or histidine (H234S) had practically no activity (Table 4.6). The E101Q mutated enzyme had about 2% of activity compared with the unmutated recombinant enzyme, with a K_m for pGlu-pNA about 30 times higher than that of the unmutated enzyme. In contrast, the E107Q mutated enzyme still had 50% of the activity of the native enzyme and the K_m was only 2.5 fold higher. These results confirmed that glutamic acid 101 is part of the active site, whereas glutamic acid 107 probably plays a minor role in the catalytic reaction.

Table. 4.6. **Biochemical analysis of the recombinant PPI active site mutants of *L. major* expressed in *E. coli* and purified on Nickel-agarose column.** The amino acids are represented using the single letter code. Numbering refers to the native PPI of *L. major*. Activity is expressed as a percentage of the activity of the recombinant non-mutated PPI of *L. major* in the same conditions. Assays were in 50 mM HEPES pH 8.0, 1 mM EDTA, 2 mM TCEP, 200 μM pGlu-pNA, at 32°C and measured continuously for 10 min at 410 nm. The K_m of the different enzymes was measured for the substrate pGlu-βNA. N/A: non applicable.

Mutant	Activity remaining	K_m (pGlu-pNA)
C210A	<0.5%	N/A
H234S	<0.3%	N/A
E101Q	2%	936 ± 5 μM
E107Q	50%	88 ± 26 μM

4.2.5. Cloning of the glutaminyl cyclase gene of *L. major*

Database searches on the genomic database of *L. major* using the human glutaminyl cyclase (QC) as a query (GenBank accession number NP 036545) detected a 2724 bp open reading frame translating into a 907 amino acid protein (GeneDB; LmjF05.0950) with QC features, and which was annotated as a putative glutaminyl cyclase. No other sequence with QC features could be detected in the *L. major* database (now totally sequenced), which suggested that the putative QC is a single copy gene in *L. major*, localised on chromosome 5. The amino acid sequence was analysed with the SignalP algorithm (Nielsen *et al.*, 1997), and TMpred (Hofmann and Stoffel, 1992), and a trans-membrane domain was predicted near the N-terminus (position 120-144 on the predicted protein).

The results of the 5' RACE showed a mRNA comprising a 108 bp 5' untranslated region (UTR), confirming the start methionine attributed (Fig. 4.15).

```
  1 TCACGCTAGC GAGGGCGGGA AATACTCATT CACGCACGCG CAGACATACG
    AGTGCGATCG CTCCCGCCCT TTATGAGTAA GTGCGTGCGC GTCTGTATGC

 51 GAGAGGGTGT GCGCGCTTCT CAACGACGAA GGTGACACGC CGTAGAAGCA
    CTCTCCCACA CGCGCGAAGA GTTGCTGCTT CCACTGTGCG GCATCTTCGT

101 CTGCCTTTTC TTTTCTGATG CAACATCAAT
    GACGGAAAAG AAAAGACTAC GTTGTAGTTA
```

FIG. 4.15. ***L. major* QC 5' untranslated region (UTR) determined by 5' RACE.** 5' UTR is on grey background. The predicted start ATG is underlined.

The protein was aligned with the human QC (Fig. 4.16). A schematic representation of the alignment is shown in Fig. 4.17. The complete *L. major* protein and the *L. major* QC protein fragment from position 408 to 907 showed little overall sequence identity with to the human QC (11% and 16% identity, respectively). However, blocks of sequence identity were observed around the active site residues of the human QC. It has a very long N-terminal extension of 407 amino acids, which contains a trans-membrane domain, whereas the human enzyme possesses a signal peptidase for localisation in the secretory system. This suggests a different location for the *L. major* QC. From the 6 key catalytic residues in the human QC, 5 were conserved in *L. major* QC: His-140, Glu-201, Glu-202, Asp-248 and His-330 (human enzyme numbering). The Asp-159 was not conserved, but in the human QC the Asp residues are not thought to be involved in the catalytic reaction (Bateman *et al.*, 2001). The *L. major* QC has a long insertion from positions 733-757 (*L. major* QC numbering) of unknown function. By computational analysis, it is unclear if the *L. major* QC may be functional, but it possesses the necessary residues for activity.

```
               1                                                                              80
HsQC     (1)   --------------------------------------------------------------------------------
LmQC     (1)   MQHQSLAARRRGRRLARRGESAFWCSLLKRCLPCLPLRWKSVHATSRTRRILMGILLCLITTLIFLIGLLVCLTWKSNRS
               81                                                                            160
HsQC     (1)   --------------------------------------------------------------------------------
LmQC    (81)   NTPLEPAAVQSTHAPRVPPEVGVHDPWAEDPVDVAVIRQREEAVVRKWKENGFRPTYWEVERRTLPRLTKRALESWFEHG
               161                                                                           240
HsQC     (1)   --------------------------------------------------------------------------------
LmQC   (161)   RRAAASATAVDVAEDVLTAHAWDAVDGGDADADLFDSLAFLYPPAQQRQHYRAAMDAILRHGPRVGGTKRDALLHHLVDE
               241                                                                           320
HsQC     (1)   --------------------------------------------------------------------------------
LmQC   (241)   GLGRRRYDAARHAYHPAPPHLSPRGKELLEKAAASWPDAPHLAAAPQDDVAGREPGWRWSLSWDNFTADIPIRIDGSSRV
               321                                                                           400
HsQC     (1)   --------------------------------------------------------------------------------
LmQC   (321)   EMQNLVFQFPGGLQFRRKQAAAASAAATGAKDALDEFVGGVPNRRRTRADDRKDVLEAEYHPADMTGETVYTQENFLQWA
               401                                                                           480
HsQC     (1)   -------MAGGRHRRVVGTLHLLLLVAALPWASRGVSPSASAWP-EEKNYHQPAILNSSALRQIAEGTSISEMWQNDLQP
LmQC   (401)   QDGVPFNPRPALVRGVPGTRVVLSPDAKGARVPLKPSLSAPAAEPATQQPVKHAVLAAHWDSKYFADIPFLAACDSAMPV
               481                                                                           560
HsQC    (73)   LLIERYPGSPGSYAARQHIMQRIQRLQADWVLEIDTFLSQTPYGYRSFSNIISTLNPTAKRHLVLACHYDSKYFSHWNNR
LmQC   (481)   VFLLRTIKNIAVLTDVAEALTESYKAERSGVVGGDASAAVPGLTSTFRNATTEAEVRERLASLLSPAHHALLYQYFFARP
               561                                                                           640
HsQC   (153)   VFVGATDSAVPCAMMLELARALDKKLLSLKTVSDSKPDLSLQLIFFDGEEAFLHWSPQDSLYGSRHLAAKMASTPHPPGA
LmQC   (561)   YSIGGEQQNFVAAGIRNHPTKVEVDVRTWLDWVQHLP--IISVILFDGEEAYKHWSGDDNTYGSRHLAQRWRSTPSVMRS
               641                                                                           720
HsQC   (233)   RGTS----QLHGMDLLVLLDLIGAPNPTFPNFFPNSARWFER-LQAIEHELHELGLLKDHSLEGRYFQNYSYG-----GV
LmQC   (639)   RHTDGPQSLYDSVDLFALYDLMGPAGTTFRNMYPTQSGIFYAGLSQREWELRHRAMQHSSAISAELLWRYDEATALPPAE
               721                                                                           800
HsQC   (303)   IQDDHIPFLRRGVP------------------------------------------------------------------
LmQC   (719)   AREKADPFLRLLGPSPHEAVASLNAAYTRRLLANNASRPSVASLPRSWLMYGSPHEMLTLHSVARTTHDVRFVENFDQLR
               801                                                                           880
HsQC   (317)   -----------------------------------------------------------VLHLIPSPFPEVWHTMDDNEE
LmQC   (799)   VYDRLDREIAAESSFDAERYLNTINHNIFFSPSQQEDLRRNRDVFVADDDHKHWLDTQRVLHLIPIPFPKSWHTANDNGS
               881                        907
HsQC   (338)   NLDESTIDNLNKILQVFVLEYLHL-----
LmQC   (879)   NVHDGTSTDLAGLLWSTVLELGDYWTRKE
```

FIG. 4.16. **Alignment of the glutaminyl cyclase (QC) of *L. major* with the human QC.** Dashes indicate gaps in the alignment. Active site residues are starred. Conserved residues are on black background. The predicted signal peptide is underlined. Lm, *Leishmania major* (GeneDB CDS: LmjF05.0950); Hs, *Homo sapiens* (Song *et al.*, 1994).

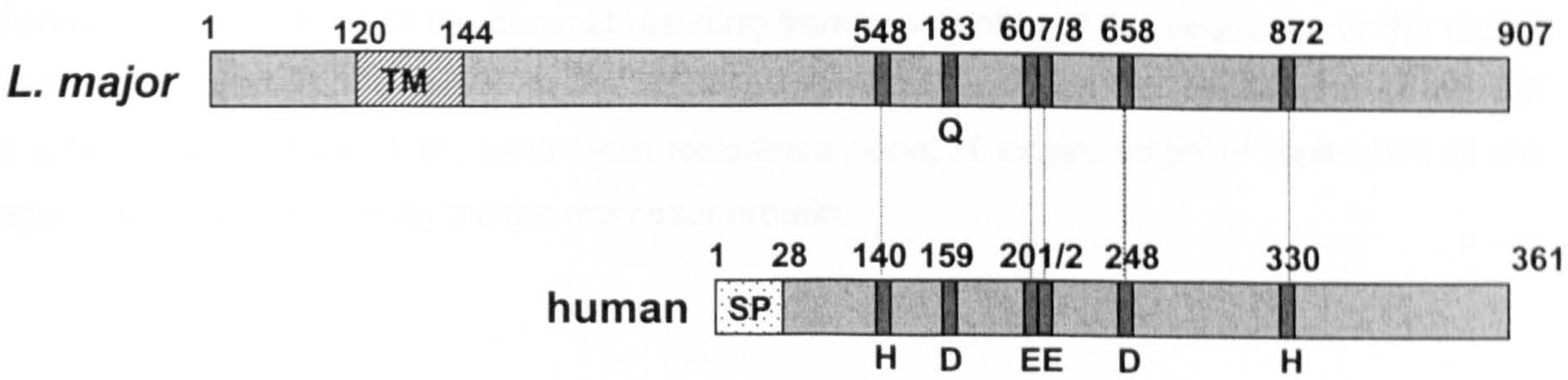


FIG. 4.17. **Schematic comparison of the QC of *L. major* and the human QC.** TM, trans-membrane domain; SP, signal peptide. Amino acids are represented using the single letter code. Numbers indicate the positions for the various domains/amino acids.

4.2.6. Recombinant expression of the glutaminyl cyclase gene of *L. major* in *E. coli*

The sequence encoding the *L. major* QC without the trans-membrane domain from amino acid position 145 to 907 (*L. major* QC numbering), and the sequence encoding for a fragment of the *L. major* QC from amino acid position 408 to 907 (*L. major* QC numbering), aligning with human QC from position 29 (the start of the human QC without the signal peptide) were amplified and cloned into the pET28a(+) vector for recombinant expression in *E.coli*. The corresponding plasmids are represented in Fig. 4.18 and Fig. 4.19.

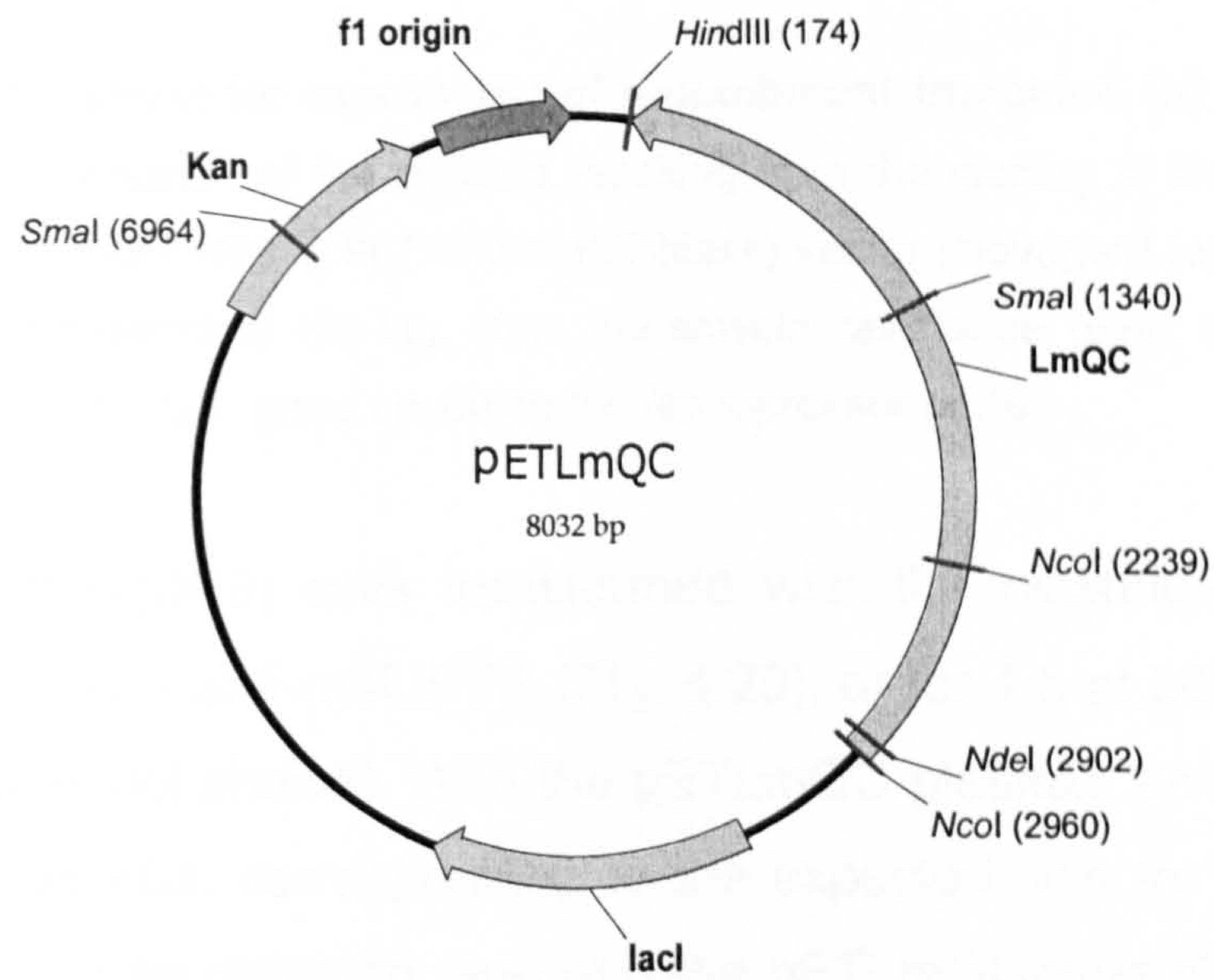


FIG. 4.18. **Construct for expression of recombinant QC of *L. major* in *E. coli*.** Schematic representation of the plasmid resulting from the cloning of the sequence of the QC of *L. major* from position 145 to 907 in the pET28a(+) vector (Novagen) for expression of the QC with a N-terminal His-tag. Kan, kanamycin resistance gene; f1 origin, origin of replication of the phage f1; lacI, gene encoding the *lac* repressor protein.

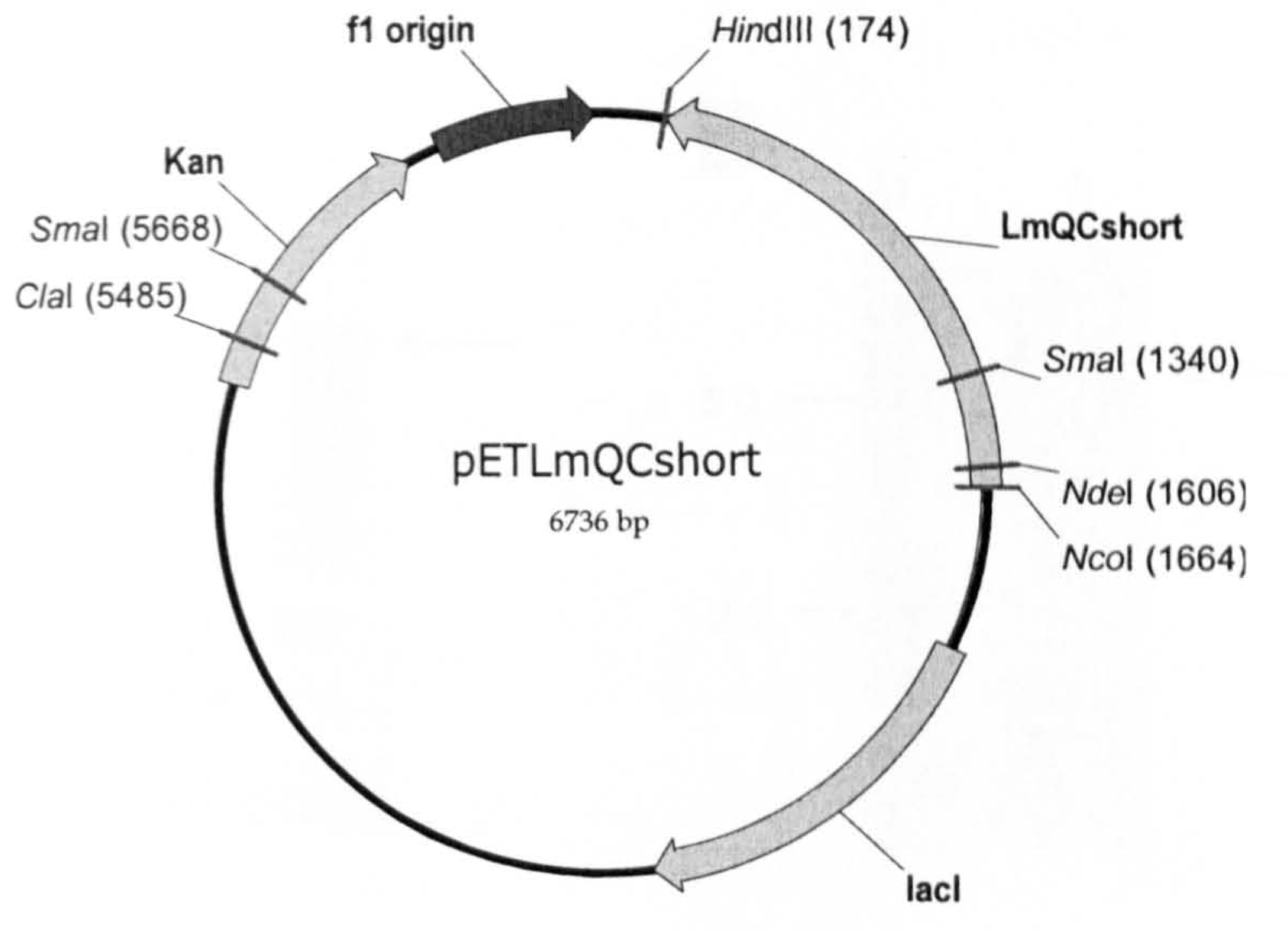


FIG. 4.19. **Construct for expression of recombinant truncated QC of *L. major* in *E. coli*.** Schematic representation of the plasmid resulting from the cloning of the sequence of the QC of *L. major* from position 408 to 907 in the pET28a(+) vector (Novagen) for expression of the truncated QC with a N-terminal His-tag. Kan, kanamycin resistance gene; f1 origin, origin of replication of the phage f1; lacI, gene encoding the *lac* repressor protein.

E. coli BL21(DE3) cells transformed with the plasmids were induced either for 4 h at 37°C with 1 mM IPTG (Fig. 4.20), or for 4 h at 20°C or 15°C with 1 mM IPTG (results not shown). With the pETLmQC plasmid, an over-expressed protein of about 98 kDa, corresponding to the expected size for N-terminal His-tagged protein could be detected, and with the pETLmQCshort plasmid, an over-expressed protein of 56 kDa, corresponding to the expected size for N-terminal His-tagged protein (55.7 kDa) could be detected, but it was found in the insoluble fraction of *E. coli* lysates in all cases (results shown for induction at 37°C only, Fig. 4.20). The expression in *E. coli* did not result in production of soluble protein for enzyme characterisation.

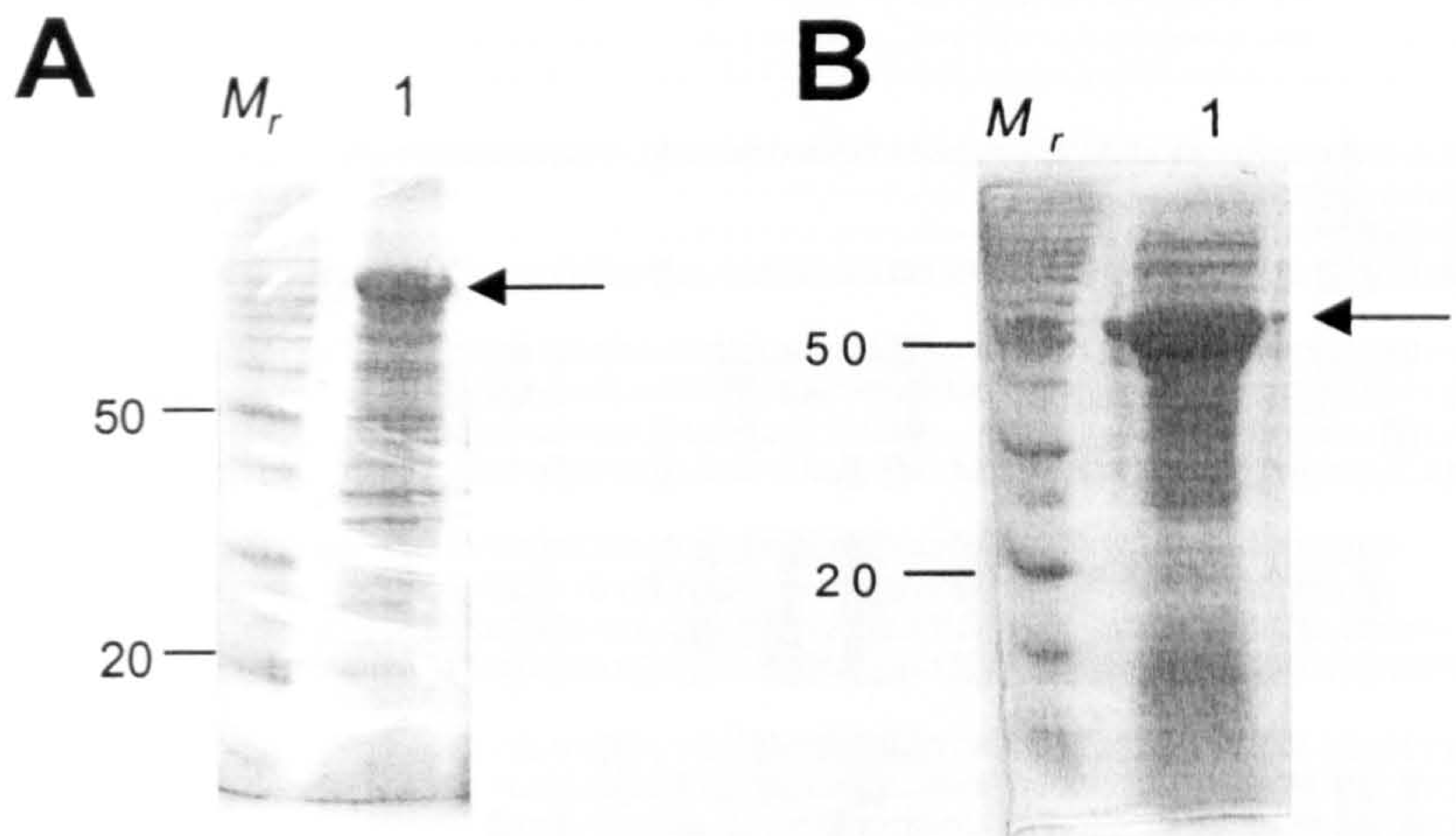
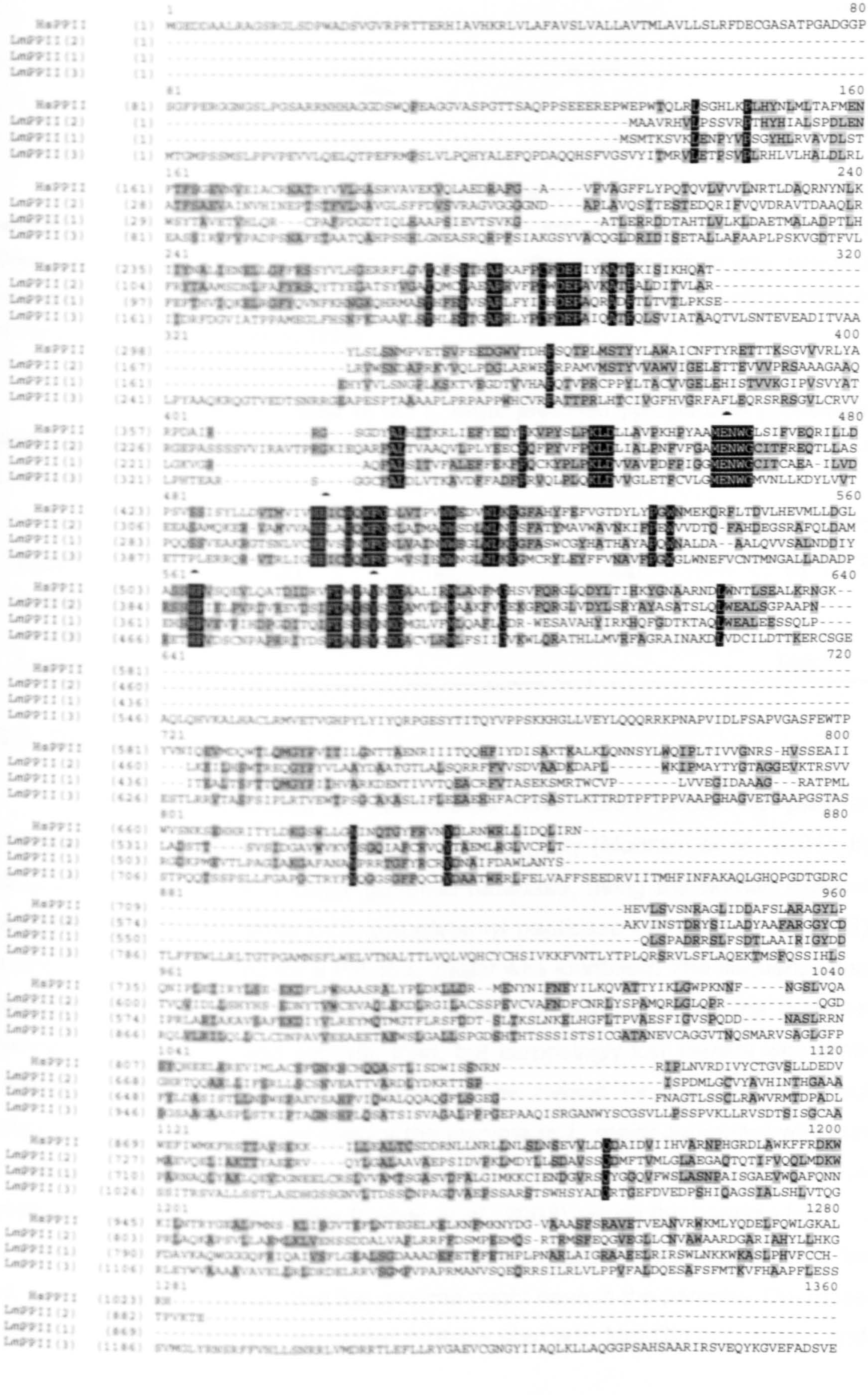


FIG. 4.20. **Analysis of expression of N-terminal His-tagged QC of *L. major* in *E. coli* BL21(DE3).** Ten µl samples were separated on a SDS-PAGE gel (12% acrylamide) and then stained with Coomassie blue. Lane M_r, protein standards. 10^{-3} x M_r is indicated on the left of the figure. (A) Analysis of expression of the full length QC of *L. major*. Lane 1, *E. coli* total cell lysate after 4 h induction at 37°C with 1 mM IPTG. (B) Analysis of expression of the truncated QC of *L. major*. Lane 1, *E. coli* total cell lysate after 4 h induction at 37°C with 1 mM IPTG. Arrows indicate the over-expressed QC.

4.2.7. Analysis of potential genes for pyroglutamyl peptidases II of *L. major*

Database searches on the genomic sequence database of *L. major* using the human PPII as a query (accession number NP 037513) detected 3 potential genes for metallopeptidase (clan MA(E), family M1) PPIIs in *L. major*, LmjF29.2240, LmjF26.0300, LmjF12.1250, all annotated as aminopeptidase-like proteins, and respectively located on chromosomes 29, 26 and 12.

The proteins were aligned with the human PPII (Fig. 4.21). The similarity table is presented in Table 4.7. Two of the *L. major* proteins showed good similarity (42-44%) to the human enzyme, whereas the last one was not very similar (11% similarity). However, the predicted active site residues (Schomburg *et al.*, 1999) were all conserved between the different proteins. The 3 *L. major* putative enzymes had the typical metal-binding motif HEXXH (Schomburg *et al.*, 1999) at position 440-444 (human numbering). Studies for the presence of PPII-like activities in *L. major* are described in chapter 5.



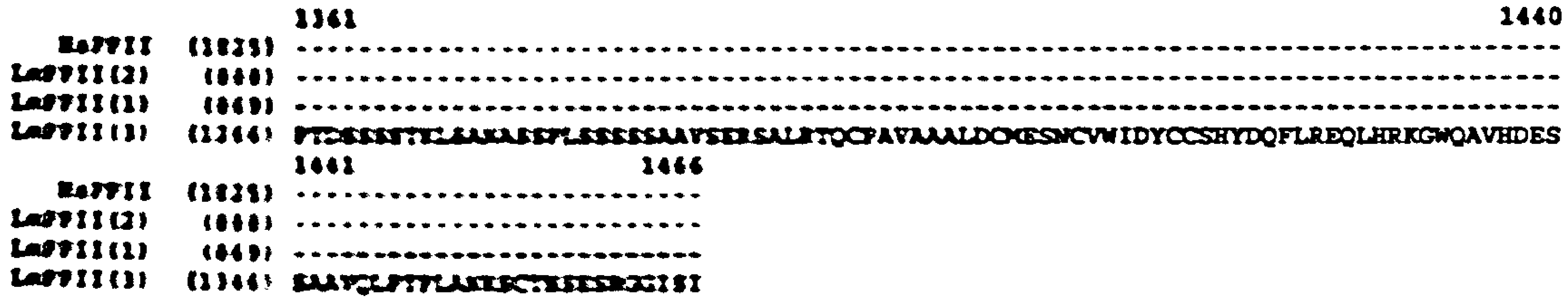


FIG. 4.21. Alignment of the potential pyroglutamyl peptidases II (PPII) of *L. major* with the human PPII. Dashes indicate gaps in the alignment. Active site residues are starred. Conserved residues are on black background. Conservative residues are on grey background Lm, *Leishmania major* (GeneDB CDS: (1) LmjF29.2240, (2) LmjF26.0300, (3) LmjF12.1250); Hs, *Homo sapiens* (Schomburg *et al*., 1999).

Table. 4.7. Similarity table of the *L. major* predicted pyroglutamyl peptidases II (PPII) with the human PPII. Numbers represent percentage of identity. Hs, *Homo sapiens*; Lm, *L. major*.

	HsPPII	LmPPII (1)	LmPPII (2)	LmPPII (3)
HsPPII	100	42	44	11
LmPPII (1)		100	51	17
LmPPII (2)			100	15
LmPPII (3)				100

4.3. Discussion

Cysteine peptidases in *Leishmania* have been shown to play an important role in parasite's survival and pathogenicity (Mottram *et al*., 2004). By database mining of the *L. major* genomic database, using the human cysteine peptidase of clan CF, family C13, pyroglutamyl peptidase I (Dando *et al*., 2003), a single sequence coding for a PPI in *L. major* was identified. The sequence was amplified from genomic DNA and the predicted protein contained the cysteine, histidine and glutamic acid catalytic triad, but surprisingly low identity with other PPIs from different organisms (Table 4.3). The reason for this is unclear, but might suggest a very different role for the enzyme in *L. major* compared with that in humans. In humans, the enzyme may be involved in the cleavage of bioactive peptides, such as neuropeptides (Dando *et al*., 2003), which the parasite lacks. It is unknown what may be the *L. major* PPI natural substrates, but this would be very informative on the role that the enzyme might have in the parasite. Database searches revealed that 352 annotated proteins start with Met-Gln in *L. major*, and can potentially be pyroglutamyl-modified after cleavage by a methionine aminopeptidase. This could be investigated by a proteomic approach, in which *L. major* cell lysates would be treated with a broad substrate specificity aminopeptidase, in order to degrade unblocked peptides, and compare with cell lysates first treated with PPI before treatment with the same aminopeptidase. The difference between the 2 samples would theoretically correspond to pGlu-modified peptides, which could be identified by mass-spectrometry.

The recombinant *L. major* PPI was produced in *E. coli*, and about 4 mg of soluble enzyme could be purified from a 1 l culture. The characteristics of the enzyme conform to what is expected for a mammalian PPI; it is strictly dependent on reducing agent (Fig. 4.7), it has an alkaline pH optimum (Fig. 4.10), and it is active as a monomer (Fig. 4.12). The very high pH optimum might be real or artefactual of recombinant expression, as it is unclear what biological significance this may have. The PPI did not have a signal peptide or a trans-membrane domain predicted, making it suggestive the enzyme would be cytosolic, as expected for the other PPIs. The localisation of the PPI in *L. major* is investigated in chapter 5. The free-living promastigotes propagate in the alimentary tract of the sandfly, primarily in the midgut where the pH is rather alkaline (7.0-9.0), and the non-motile amastigotes reside in phagolysosomes of mammalian macrophages, whose environment is acidic (4.5-5.5), but the intracellular pH is maintained close to neutrality (6.4-6.7) at all times (Zilberstein *et al.*, 1989). It is unclear why a cytosolic enzyme might have such a high pH optimum. In terms of substrate specificity, it cleaves preferentially pGlu synthetic substrates, but can also cleave glutamic acid as a N-terminal residue (Fig. 4.13), which is unreported for mammalian enzymes. This suggests that leishmanial PPI may have a slightly different conformation, which allows Glu to bind to the active site pocket. However, it cannot be excluded that this activity might be due to the presence of a small amount of pGlu in the Glu substrate, as cyclisation of Glu to pGlu is also possible (Schilling *et al.*, 2004). Activity against N-terminal glutamine may be zero (Fig. 4.14), as spontaneous chemical conversion of the synthetic substrate to the pGlu derivative would invalidate the assay. The recombinant enzyme is currently being tested for substrate specificity by our collaborators (Juliano and co-workers, Sao Paolo, Brazil). The K_m against the pGlu substrate falls within the range of expected K_m values reported for mammalian PPIs against synthetic pGlu substrate (20 to 80 μM) (Mantle *et al.*, 1995; Cummins and O'Connor, 1998; Dando *et al.*, 2003), and not the range for the bacterial enzymes (180 to 250 μM) (Le Saux *et al.*, 1996; Ito *et al.*, 2001). The inhibition pattern was similar to the human PPI (Dando *et al.*, 2003); *L. major* PPI was inhibited by transition metal ions and sylfhydryl blocking agents, but not by the cysteine peptidase inhibitor E64 (Tables 4.4 and 4.5). E64 is a potent inhibitor of the cysteine peptidase of the clan CA; the PPI belongs to the clan CF and possesses very few common features with papain, except from the catalytic mechanism involving an active site cysteine and histidine. The clan CF enzymes are not produced as inactive zymogens that require activation, as it is the case for the clan CA enzymes, but directly as mature proteins. The *L. major* PPI possesses cysteine residues, excluding the active site, which could potentially form 2 disulfide bridges. But the position of these cysteine residues among the different organisms is not conserved, making the presence of the disulfide bridges less likely to be important for activity.

The biochemical characteristics of *L. major* PPI seemed to conform to the data reported for the mammalian PPIs, and this is consistent with the phylogenetic analysis of the enzyme. A phylogenetic tree generated using the alignment of the *L. major* PPI with various PPs from different organisms (see Appendix 1) showed that *L. major* PPI is more closely related to eukaryotic PPIs than bacterial PPs (Fig. 4.22).

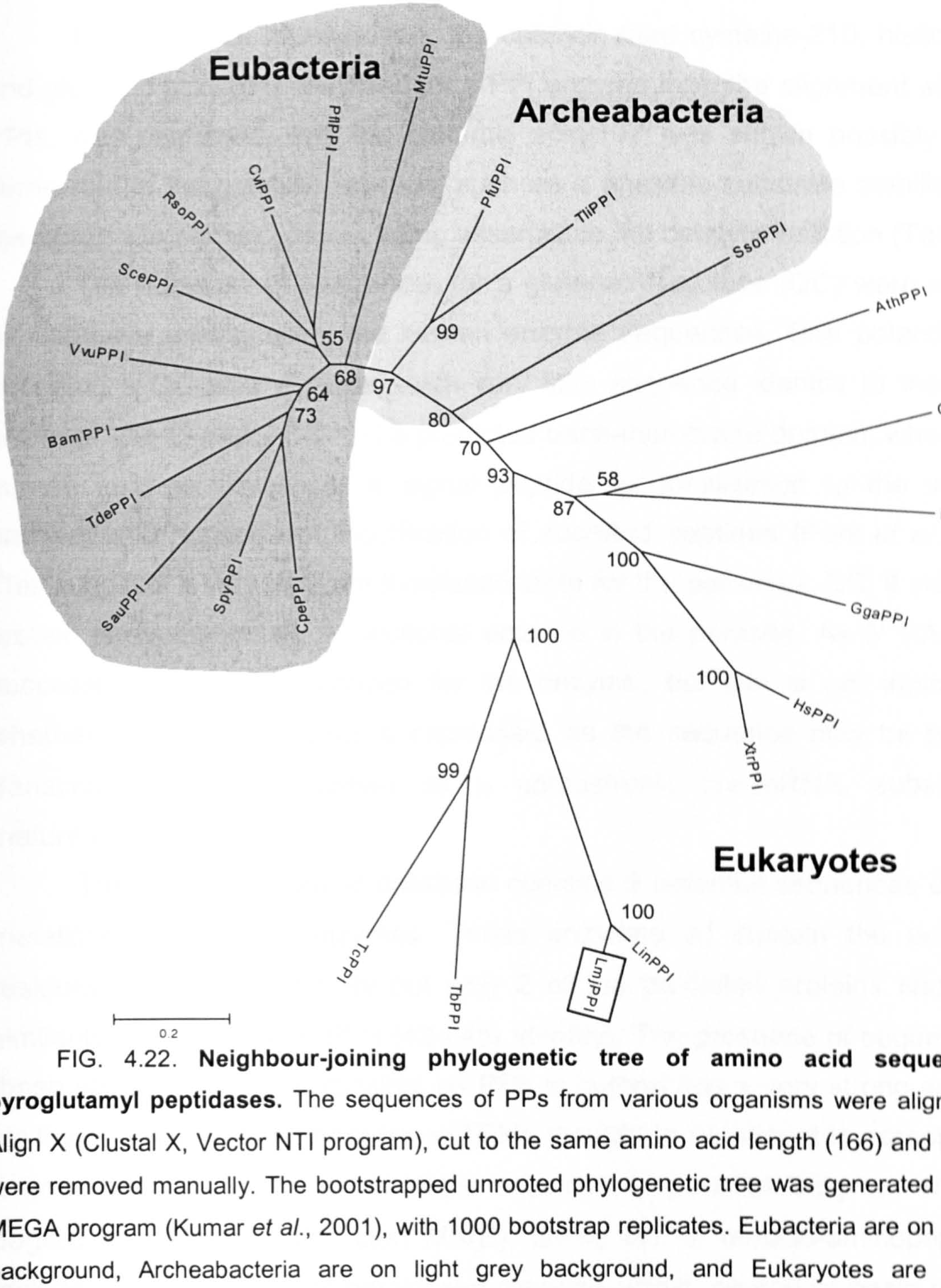


FIG. 4.22. **Neighbour-joining phylogenetic tree of amino acid sequences of pyroglutamyl peptidases.** The sequences of PPs from various organisms were aligned using Align X (Clustal X, Vector NTI program), cut to the same amino acid length (166) and insertions were removed manually. The bootstrapped unrooted phylogenetic tree was generated using the MEGA program (Kumar *et al.*, 2001), with 1000 bootstrap replicates. Eubacteria are on dark grey background, Archeabacteria are on light grey background, and Eukaryotes are on white background. *L. major* PPI is boxed. The scale bar indicates the number of estimated amino acid substitutions per site. Abbreviations: Cpe, *Clostridium perfringens*; Spy, *Streptococcus pyogenes*; Sau, *Staphylococcus aureus*; Tde, *Treponema denticola*; Bam, *Bacillus amyloliquefaciens*, Vvu, *Vibrio vulnificus*, Sce, *Streptomyces ceolicor*, Rso, *Ralstonia solanacearum*; Cvi, *Chromobacterium violaceum*; Pfl, *Pseudomonas fluorescens*; Mtu, *Mycobacterium tuberculosis*; Pfu, *Pyrococcus furiosis*; Tlit, *Thermococcus litoralis*; Sso, *Sulfolobus solfatanicus*; Ath, *Arabidopsis thaliana*; Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*, Gga, *Gallus gallus*; Hs, *Homo sapiens*; Xtr, *Xenopus tropicalis*; Lin, *Leishmania infantum*; Lmj, *Leishmania major*, Tb, *Trypanosoma brucei*; Tc, *Trypanosoma cruzi*.

By active site mutagenesis, the catalytic triad cysteine-210, histidine-234 and glutamic acid-101, predicted for a PPI enzyme from the alignment with other PPIs, was confirmed, and the glutamic acid-107 was shown possibly to play some role in the catalytic reaction, such as a possible substrate stabilisation in the active site pocket, without being essential to the catalytic reaction (Table 4.6).

The presence of sequences for a glutaminyl cyclase (QC) were analysed by database mining using the human enzyme sequences. One potential gene encoding a QC was detected, with very little sequence identity to the human enzyme. The *L. major* QC had a predicted trans-membrane domain, whereas the human enzyme possesses a signal peptide for localisation to the secretory pathway and subsequent modification of secreted peptides (Pohl *et al.*, 1991). This suggests a very different localisation/role for the parasite's QC, if indeed the protein is expressed as a functional enzyme in the parasite. As 5' RACE was successful, mRNA is produced for the enzyme, but this is not indicative of whether or not the enzyme is expressed, as the sequence may be part of a transcription unit, transcribed as a polycistronic pre-mRNA, subsequently matured.

The *L. major* genomic database contains 3 potential sequences encoding metallopeptidase PPII enzymes. These enzymes all contain the active site residues required for activity but only 2 of the predicted proteins show good similarity with the human PPII (42-44% identity). The presence of sequences for these proteins is quite surprising, as PPII in human has a very strong specificity for thyrotropin-releasing hormones (TRH), thought to be absent in parasites, and it would unexpected for a unicellular organism to possess enzymes with such degree of specificity. The clan MA(E), family M1 of metallo-aminopeptidases includes a whole range of peptidases cleaving acidic, neutral or basic residues, and it could be that the *L. major* enzymes with PPII similarities just happen to share the features classifying them in this clan, without being able to cleave a pGlu residue at the N-terminus. The presence of PPI, QC and PPII in *L. major* cell extracts is assessed in chapter 5, as well as localisation and functional analysis of the PPI.

Chapter 5

Functional study of the pyroglutamyl peptidase I of *L. major*

5.1. Introduction

It was shown in Chapter 4 that *L. major* possesses a single gene encoding pyroglutamyl peptidase I (PPI), 3 potential genes for pyroglutamyl peptidases II (PPII), and 1 potential gene for a glutaminyl cyclase (QC). Functional recombinant expression of the PPI of *L. major* in *E. coli* was successful. The presence of such an enzyme in *L. major* is very intriguing, as very few PPIs of mammals have been fully characterised so far and a definitive physiological role has yet to be attributed. Bacterial PPs have been widely used for unblocking of the N-terminus of proteins prior to N-terminal sequencing, but, as with the mammalian PPIs, the physiological function of bacterial PPs is still unknown. The leishmanial enzyme may be involved in protein metabolism, as reported for bacteria (Awade *et al.*, 1994; Ito *et al.*, 2001), although it is unclear if *L. major* is able to produce its own pGlu-modified peptides, as the gene encoding for a QC is only putative. Therefore, it would be of interest to determine the enzyme's natural substrate(s), and if it interacts with proteins in the insect vector or the mammalian host, and consequently contribute to the parasite's pathogenicity. If so, the parasite's PPI could be a new drug target.

In order to investigate the function of the PPI in the parasite, the first aim was to generate gene deletion lines, where the PPI gene would be replaced by selectable markers, and lines over-expressing PPI. These approaches would give information on whether the modification in PPI expression affected the survival of the parasite *in vitro*, metacyclogenesis and infectivity to mice. Generation of PPI knock-out mutants would also facilitate study of whether other PP activities are present in *L. major* (for instance PPII activities), and therefore, if the genes encoding putative PPIIs are expressed. In addition, testing the effects of pGlu-modified peptides such as gomesin on the different cell lines should give insight into the role of PPI in the host-parasite interaction, and whether the enzyme might confer protection against pGlu-modified peptides from the host or the insect vector.

It was also intended to carry out immuno-localisation studies to investigate the enzyme's location and also look for secretion.

5.2. Results

5.2.1. Over-expression of LmPPI in *L. major* promastigotes

5.2.1.1. Over-expression

The sequence encoding PPI of *L. major* was amplified from genomic DNA and cloned into the derivative of the pXG vector pGL102, containing a selectable marker (*Neo* gene from transposon Tn5 encoding an aminoglycoside 3'-phosphotransferase) for use with neomycin, for over-expression of a N-terminally His-tagged protein in the parasites. The sequence encoding PPI, in which the active site cysteine had been changed to an alanine, was cloned in the same way. Both were transfected into *L. major* promastigotes. As a control, the empty pGL102 vector was transfected into *L. major* promastigotes, to check that the over-expression of the neomycin resistance gene did not have a detrimental effect on the parasite's growth or survival. All vectors were transfected for expression as episomes (free plasmids), and the maps are presented in Fig. 5.1.

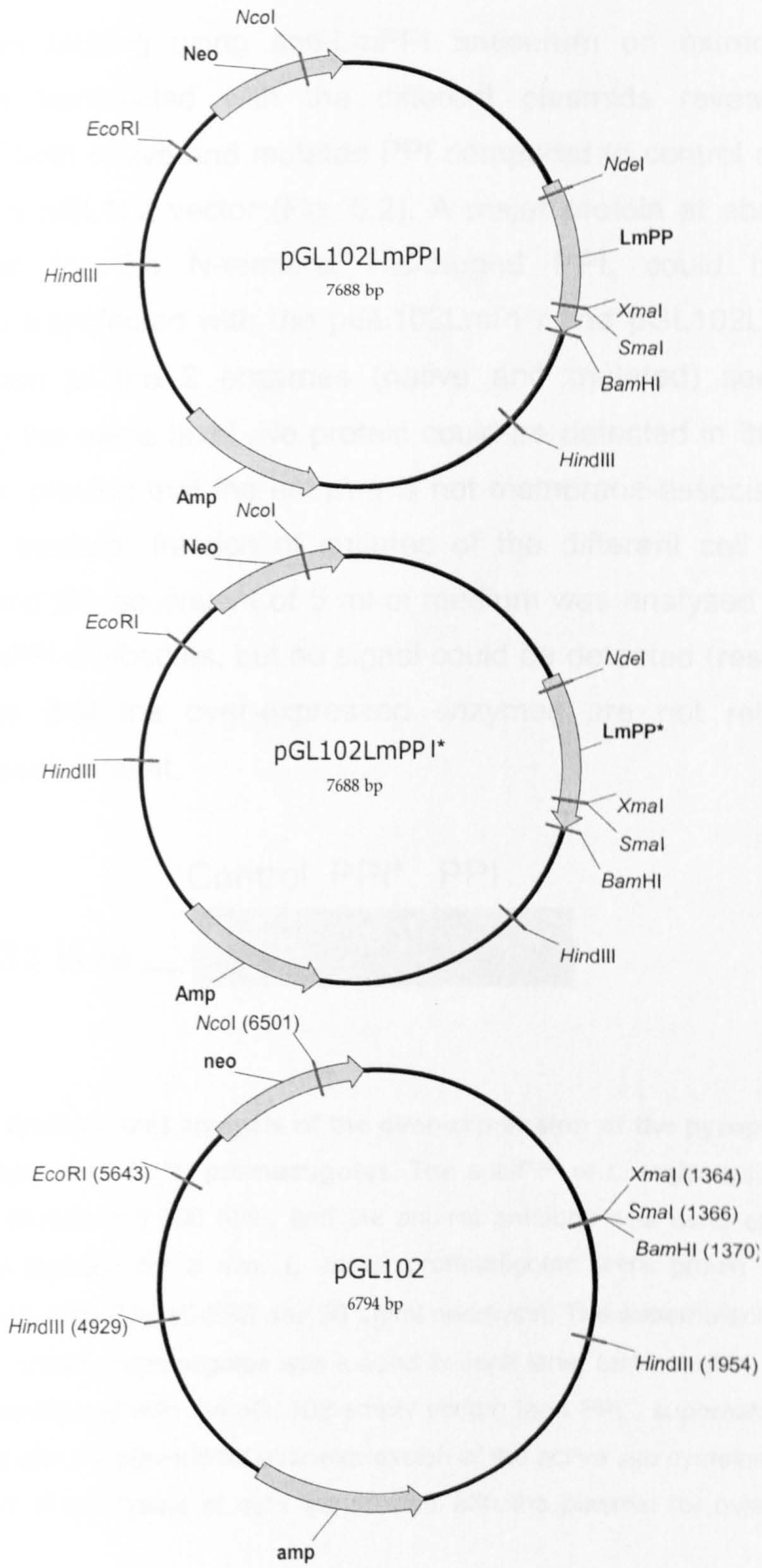


FIG. 5.1. **Constructs for over-expression of LmPPI in *L. major* promastigotes.** (A) Plasmid for over-expression of the native N-terminal His-tagged enzyme. (B) Plasmid for over-expression of the active site cysteine mutated N-terminal His-tagged enzyme. (C) Empty vector. Neo, neomycin resistance gene; amp, ampicillin resistance gene; Lm, *L. major*.

5.2.1.2. Analysis of over-expression

Western blotting using anti-LmPPI antiserum on extracts of *L. major* promastigotes transfected with the different plasmids revealed high over-expression of both native and mutated PPI compared to control cells transfected with the empty pGL102 vector (Fig. 5.2). A major protein at about 31 kDa, the size expected for the N-terminal His-tagged PPI, could be detected in promastigotes transfected with the pGL102LmPPI and pGL102LmPPI* plasmid. Over-expression of the 2 enzymes (native and mutated) seemed to be at approximately the same level. No protein could be detected in the pellet fraction of cell lysates, proving that the enzyme is not membrane-associated (results not shown). The acellular fraction of cultures of the different cell lines was TCA precipitated and the equivalent of 5 ml of medium was analysed by western blot using anti-LmPPI antibodies, but no signal could be detected (results not shown). This suggests that the over-expressed enzymes are not released into the extracellular environment.

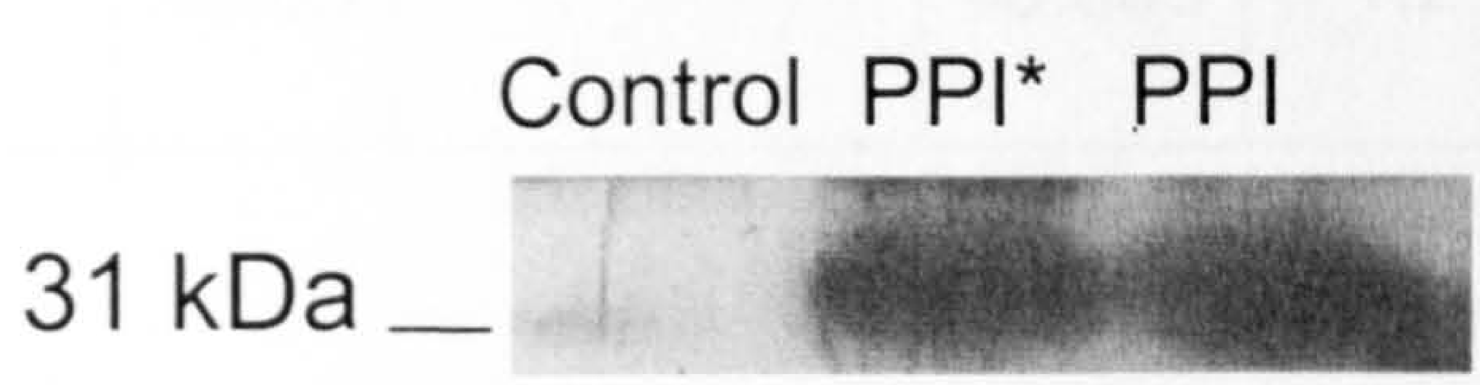


FIG. 5.2. **Western blot analysis of the over-expression of the pyroglutamyl peptidase I (PPI) of *L. major* in *L. major* promastigotes.** The anti-PPI of *L. major* antiserum raised in rat was used at a dilution of 1:500 (v/v), and the anti-rat antibody was used at 1:1500 (v/v). The membrane was exposed for 3 min. *L. major* promastigotes were grown in HOMEM media supplemented with 10% (v/v) HI-FBS and 50 μg/ml neomycin. The supernatant extract of $5x10^7$ *L. major* stationary phase promastigotes was loaded in each lane. Lane control, supernatant of cell lysate of cells transfected with the pGL102 empty vector; lane PPI*, supernatant of cell lysate of cells transfected with the plasmid for over-expression of the active site cysteine mutated PPI; lane PPI, supernatant of cell lysate of cells transfected with the plasmid for over-expression of the native PPI.

The PPI activities present in the different fractions of cell lystates were investigated using the substrate pGlu-βNA, and the activity of the over-expression lines was compared to the control cells transfected with the empty pGL102 vector. The PPI activity was exclusively associated with the soluble fraction of cell lysate, and the supernatant of cells over-expressing the dead mutant of PPI had very similar activity to control cells. However, the supernatant of cells over-expressing the active enzyme had about 7 times greater activity than the control (Table 5.1).

TABLE. 5.1. **Comparison of the activity towards pGlu-βNA of *L. major* promastigotes over-expressing native PPI, the active site cysteine mutant of the enzyme, or the empty vector.** Cells were grown in presence of 50 μg/ml of neomycin. The equivalent of 8×10^7 stationary phase cells were tested per assay. Assays were in 50 mM HEPES pH 8.0, 2 mM TCEP, 1 mM EDTA, 200 μM pGlu-βNA, at 32°C. Measurements were continuous for 10 min (excitation 320 nm, emission 410 nm). Activity is represented in comparison with the control.

Plasmid transfected	**pGL102 empty**		**pGL102LmPPI***		**pGL102LmPPI**	
Cell lysate fraction	*Pellet*	*Supernatant*	*Pellet*	*Supernatant*	*Pellet*	*Supernatant*
Comparative activity	<0.05	1	<0.005	1.2 ± 0.3	<0.05	7 ± 2

The dependence of the over-expression on neomycin concentration was tested by growing the cell lines in the presence of different concentrations of the selection drug. The supernatant of cell lysates was analysed by western-blot using anti-LmPPI antibodies, but no difference could be detected in the expression after 1 growth cycle (1 passage) in the different neomycin concentrations (Fig. 5.3), suggesting either that the accumulation of episomes is a slow process and several cycles of growths would be necessary to observe a difference, or that the cells reached the maximum level of expression tolerable even at the lowest neomycin concentrations.

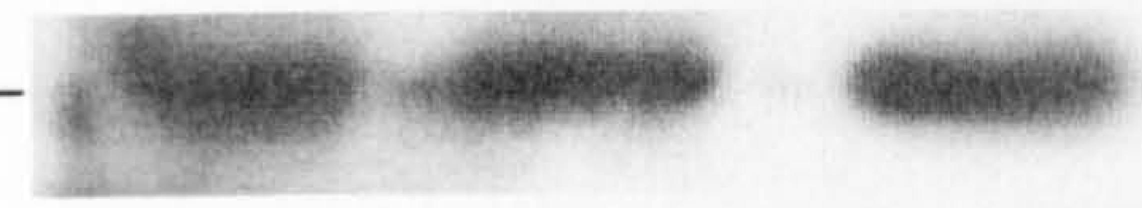


FIG. 5.3. **Dependence of the over-expression of the native LmPPI in *L. major* promastigotes cultures on the amount of selection drug.** The anti-PPI of *L. major* antiserum raised in rat was used at a dilution of 1:500 (v/v), and the anti-rat antibody was used at 1:1500 (v/v). The membrane was exposed for 3 min. The supernatant of $5x10^7$ *L. major* cell lysate was loaded per lane. Lane 1, supernatant of cell lysate of cells grown with 50 μg/ml neomycin; lane 2, supernatant of cell lysate of cells grown with 100 μg/ml neomycin; lane 3, supernatant of cell lysate of cells grown with 150 μg/ml neomycin.

To test whether the enzyme was released into the extracellular milieu, supernatants of cultures of the different cell lines were TCA precipitated and the equivalent of 1 ml was loaded on a SDS-PAGE gel and used for western blot analysis, using anti-PPI antibodies. No signal could be detected in any of the supernatants of cultures. Furthermore, 100 μl of crude supernatant of cultures of each line was assayed for activity towards 200 μM pGlu-βNA. No activity could be detected in any case (less than 10 μU/ml were detected, which corresponds to the maximum level of activity detected with buffer and substrate alone). This was indicative that either the PPI was released into the supernatant of culture at a level that was not detectable, or was not released at all.

5.2.1.3. Phenotype analysis of PPI over-expressing cell lines

5.2.1.3.1. Morphology and growth

The over-expression of the PPI in *L. major* promastigotes did not result in any difference detectable at the morphological level. In terms of growth, all cell lines were able to multiply, though the cells transfected with the empty pGL102 vector grew significantly faster from day 3 to day 6 and the final cell density higher (about 1.5 fold higher) (Fig. 5.4). The cells over-expressing the dead mutant enzyme grew at similar rates as the cells over-expressing the active enzyme but reached a slightly higher density (1.6×10^7/ml compared to 1.3×10^7/ml). Thus, the decrease of the growth rate might be related to the over-expression, whereas the final density may be influenced by the active enzyme over-expressed.

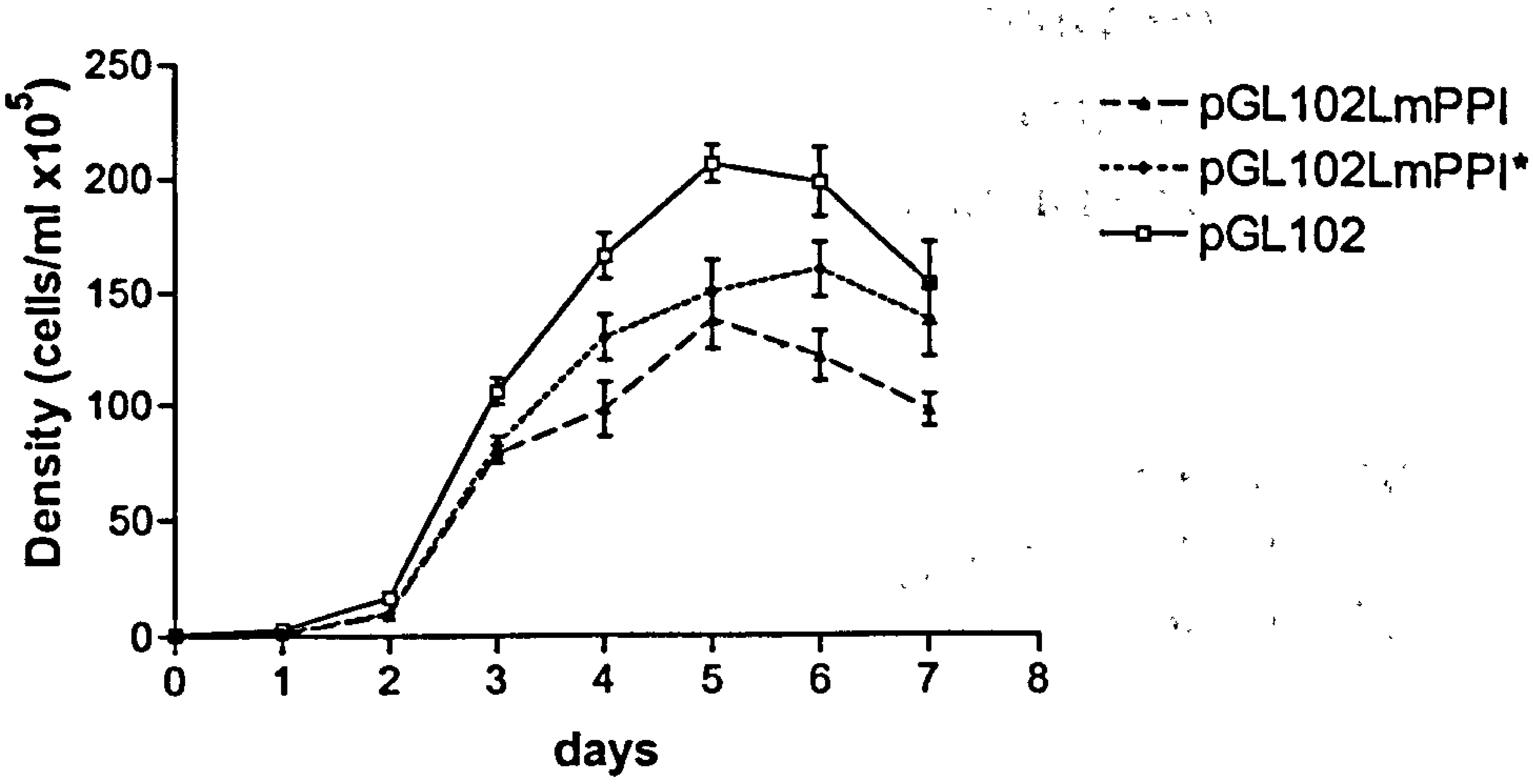


FIG. 5.4. **Growth curve of *L. major* promastigote cell lines over-expressing the native LmPPI, the active site cysteine mutant of the enzyme and the empty vector.** Cells were grown with 50 µg/ml of neomycin and counted at regular intervals. pGL102LmPPI, cells expressing the native PPI; pGL102LmPPI*, cells over-expressing the active site cysteine mutant; pGL102, cells transfected with the empty vector. Results are the means ± SD from 3 independent experiments. The cells transfected with the empty pGL102 vector grew significantly faster from day 3 to day 6 compared to the 2 other cell lines ($P<0.05$) and the cells transfected with the pGL102LmPPI* plasmid grew significantly faster than the cells transfected with the pGL102LmPPI plasmid from day 4 to day 7 ($P<0.05$), except on day 5.

5.2.1.3.2. Differentiation to metacyclic promastigotes

It was interesting to determine if there was a difference in the kinetics of formation of infective metacyclic promastigotes in the different cultures. Metacyclic promastigotes were separated from the procyclic promastigotes using the peanut agglutinin method (Sacks *et al.*, 1985). The method is based on the observation that concentrations of peanut agglutinin necessary to agglutinate the totality of log-phase *Leishmania* promastigotes do not agglutinate metacyclic promastigotes. The difference in agglutination is due to developmental regulation of surface lipophosphoglycans, shed by procyclic promastigotes, and acquisition of new carbohydrate determinants during metacyclogenesis (Sacks and da Silva, 1987). It was shown that the number of metacyclic promastigotes in the culture of cells over-expressing the active enzyme increased significantly more slowly than in the 2 other cultures (Fig. 5.5), though these metacyclic promastigotes were morphologically identical to wild type cells. The measurements were taken in duplicate from the time point at which the cells entered stationary phase of growth (day 4 after initiation of the cultures at 10^6/ml). It seemed that the protein over-expression was not responsible for this slow conversion, as the cells over-expressing the PPI dead mutant produced the same number of metacyclic promastigotes as the control cells. Thus it seemed that over-expression of the active PPI resulted in slower metacyclogenesis.

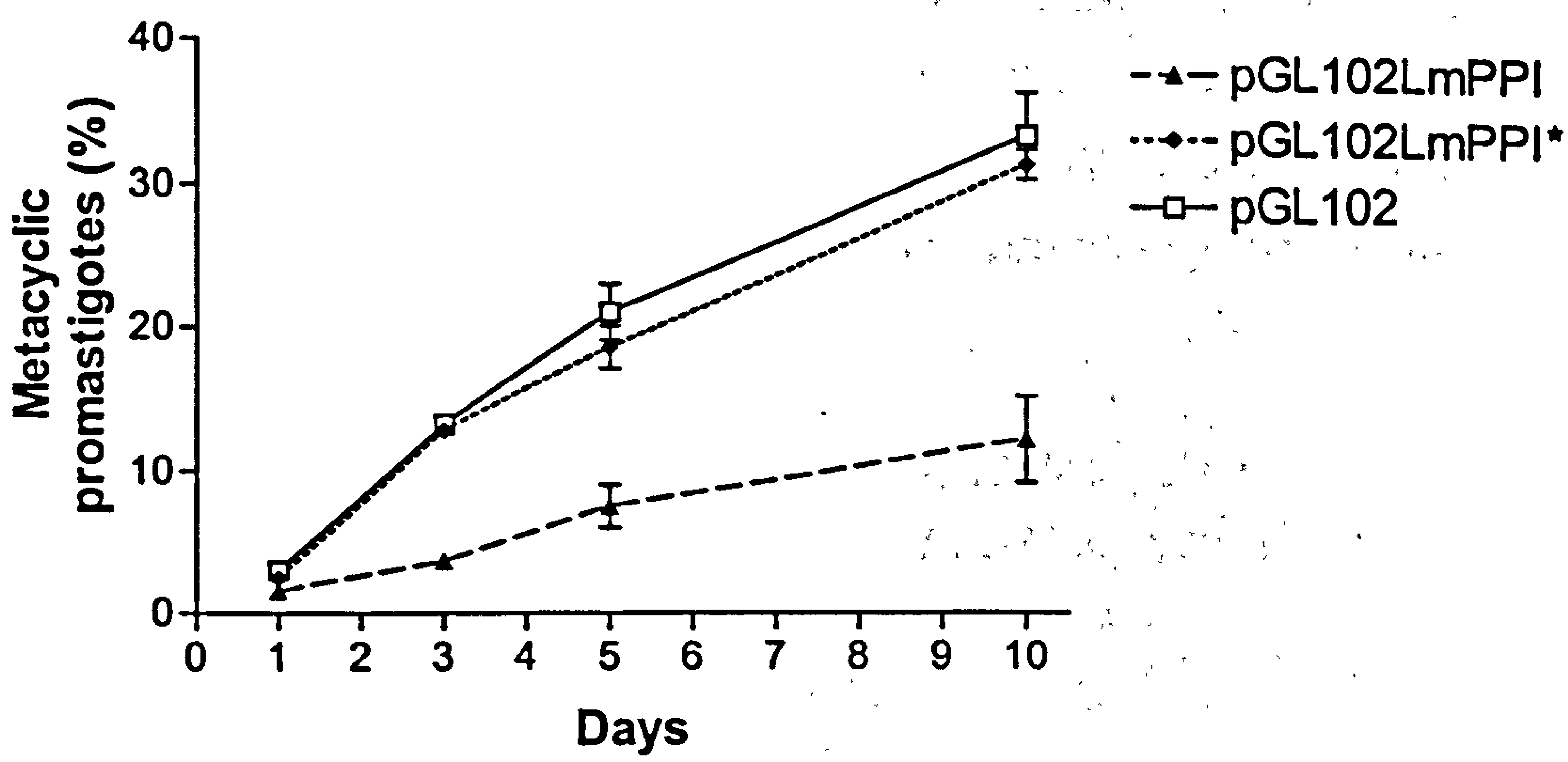


FIG. 5.5. **Production of metacyclic promastigotes in *L. major*.** Cells were grown with 50 µg/ml of neomycin and metacyclic proportions were assessed using peanut agglutinin agglutination at regular time intervals. pGL102LmPPI, cells expressing the native PPI; pGL102LmPPI*, cells over-expressing the active site cysteine mutant; pGL102, cells transfected with the empty vector. Results are the means ± SE from 2 independent measurements. pGL102LmPPI transfected cells produced significantly less metacyclic promastigotes from day 3 compared to the 2 other cell lines ($P<0.05$, $P<0.03$ at day 10).

5.2.1.3.3. *In vitro* infectivity to macrophages

Infection *in vitro* was assessed using peritoneal macrophages, with stationary phase promastigotes and purified metacyclic promastigotes. Infection with stationary phase promastigotes over-expressing the active PPI showed a significantly lower level of infection compared to the promastigotes over-expressing the dead PPI or the empty episome (about 2 times less) (Fig. 5.6). But when the macrophages were infected with purified metacyclic promastigotes, the infections rates observed with the different cell lines were very similar (all around 40%). The culture of cells over-expressing the active PPI contained about 2 times fewer metacyclic promastigotes compared to the other 2 lines, which may explain the lower level of infection when total populations of promastigotes were used. Using a ratio of metacyclic promastigotes of 0.5:1, the maximum infection rate expected would be around 50%, which fits well with the results obtained; and when using promastigotes at a 2:1 ratio, the infection expected would be around 2 times the percentage of metacyclic promastigotes in the culture. In this case, the infection rates obtained are slightly higher; this may be due to loss of some macrophages during the washing steps. Comparing the percentages of metacyclic promastigotes in the different cultures, a 2-fold lower infection rate would be expected with the line over-expression the active PPI, which is the case. So the over-expression of the active PPI seems to hinder the production of metacyclic promastigotes, but the metacyclic promastigotes produced are infective for macrophages.

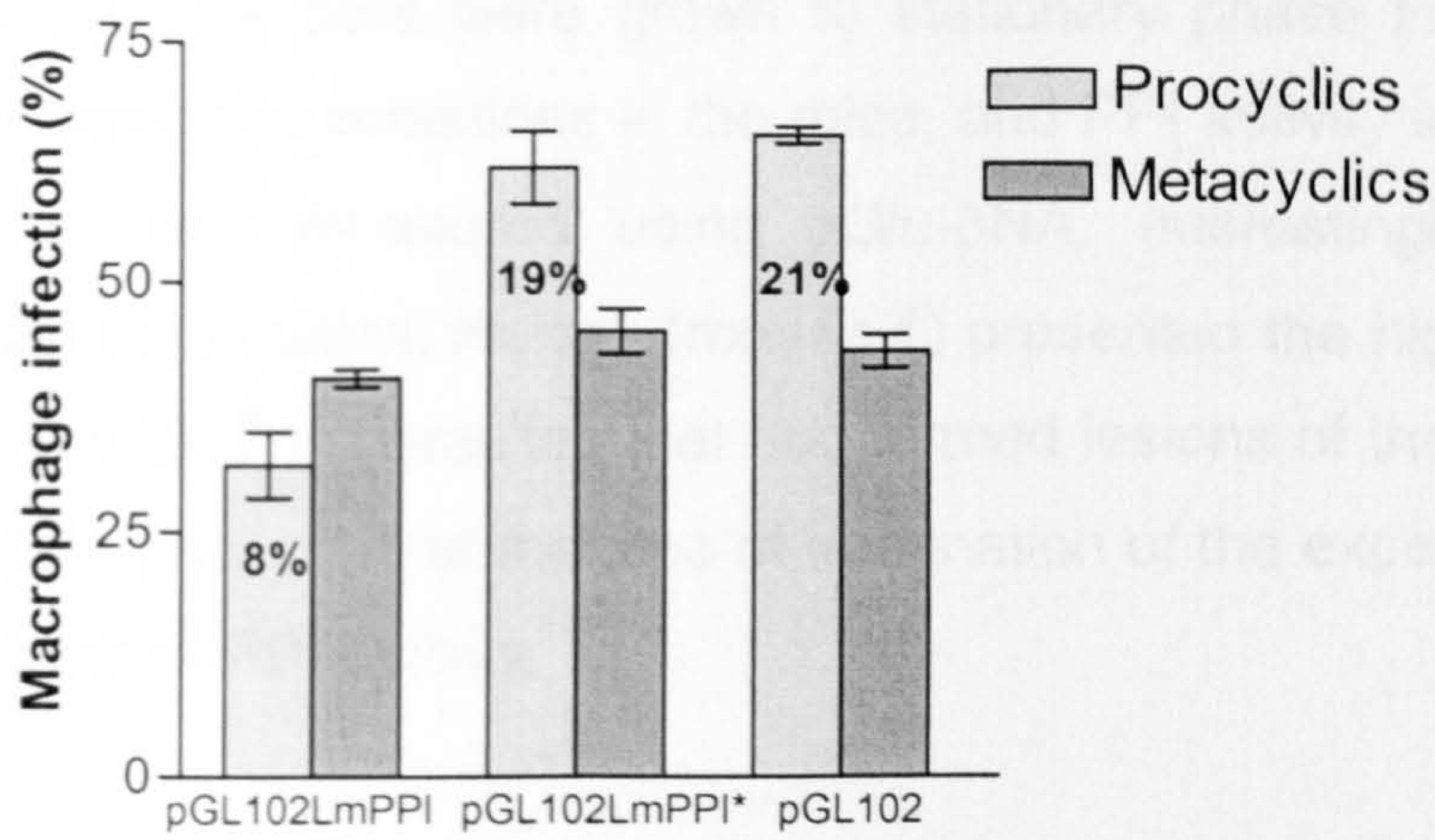


FIG. 5.6. ***In vitro* macrophage infection by *L. major* stationary phase promastigotes or purified metacyclic promastigotes.** Promastigotes were infected at a ratio of 2:1, metacyclic promastigotes at a ratio of 0.5:1. Percentages on the bars correspond to the percentage of metacyclic promastigotes in the stationary phase culture. Incubations were for 6 days. pGL102LmPPI, cells expressing the native PPI; pGL102LmPPI*, cells over-expressing the active site cysteine mutant; pGL102, cells transfected with the empty vector. Results are the means ± SD from 3 independent experiments. The infection rates obtained using pGL102LmPPI transfected promastigotes were significantly lower than the ones obtained with the 2 others cell lines ($P<0.004$), whereas the infection rates obtained with pGL102LmPPI* and pGL102 transfected promastigotes were not significantly different from each other ($P>0.05$). Infection rates obtained with metacyclic promastigotes from the different lines were similar ($P>0.05$).

5.2.1.3.4. *In vivo* infectivity in mice

5 x 10^5 stationary phase promastigotes of the cell lines transfected with the pGL102LmPPI, pGL102LmPPI* or pGL102 plasmids were inoculated into the footpad of BALB/c mice. The evolution of the infection was assessed by

measurement of the footpad widths over time. The results showed that cells over-expressing the dead mutant were infecting as well as the cells transfected with the empty vector, whereas the cells over-expressing the active enzyme showed a delay in infection of 5 weeks (Fig. 5.7). The over-expression was via an episomal plasmid, and the selection for neomycin resistance was not possible during the experiment, thus it was expected the plasmid would probably be slowly lost over time. Whether this process would take 5 weeks to occur and so the later growth of the lesions correlated with loss of the over-expression of PPI is unknown. To test if the level of PPI activity might be related to the size of lesions formed, amastigotes from lesions of different sizes were extracted from different mice from this group. The cells were grown to stationary phase in medium without neomycin to mimic the conditions in the mice, and PPI activity in the supernatant of cell lysates was measured using pGlu-βNA. Interestingly, the parasites extracted from the smallest lesions (mouse 1) presented the highest level of PPI activity (Table 5.2). The parasites that had formed lesions of the same size as of the control cells (mouse 2) at the time of termination of the experiment, showed a very similar level of PPI activity.

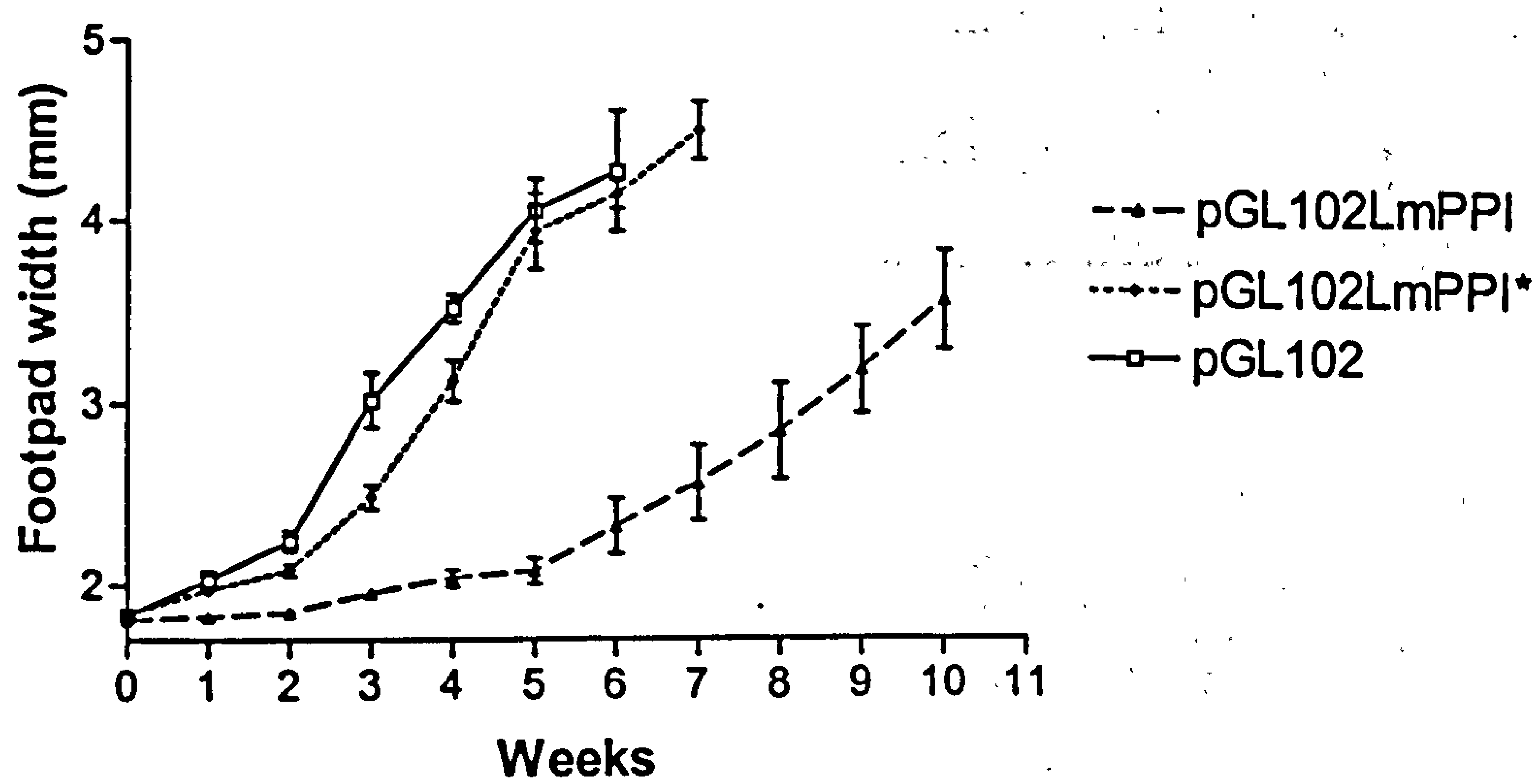


FIG. 5.7. Mice infectivity of *L. major* cell lines. $5x10^5$ cells resuspended in 20 µl of PBS were infected into one footpad of BALB/C mice. Measurements of the width of the footpads were taken at regular intervals. pGL102LmPPI, cells expressing the native PPI; pGL102LmPPI*, cells over-expressing the active site cysteine mutant; pGL102, cells transfected with the empty vector. Results are the means ± SD from 6 mice. The lesions provoked by pGL102LmPPI transfected cells were significantly smaller from week 2 compared to the 2 other cells lines ($P<0.0001$), and the lesions provoked by pGL102 transfected cells were only significantly bigger than the ones provoked by pGL102LmPPI* transfected cells from week 2 to 4 ($P<0.02$).

TABLE. 5.2. **Comparison between infected mice footpad sizes and PPI activity of promastigotes generated from extracted amastigotes.** At the termination time of the mice infectivity experiment, *L. major* amastigotes from 2 of the 6 mice infected with the *L. major* cell line over-expressing the native PPI of *L. major*, and from 1 of the 6 mice infected with cells transfected with the empty pGL102 vector were taken from the footpad lesions and grown in medium without neomycin. The PPI activity in the supernatants of cell lysates were measured in 50 mM HEPES pH 8.0, 2 mM TCEP, 1 mM EDTA, 200 μM pGlu-βNA, continuously for 5 min at 32°C (excitation 320 nm, emission 410 nm). $8x10^7$ cell equivalents were used per assay. Activity is represented in comparison to activity in control. pGL102LmPPI (mouse 1), cells expressing the native PPI and extracted from the footpad of the mouse with the smallest footpad size from the group; pGL102LmPPI (mouse 2), cells expressing the native PPI and extracted from the footpad of the mouse with the biggest footpad size from the group; pGL102, cells transfected with the empty vector.

Cell line	pGL102LmPPI (mouse 1)	pGL102LmPPI (mouse 2)	pGL102
Duration of mice footpad infection (weeks)	10	10	6
Size of footpad at end of experiment (mm)	2.40	4.20	4.50
Comparative PPI activity in supernatant of cell lysate	2.03	1.11	1

5.2.2. Knock-out of *LmPPI* gene in *L. major* promastigotes

5.2.2.1. Analysis of knock-out

Using 5' and 3' specific flanks to the PPI, a plasmid was designed for gene knock-out (Fig. 5.8). The selection markers used for the first allele knock-out was the blasticidin S resistance gene, and the selection marker used for the second allele knock-out was the hygromicin B resistance gene. Two knock-out lines were generated from independent sequential transfection events (lines KOPPa and KOPPb).

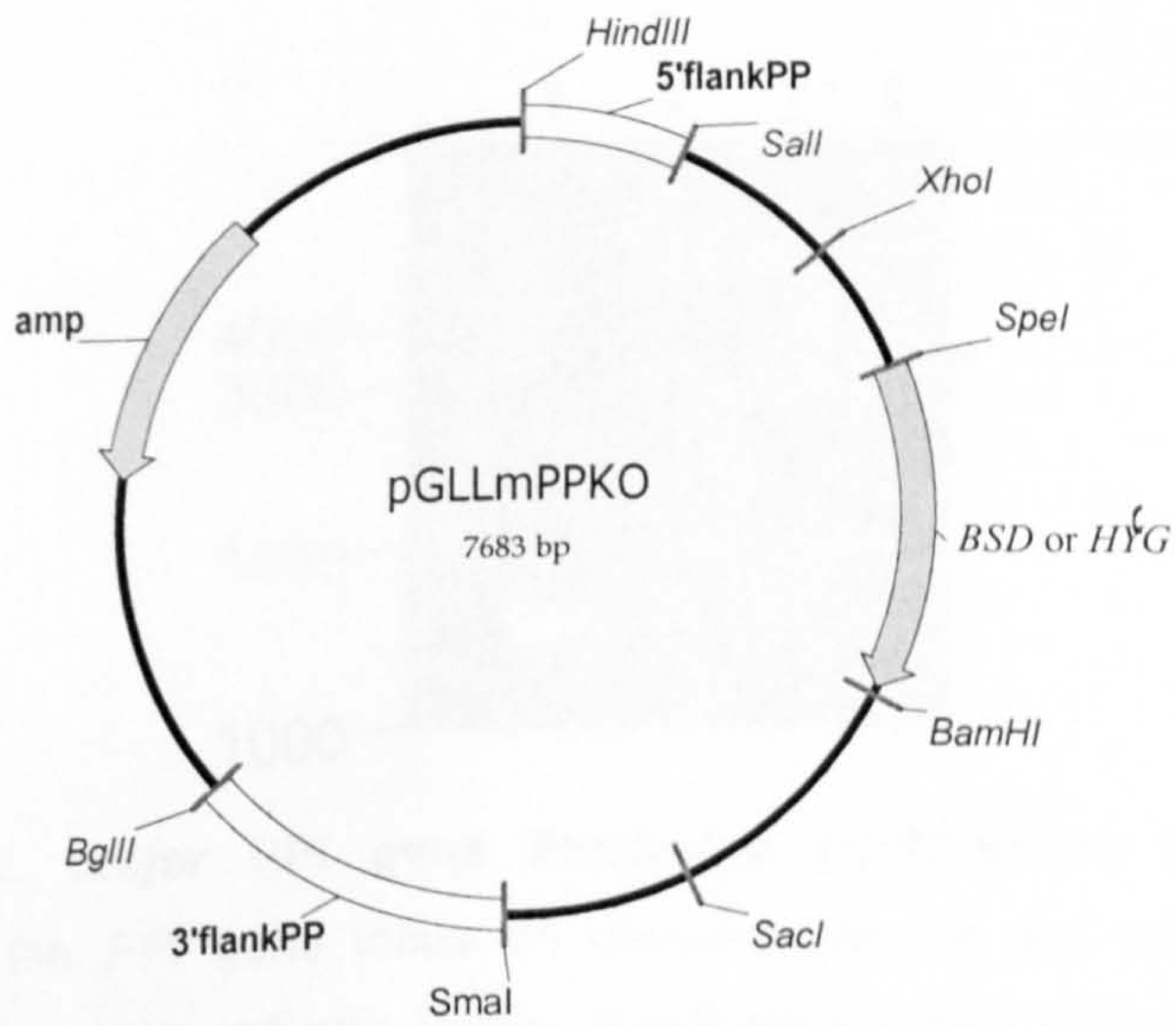


FIG. 5.8. **Construct for gene knock-out of the pyroglutamyl peptidase I (PPI) of *L. major*.** Schematic representation of the plasmid resulting from the cloning of the sequences for the 5' and 3' flanks of the PPI of *L. major* in pGL345 (*HYG*) or pGL842 (*BSD*) for gene replacement of the PPI. amp, ampicillin resistance gene; *BSD*, blasticidin resistance gene; *HYG*, hygromycin resistance gene.

Gene replacement was first confirmed by PCR (Fig. 5.9B) (a scheme of the locus and the position of the primers used is shown on Fig. 5.9A). A single DNA fragment at ~1.4 kb corresponding to the wild type *PPI* gene was amplified from the wild type cells. With the single allele knock-out line, 2 DNA fragments were amplified; one at ~1.4 kb corresponding to the wild type *PPI*, and one at ~3.2 kb corresponding to the blasticidin resistance gene. In the knock-out lines, 2 DNA fragments were amplified; one at ~3.2 kb corresponding the blasticidin resistance gene, and one at ~4 kb corresponding to the hygromycin resistance gene.

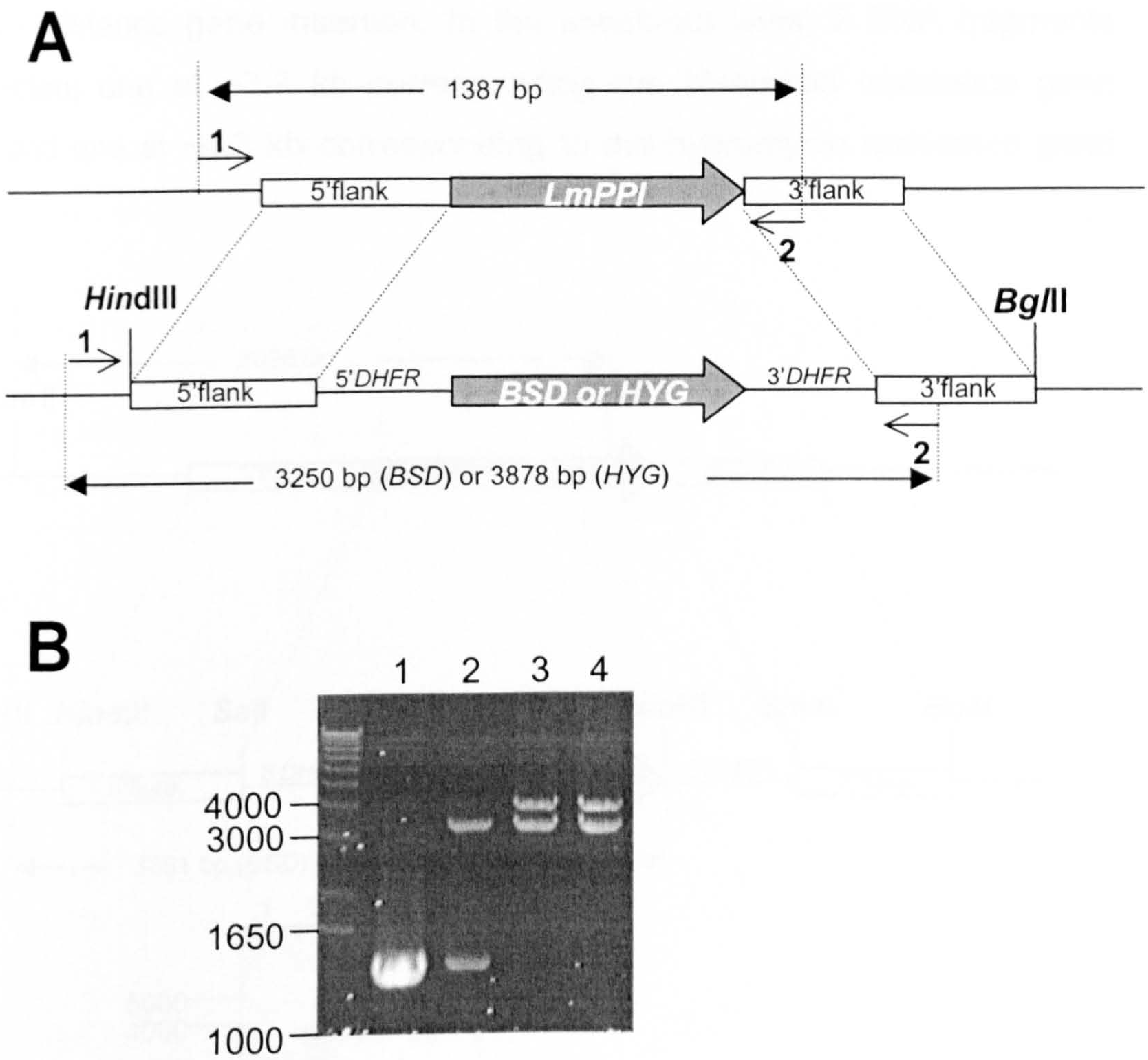


FIG. 5.9. ***L. major* PPI gene knock-out confirmation by PCR.** (A) Schematic representation of the *PPI* gene locus on chromosome 34 and the result of the *PPI* gene replacement. Block arrows indicate ORFs. Small arrows indicate position of PCR primers. 1: checkKOPPFor; 2: checkKOPPRev (see Table 2.1). The predicted sizes of fragments amplified by PCR are indicated. *LmPPI*, *L. major* pyroglutamyl peptidase gene; *DHFR*, dihydrofolate reductase gene; *BSD*, blasticidin resistance gene; *HYG*, hygromycin resistance gene. (B) PCR analysis. PCR was performed on genomic DNA with primers 1 and 2 and using the Long Template PCR System (Roche) and 10 μl were run on a 1% (w/v) agarose gel. Lane 1, wild type; lane 2, single allele knock-out (1 allele replaced with *BSD* gene); lane 3, double *PPI* gene knock-out, clone KOPPa; lane 4, double *PPI* gene knock-out, clone KOPPb. Clones KOPPa and b were generated from independent first and second allele replacements. The positions of the molecular size markers in bp are shown on the left of the figure.

Knock-out cells lines were also confirmed by Southern-blot, using the 5' flank sequence as a probe (Fig. 5.10). On the wild type cells, a single DNA fragment at ~3 kb corresponding to the wild type *PPI* gene was detected. On the single allele knock-out line, 2 DNA fragments were detected; one at ~3 kb corresponding to the wild type *PPI*, and one at ~3.7 kb corresponding the blasticidin resistance gene insertion. In the knock-out lines, 2 DNA fragments were detected; one at ~3.7 kb corresponding the blasticidin resistance gene insertion, and one at ~4.3 kb corresponding to the hygromycin resistance gene insertion.

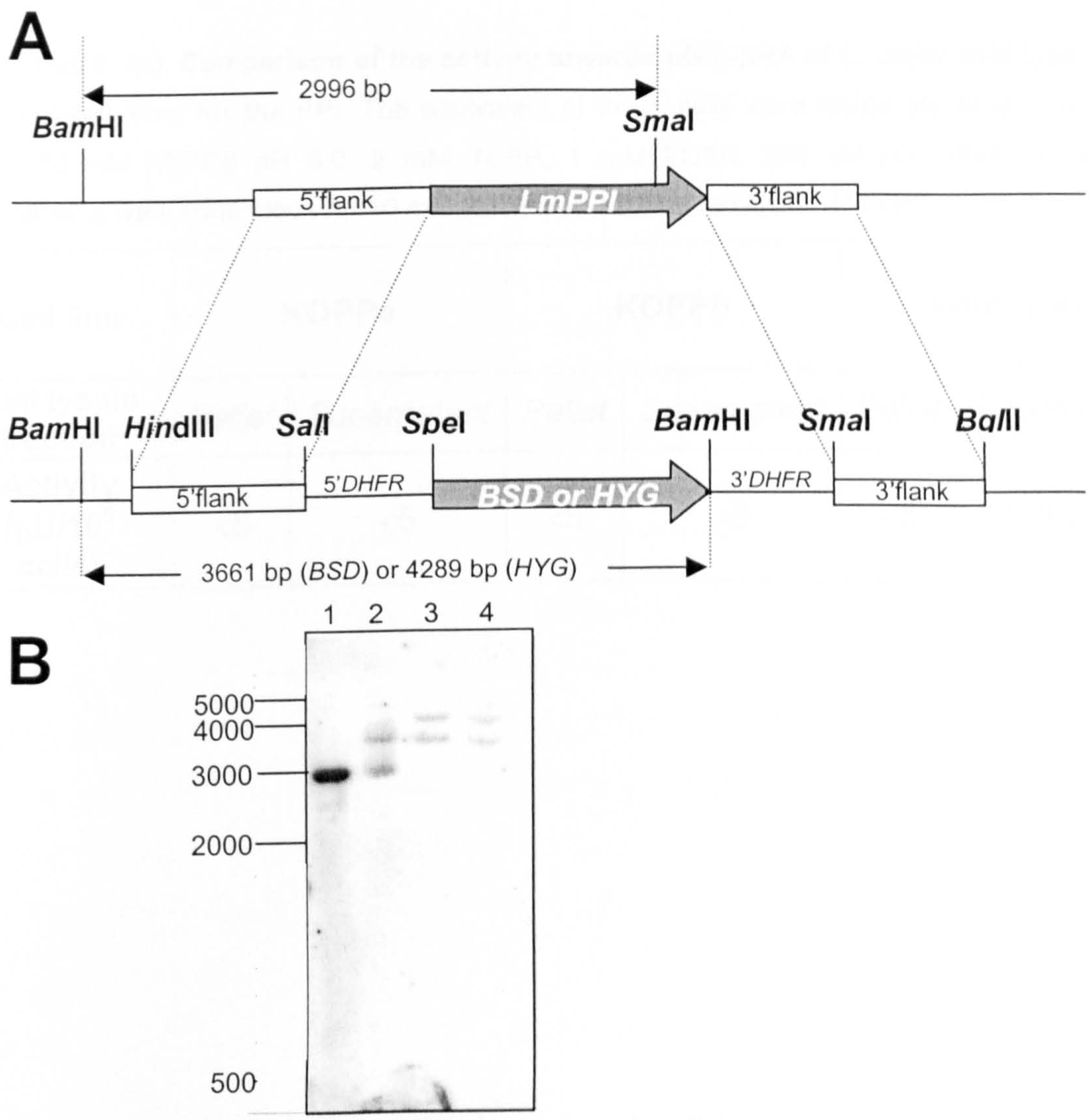


FIG. 5.10. ***L. major* PPI gene knock-out confirmation by Southern blot.** (A) Schematic representation of the *PPI* gene locus on chromosome 34 and the result of the *PPI* gene replacement. Arrows indicate ORFs. The predicted sizes of the *Bam*HI/*Sma*I fragments detected by Southern blot using the 5'flank as a probe are indicated. *LmPPI*, *L. major* pyroglutamyl peptidase gene; *DHFR*, dihydrofolate reductase gene; *BSD*, blasticidin resistance gene; *HYG*, hygromycin resistance gene. (B) Southern blot analysis. Genomic DNA was digested with *Bam*HI and *Sma*I, run on a 1% (w/v) agarose gel, blotted on nylon membrane and hybridised with ^{32}P-labelled 5' flank sequence. Lane 1, wild type; lane 2, single allele knock-out (1 allele replaced with *BSD* gene); lane 3, double *PPI* gene knock-out, clone KOPPa; lane 4, double *PPI* gene knock-out, clone KOPPb. Clones KOPPa and b were generated from independent first and second allele replacements. Blot was exposed for 8 days. The positions of the molecular size markers (bp) are shown on the left of the figure.

The PPI activity present in the different fractions of cell lysates was measured using the substrate pGlu-βNA, and the activity of knock-out lines was compared to the wild type cells. PPI activity was totally absent from the knock-out lines compared to the wild type cells (Table 5.3).

TABLE. 5.3. **Comparison of the activity towards pGlu-βNA of *L. major* wild type and knock-out cell lines for the PPI.** The equivalent of 8×10^7 cells were tested per assay. Assays were in 50 mM HEPES pH 8.0, 2 mM TCEP, 1 mM EDTA, 200 μM pGlu-βNA, at 32°C. Measurements were continuous for 10 min (excitation 320 nm, emission 410 nm).

Cell line	**KOPPa**		**KOPPb**		**Wild type**	
Cell lysate fraction	*Pellet*	*Supernatant*	*Pellet*	*Supernatant*	*Pellet*	*Supernatant*
Activity ($\mu U/10^9$ cells)	<5	<5	<5	<5	<5	100 ± 20

5.2.2.2. Phenotype analysis of PPI knock-out cell lines

5.2.2.2.1. Morphology and growth

Promastigotes of PPI knock-out lines of *L. major* did not show any difference detectable at the morphological level compared to wild type cells. In terms of growth, all cell lines multiplied at the same rate and reached the same final density as the wild type parasites (Fig. 5.11). Thus, the knock-out of *PPI* did not influence the growth of promastigotes *in vitro*.

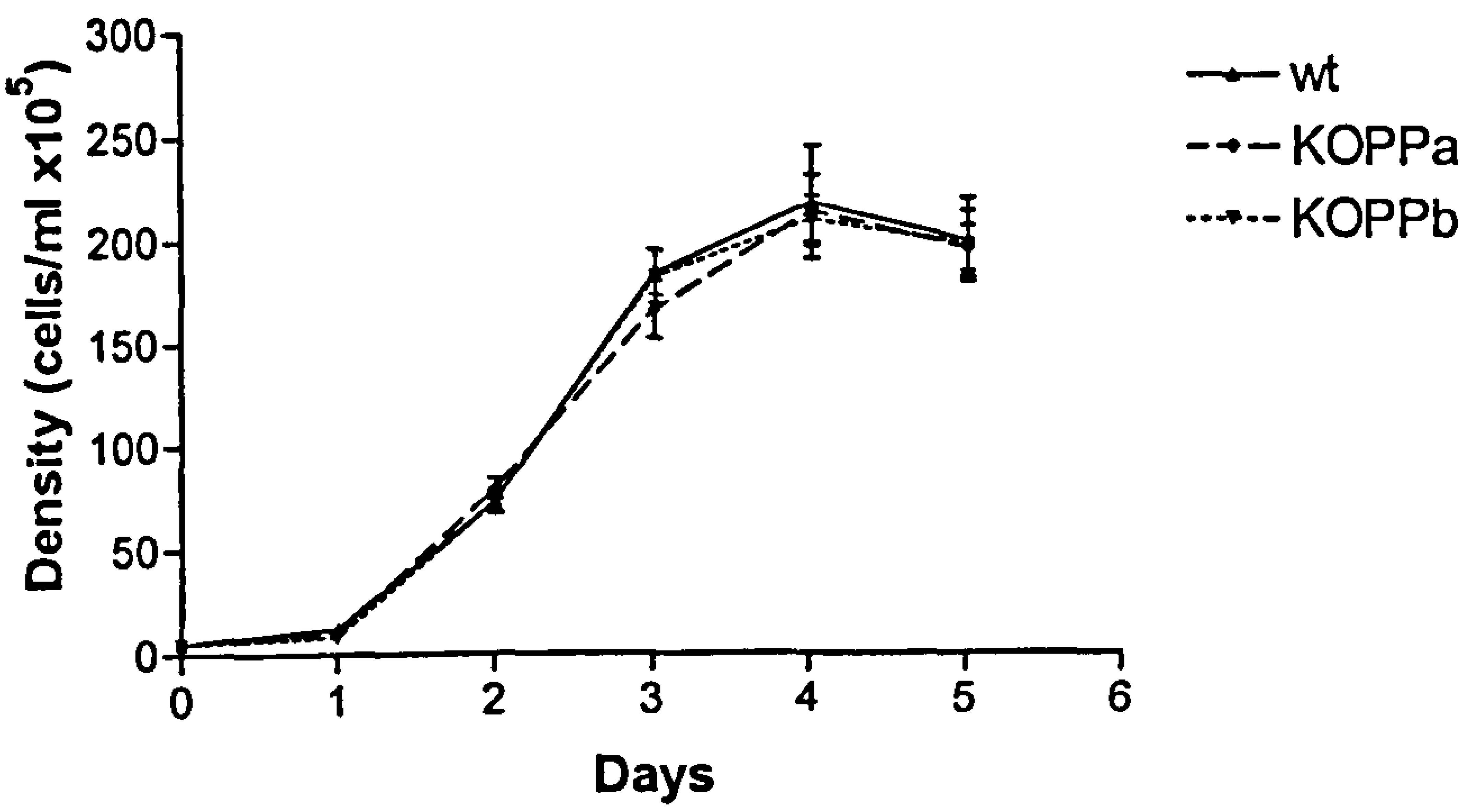


FIG. 5.11. **Growth of *L. major* wild type and PPI knock-out line as promastigotes.** Cells were grown without selection drug and counted at regular intervals. KOPPa and KOPPb correspond to 2 different cell lines obtained from individual gene knock-out events; PPI knock-out cell lines; wt, wild type cells. Results are the means ± SD from 3 independent experiments.

5.2.2.2.2. Conversion to metacyclic promastigotes

The percentage of metacyclic promastigotes in the populations were determined from the time point at which the cells entered stationary phase of growth (day 4 after initiation of cultures at 10^6 cells/ml). There did not appear to be a difference either in the morphology, or the kinetics of formation of metacylic promastigotes over time between the knock-out cell lines and wild type cells (Fig. 5.12). Thus, the knock-out of the *PPI* did not influence the conversion.

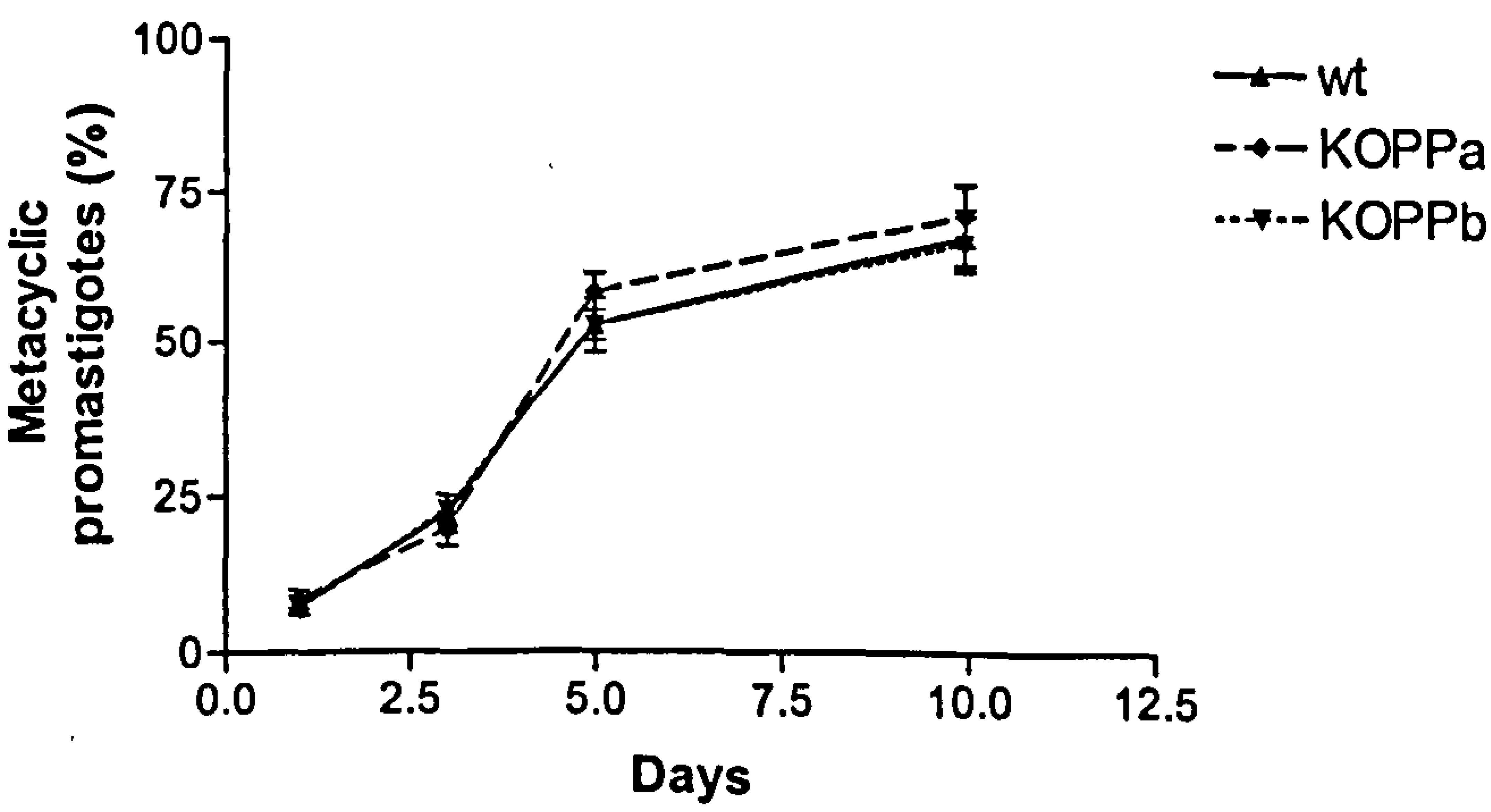


FIG. 5.12. **Metacyclogenesis in *L. major* wild type and PPI knock-out cell lines.** Cells were grown to stationary phase and metacyclic promastigotes proportions were assessed using the peanut agglutinin agglutination test at regular time intervals. wt, wild type; KOPPa and KOPPb correspond to 2 different cell lines obtained from individual gene knock-out events. Results are the means ± SE from 2 independent measurements.

5.2.2.2.3. *In vitro* infectivity in macrophages

Infectivity *in vitro* towards peritoneal macrophages was assessed, using stationary phase promastigotes and purified metacyclic promastigotes. Promastigotes and purified metacyclic promastigotes from wild type and the 2 independent knock-out lines resulted in infection rates that were very similar, of about 60% and 40% respectively (Fig. 5.13). The numbers of metacyclics in the different promastigote cultures were very similar, so these results suggest that the knock-out of the PPI did not affect infectivity *in vitro*.

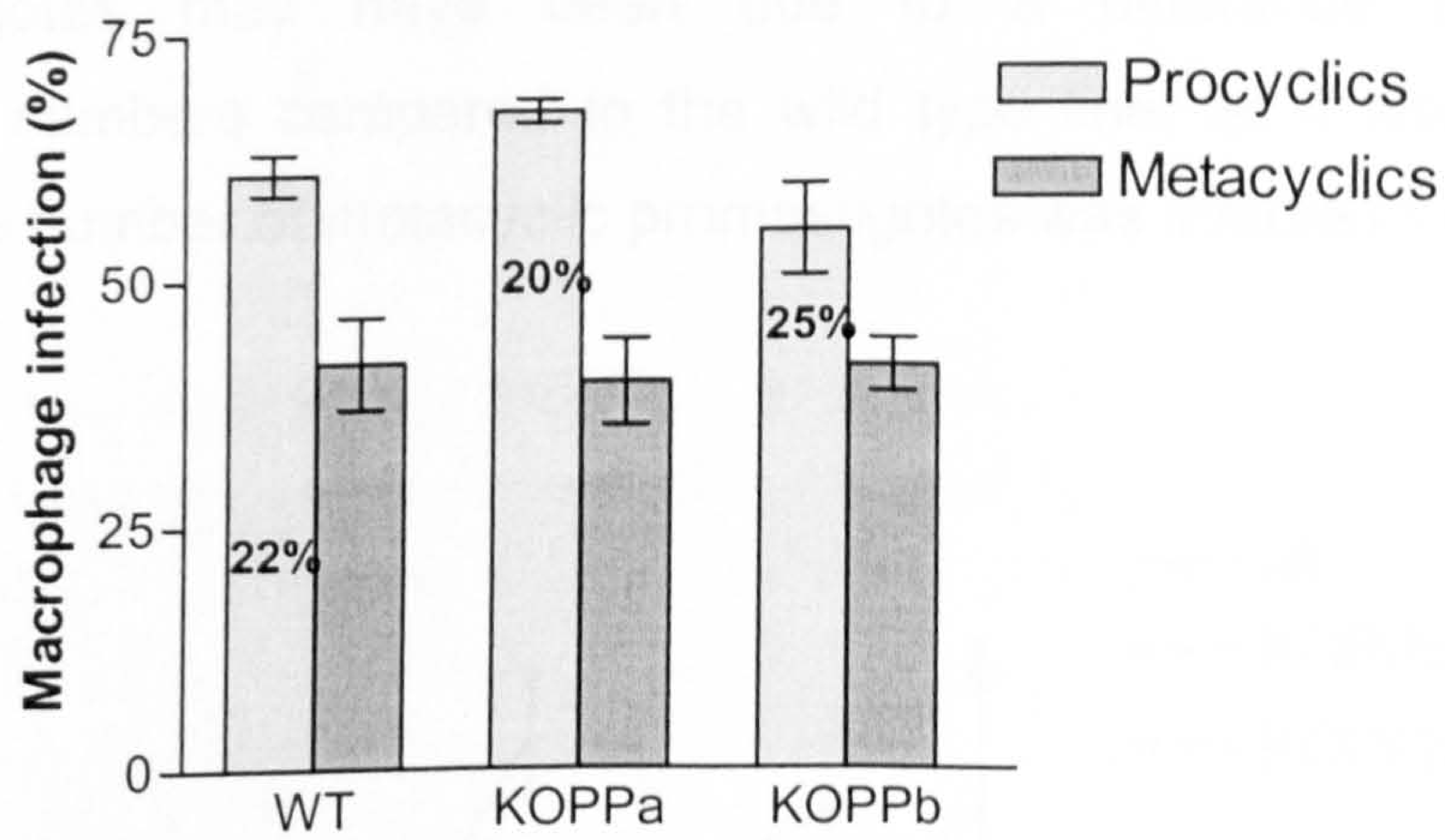


FIG. 5.13. ***In vitro* macrophage infection by *L. major* stationary phase promastigotes or purified metacyclic promastigotes of wild type or *PPI* knock-out cell lines.** Promastigotes were infected at a ratio of 2:1, metacyclic promastigotes at a ratio of 0.5:1. Percentages on the bars correspond to the percentage of metacyclic promastigotes in the stationary phase cultures. Incubations were for 6 days. WT, wild type; KOPPa and KOPPb correspond to 2 different cell lines obtained from individual gene knock-out events. Results are the means ± SD from 3 independent experiments. The different cells lines did not induce significantly different infection levels ($P>0.05$).

5.2.2.2.4. *In vivo* infectivity in mice

Lesions resulting from inoculation of mice footpads with 5 x 10^5 promastigotes of the 2 knock-out cells lines grew more slowly than those of wild type parasites, showing a delay of about 2-3 weeks. Subsequent growth was also somewhat slower than the lesions caused by wild type parasites (Fig. 5.14). The proportions of metacyclic promastigotes in the populations of promastigotes used to infect were determined and it was shown that while wild type culture contained 12% metacyclic promastigotes, the PPI knock-out lines contained 6%. It was unclear if the difference in growth of the lesions was due to the difference in numbers of metacyclic promastigotes injected into the footpads, to difference in infectivity of the metacyclic promastigotes in the different cultures, or to a difference in virulence of the amastigotes of the different lines. Therefore the experiment was repeated using the same number of metacyclic promastigotes per injection for all lines. In this case, the infection rates observed for all 3 lines (wild type and 2 lines of knock-out) were the same (Fig. 5.15). So it seemed that the knock-out of the *PPI* gene did not have an effect on infectivity to mice, and that the delay in infectivity observed in the case of mice infection with *PPI* knock-out promastigotes may have been due to a difference in metacyclic promastigotes numbers compared to the wild type line, as it was suppressed when the same number of metacyclic promastigotes was infected for each lines.

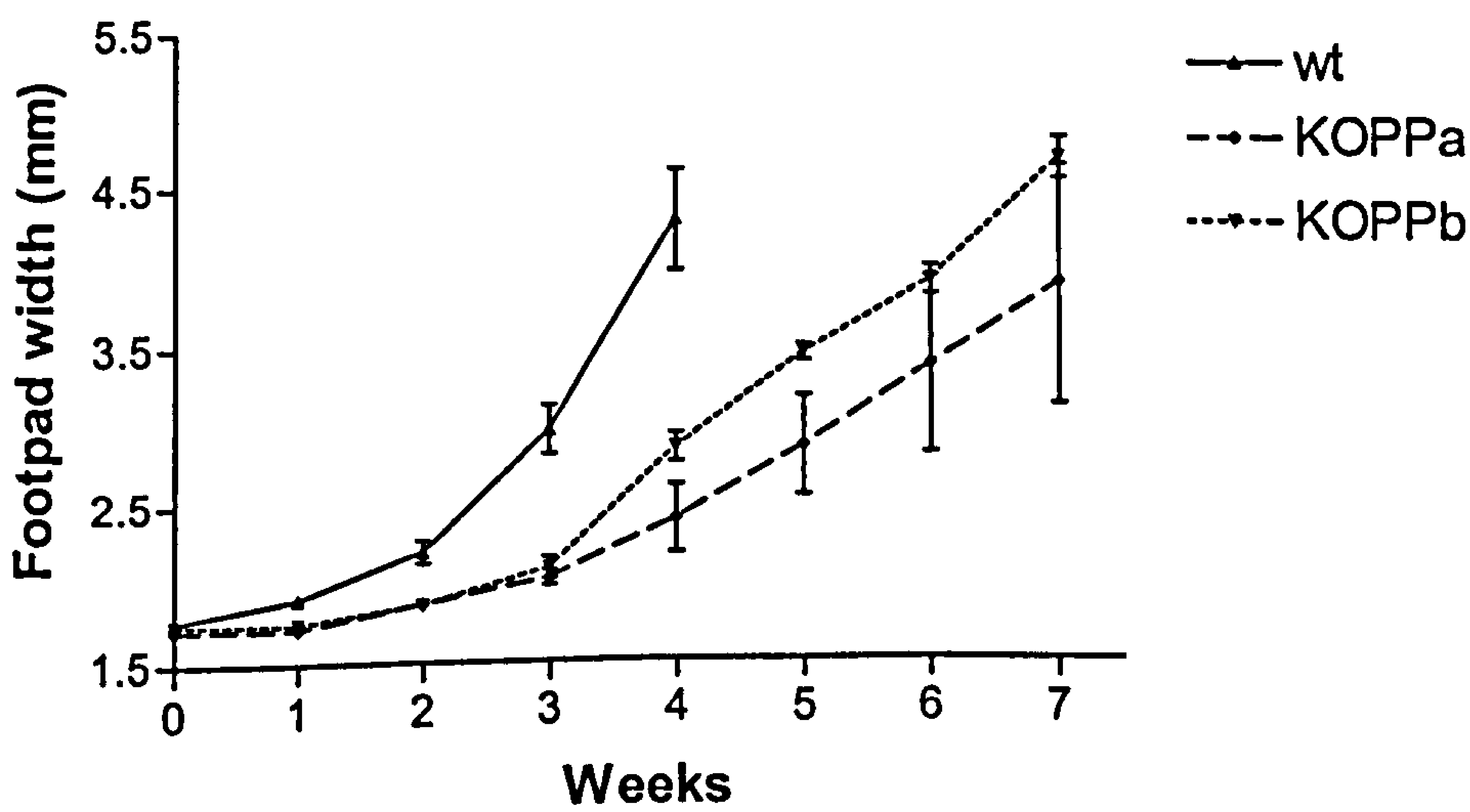


FIG. 5.14. **Mice infectivity of *L. major PPI* knock-out and wild type cell lines.** $5x10^5$ cells resuspended in 20 µl of PBS were infected into one footpad of BALB/c mice. Six mice per cell line were infected. Measurements of the width of the footpads were taken at regular intervals. KOPP, knock-out cell lines for the PPI; wt, wild type cells. KOPPa and KOPPb correspond to 2 different cell lines obtained from individual gene knock-out events. Results are the means ± SD from 6 mice. The infections generated by wild type parasite were significantly greater than the 2 others from week 1 (P=0.0002), and the infections generated by the 2 individual knock-out lines were only significantly different from each other at week 4 (P=0.005) and week 5 (P=0.0017).

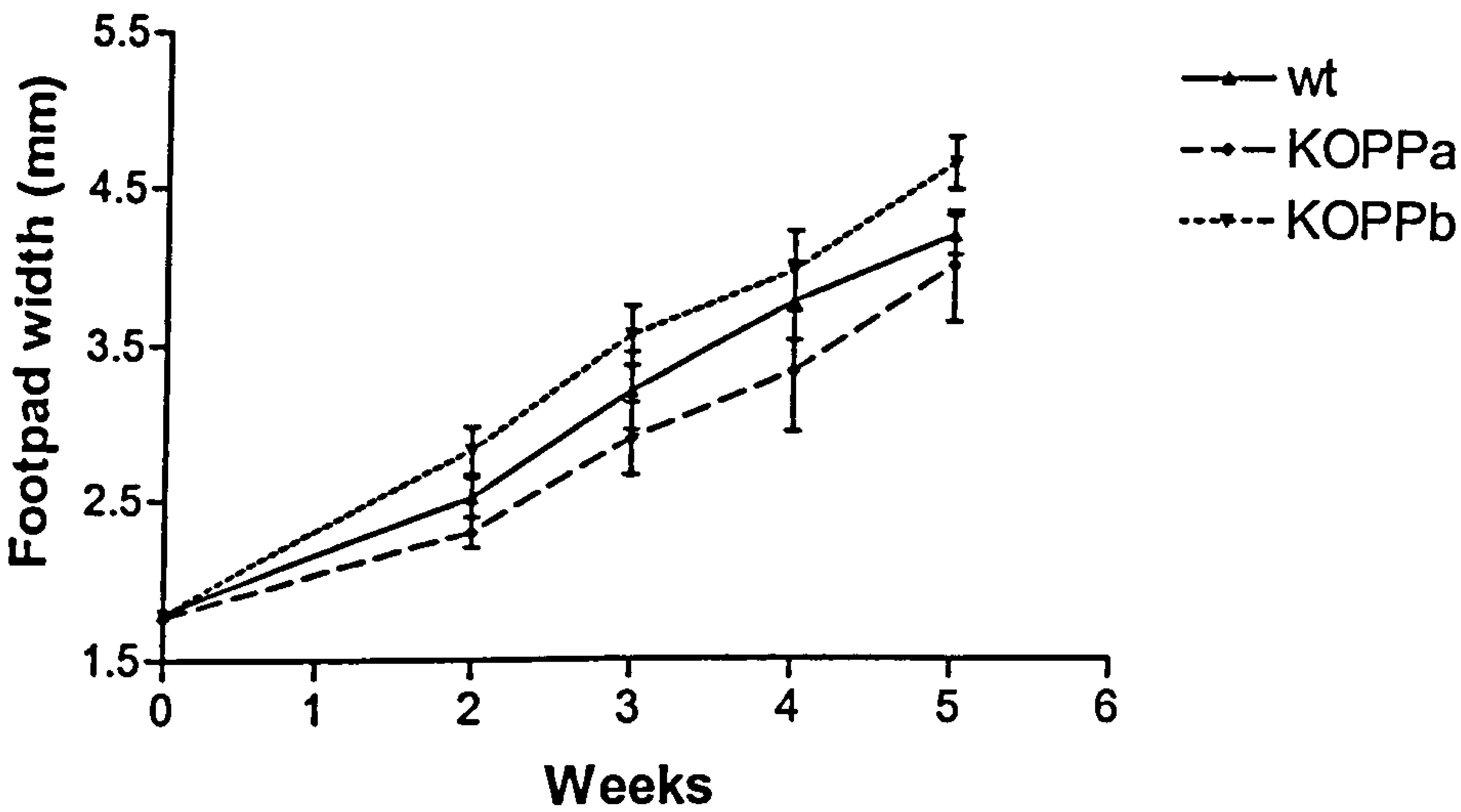


FIG. 5.15. **Mice infectivity of *L. major PPI* knock-out and wild type purified metacyclic promastigotes.** 10^5 purified metacyclic promastigotes resuspended in 20 μl of PBS were infected into one footpad of BALB/c mice. Four mice per cell line were infected. Measurements of the width of the footpads were taken at regular intervals. KOPP, knock-out cell lines for the PPI; wt, wild type cells. KOPPa and KOPPb correspond to 2 different cell lines obtained from individual gene knock-out events. Results are the means ± SD from 6 mice. The infections generated by the different cell lines were not significantly different (P>0.05).

5.2.3. PPI localisation in *L. major* promastigotes

Immunolocalisation studies using the anti-LmPPI antiserum on *L. major* promastigote cell lines (knock-out for the *PPI* gene, wild type and over-expressing the PPI enzyme) showed the enzyme may be cytosolic (Fig. 5.16). The staining was not associated with the nucleus, and was diffuse throughout the cytoplasm.

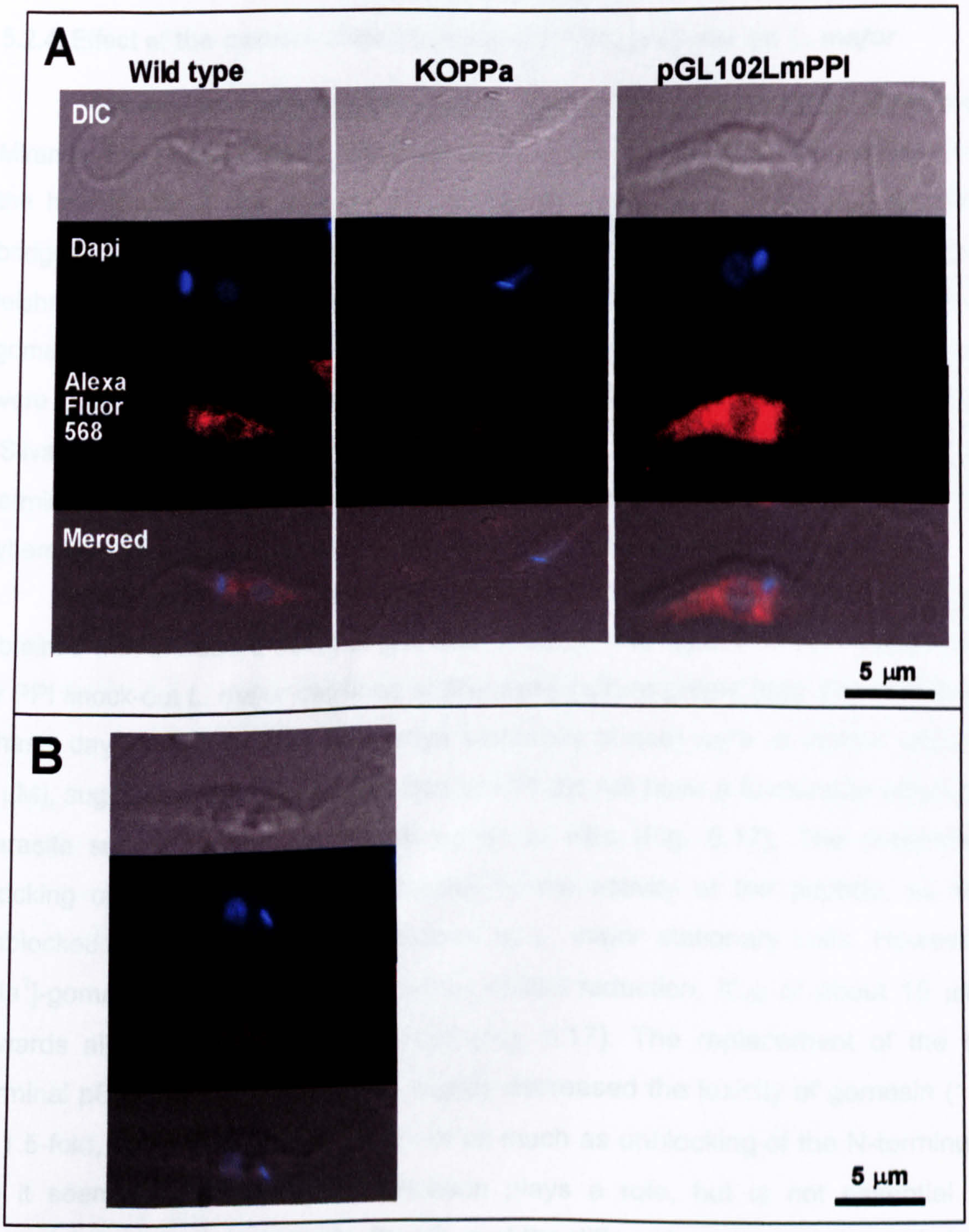


FIG. 5.16. **Immunolocalisation of the LmPPI in *L. major* promastigotes using anti-LmPPI antiserum raised in rat.** The secondary antibody was the Alexa Fluor® 568 goat anti-rat IgG (Molecular Probes) at a dilution of 1:500 (v/v). Blue staining corresponds to DAPI staining of nucleus and kinetoplast; red staining corresponds to PPI localisation. From top to bottom, pictures are: differential interference contrast (DIC), DAPI, Alexa Fluor® 568 and merged. Wild type: *L. major* wild type; KOPPa: *L. major* cell line having both alleles for the *PPI* gene removed; pGL102LmPPI: *L. major* cell line over-expressing the active PPI of *L. major*. (A) Anti-LmPPI antiserum was used at 1:500 (v/v). (B) No incubation with anti-LmPPI antiserum, wild type cells.

5.2.4. Effect of the natural antimicrobial peptide gomesin on *L. major*

Gomesin and gomesin analogues were kindly provided by Dr Antonio de Miranda (Sao Paolo, Brazil). Gomesin is an 18-residue cysteine-rich peptide from the hemolymph of the spider *Acanthoscurria gomesiana*, carrying 2 disulfide bridges, a N-terminal pGlu and a C-terminal α-amidation, which showed leishmanicidal properties (Silva *et al.*, 2000). The primary amino acid structure of gomesin is ZCRRLCYKQRCVTYCRGR* (using the single letter amino acid code, were Z stands for pyroglutamate and the asterisk indicates a C-terminal α-amide) (Silva *et al.*, 2000). The gomesin analogues were [Ala^1]-gomesin, where the N-terminal pyroglutamate was replaced by an alanine, and Ac-[Ala^1]-gomesin, where the N-terminal pyroglutamate was replaced by an acetylated alanine.

The IC_{50} (inhibitor concentration required to reduce by 50% the response obtained with untreated cells) of gomesin towards wild type, PPI over-expression or PPI knock-out *L. major* cell lines at the same culture phase (day 3 of stationary phase, day 1 being the day cells enter stationary phase) were all similar (around 5 μM), suggesting the over-expression of PPI did not have a favourable effect on parasite survival in the conditions tested *in vitro* (Fig. 5.17). The N-terminal blocking of gomesin was not essential to the activity of the peptide, as the unblocked [Ala^1]-gomesin was still toxic to *L. major* stationary cells. However, [Ala^1]-gomesin showed reduced toxicity (3-fold reduction, IC_{50} of about 15 μM) towards all *L. major* cell lines tested (Fig. 5.17). The replacement of the N-terminal pGlu by an acetyl-alanine slightly decreased the toxicity of gomesin (1.2 to 1.5-fold, IC_{50} of about 7 μM), but not as much as unblocking of the N-terminus. So it seems that the pGlu modification plays a role, but is not essential to gomesin's toxicity. Interestingly, the IC_{50}s of the different compounds towards late log-phase *L. major* promastigotes (wild type, PPI over-expressing and PPI knock-out cell lines) were about 5-fold lower than the IC_{50} towards stationary phase promastigotes at day 3 of stationary phase (data for wild type cells are shown on Fig. 5.18). The wild type culture contained about 15-20% metacyclics promastigotes at day 3 of stationary phase (see Fig. 5.13). The decreased susceptibility of stationary phase promastigotes compared to log-phase promastigotes was probably due to surface remodelling during stationary phase and transformation to metacyclic forms (Sacks and da Silva, 1987).

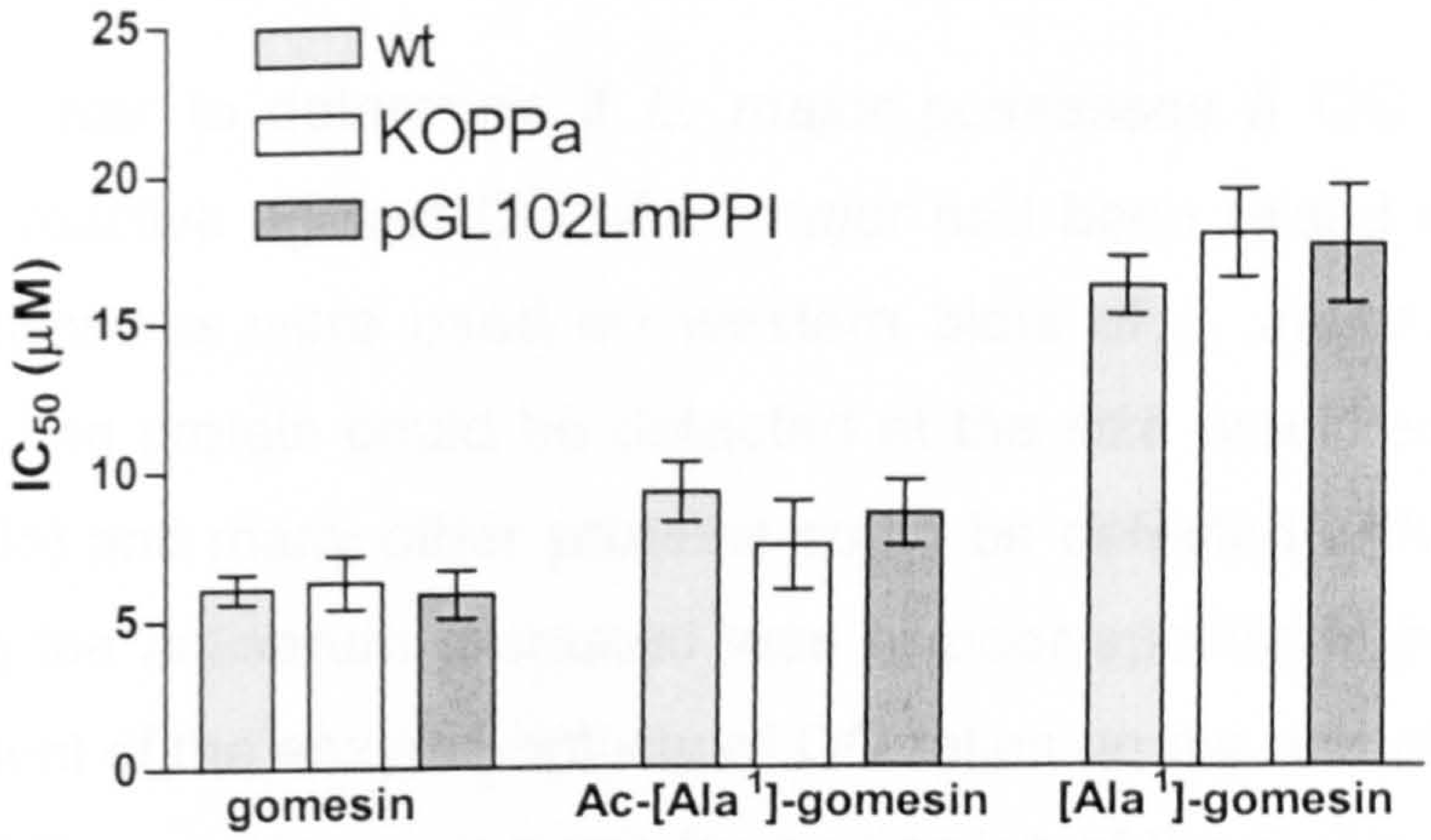


Fig. 5.17. **IC_{50} of gomesin and gomesin analogues on *L. major* cell lines.** 5 x 10^6 stationary phase *L. major* promastigotes (day 3 of stationary phase) were incubated with various concentrations of gomesin (Silva *et al.*, 2000) or gomesin analogues. After 1 h incubation at 25°C, MTT was added to a final concentration of 1 mg/ml. The samples were incubated for a further 1 h at 25°C, before the absorbance at 620 nm was read in a microtiter plate reader. wt, *L. major* wild type; pGL102LmPPI, *L. major* cell line over-expressing the LmPPI; KOPPa, *PPI* knock-out *L. major* cell line. [Ala^1]-gomesin, the N-terminal pyroglutamate of gomesin was replaced by an alanine; Ac-[Ala^1]-gomesin, the N-terminal pyroglutamate of gomesin was replaced by an acetylated alanine. Values are the means ± SD of 3 independent experiments. The IC_{50}s of each compound towards the different cell lines were not statistically different from each other ($P>0.4$), but the IC_{50}s of the 3 compounds towards the same cell lines were statistically different from each other ($P<0.025$).

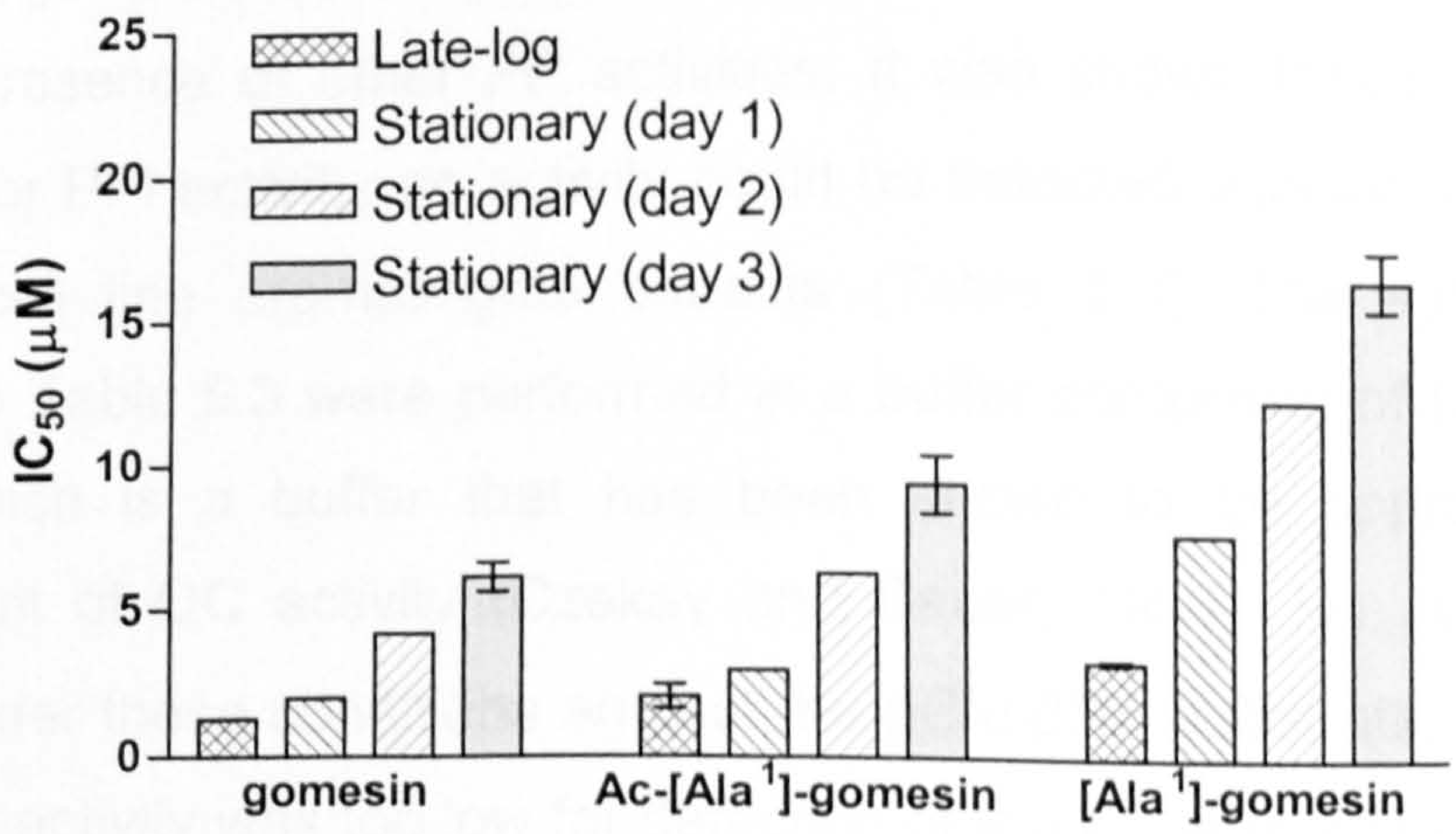


Fig. 5.18. **Dependence of the IC_{50} of gomesin on *L. major* culture phase.** 5 x 10^6 wild type late log or stationary phase *L. major* promastigotes were incubated with various concentrations of gomesin (Silva *et al.*, 2000) or gomesin analogues. After 1 h incubation at 25°C, MTT was added to a final concentration of 1 mg/ml. The samples were incubated for a further 1 h at 25°C, before the absorbance at 620 nm was read in a microtiter plate reader. [Ala^1]-gomesin, the N-terminal residue is an alanine; Ac-[Ala^1]-gomesin, the N-terminal residue is an acetylated alanine. Values are single measurements or means ± SD of 3 independent experiments.

5.2.5. Glutaminyl cyclase (QC) and PPII activities

In order to determine if *L. major* possesses a QC activity, polyclonal antibodies reactive against QC of *L. major* had been raised in rat (see Chapter 4). The antibodies were used on western blots of *L. major* promastigote cells lysates, but no protein could be detected at the size predicted for the full length QC (98 kDa) and many other proteins could be detected with similar intensities, suggesting the antiserum produced was of poor specificity (results not shown). Measurement of the enzyme activity of QC relies on the use of an indirect assay, using the QC as a primary enzyme for cyclisation of the H-Gln substrate, followed by cleavage of the cyclised substrate (pGlu) by PPI (Chikuma *et al.*, 2004; Schilling *et al.*, 2002). But as shown in Chapter 4, spontaneous cyclisation of H-Gln substrate does occur, making the measurement of any QC activity problematic. These fluorometric assays do not permit one to distinguish the signal generated from spontaneously generated pGlu substrate from that of the enzymatically produced pGlu substrate (Chikuma *et al.*, 2004). Precise quantification of pGlu formed from the H-Gln substrate in presence or absence of QC requires HPLC analysis, and this was not attempted. Thus, it is still unknown if *L. major* possesses its own QC activity and is able to produce pGlu-modified peptides.

The generation of knock-out cells lines for PPI provided a good means to study the presence of other PP activities. It was shown that under the assay conditions for PPI activity, no activity could be detected against pGlu-βNA in the knock-out cell line promastigote extracts (Table 5.3). The same assays as described in Table 5.3 were performed in a buffer composed of 50 mM Tris/HCl pH 8.0, which is a buffer that has been shown to be appropriate for the measurement of QC activity (Czekay and Bauer, 1993). No activity could be detected under these conditions against the pGlu-βNA substrate, suggesting the level of PPII activity was too low for detection or the enzyme was absent from the cells.

5.3. Discussion

The presence of a pyroglutamyl peptidase I (PPI) sequence in *L. major* is most intriguing, as one presumed role of PPI is to inactivate bioactive peptides, such as neuropeptides in human, which the parasite apparently lacks. However, this study has shown that the parasite indeed possesses such a PPI activity (Table 5.1) and that the PPI is the major, if not the only PP activity in *L. major* promastigotes. The anti-QC antiserum of *L. major* was found to be poorly specific, and the absence of an easy straightforward assay for the QC activity meant that I was unable to determine if *L. major* is indeed able to produce its own pGlu-modified peptides.

Over-expression of the PPI in *L. major* was successfully achieved using an episomal vector, and an increase of 5-9 fold in activity in the cell extract against the pGlu substrate could be detected in the cells over-expressing the active enzyme. The activity was exclusively associated with the soluble fraction of cell lysates, which would be consistent with the enzyme being cytosolic, and no PPI protein nor PPI activity could be detected in the supernatant of culture, being suggestive that the enzyme is not released. Immunolocalisation study of the PPI in *L. major* promastigotes showed the enzyme is cytosolic and is not found associated with a particular organelle in the cell (Fig. 5.16), which coincides with what was found for other PPIs in different organisms (Mantle *et al.*, 1991).

The over-expression of the active PPI in *L. major* promastigote did not seem to influence much the growth of the parasite, though lower cell density was reached at stationary stage (Fig. 5.4). A difference was detected in the ability of the parasite to transform to infective metacyclic promastigotes at stationary phase of growth (Fig 5.5), though the morphology of the metacyclics promastigotes produced was identical to that of wild type cells. The over-expression of the active PPI seemed to render the rate of production of metacyclic promastigotes very slow, whereas over-expression of the dead PPI and the empty episome resulted in cell lines with comparable metacyclogenesis rates. Both formed about 3-fold more metacyclic promastigotes by 10 days than did the cells over-expressing the active PPI. In terms of infectivity to macrophages *in vitro*, the cells over-expressing the active PPI showed a 2 times lower infectivity rate when total promastigotes were used, but no difference when purified metacyclic promastigotes were used (Fig. 5.6). The lower infection rate may be explained by the lower number of metacyclics present in the promastigote culture, and though the over-expressing line (for the active PPI) does not produce as many metacyclic promastigotes as the control lines, these metacyclic promastigotes are as infective for macrophages as metacyclics purified from the cells over-expressing the dead PPI or the empty episome. *In vivo*, the cells over-expressing the active PPI showed a 5 weeks delay in infection of mice footpads, whereas cells over-expressing the dead PPI showed the same pattern of infection as the cells over-expressing only the selection marker (Fig. 5.7). It was unclear if the delay in infection might correlate with the amount of PPI activity present in cells, which would indicate that over-expression of the PPI in an active form is detrimental to infection and that the lesion growth with this line was due to slow loss of the episome in the absence of the selection drug in the mice. An alternative possible explanation is that the number of metacyclic promastigotes injected differed, being significantly lower in the culture of cells over-expressing the active PPI, though the 5 week delay in infection observed is longer than what would be anticipated; a 2 week delay was observed when 5 x 10^4 as opposed to 5 x 10^5 promastigotes were injected (Susan Baillie, personal communication). So these results would suggest that to explain a 5 week delay the number of infective metacyclics promastigotes in the culture of cells over-expressing the active PPI at the time of infection would have to be more than 10-fold decreased compared to the other cultures. Based on the data in Fig. 5.5, this was not the case as the number of metacyclics promastigotes was at the most 4-fold decreased in the PPI over-expressing cell line compared to the other 2 lines in cultures at the same phase of growth, and cultures at the same stage of growth were used for mice infections. It would be very interesting to analyse the effect of infecting the same number of metacyclic promastigotes for the different cell lines in the mice.

The results *in vivo* differ slightly from the results *in vitro*, but this is not unexpected as with the *in vitro* studies only the ability to infect isolated macrophages is tested, whereas with the *in vivo* investigations, the host's immune response may have an influence on how the parasites infect (Liew *et al*., 1990; Alexander and Russell, 1992). As an example, the targeted gene disruption of the *CPB* gene array of *L. mexicana*, encoding for cathepsin L-like enzymes, resulted in a >80% decreased infectivity to macrophages compared to wild-type parasites, but lesions in mice, though smaller than those induced by wild type parasites, still occurred (Mottram *et al.*, 1996).

Two individual knock-out cell lines for the *PPI* were successfully produced (Fig. 5.9 and 5.10). The absence of PPI activity did not result in a difference in morphology or growth compared to the wild type cells (Fig. 5.11) and metacyclic promastigotes were formed at the same rate in stationary phase cultures of the knock-out line compared to the cultures of wild type parasites (Fig. 5.12). The maximum proportion of metacyclic promastigotes seemed to also depend on the time the parasites had been maintained in culture, the cells freshly transformed from amastigotes giving higher numbers of metacyclic promastigotes (in this case, wild type, and knock-out lines have been maintained in culture for 2 weeks, whereas the over-expression lines had been kept for 16 weeks). *In vitro* infection in macrophages showed both wild type and knock-out lines were generating the same levels of infection (Fig. 5.13). *In vivo* infection in mice using the same number of promastigotes (but about half the number of metacyclic promastigotes in the knock-out lines compared to the wild type) showed a delay in infection of about 3-4 weeks in the knock-out lines (Fig. 5.14). To test if this delay was due to the different number of metacyclic promastigotes infected or difference of infectivity of the metacyclic promastigotes, the same number of metacyclic promastigotes was injected for knock-out and wild type lines. In this case, the infection rates for all lines were comparable, indicating the metacyclic promastigotes produced in the *PPI* knock-out lines were as infective to mice as the wild type parasites (Fig. 5.15).

So it appears that the over-expression rather than the knock-out of the PPI activity is detrimental to the parasites, in terms of infectivity in mice. It is still unclear what may be the significance of the presence of such an enzyme in *L. major.*

Upon microbial infection, the mammalian host has at its disposal various mechanisms to kill the pathogen. The first barrier involves the innate immune system, machinery present in all other animals and also plants and mainly composed of peptides with antimicrobial activity. These peptides are cationic potent antibiotics of 15-45 amino acids, active against a wide range of organisms, including bacteria, fungi (Tang *et al.*, 2002), viruses, protozoa (Aley *et al.*, 1994; Lugli *et al.*, 2004), and tumour cells.

The peptides are classified in different families, depending on their structure. In human, there are 2 main families: the defensins, divided in the 2 classes α- and β-defensins, and the cathelicidins. Many of the β-defensins and the pro-mature human cathelicidin carry a N-terminal pGlu modification (Selsted *et al.*, 1993; Andersson *et al.*, 2002; Sahl *et al.*, 2004). Defensins are peptides with 3 disulfide bonds and a flat dimeric β-sheet structure, with a difference in disulfide bridges pattern between the α- and β-class, whereas cathelicidins are linear peptides free of cysteines and with a α-helical and amphipatic solution structure. Other antimicrobial peptides that occur, but are less widespread, have a bias in certain amino acids, like proline, arginine, tryptophane or histidine (Boman, 2003). α-defensins are produced by neutrophils in a constitutive manner, while β-defensins are inducible and produced *in neutrophils*, macrophages and epithelial cells, where they act as anti-microbial agents (Hoover *et al.*, 2001). The disulfide bridges are essential to the activity (Torres and Kuchel, 2004) and their mechanism of action relies initially on electrostatic interaction of the cationic poles of the peptide with negatively charged moieties on the target cell surface; the peptide is then inserted into the phospholipid bilayer, and diffusion of the peptides in the membrane eventually promotes the formation of pores, provoking cell lysis (Boman, 2003). Cathelicidin production is induced in the epithelium following epithelial wounding and infectious challenge (Schaller-Bals, *et al.*, 2002). In human, the only cathelicidin, hCAP-18, is produced as a 19 kDa pre-pro-peptide, with a conserved cathelin-like sequence in the pro-domain and a variable antimicrobial C-terminal sequence (Gennaro *et al.*, 1998). Cathelin is a 12 kDa peptide isolated from porcine neutrophils, that shares similarities with the cystatin superfamily of cysteine peptidases inhibitors (Scocchi *et al.*, 1997). Two distinct activities have been attributed to the pro-domain; a role of inhibition of bacterial growth, and a role of inhibition of cysteine peptidase-mediated tissue damage (Zaiou *et al.*, 2003). The C-terminal fragment, namely LL-37, is thought to provoke cell lysis using a similar mechanism to the defensins (Sahl *et al.*, 2004). The pro-mature precursor is targeted to granules for storage, and upon activation, the content of the granules is released or delivered to phagolysosomes (Sorensen *et al.*, 2001). But hCAP-18 is not processed to the active LL-37 form in the phagolysosomes. The main function of hCAP-18 seems to be extracellular, where LL-37 is formed and acts as a chemotactic agent for neutrophils, monocytes and T cells; LL-37 seems therefore to be an important mediator of the extracellular immune activity (Sorensen *et al.*, 2001).

So how might the PPI activity in *L. major* act on these compounds? First, as the activity of the defensins is mainly due to the arrangement of disulfide bridges, and for LL-37 to the amphipatic properties, it is unlikely the pGlu modification at the N-terminus of some of these compounds might play a role in the activity, but rather in the stabilisation of these compounds. Furthermore, the LL-37 compound is most probably not active in the phagolysosomes, but present as the inactive hCAP-18 precursor.

On hypothesis is that the PPI might protect *L. major* from pGlu-modified natural peptides from the host, by either cleaving off the pGlu residue essential for the activity of the peptide, or destabilising the peptides by making them prone to exopeptidase degradation, one would expect an increase in the infectivity when the active PPI is over-expressed, which is not the case here. In the case of hCAP-18, the cleavage of the N-terminal pGlu might destabilise the protein and make it more likely to be processed to its active LL-37 form in the phagolysosomes. Although release of PPI activity from amastigote cells was not tested, the PPI enzyme did not seem to be released from the parasite cell in promastigote cultures, and would therefore have to act on internalised compounds. PPI seems to be the main, if not the only, PP activity present in *L. major* but the expression of other PP activities, such as PPII, in the amastigote stage cannot be excluded.

A way by which *Leishmania* has been shown to evade the immune response from the host is to promote phagocytosis by macrophages and so avoid humoral immunoglobulin mediated response. The presence of lipophosphoglycans (LPS) on the parasite cell surface makes it rather resistant to acid hydrolases released from the lysosomes upon fusion with the phagosome (Desjardins, 2003). Resistance to *Leishmania* has been shown to rely on the development of CD4+ Th1 cells, which promote an effective cellular mediated immune response, by increasing the release of the cytokine IFN-γ, which promotes macrophage activation and parasite killing (Jones *et al.*, 1998). On the contrary, susceptibility to *Leishmania* infection is characterised by a mainly Th2 response, promoting the release of IL-4, which stimulates the production of immunoglobulins, ineffective against phagocytosed parasites. Adaptive immunity relies on the degradation of the pathogen proteins into peptides, which can be presented on the surface on the macrophage as antigens in conjunction with MHC-I or –II molecules, and trigger a cell mediated immune response. It is unclear though why the over-expression of the active PPI enzyme may lead to decreased infectivity compared to over-expression of the inactive PPI, and the results of the experiments *in vivo* cannot therefore be clearly explained by an increase in antigenicity when the enzyme is over-expressed in the parasite in its active form, unless major structural perturbation has occurred in the mutant enzyme. Furthermore, the PPI enzyme is not secreted, and immune response would have to be triggered by released PPI from dead cells. It would be interesting to determine if indeed there is a difference in the immune response generated with the different cell lines in the mice.

Another hypothesis is that the over-expression of the active PPI may perturbate some essential functions in the parasite, such as formation of infective metacyclic promastigotes and infectivity of these cells. The presence of a glutaminyl cyclase could not be determined in the promastigote form of *L. major*, due to lack of assay for the enzyme. The gene seemed to be expressed, but no evidence of the expression of the protein could be found. Again it cannot be ruled out that the cyclase might be expressed in the promastigote form or/and in the amastigote form. As the over-expression of active PPI seemed to impair the formation of metacyclic promastigotes, important regulatory proteins, which might carry a pGlu modification may be excessively degraded and perturb the normal progression to metacyclic form. Also, there might be a perturbation of the signals involved in the modulation of the immune response in the host; in the hypothesis pGlu-modified peptides might be involved. So the over-expression of the active PPI might influence the composition of pGlu modified peptides in the parasite more than in the hosts, hence the relevance of trying to determine what are the natural substrates of PPI of *L. major*, and if indeed they might play a role in the parasite's infectivity. Mass-spectrometry approaches, involving the complete N-terminal degradation of whole cells peptide extracts from wild-type and PPI knock-out lines, should allow the isolation of the totality of N-terminal modified peptides, and the subsequent unblocking using the PPI would allow selection of all modified peptides except the ones carrying the N-terminal pGlu. By proteomic analysis on 2-dimensions (2-D) gels of the different treated extracts, N-terminally pGlu-modified proteins might be isolated and identified. This would identify the natural substrates of PPI in the parasites, and if the parasites are indeed able to generate pGlu-modified peptides on their own.

Another possible explanation might be that the over-expression of the active PPI may perturb the parasite cell in a non-specific way, as a result of excessive processing by the PPI. A complete analysis of the substrate specificity may give an insight on what other residues might be susceptible to cleavage by the enzyme, especially in conditions where the enzyme is used in excess. The over-expression of the active enzyme may have some detrimental effect on the cell, not only specific to the cleavage of pGlu residues on the proteins, but also to other possible non-specific cleavage of proteins.

Could it be that the enzyme might have a more important role in the insect vector? The genome of the *Plasmodium* vector, the mosquito *Anopheles gambia*, has been extensively searched for the presence of anti-microbial peptides. Four defensins and 4 cecropin genes have been shown to be induced in the insect vector by a malaria infection (Christophides *et al.*, 2002). But none of these compounds carry a pGlu modification at the N-terminus. A defensin-like antimicrobial peptide was isolated from the hemolymph of the sandfly *Phlebotomus duboscqi*, a vector of *Leishmania*. The peptide showed anti-parasite properties on *L. major*, and its expression was induced by *L. major* infection in the insect (Boulanger *et al.*, 2004). The peptide was synthesised as a pre-pro-peptide, the sequence of the pro-peptide starting with a glutamic acid, which could potentially be cyclised to pGlu. Gomesin, the antimicrobial peptide from the spider *Acanthoscurria gomesiana*, which showed leishmanicidal properties, is an 18-residue cysteine-rich peptide, carrying 2 disulfide bridges and a N-terminal pGlu (Silva *et al.*, 2000). The IC_{50} of gomesin towards wild type, PPI over-expression or PPI knock-out *L. major* cell lines were all similar, suggesting the over-expression of PPI did not have a favourable effect on parasite survival in the conditions tested *in vitro* (Fig. 5.17). This suggests a different compartmentation between the site of action of gomesin and the localisation of PPI in *L. major* promastigotes. This is consistent with the fact that cysteine-rich antimicrobial peptides promote the formation of pores into the parasite's membrane, provoking cell lysis (Boman, 2003), and that PPI is a cytosolic enzyme that is not released by *L. major* promastigotes. The N-terminal blocking of gomesin was not essential to the activity of the peptide, but the unblocked [Ala^1]-gomesin showed reduced toxicity (3-fold reduction) towards all *L. major* cell lines tested (Fig. 5.17). The replacement of the N-terminal pGlu by an acetyl-alanine also decreased the toxicity of gomesin (about 1.2-1.5-fold, Fig. 5.17). So it seems that the N-terminal blocking of gomesin is not essential to the peptide toxicity, and that the toxic activity of gomesin mainly relies upon features that are independent from this N-terminal blocking. However, the N-terminal pGlu modification, or N-terminal blocking by acetylation, is favourable to the leishmanicidal activity of gomesin, probably by increasing the peptide's stability. The leishmanial PPI might play a role in the protection against pGlu-modified peptides in the sandfly vector, provided PPI can act on these peptides. The study of the propagation of PPI knock-out and over-expressing cell lines in the insect vector would be of high interest.

So in conclusion, no phenotype could be observed in the PPI knock-out cell lines, but an impairment in the production of metacyclic promastigotes and infectivity *in vivo* could be observed with the active PPI over-expressing lines. What significance could this have in terms of therapy? The over-expression of PPI is not likely to occur in the field, nor can the induction of such over-expression be easily achieved. Many naturally occurring anti-microbial peptides, in mammals and insect, carry a pGlu modification at the N-terminus, but also an amidation at the C-terminus (Silva *et al.*, 2000), which is not thought to play a role in the activity of the peptides. Post-transcriptional modifications are more likely to stabilise the compounds. Whether the cleavage of such modification in the wild may impair their efficiency in a significant way, is unknown. The mechanism by which an intracellular enzyme, not secreted, may influence the effect of the extracellular peptides, thought to provoke membrane destabilisation of the microbial agent, is unclear. The over-expression of the active PPI might influence the composition of pGlu-modified peptides in the parasite more than in the hosts, provided *L. major* is able to catalyse the N-terminal glutamic acid cyclisation. Thus, a gene knock-out of the glutaminyl cyclase might be more informative. If indeed pGlu modified peptides are among regulatory peptides that participate in metacyclogenesis or infectivity *in vivo*, the absence of cyclisation may have some detrimental effect on the parasite. Very low sequence similarity was detected between the putative glutaminyl cyclase of *L. major* and the human enzyme (11%), which could potentially make the design of specific inhibitors to the *L. major* glutaminyl cyclase possible. This may be more realistic in terms of therapy. But no reports has been made previously on pyroglutamyl peptidases in parasites, and no biological function has been attributed to the PP in bacteria yet, though they have been known for almost 30 years (Exterkate, 1977). Hopefully the years to come will bring more knowledge in this field and provide insight into the roles of PPs.

References

A

Abrahamson, M., Alvarez-Fernandez, M., & Nathanson, C. M. (2003). Cystatins. *Biochem. Soc. Symp.* 179-199.

Adams, J. H. & Bushell, G. R. (1988). The effect of protease inhibitors on *Eimeria vermiformis* invasion of cultured cells. *Int. J. Parasitol.* **18:** 683-685.

Alexander, J. & Russell, D. G. (1992). The interaction of *Leishmania* species with macrophages. *Adv. Parasitol.* **31:** 175-254.

Alexander, J., Coombs, G. H., & Mottram, J. C. (1998). *Leishmania mexicana* Cysteine Proteinase-Deficient Mutants Have Attenuated Virulence for Mice and Potentiate a Th1 Response. *J. Immuno.l* **161:** 6794-6801.

Aley, S. B., Zimmerman, M., Hetsko, M., Selsted, M. E., & Gillin, F. D. (1994). Killing of *Giardia lamblia* by cryptdins and cationic neutrophil peptides. *Infect. Immun.* **62:** 5397-5403.

Allocco, J. J., Profous-Juchelka, H., Myers, R. W., Nare, B., & Schmatz, D. M. (1999). Biosynthesis and catabolism of mannitol is developmentally regulated in the protozoan parasite *Eimeria tenella*. *J. Parasitol.* **85:** 167-173.

Alvar, J., Canavate, C., Molina, R., Moreno, J., & Nieto, J. (2004). Canine leishmaniasis. *Adv. Parasitol.* **57:** 1-88.

Alvarez, F., Robello, C., & Vignali, M. (1994). Evolution of codon usage and base contents in kinetoplastid protozoans. *Mol. Biol. Evol.* **11:** 790-802.

Alves, L. C., Melo, R. L., Cezari, M. H., Sanderson, S. J., Mottram, J. C., Coombs, G. H., Juliano, L., & Juliano, M. A. (2001). Analysis of the S(2) subsite specificities of the recombinant cysteine proteinases CPB of *Leishmania mexicana*, and cruzain of *Trypanosoma cruzi*, using fluorescent substrates containing non-natural basic amino acids. *Mol. Biochem. Parasitol.* **117:** 137-143.

Amato, V., Amato, J., Nicodemo, A., Uip, D., Amato-Neto, V., & Duarte, M. (1998). Treatment of mucocutaneous leishmaniasis with pentamidine isothionate. *Ann. Dermatol. Venereol.* **125:** 492-495.

Ambroise-Thomas, P. (2001). Parasitic diseases and immunodeficiencies. *Parasitology* **122 Suppl:** S65-S71.

Andersson, E., Sorensen, O. E., Frohm, B., Borregaard, N., Egesten, A., & Malm, J. (2002). Isolation of human cationic antimicrobial protein-18 from seminal plasma and its association with prostasomes. *Hum. Reprod.* **17:** 2529-2534.

Aparicio, I. M., Scharfstein, J., & Lima, A. P. (2004). A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infect. Immun.* **72:** 5892-5902.

Arnoult, D., Akarid, K., Grodet, A., Petit, P. X., Estaquier, J., & Ameisen, J. C. (2002). On the evolution of programmed cell death: apoptosis of the unicellular eukaryote *Leishmania major* involves cysteine proteinase activation and mitochondrion permeabilization. *Cell Death. Differ.* **9:** 65-81.

Augustine, P. C., Smith, C. K., Danforth, H. D., & Ruff, M. D. (1987). Effect of ionophorous anticoccidials on invasion and development of *Eimeria*: comparison of sensitive and resistant isolates and correlation with drug uptake. *Poult. Sci.* **66:** 960-965.

Authie, E., Muteti, D. K., Mbawa, Z. R., Lonsdale-Eccles, J. D., Webster, P., & Wells, C. W. (1992). Identification of a 33-kilodalton immunodominant antigen of *Trypanosoma congolense* as a cysteine protease. *Mol. Biochem. Parasitol* **56:** 103-116.

Authie, E., Duvallet, G., Robertson, C., & Williams, D. J. (1993). Antibody responses to a 33 kDa cysteine protease of *Trypanosoma congolense*: relationship to 'trypanotolerance' in cattle. *Parasite Immunol.* **15:** 465-474.

Authie, E., Boulange, A., Muteti, D., Lalmanach, G., Gauthier, F., & Musoke, A. J. (2001). Immunisation of cattle with cysteine proteinases of *Trypanosoma congolense*: targetting the disease rather than the parasite. *Int. J. Parasitol.* **31:** 1429-1433.

Awade, A., Cleuziat, P., Gonzales, T., & Robert-Baudouy, J. (1992)a. Characterization of the pcp gene encoding the pyrrolidone carboxyl peptidase of *Bacillus subtilis*. *FEBS Lett.* **305:** 67-73.

Awade, A., Gonzales, T., Cleuziat, P., & Robert-Baudouy, J. (1992)b. One step purification and characterization of the pyrrolidone carboxyl peptidase of *Streptococcus pyrogenes* over-expressed in *Escherichia coli*. *FEBS Lett.* **308:** 70-74.

B

Barinka, C., Sacha, P., Sklenar, J., Man, P., Bezouska, K., Slusher, B. S., & Konvalinka, J. (2004). Identification of the N-glycosylation sites on glutamate carboxypeptidase II necessary for proteolytic activity. *Protein Sci.* **13:** 1627-1635.

Barrett, A. J. & McDonald, J. K. (1985). Nomenclature: a possible solution to the 'peptidase anomaly'. *Biochem. J.* **231:** 807.

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

Barrett, A. J. & Rawlings, N. D. (2001). Evolutionary lines of cysteine peptidases. *Biol. Chem.* **382:** 727-733.

Bart, G., Frame, M. J., Carter, R., Coombs, G. H., & Mottram, J. C. (1997). Cathepsin B-like cysteine proteinase-deficient mutants of *Leishmania mexicana. Mol. Biochem. Parasitol.* **88:** 53-61.

Basselin, M., Denise, H., Coombs, G. H., & Barrett, M. P. (2002). Resistance to pentamidine in *Leishmania mexicana* involves exclusion of the drug from the mitochondrion. *Antimicrob. Agents Chemother.* **46:** 3731-3738.

Bateman, R. C., Jr., Temple, J. S., Misquitta, S. A., & Booth, R. E. (2001). Evidence for essential histidines in human pituitary glutaminyl cyclase. *Biochemistry* **40:** 11246-11250.

Bauer, K. & Nowak, P. (1979). Characterization of a thyroliberin-degrading serum enzyme catalyzing the hydrolysis of thyroliberin at the pyroglutamyl-histidine bond. *Eur. J. Biochem* **99:** 239-246.

Bauer, K., Carmeliet, P., Schulz, M., Baes, M., & Denef, C. (1990). Regulation and cellular localization of the membrane-bound thyrotropin-releasing hormone-degrading enzyme in primary cultures of neuronal, glial and adenohypophyseal cells. *Endocrinology* **127:** 1224-1233.

Bauer, K. (1994). Purification and characterization of the thyrotropin-releasing-hormone-degrading ectoenzyme. *Eur. J. Biochem.* **224:** 387-396.

Belli, S. I., Wallach, M. G., & Smith, N. C. (2003)a. Cloning and characterization of the 82 kDa tyrosine-rich sexual stage glycoprotein, GAM82, and its role in oocyst wall formation in the apicomplexan parasite, *Eimeria maxima. Gene* **307:** 201-212.

Belli, S. I., Wallach, M. G., Luxford, C., Davies, M. J., & Smith, N. C. (2003)b. Roles of tyrosine-rich precursor glycoproteins and dityrosine- and 3,4-dihydroxyphenylalanine-mediated protein cross-linking in development of the oocyst wall in the coccidian parasite *Eimeria maxima. Eukaryot. Cell* **2:** 456-464.

Benne, R., Van den, B. J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H., & Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46:** 819-826.

Boman, H. G. (2003). Antibacterial peptides: basic facts and emerging concepts. *J. Intern. Med.* **254:** 197-215.

Booth, R. E., Lovell, S. C., Misquitta, S. A., & Bateman, R. C., Jr. (2004). Human glutaminyl cyclase and bacterial zinc aminopeptidase share a common fold and active site. *BMC. Biol.* **2:** 2.

Boulange, A., Serveau, C., Brillard, M., Minet, C., Gauthier, F., Diallo, A., Lalmanach, G., & Authie, E. (2001). Functional expression of the catalytic domains of two cysteine proteinases from *Trypanosoma congolense.* *Int. J. Parasitol.* **31:** 1435-1440.

Boulanger, N., Lowenberger, C., Volf, P., Ursic, R., Sigutova, L., Sabatier, L., Svobodova, M., Beverley, S. M., Spath, G., Brun, R., Pesson, B., & Bulet, P. (2004). Characterization of a defensin from the sand fly *Phlebotomus duboscqi* induced by challenge with bacteria or the protozoan parasite *Leishmania major. Infect. Immun.* **72:** 7140-7146.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72:** 248-254.

Bromley, E., Leeds, N., Clark, J., McGregor, E., Ward, M., Dunn, M. J., & Tomley, F. (2003). Defining the protein repertoire of microneme secretory organelles in the apicomplexan parasite *Eimeria tenella. Proteomics.* **3:** 1553-1561.

Bromme, D. & Kaleta, J. (2002). Thiol-dependent cathepsins: pathophysiological implications and recent advances in inhibitor design. *Curr. Pharm. Des* **8:** 1639-1658.

Brooks, D. R., Tetley, L., Coombs, G. H., & Mottram, J. C. (2000). Processing and trafficking of cysteine proteases in *Leishmania mexicana*. *J. Cell Sci.* **113:** 4035-4041.

Browne, P. & O'Cuinn, G. (1983). An evaluation of the role of a pyroglutamyl peptidase, a post-proline cleaving enzyme and a post-proline dipeptidyl amino peptidase, each purified from the soluble fraction of guinea-pig brain, in the degradation of thyroliberin in vitro. *Eur. J. Biochem.* **137:** 75-87.

Busby, W. H., Jr., Quackenbush, G. E., Humm, J., Youngblood, W. W., & Kizer, J. S. (1987). An enzyme(s) that converts glutaminyl-peptides into pyroglutamyl-peptides. Presence in pituitary, brain, adrenal medulla, and lymphocytes. *J. Biol. Chem.* **262:** 8532-8536.

Buxbaum, L. U., Denise, H., Coombs, G. H., Alexander, J., Mottram, J. C., & Scott, P. (2003). Cysteine protease B of *Leishmania mexicana* inhibits host Th1 responses and protective immunity. *J. Immunol.* **171:** 3711-3717.

C

Caffrey, C. R., Salter, J. P., Lucas, K. D., Khiem, D., Hsieh, I., Lim, K. C., Ruppel, A., McKerrow, J. H., & Sajid, M. (2002). SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **121:** 49-61.

Capecchi, J. T. & Loudon, G. M. (1985). Substrate specificity of pyroglutamyl aminopeptidase. *J. Med. Chem.* **28:** 140-143.

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

Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., & Menard, R. (1996). Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry* **35:** 8149-8157.

Carruthers, V. B., Sherman, G. D., & Sibley, L. D. (2000). The *Toxoplasma* adhesive protein MIC2 is proteolytically processed at multiple sites by two parasite-derived proteases. *J. Biol. Chem.* **275:** 14346-14353.

Chagas, J. R., Authie, E., Serveau, C., Lalmanach, G., Juliano, L., & Gauthier, F. (1997). A comparison of the enzymatic properties of the major cysteine proteinases from *Trypanosoma congolense* and *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **88:** 85-94.

Chan, S. J., San Segundo, B., McCormick, M. B., & Steiner, D. F. (1986). Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proc. Natl. Acad Sci. U. S. A.* **83:** 7721-7725.

Chan, V. J., Selzer, P. M., McKerrow, J. H., & Sakanari, J. A. (1999). Expression and alteration of the S2 subsite of the *Leishmania major* cathepsin B-like cysteine protease. *Biochem. J.* **340:** 113-117.

Chapman, H. A., Riese, R. J., & Shi, G. P. (1997). Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.* **59:** 63-88.

Chapman, H. D., Cherry, T. E., Danforth, H. D., Richards, G., Shirley, M. W., & Williams, R. B. (2002). Sustainable coccidiosis control in poultry production: the role of live vaccines. *Int. J. Parasitol.* **32:** 617-629.

Charli, J. L., Mendez, M., Vargas, M. A., Cisneros, M., Assai, M., Joseph-Bravo, P., & Wilk, S. (1989). Pyroglutamyl peptidase II inhibition specifically increases recovery of TRH released from rat brain slices. *Neuropeptides* **14:** 191-196.

Chibale, K. & Musonda, C. C. (2003). The synthesis of parasitic cysteine protease and trypanothione reductase inhibitors. *Curr Med. Chem.* **10:** 1863-1889.

Chikuma, T., Taguchi, K., Yamaguchi, M., Hojo, H., & Kato, T. (2004). Improved determination of bovine glutaminyl cyclase activity using precolumn derivatization and reversed-phase high-performance liquid chromatography with ultraviolet detection. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **806:** 113-118.

Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P. T., Collins, F. H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N. T., Hoffmann, J. A., Kanzok, S. M., Letunic, I., Levashina, E. A., Loukeris, T. G., Lycett, G., Meister, S., Michel, K., Moita, L. F., Muller, H. M., Osta, M. A., Paskewitz, S. M., Reichhart, J. M., Rzhetsky, A., Troxler, L., Vernick, K. D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., & Kafatos, F. C. (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298:** 159-165.

Cociancich, S., Goyffon, M., Bontems, F., Bulet, P., Bouet, F., Menez, A., & Hoffmann, J. (1993). Purification and characterization of a scorpion defensin, a 4kDa antibacterial peptide presenting structural similarities with insect defensins and scorpion toxins. *Biochem Biophys. Res. Commun.* **194:** 17-22.

Consalvo, A. P., Young, S. D., Jones, B. N., & Tamburini, P. P. (1988). A rapid fluorometric assay for N-terminal glutaminyl cyclase activity using high-performance liquid chromatography. *Anal. Biochem.* **175:** 131-138.

Coombs G.H. & Mottram J.C. (1997). Parasite proteinases and amino acid metabolism: possibilities for chemotherapeutic exploitation. *Parasitology* **114:** S61-80.

Coombs, G. H., Denton, H., Brown, S. M., & Thong, K. W. (1997). Biochemistry of the coccidia. *Adv. Parasitol.* **39:** 141-226.

Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J., & Mottram, J. C. (2001). Aspartic proteases *of Plasmodium falciparum* and other parasitic protozoa as drug targets. *Trends Parasitol.* **17:** 532-537.

Coombs, G. H., Tetley, L., Moss, V. A., & Vickerman, K. (1986). Three dimensional structure of the *Leishmania* amastigote as revealed by computer-aided reconstruction from serial sections. *Parasitology* **92:** 13-23.

Cornet, B., Bonmatin, J. M., Hetru, C., Hoffmann, J. A., Ptak, M., & Vovelle, F. (1995). Refined three-dimensional solution structure of insect defensin A. *Structure.* **3:** 435-448.

Croft, S. L. & Coombs, G. H. (2003). Leishmaniasis--current chemotherapy and recent advances in the search for novel drugs. *Trends Parasitol.* **19:** 502-508.

Croft, S. L., Yardley, V., & Kendrick, H. (2002). Drug sensitivity of *Leishmania* species: some unresolved problems. *Trans. R. Soc. Trop. Med. Hyg.* **96:** S127-S129.

Cruz, A., Coburn, C., M., Beverley, S. M. (1991)a. Double targeted gene replacement for creating null mutants. *Proc. Natl. Acad. Sci. U. S. A.* **88:** 7170–7174.

Cruz, C., Charli, J. L., Vargas, M. A., & Joseph-Bravo, P. (1991)b. Neuronal localization of pyroglutamate aminopeptidase II in primary cultures of fetal mouse brain. *J. Neurochem.* **56:** 1594-1601.

Cummins, P. M. & O'Connor, B. (1996). Bovine brain pyroglutamyl aminopeptidase (type-1): purification and characterisation of a neuropeptide-inactivating peptidase. *Int. J. Biochem. Cell Biol.* **28:** 883-893.

Cummins, P. M. & O'Connor, B. (1998). Pyroglutamyl peptidase: an overview of the three known enzymatic forms. *Biochim. Biophys. Acta* **1429:** 1-17.

Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., & Mort, J. S. (1996). Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure.* **4:** 405-416.

Czekay, G., & Bauer, K. (1993). Identification of the thyrotropin-releasing-hormone-degrading ectoenzyme as a metallopeptidase. *Biochem J.* **290:** 921-926.

D

Da Silva, R. & Sacks, D. L. (1987). Metacyclogenesis is a major determinant of *Leishmania* promastigote virulence and attenuation. *Infect. Immun.* **55:** 2802-2806.

Dando, P. M., Fortunato, M., Strand, G. B., Smith, T. S., & Barrett, A. J. (2003). Pyroglutamyl-peptidase I: cloning, sequencing, and characterisation of the recombinant human enzyme. *Protein Expr. Purif.* **28:** 111-119.

Del Nery, E., Juliano, M. A., Lima, A. P., Scharfstein, J., & Juliano, L. (1997). Kininogenase activity by the major cysteinyl proteinase (cruzipain) from *Trypanosoma cruzi. J. Biol. Chem.* **272:** 25713-25718.

Denton, H., Thong, K. W., & Coombs, G. H. (1994). *Eimeria tenella* contains a pyrophosphate-dependent phosphofructokinase and a pyruvate kinase with unusual allosteric regulators. *FEMS Microbiol. Lett.* **115:** 87-91.

Desjardins, M. (2003). ER-mediated phagocytosis: a new membrane for new functions. *Nat. Rev. Immunol.* **3:** 280-291.

Desjeux, P. (2004). Leishmaniasis: current situation and new perspectives. *Comp. Immunol. Microbiol. Infect. Dis.* **27:** 305-318.

Destoumieux, D., Bulet, P., Loew, D., Van Dorsselaer, A., Rodriguez, J., & Bachere, E. (1997). Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). *J. Biol. Chem.* **272:** 28398-28406.

Dickinson, D. P. (2002). Salivary (SD-type) cystatins: over one billion years in the making--but to what purpose? *Crit. Rev. Oral Biol. Med.* **13:** 485-508.

Doolittle, R. F. & Armentrout, R. W. (1968). Pyrrolidonyl peptidase. An enzyme for selective removal of pyrrolidonecarboxylic acid residues from polypeptides. *Biochemistry* **7:** 516-521.

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

Dubremetz, J. F. (1998). Host cell invasion by *Toxoplasma gondii. Trends Microbiol.* **6:** 27-30.

Duffy, M. J. (1992). The role of proteolytic enzymes in cancer invasion and metastasis. *Clin. Exp. Metastasis* **10:** 145-155.

Dutton, C. J., Banks, B. J., & Cooper, C. B. (1995). Polyether ionophores. *Nat. Prod. Rep.* **12:** 165-181.

E

Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., & Craik, C. S. (1992). The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi. J. Biol. Chem.* **267:** 7411-7420.

Eksi, S., Czesny, B., Greenbaum, D. C., Bogyo, M., & Williamson, K. C. (2004). Targeted disruption of *Plasmodium falciparum* cysteine protease, falcipain 1, reduces oocyst production, not erythrocytic stage growth. *Mol. Microbiol.* **53:** 243-250.

Ellis, J. & Revets, H. (1990). *Eimeria* species which infect the chicken contain virus-like RNA molecules. *Parasitology* **101:** 163-169.

Ellis, J. T., Morrison, D. A., Avery, D., & Johnson, A. M. (1994). Codon usage and bias among individual genes of the coccidia and piroplasms. *Parasitology* **109:** 265-272.

Ellis, M., Sharma, D. K., Hilley, J. D., Coombs, G. H., & Mottram, J. C. (2002). Processing and trafficking of *Leishmania mexicana* GP63. Analysis using GP18 mutants deficient in glycosylphosphatidylinositol protein anchoring. *J. Biol. Chem.* **277:** 27968-27974.

Elmore, M. A., Griffiths, E. C., O'Connor, B., & O'Cuinn, G. (1990). Further characterization of the substrate specificity of a TRH hydrolysing pyroglutamate aminopeptidase from guinea-pig brain. *Neuropeptides* **15:** 31-36.

Entzeroth, R., Mattig, F. R., & Werner-Meier, R. (1998). Structure and function of the parasitophorous vacuole in *Eimeria* species. *Int. J. Parasitol.* **28:** 1015-1018.

Exterkate, F. A. (1977). Pyrrolidone carboxylyl peptidase in *Streptococcus cremoris*: dependence on an interaction with membrane components. *J. Bacteriol.* **129:** 1281-1288.

F

Fairlamb, A. H. & Cerami, A. (1992). Metabolism and functions of trypanothione in the Kinetoplastida. *Annu. Rev. Microbiol.* **46:** 695-729.

Ferguson, D. J., Brecht, S., & Soldati, D. (2000). The microneme protein MIC4, or an MIC4-like protein, is expressed within the macrogamete and associated with oocyst wall formation in *Toxoplasma gondii. Int. J. Parasitol.* **30:** 1203-1209.

Ferguson, D. J., Belli, S. I., Smith, N. C., & Wallach, M. G. (2003). The development of the macrogamete and oocyst wall in *Eimeria maxima*: immuno-light and electron microscopy. *Int. J. Parasitol.* **33:** 1329-1340.

Fish, W. R., Nkhungulu, Z. M., Muriuki, C. W., Ndegwa, D. M., Lonsdale-Eccles, J. D., & Steyaert, J. (1995). Primary structure and partial characterization of a life-cycle-regulated cysteine protease from *Trypanosoma* (Nannomonas) *congolense. Gene* **161:** 125-128.

Forney, J. R., Yang, S., & Healey, M. C. (1996). Protease activity associated with excystation of *Cryptosporidium parvum* oocysts. *J. Parasitol.* **82:** 889-892.

Frame, M. J., Mottram, J. C., & Coombs, G. H. (2000). Analysis of the roles of cysteine proteinases of *Leishmania mexicana* in the host-parasite interaction. *Parasitology* **121:** 367-377.

Friedman, M., Grosjean, O. K., & Zahnley, J. C. (1986). Inactivation of metalloenzymes by lysinoalanine, phenylethylaminoalanine, alkali-treated food proteins, and sulfur amino acids. *Adv. Exp. Med. Biol.* **199:** 531-560.

Fujii, N., Mallari, J. P., Hansell, E. J., Mackey, Z., Doyle, P., Zhou, Y. M., Gut, J., Rosenthal, P. J., McKerrow, J. H., & Guy, R. K. (2005). Discovery of potent thiosemicarbazone inhibitors of rhodesain and cruzain. *Bioorg. Med. Chem. Lett.* **15:** 121-123.

Fujiwara, K., Kobayashi, R., & Tsuru, D. (1979). The substrate specificity of pyrrolidone carboxylyl peptidase from *Bacillus amyloliquefaciens. Biochim. Biophys. Acta* **570:** 140-148.

Fujiwara, K., Matsumoto, E., Kitagawa, T., & Tsuru, D. (1982). N-alpha-carbobenzoxy pyroglutamyl diazomethyl ketone as active-site-directed inhibitor for pyroglutamyl peptidase. *Biochim. Biophys. Acta* **702:** 149-154.

Fuller, A. L. & McDougald, L. R. (1990). Reduction in cell entry of *Eimeria tenella* (Coccidia) sporozoites by protease inhibitors, and partial characterization of proteolytic activity associated with intact sporozoites and merozoites. *J. Parasitol.* **76:** 464-467.

G

Gallagher, S. P., O'Leary, R. M., & O'Connor, B. (1997). The development of two fluorimetric assays for the determination of pyroglutamyl aminopeptidase type-II activity. *Anal. Biochem.* **250:** 1-9.

Ganz, T. & Lehrer, R. I. (1999). Antibiotic peptides from higher eukaryotes: biology and applications. *Mol. Med. Today* **5:** 292-297.

Gelb, B. D., Shi, G. P., Chapman, H. A., & Desnick, R. J. (1996). Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. *Science* **273:** 1236-1238.

Gennaro, R., Scocchi, M., Merluzzi, L., & Zanetti, M. (1998). Biological characterization of a novel mammalian antimicrobial peptide. *Biochim. Biophys. Acta* **1425:** 361-368.

Gibson, W., Bingle, L., Blendeman, W., Brown, J., Wood, J., & Stevens, J. (2000). Structure and sequence variation of the trypanosome spliced leader transcript. *Mol. Biochem. Parasitol.* **107:** 269-277.

Gluzman, I. Y., Francis, S. E., Oksman, A., Smith, C. E., Duffin, K. L., & Goldberg, D. E. (1994). Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway. *J. Clin. Invest.* **93:** 1602-1608.

Gololobov, M. Y., Wang, W., & Bateman, R. C., Jr. (1996). Substrate and inhibitor specificity of glutamine cyclotransferase (QC). *Biol. Chem. Hoppe Seyler* **377:** 395-398.

Good, N. E., Winget, G. D., Winter, W., Connolly, T. N., Izawa, S., & Singh, R. M. (1966). Hydrogen ion buffers for biological research. *Biochemistry* **5:** 467-477.

Graham, J. E., Spanier, J. G., & Jarvik, J. W. (1995). Isolation and characterization of Pioneer1, a novel *Chlamydomonas* transposable element. *Curr. Genet.* **28:** 429-436.

Greenbaum, D. C., Baruch, A., Grainger, M., Bozdech, Z., Medzihradszky, K. F., Engel, J., DeRisi, J., Holder, A. A., & Bogyo, M. (2002). A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298:** 2002-2006.

Greenbaum, D. C., Mackey, Z., Hansell, E., Doyle, P., Gut, J., Caffrey, C. R., Lehrman, J., Rosenthal, P. J., McKerrow, J. H., & Chibale, K. (2004). Synthesis and structure-activity relationships of parasiticidal thiosemicarbazone cysteine protease inhibitors against *Plasmodium falciparum, Trypanosoma brucei, and Trypanosoma cruzi. J. Med. Chem.* **47:** 3212-3219.

Grioli, S., Lomeo, C., Quattropani, M. C., Spignoli, G., & Villardita, C. (1990). Pyroglutamic acid improves the age associated memory impairment. *Fundam. Clin. Pharmacol.* **4:** 169-173.

H

Hackstein, J. H., Mackenstedt, U., Mehlhorn, H., Meijerink, J. P., Schubert, H., & Leunissen, J. A. (1995). Parasitic apicomplexans harbor a chlorophyll a-D1 complex, the potential target for therapeutic triazines. *Parasitol. Res.* **81:** 207-216.

Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J. P., Rigden, D. J., Michels, P. A., & Opperdoes, F. R. (2003). Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl. Acad Sci. U. S. A.* **100:** 1067-1071.

Hanspal, M., Dua, M., Takakuwa, Y., Chishti, A. H., & Mizuno, A. (2002). *Plasmodium falciparum* cysteine protease falcipain-2 cleaves erythrocyte membrane skeletal proteins at late stages of parasite development. *Blood* **100:** 1048-1054.

Hart, D. T., Vickerman, K., & Coombs, G. H. (1981). A quick, simple method for purifying *Leishmania mexicana* amastigotes in large numbers. *Parasitology* **82:** 345-355.

Harth, G., Andrews, N., Mills, A. A., Engel, J. C., Smith, R., & McKerrow, J. H. (1993). Peptide-fluoromethyl ketones arrest intracellular replication and intercellular transmission of *Trypanosoma cruzi. Mol. Biochem. Parasitol.* **58:** 17-24.

Hasnain, S., Hirama T., Huber C. P., Mason P., Mort J. S. (1993). Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu245 in the S2-P2 specificity for arginine and its role in transition state stabilization. *J. Biol. Chem.* **268:** 235-240.

Henderson, G. B. & Fairlamb, A. H. (1987). Trypanothione metabolism: a chemotherapeutic target in trypanosomatids. *Parasitol. Today* **3:** 312-315.

Hilley, J. D., Zawadzki, J. L., McConville, M. J., Coombs, G. H., & Mottram, J. C. (2000). *Leishmania mexicana* mutants lacking glycosylphosphatidylinositol (GPI):protein transamidase provide insights into the biosynthesis and functions of GPI-anchored proteins. *Mol. Biol. Cell* **11:** 1183-1195.

Hinke, S. A., Pospisilik, J. A., Demuth, H. U., *Mannhart*, S., Kuhn-Wache, K., Hoffmann, T., Nishimura, E., Pederson, R. A., & McIntosh, C. H. (2000). Dipeptidyl peptidase IV (DPIV/CD26) degradation of glucagon. Characterization of glucagon degradation products and DPIV-resistant analogs. *J. Biol. Chem.* **275:** 3827-3834.

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., & Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. U. S. A* **86:** 8247-8251.

Hofmann, K. & Stoffel, W. (1992). PROFILEGRAPH: an interactive graphical tool for protein sequence analysis. *Comput. Appl. Biosci.* **8:** 331-337.

Hoover, D. M., Chertov, O., & Lubkowski, J. (2001). The structure of human beta-defensin-1: new insights into structural properties of beta-defensins. *J. Biol. Chem.* **276:** 39021-39026.

Huete-Perez, J. A., Engel, J. C., Brinen, L. S., Mottram, J. C., & McKerrow, J. H. (1999). Protease trafficking in two primitive eukaryotes is mediated by a prodomain protein motif. *J Biol. Chem.* **274:** 16249-16256.

Hughes, A. L. (1994). Evolution of cysteine proteinases in eukaryotes. *Mol Phylogenet. Evol.* **3:** 310-321.

I

Isogai, E., Isogai, H., Matuo, K., Hirose, K., Kowashi, Y., Okumuara, K., & Hirata, M. (2003). Sensitivity of genera *Porphyromonas* and *Prevotella* to the bactericidal action of C-terminal domain of human CAP18 and its analogues. *Oral Microbiol. Immunol.* **18:** 329-332.

Ito, K., Inoue, T., Takahashi, T., Huang, H. S., Esumi, T., Hatakeyama, S., Tanaka, N., Nakamura, K. T., & Yoshimoto, T. (2001). The mechanism of aubstrate eecognition of pyroglutamyl-peptidase I from Bacillus amyloliquefaciens as determined by X-ray crystallography and site-directed mutagenesis. *J. Biol. Chem.* **276:** 18557-18562.

J

Jacobsen, W., Christians, U., & Benet, L. Z. (2000). In vitro evaluation of the disposition of A novel cysteine protease inhibitor. *Drug Metab. Dispos.* **28:** 1343-1351.

Jean, L., Grosclaude, J., Labbe, M., Tomley, F., & Pery, P. (2000). Differential localisation of an *Eimeria tenella* aspartyl proteinase during the infection process. *Int. J. Parasitol.* **30:** 1099-1107.

Jean, L., Pery, P., Dunn, P., Bumstead, J., Billington, K., Ryan, R., & Tomley, F. (2001). Genomic organisation and developmentally regulated expression of an apicomplexan aspartyl proteinase. *Gene* **262:** 129-136.

Jones, D. E., Elloso, M. M., & Scott, P. (1998). Host susceptibility factors to cutaneous leishmaniasis. *Front. Biosci.* **3:** D1171-D1180.

Joseph-Bravo, P., Uribe, R. M., Vargas, M. A., Perez-Martinez, L., Zoeller, T., & Charli, J. L. (1998). Multifactorial modulation of TRH metabolism. *Cell Mol. Neurobiol.* **18:** 231-247.

Judice, W. A., Puzer, L., Cotrin, S. S., Carmona, A. K., Coombs, G. H., Juliano, L., & Juliano, M. A. (2004). Carboxydipeptidase activities of recombinant cysteine peptidases. Cruzain of *Trypanosoma cruzi* and CPB of *Leishmania mexicana*. *Eur. J. Biochem.* **271:** 1046-1053.

K

Kaga, M. M., Laurent, F., Doh, A., Luffau, G., Yvore, P., & Pery, P. (1998). Purification of a leucine aminopeptidase from *Eimeria falciformis*. *Vet. Res.* **29:** 107-111.

Karrer, K. M., Peiffer, S. L., & 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.

Katunuma, N., Tsuge, H., Nukatsuka, M., & Fukushima, M. (2002). Structure-based development of cathepsin L inhibitors and therapeutic applications for prevention of cancer metastasis and cancer-induced osteoporosis. *Adv. Enzyme Regul.* **42:** 159-172.

Kelleher, M. & Tomley, F. M. (1998). Transient expression of beta-galactosidase in differentiating sporozoites of *Eimeria tenella*. *Mol. Biochem. Parasitol.* **97:** 21-31.

Kembhavi, A. A., Buttle, D. J., Knight, C. G., & Barrett, A. J. (1993). The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* **303:** 208-213.

Kim, K. (2004). Role of proteases in host cell invasion by *Toxoplasma gondii* and other Apicomplexa. *Acta Trop.* **91:** 69-81.

Klemba, M. & Goldberg, D. E. (2002). Biological roles of proteases in parasitic protozoa. *Annu. Rev. Biochem.* **71:** 275-305.

Klinkert, M. Q., Cioli, D., Shaw, E., Turk, V., Bode, W., & Butler, R. (1994). Sequence and structure similarities of cathepsin B from the parasite *Schistosoma mansoni* and human liver. *FEBS Lett.* **351:** 397-400.

Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D., & Roos, D. S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science* **275:** 1485-1489.

Kornfeld, S. & Mellman, I. (1989). The biogenesis of lysosomes. *Annu. Rev. Cell Biol.* **5:** 483-525.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234:** 187-208.

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

L

Lalmanach, G., Boulange, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., & Authie, E. (2002). Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. *Biol. Chem.* **383:** 739-749.

Lauffart, B. & Mantle, D. (1988). Rationalization of aminopeptidase activities in human skeletal muscle soluble extract. *Biochim. Biophys. Acta* **956:** 300-306.

Laurent, F., Bourdieu, C., Kaga, M., Chilmonczyk, S., Zgrzebski, G., Yvore, P., & Pery, P. (1993). Cloning and characterization of an *Eimeria acervulina* sporozoite gene homologous to aspartyl proteinases. *Mol. Biochem. Parasitol.* **62:** 303-312.

Lazdunski, A. M. (1989). Peptidases and proteases of *Escherichia coli* and *Salmonella typhimurium*. *FEMS Microbiol. Rev.* **5:** 265-276.

Le Saux, O., Gonzales, T., & Robert-Baudouy, J. (1996). Mutational analysis of the active site of *Pseudomonas fluorescens* pyrrolidone carboxyl peptidase. *J. Bacterio.l* **178:** 3308-3313.

Le Saux, O. & Robert-Baudouy, J. (1997). Pyroglutamic acid and iron regulate the expression of the pcp gene in *Pseudomonas fluorescens* MFO. *FEMS Microbiol. Lett.* **155:** 209-215.

Liao, Y. D., Wang, S. C., Leu, Y. J., Wang, C. F., Chang, S. T., Hong, Y. T., Pan, Y. R., & Chen, C. (2003). The structural integrity exerted by N-terminal pyroglutamate is crucial for the cytotoxicity of frog ribonuclease from *Rana pipiens*. *Nucleic Acids Res.* **31:** 5247-5255.

Liew, F. Y., Li, Y., & Millott, S. (1990). Tumor necrosis factor-alpha synergizes with IFN-gamma in mediating killing of *Leishmania* major through the induction of nitric oxide. *J. Immunol.* **145:** 4306-4310.

Lugli, E. B., Pouliot, M., Portela, M. P., Loomis, M. R., & Raper, J. (2004). Characterization of primate trypanosome lytic factors. *Mol. Biochem. Parasitol.* **138:** 9-20.

M

Mackey, Z. B., O'Brien, T. C., Greenbaum, D. C., Blank, R. B., & McKerrow, J. H. (2004). A cathepsin B-like protease is required for host protein degradation in *Trypanosoma brucei. J. Biol. Chem.* **279:** 48426-48433.

Mantle, D., Lauffart, B., & Gibson, A. (1991). Purification and characterization of leucyl aminopeptidase and pyroglutamyl aminopeptidase from human skeletal muscle. *Clin. Chim. Acta* **197:** 35-45.

Mantle, D., Falkous, G., Ishiura, S., Perry, R. H., & Perry, E. K. (1995). Comparison of cathepsin protease activities in brain tissue from normal cases and cases with Alzheimer's disease, Lewy body dementia, Parkinson's disease and Huntington's disease. *J. Neurol. Sci.* **131:** 65-70.

Mathieu, M. A., Bogyo, M., Caffrey, C. R., Choe, Y., Lee, J., Chapman, H., Sajid, M., Craik, C. S., & McKerrow, J. H. (2002). Substrate specificity of schistosome versus human legumain determined by P1-P3 peptide libraries. *Mol Biochem Parasitol* **121:** 99-105.

McConkey, G. A. (1999). Targeting the shikimate pathway in the malaria parasite *Plasmodium falciparum. Antimicrob. Agents Chemother.* **43:** 175-177.

McDonald, J. K. (1985). An overview of protease specificity and catalytic mechanisms: aspects related to nomenclature and classification. *Histochem. J.* **17:** 773-785.

McIntyre, G. F. & Erickson, A. H. (1993). The lysosomal proenzyme receptor that binds procathepsin L to microsomal membranes at pH 5 is a 43-kDa integral membrane protein. *Proc. Natl. Acad. Sci. U. S. A.* **90:** 10588-10592.

McKerrow, J. H., Sun, E., Rosenthal, P. J., & Bouvier, J. (1993). The proteases and pathogenicity of parasitic protozoa. *Annu. Rev. Microbiol.* **47:** 821-853.

McKerrow, J. H., Engel, J. C., & Caffrey, C. R. (1999). Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg. Med. Chem.* **7:** 639-644.

Medina-Acosta, E. & Cross, G. A. (1993). Rapid isolation of DNA from trypanosomatid protozoa using a simple 'mini-prep' procedure. *Mol. Biochem. Parasitol* .**59:** 327-329.

Michalski, W. P., Edgar, J. A., & Prowse, S. J. (1992). Mannitol metabolism in *Eimeria tenella. Int. J. Parasitol.* **22:** 1157-1163.

Michalski, W. P., Crooks, J. K., & Prowse, S. J. (1994). Purification and characterisation of a serine-type protease from *Eimeria tenella* oocysts. *Int. J. Parasitol.* **24:** 189-195.

Morrot, A., Strickland, D. K., Higuchi, M. L., Reis, M., Pedrosa, R., & Scharfstein, J. (1997). Human T cell responses against the major cysteine proteinase (cruzipain) of *Trypanosoma cruzi*: role of the multifunctional alpha 2- macroglobulin receptor in antigen presentation by monocytes. *Int. Immunol.* **9:** 825-834.

Mottram, J. C., Brooks, D. R., & Coombs, G. H. (1998). Roles of cysteine proteinases of trypanosomes and *Leishmania* in host- parasite interactions. *Curr. Opin. Microbiol.* **1:** 455-460.

Mottram, J. C., Helms, M. J., Coombs, G. H., & Sajid, M. (2003). Clan CD cysteine peptidases of parasitic protozoa. *Trends Parasitol.***19:** 182-187.

Mottram, J. C., Coombs, G. H., & Alexander, J. (2004). Cysteine peptidases as virulence factors of *Leishmania. Curr Opin Microbiol.* **7:** 375-381.

Mottram, J. C., Souza, A. E., Hutchison, J. E., Carter, R., Frame, M. J., & Coombs, G. H. (1996). Evidence from disruption of the lmcpb gene array of *Leishmania mexicana* that cysteine proteinases are virulence factors. *Proc. Natl. Acad Sci. U. S. A.* **93:** 6008-6013.

Mottram, J. C., Frame, M. J., Brooks, D. R., Tetley, L., Hutchison, J. E., Souza, A. E., & Coombs, G. H. (1997). The Multiple cpb Cysteine Proteinase Genes of *Leishmania mexicana* Encode Isoenzymes That Differ in Their Stage Regulation and Substrate Preferences. *J. Biol. Chem.* **272:** 14285-14293.

Muller, M., Lee, J. A., Gordon, P., Gaasterland, T., & Sensen, C. W. (2001). Presence of prokaryotic and eukaryotic species in all subgroups of the PP(i)-dependent group II phosphofructokinase protein family. *J. Bacteriol.* **183:** 6714-6716.

Mullin, K. A., Foth, B. J., Ilgoutz, S. C., Callaghan, J. M., Zawadzki, J. L., McFadden, G. I., & McConville, M. J. (2001). Regulated degradation of an endoplasmic reticulum membrane protein in a tubular lysosome in *Leishmania mexicana. Mol. Biol. Cell* **12:** 2364-2377.

Myler, P. J., Sisk, E., McDonagh, P. D., Martinez-Calvillo, S., Schnaufer, A., Sunkin, S. M., Yan, S., Madhubala, R., Ivens, A., & Stuart, K. (2000). Genomic organization and gene function in *Leishmania. Biochem Soc. Trans.* **28:** 527-531.

N

Na, B. K., Shenai, B. R., Sijwali, P. S., Choe, Y., Pandey, K. C., Singh, A., Craik, C. S., & Rosenthal, P. J. (2004). Identification and biochemical characterization of vivapains, cysteine proteases of the malaria parasite *Plasmodium vivax*. *Biochem. J.* **378:** 529-538.

Nagler, D. K., Storer, A. C., Portaro, F. C., Carmona, E., Juliano, L., & Menard, R. (1997). Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry* **36:** 12608-12615.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.* **8:** 581-599.

O

O'Connor, B. & O'Cuinn, G. (1984). Localization of a narrow-specificity thyroliberin hydrolyzing pyroglutamate aminopeptidase in synaptosomal membranes of guinea-pig brain. *Eur. J. Biochem.* **144:** 271-278.

Oberg, K. A., Ruysschaert, J. M., Azarkan, M., Smolders, N., Zerhouni, S., Wintjens, R., Amrani, A., & Looze, Y. (1998). Papaya glutamine cyclase, a plant enzyme highly resistant to proteolysis, adopts an all-beta conformation. *Eur. J. Biochem.* **258:** 214-222.

Ogasahara, K., Khechinashvili, N. N., Nakamura, M., Yoshimoto, T., & Yutani, K. (2001). Thermal stability of pyrrolidone carboxyl peptidases from the hyperthermophilic Archaeon, *Pyrococcus furiosus*. *Eur. J. Biochem.* **268:** 3233-3242.

Ojcius, D. M., Perfettini, J. L., Bonnin, A., & Laurent, F. (1999). Caspase-dependent apoptosis during infection with *Cryptosporidium parvum*. *Microbes Infect.* **1:** 1163-1168.

Olliaro, P. L. & Bryceson, A. D. (1993). Practical progress and new drugs for changing patterns of leishmaniasis. *Parasitol Today* **9:** 323-328.

Otto, H. H. & Schirmeister, T. (1997). Cysteine Proteases and Their Inhibitors. *Chem. Rev.* **97:** 133-172.

Ouellette, M. & Papadopoulou, B. (1993). Mechanisms of drug resistance in *Leishmania*. *Parasitol. Today* **9:** 150-153.

P

Pamer, E. G., So, M., & Davis, C. E. (1989). Identification of a developmentally regulated cysteine protease of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **33:** 27-32.

Pandey, K. C., Sijwali, P. S., Singh, A., Na, B. K., & Rosenthal, P. J. (2004). Independent intramolecular mediators of folding, activity, and inhibition for the *Plasmodium falciparum* cysteine protease falcipain-2. *J. Biol. Chem.* **279:** 3484-3491.

Parsons, M. (2004). Glycosomes: parasites and the divergence of peroxisomal purpose. *Mol. Microbiol.* **53:** 717-724.

Parussini, F., Duschak, V. G., & Cazzulo, J. J. (1998). Membrane-bound cysteine proteinase isoforms in different developmental stages of *Trypanosoma cruzi*. *Cell Mol Biol. (Noisy -le-grand)* **44:** 513-519.

Pearson, R. D., Romito, R., Symes, P. H., & Harcus, J. L. (1981). Interaction of *Leishmania donovani* promastigotes with human monocyte-derived macrophages: parasite entry, intracellular survival, and multiplication. *Infect. Immun.* **32:** 1249-1253.

Perlman, J. H., Thaw, C. N., Laakkonen, L., Bowers, C. Y., Osman, R., & Gershengorn, M. C. (1994). Hydrogen bonding interaction of thyrotropin-releasing hormone (TRH) with transmembrane tyrosine 106 of the TRH receptor. *J. Biol. Chem.* **269:** 1610-1613.

Pertuz, B. S., Ostoa, S. P., Benitez, I., Soldevila, G., Olivos, A., & Garcia-Zepeda, E. (2004). *Entamoeba histolytica* cysteine protease 2 (EhCP2) modulates leucocyte migration by proteolytic cleavage of chemokines. *Parasite Immunol.* **26:** 237-241.

Pohl, T., Zimmer, M., Mugele, K., & Spiess, J. (1991). Primary structure and functional expression of a glutaminyl cyclase. *Proc. Natl. Acad. Sci. U. S. A.* **88:** 10059-10063.

Polgar, L. (2004). Catalytic mechanisms of cysteine peptidases, Handbook of Proteolytic Enzymes, 2nd edition, Elsevier, London, UK: 1072-1079.

Prasad, C., Wilber, J. F., & Amborski, R. L. (1982). Thyrotropin-releasing hormone: its distribution and metabolism during development in bullfrog (*Rana catesbeiana*). *Dev. Neurosci.* **5:** 293-297.

Predel, R., Brandt, W., Kellner, R., Rapus, J., Nachman, R. J., & Gade, G. (1999). Post-translational modifications of the insect sulfakinins: sulfation, pyroglutamate-formation and O-methylation of glutamic acid. *Eur. J. Biochem.* **263:** 552-560.

Pupkis, M. F., Tetley, L., & Coombs, G. H. (1986). *Leishmania mexicana*: amastigote hydrolases in unusual lysosomes. *Exp. Parasitol.* **62:** 29-39.

Q

Que, X. & Reed, S. L. (2000). Cysteine proteinases and the pathogenesis of amebiasis. *Clin. Microbiol. Rev.* **13:** 196-206.

Que, X., Ngo, H., Lawton, J., Gray, M., Liu, Q., Engel, J., Brinen, L., Ghosh, P., Joiner, K. A., & Reed, S. L. (2002). The cathepsin B of Toxoplasma gondii, toxopain-1, is critical for parasite invasion and rhoptry protein processing. *J. Biol. Chem.* **277:** 25791-25797.

Que, X., Kim, S. H., Sajid, M., Eckmann, L., Dinarello, C. A., McKerrow, J. H., & Reed, S. L. (2003). A surface amebic cysteine proteinase inactivates interleukin-18. *Infect. Immun.* **71:** 1274-1280.

Que, X., Wunderlich, A., Joiner, K. A., & Reed, S. L. (2004). Toxopain-1 is critical for infection in a novel chicken embryo model of congenital toxoplasmosis. *Infect. Immun.* **72:** 2915-2921.

R

Rabel, D., Charlet, M., Ehret-Sabatier, L., Cavicchioli, L., Cudic, M., Otvos, L., Jr., & Bulet, P. (2004). Primary structure and in vitro antibacterial properties of the *Drosophila melanogaster* attacin C Pro-domain. *J. Biol. Chem.* **279:** 14853-14859.

Rafati, S., Salmanian, A. H., Hashemi, K., Schaff, C., Belli, S., & Fasel, N. (2001). Identification of *Leishmania major* cysteine proteinases as targets of the immune response in humans. *Mol. Biochem. Parasitol.* **113:** 35-43.

Rawlings, N. D. & Barrett, A. J. (1994). Families of serine peptidases. *Methods Enzymol.* **244:** 19-61.

Rawlings, N. D. & Barrett, A. J. (1995)a. Evolutionary families of metallopeptidases. *Methods Enzymol.* **248:** 183-228.

Rawlings, N. D. & Barrett, A. J. (1995)b. Families of aspartic peptidases, and those of unknown catalytic mechanism. *Methods Enzymol.* **248:** 105-120.

Rawlings, N. D. & Barrett, A. J. (1999). MEROPS: the peptidase database. *Nucl. Acids. Res.* **27:** 325-331.

Rawlings, N. D., O'Brien, E., & Barrett, A. J. (2002). MEROPS: the protease database. *Nucl. Acids Res.* **30:** 343-346.

Reiss, M., Viebig, N., Brecht, S., Fourmaux, M. N., Soete, M., Di Cristina, M., Dubremetz, J. F., & Soldati, D. (2001). Identification and characterization of an escorter for two secretory adhesins in *Toxoplasma gondii*. *J. Cell Biol.* **152:** 563-578.

Reithinger, R., Quinnell, R. J., Alexander, B., & Davies, C. R. (2002). Rapid detection of *Leishmania infantum* infection in dogs: comparative study using an immunochromatographic dipstick test, enzyme-linked immunosorbent assay, and PCR. *J. Clin. Microbiol.* **40:** 2352-2356.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34:** 167-339.

Roberts, F., Roberts, C. W., Johnson, J. J., Kyle, D. E., Krell, T., Coggins, J. R., Coombs, G. H., Milhous, W. K., Tzipori, S., Ferguson, D. J., Chakrabarti, D., & McLeod, R. (1998). Evidence for the shikimate pathway in apicomplexan parasites. *Nature* **393:** 801-805.

Robertson, C. D., North, M. J., Lockwood, B. C., & Coombs, G. H. (1990). Analysis of the proteinases of *Trypanosoma brucei. J. Gen. Microbiol.* **136:** 921-925.

Robertson, C. D. & Coombs, G. H. (1994). Multiple high activity cysteine proteases of Leishmania mexicana are encoded by the lmcpb gene array. *Microbiology* **140:** 417-424.

Robertson, C. D. (1999). The *Leishmania mexicana* proteasome. *Mol. Biochem. Parasitol.* **103:** 49-60.

Robertson, C. D., Coombs, G. H., North M.J., & Mottram, J. C. (1996). Parasite cysteine proteinases. *Perspectives in Drug Discovery and Design* **6:** 99-118.

Rosenthal, P. J. (1996). Conservation of key amino acids among the cysteine proteinases of multiple malarial species. *Mol. Biochem. Parasitol.* **75:** 255-260.

Rosenthal, P. J. (1999). Proteases of protozoan parasites. *Adv. Parasitol.* **43:** 105-159.

Rosenthal, P. J. (2002). Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Curr. Opin. Hematol.* **9:** 140-145.

Rzychon, M., Chmiel, D., & Stec-Niemczyk, J. (2004). Modes of inhibition of cysteine proteases. *Acta Biochim. Pol.* **51:** 861-873.

S

Sacks, D. L., Hieny, S., & Sher, A. (1985). Identification of cell surface carbohydrate and antigenic changes between noninfective and infective developmental stages of *Leishmania major* promastigotes. *J. Immunol.* **135:** 564-569.

Sacks, D. L. & da Silva, R. P. (1987). The generation of infective stage *Leishmania major* promastigotes is associated with the cell-surface expression and release of a developmentally regulated glycolipid. *J. Immunol.* **139:** 3099-3106.

Sacks, D. L. (1989). Metacyclogenesis in *Leishmania* promastigotes. *Exp. Parasitol.* **69:** 100-103.

Sahl, H. G., Pag, U., Bonness, S., Wagner, S., Antcheva, N., & Tossi, A. (2004). Mammalian defensins: structures and mechanism of antibiotic activity. *J. Leukoc. Biol.* **74:** 466-475

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

Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I., Greenbaum, D., Bogyo, M., Salter, J. P., Lim, K. C., Franklin, C., Kim, J. H., & Caffrey, C. R. (2003). Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol.* **131:** 65-75.

Sakanari, J. A., Nadler, S. A., Chan, V. J., Engel, J. C., Leptak, C., & Bouvier, J. (1997). *Leishmania major*: comparison of the cathepsin L- and B-like cysteine protease genes with those of other trypanosomatids. *Exp. Parasitol.* **85:** 63-76.

Salers, P., Ouafik, L. H., Giraud, P., Dutour, A., Maltese, J. Y., & Oliver, C. (1991). Evidence for pyroglutamyl peptidase I and prolyl endopeptidase activities in the rat insulinoma cell line RINm 5F: lack of relationship with TRH metabolism. *Mol. Cell Biochem.* **106:** 15-24.

Sambrook, J., Tritsch, E. F., & Maniatis, T. (1999). Molecular cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanderson, S. J., Pollock, K. G., Hilley, J. D., Meldal, M., Hilaire, P. S., Juliano, M. A., Juliano, L., Mottram, J. C., & Coombs, G. H. (2000). Expression and characterization of a recombinant cysteine proteinase of *Leishmania mexicana*. *Biochem. J.* **347:** 383-388.

Sanderson, S. J., Westrop, G. D., Scharfstein, J., Mottram, J. C., & Coombs, G. H. (2003). Functional conservation of a natural cysteine peptidase inhibitor in protozoan and bacterial pathogens. *FEBS Lett.* **542:** 12-16.

Schaller-Bals, S., Schulze, A., & Bals, R. (2002). Increased levels of antimicrobial peptides in tracheal aspirates of newborn infants during infection. *Am. J. Respir. Crit. Care Med.* **165:** 992-995.

Scheidt, K. A., Roush, W. R., McKerrow, J. H., Selzer, P. M., Hansell, E., & Rosenthal, P. J. (1998). Structure-based design, synthesis and evaluation of conformationally constrained cysteine protease inhibitors. *Bioorg. Med. Chem.* **6:** 2477-2494.

Schibli, D. J., Epand, R. F., Vogel, H. J., & Epand, R. M. (2002). Tryptophan-rich antimicrobial peptides: comparative properties and membrane interactions. *Biochem. Cell Biol.* **80:** 667-677.

Schilling, S., Hoffmann, T., Manhart, S., Hoffmann, M., & Demuth, H. U. (2004). Glutaminyl cyclases unfold glutamyl cyclase activity under mild acid conditions. *FEBS Lett.* **563:** 191-196.

Schmatz, D. M., Baginsky, W. F., & Turner, M. J. (1989). Evidence for and characterization of a mannitol cycle in *Eimeria tenella. Mol. Biochem. Parasitol.* **32:** 263-270.

Schmitmeier, S., Thole, H., Bader, A., & Bauer, K. (2002). Purification and characterization of the thyrotropin-releasing hormone (TRH)-degrading serum enzyme and its identification as a product of liver origin. *Eur. J. Biochem.* **269:** 1278-1286.

Schneider, A. (2001). Unique aspects of mitochondrial biogenesis in trypanosomatids. *Int. J. Parasitol.* **31:** 1403-1415.

Schomburg, L. & Bauer, K. (1995). Thyroid hormones rapidly and stringently regulate the messenger RNA levels of the thyrotropin-releasing hormone (TRH) receptor and the TRH-degrading ectoenzyme. *Endocrinology* **136:** 3480-3485.

Schomburg, L., Turwitt, S., Prescher, G., Lohmann, D., Horsthemke, B., & Bauer, K. (1999). Human TRH-degrading ectoenzyme cDNA cloning, functional expression, genomic structure and chromosomal assignment. *Eur. J. Biochem.* **265:** 415-422.

Scocchi, M., Wang, S., & Zanetti, M. (1997). Structural organization of the bovine cathelicidin gene family and identification of a novel member. *FEBS Lett.* **417:** 311-315.

Seemuller, E., Lupas, A., Stock, D., Lowe, J., Huber, R., & Baumeister, W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* **268:** 579-582.

Seidah, N. G., Day, R., Marcinkiewicz, M., & Chretien, M. (1998). Precursor convertases: an evolutionary ancient, cell-specific, combinatorial mechanism yielding diverse bioactive peptides and proteins. *Ann. N. Y. Acad Sci.* **839:** 9-24.

Seifert, K., Matu, S., Javier Perez-Victoria, F., Castanys, S., Gamarro, F., & Croft, S. L. (2003). Characterisation of *Leishmania donovani* promastigotes resistant to hexadecylphosphocholine (miltefosine). *Int. J. Antimicrob. Agents* **22:** 380-387.

Selsted, M. E. & Ouellette, A. J. (1995). Defensins in granules of phagocytic and non-phagocytic cells. *Trends Cell Biol.* **5:** 114-119.

Selsted, M. E., Tang, Y. Q., Morris, W. L., McGuire, P. A., Novotny, M. J., Smith, W., Henschen, A. H., & Cullor, J. S. (1993). Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J. Biol. Chem.* **268:** 6641-6648.

Selzer, P. M., Chen, X., Chan, V. J., Cheng, M., Kenyon, G. L., Kuntz, I. D., Sakanari, J. A., Cohen, F. E., & McKerrow, J. H. (1997). *Leishmania major*: molecular modeling of cysteine proteases and prediction of new nonpeptide inhibitors. *Exp. Parasitol* **87:** 212-221.

Selzer, P. M., Pingel, S., Hsieh, I., Ugele, B., Chan, V. J., Engel, J. C., Bogyo, M., Russell, D. G., Sakanari, J. A., & McKerrow, J. H. (1999). Cysteine protease inhibitors as chemotherapy: lessons from a parasite target. *Proc. Natl. Acad. Sci. U. S. A.* **96:** 11015-11022.

Shaked-Mishan, P., Ulrich, N., Ephros, M., & Zilberstein, D. (2001). Novel Intracellular SbV reducing activity correlates with antimony susceptibility in *Leishmania donovani*. *J. Biol. Chem.* **276:** 3971-3976.

Shenai, B. R., Lee, B. J., Alvarez-Hernandez, A., Chong, P. Y., Emal, C. D., Neitz, R. J., Roush, W. R., & Rosenthal, P. J. (2003). Structure-activity relationships for inhibition of cysteine protease activity and development of *Plasmodium falciparum* by peptidyl vinyl sulfones. *Antimicrob. Agents Chemother.* **47:** 154-160.

Shirley, M. W. (1989). Development of a live attenuated vaccine against coccidiosis of poultry. *Parasite Immunol* **11:** 117-124.

Shirley, M. W. & Harvey, D. A. (2000). A genetic linkage map of the apicomplexan protozoan parasite *Eimeria tenella*. *Genome Res.* **10:** 1587-1593.

Sijwali, P. S. & Rosenthal, P. J. (2004). Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 4384-4389.

Sijwali, P. S., Kato, K., Seydel, K. B., Gut, J., Lehman, J., Klemba, M., Goldberg, D. E., Miller, L. H., & Rosenthal, P. J. (2004). *Plasmodium falciparum* cysteine protease falcipain-1 is not essential in erythrocytic stage malaria parasites. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 8721-8726.

Sijwali, P. S., Shenai, B. R., Gut, J., Singh, A., & Rosenthal, P. J. (2001). Expression and characterization of the *Plasmodium falciparum* haemoglobinase falcipain-3. *Biochem. J.* **360:** 481-489.

Silva, P. I., Jr., Daffre, S., & Bulet, P. (2000). Isolation and characterization of gomesin, an 18-residue cysteine-rich defense peptide from the spider *Acanthoscurria gomesiana* hemocytes with sequence similarities to horseshoe crab antimicrobial peptides of the tachyplesin family. *J. Biol. Chem.* **275:** 33464-33470.

Simpson, L., Aphasizhev, R., Gao, G., & Kang, X. (2004). Mitochondrial proteins and complexes in *Leishmania* and *Trypanosoma* involved in U-insertion/deletion RNA editing. *RNA.* **10:** 159-170.

Singh, A. & Rosenthal, P. J. (2001). Comparison of efficacies of cysteine protease inhibitors against five strains of *Plasmodium falciparum. Antimicrob. Agents Chemother.* **45:** 949-951.

Soldati, D., Dubremetz, J. F., & Lebrun, M. (2001). Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii. Int. J. Parasitol.* **31:** 1293-1302.

Soldati, D., Lassen, A., Dubremetz, J. F., & Boothroyd, J. C. (1998). Processing of *Toxoplasma* ROP1 protein in nascent rhoptries. *Mol. Biochem. Parasitol.* **96:** 37-48.

Song, I., Chuang, C. Z., & Bateman, R. C., Jr. (1994). Molecular cloning, sequence analysis and expression of human pituitary glutaminyl cyclase. *J. Mol. Endocrinol.* **13:** 77-86.

Sorensen, O. E., Follin, P., Johnsen, A. H., Calafat, J., Tjabringa, G. S., Hiemstra, P. S., & Borregaard, N. (2001). Human cathelicidin, hCAP-18, is processed to the antimicrobial peptide LL-37 by extracellular cleavage with proteinase 3. *Blood* **97:** 3951-3959.

Souza Leao, S., Lang, T., Prina, E., *Hellio,* R., & Antoine, J. C. (1995). Intracellular *Leishmania amazonensis* amastigotes internalize and *degrade MHC* class II molecules of their host cells. *J. Cell Sci.* **108:** 3219-3231.

Souza, A. E., Waugh, S., Coombs, G. H., & Mottram, J. C. (1992). Characterization of a multi-copy gene for a major stage-specific cysteine proteinase of *Leishmania mexicana. FEBS Lett.* **311:** 124-127.

St Hilaire, P. M., Alves, L. C., Sanderson, S. J., Mottram, J. C., Juliano, M. A., Juliano, L., Coombs, G. H., & Meldal, M. (2000). The substrate specificity of a recombinant cysteine protease from *Leishmania mexicana*: application of a combinatorial peptide library approach. *Chembiochem.* **1:** 115-122.

Sykes, P. A., Watson, S. J., Temple, J. S., & Bateman, R. C., Jr. (1999). Evidence for tissue-specific forms of glutaminyl cyclase. *FEBS Lett.* **455:** 159-161.

T

Tang, Y. Q., Yeaman, M. R., & Selsted, M. E. (2002). Antimicrobial peptides from human platelets. *Infect. Immun.* **70:** 6524-6533.

Taubert, H., Riemann, D., Kehlen, A., Meye, A., Bartel, F., John, V., Brandt, J., Bache, M., Wurl, P., Schmidt, H., & Weber, E. (2002). Expression of cathepsin B, D and L protein in juvenile idiopathic arthritis. *Autoimmunity* **35:** 221-224.

Thakur, C. P., Kumar, P., Kumar, N., Singh, G. N., Singh, A. K., & Narain, S. (1998). A randomized comparison of classical mode of administration of amphotericin B with its newer modes of administration in kala-azar. *J. Assoc. Physicians India* **46:** 779-783.

Tierney, J. & Mulcahy, G. (2003). Comparative development of *Eimeria tenella* (Apicomplexa) in host cells in vitro. *Parasito.l Res.* **90:** 301-304.

Tomas, A. M., Miles, M. A., & Kelly, J. M. (1997). Overexpression of cruzipain, the major cysteine proteinase of *Trypanosoma cruzi*, is associated with enhanced metacyclogenesis. *Eur. J. Biochem.* **244:** 596-603.

Tomley, F. (1997). Techniques for isolation and characterization of apical organelles from *Eimeria tenella* sporozoites. *Methods* **13:** 171-176.

Tomley, F. M. & Soldati, D. S. (2001). Mix and match modules: structure and function of microneme proteins in apicomplexan parasites. *Trends Parasitol.* **17:** 81-88.

Torres, A. M. & Kuchel, P. W. (2004). The beta-defensin-fold family of polypeptides. *Toxicon.* **44:** 581-588.

Touz, M. C., Nores, M. J., Slavin, I., Carmona, C., Conrad, J. T., Mowatt, M. R., Nash, T. E., Coronel, C. E., & Lujan, H. D. (2002). The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote *Giardia lamblia*. *J. Biol. Chem.* **277:** 8474-8481.

Trabandt, A., Aicher, W. K., Gay, R. E., Sukhatme, V. P., Nilson-Hamilton, M., Hamilton, R. T., McGhee, J. R., Fassbender, H. G., & Gay, S. (1990). Expression of the collagenolytic and Ras-induced cysteine proteinase cathepsin L and proliferation-associated oncogenes in synovial cells of MRL/I mice and patients with rheumatoid arthritis. *Matrix* **10:** 349-361.

Troeberg, L., Pike, R. N., Morty, R. E., Berry, R. K., Coetzer, T. H., & Lonsdale-Eccles, J. D. (1996). Proteases from *Trypanosoma brucei brucei*. Purification, characterisation and interactions with host regulatory molecules. *Eur. J. Biochem.* **238:** 728-736.

Tsuru, D., Sakabe, K., Yoshimoto, T., & Fujiwara, K. (1982). Pyroglutamyl peptidase from chicken liver: purification and some properties. *J. Pharmacobiodyn.* **5:** 859-868.

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

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

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

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

V

Van Coillie, E., Proost, P., Van, A., I, Struyf, S., Polfliet, M., De, M., I, Harvey, D. J., Van Damme, J., & Opdenakker, G. (1998). Functional comparison of two human monocyte chemotactic protein-2 isoforms, role of the amino-terminal pyroglutamic acid and processing by CD26/dipeptidyl peptidase IV. *Biochemistry* **37:** 12672-12680.

Vial, H. J. (2000). Isoprenoid biosynthesis and drug targeting in the Apicomplexa. *Parasitol. Today* **16:** 140-141.

Vollmer, M., Thomsen, N., Wiek, S., & Seeber, F. (2001). Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast- localized, plant-type ferredoxin-NADP+ reductase and ferredoxin. *J. Biol. Chem.* **276:** 5483-5490.

W

Waldenstedt, L., Lunden, A., Elwinger, K., Thebo, P., & Uggla, A. (1999). Comparison between a live, attenuated anticoccidial vaccine and an anticoccidial ionophore, on performance of broilers raised with or without a growth promoter, in an initially *Eimeria*-free environment. *Acta Vet. Scand.* **40:** 11-21.

Waller, R. F. & McConville, M. J. (2002). Developmental changes in lysosome morphology and function *Leishmania* parasites. *Int. J. Parasitol.* **32:** 1435-1445.

Waller, R. F., McConville, M. J., & McFadden, G. I. (2004). More plastids in human parasites? *Trends Parasitol.* **20:** 54-57.

Walters, L. L., Irons, K. P., Chaplin, G., & Tesh, R. B. (1993). Life cycle of *Leishmania major* (Kinetoplastida: Trypanosomatidae) in the neotropical sand fly *Lutzomyia longipalpis* (Diptera: Psychodidae). *J. Med. Entomol.* **30:** 699-718.

Ward, W., Alvarado, L., Rawlings, N. D., Engel, J. C., Franklin, C., & McKerrow, J. H. (1997). A primitive enzyme for a primitive cell: the protease required for excystation of *Giardia*. *Cell* **89:** 437-444.

Werbovetz, K. A. (2002). Promising therapeutic targets for anti-leishmanial drugs. *Expert. Opin. Ther. Targets* **6:** 407-422.

White, N. J. (1998). Drug resistance in malaria. *Br. Med. Bull.* **54:** 703-715.

Wiederanders, B. (2000). The function of propeptide domains of cysteine proteinases. *Adv. Exp. Med. Biol.* **477:** 261-270.

Wiemer, E. A., IJlst, L., van Roy, J., Wanders, R. J., & Opperdoes, F. R. (1996). Identification of 2-enoyl coenzyme A hydratase and NADP(+)-dependent 3-hydroxyacyl-CoA dehydrogenase activity in glycosomes of procyclic *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **82:** 107-111.

Williams, R. B. (2001). Quantification of the crowding effect during infections with the seven *Eimeria* species of the domesticated fowl: its importance for experimental designs and the production of oocyst stocks. *Int. J. Parasitol.* **31:** 1056-1069.

Williamson, D. H., Gardner, M. J., Preiser, P., Moore, D. J., Rangachari, K., & Wilson, R. J. (1994). The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. *Mol. Gen. Genet.* **243:** 249-252.

Wilson, R. J. & Williamson, D. H. (1997). Extrachromosomal DNA in the Apicomplexa. *Microbiol. Mol. Biol. Rev.* **61:** 1-16.

Worthey, E. A., Schnaufer, A., Mian, I. S., Stuart, K., & Salavati, R. (2003). Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Res.* **31:** 6392-6408.

X

Xu, P., Widmer, G., Wang, Y., Ozaki, L. S., Alves, J. M., Serrano, M. G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A. J., Pearson, W. R., Dear, P. H., Bankier, A. T., Peterson, D. L., Abrahamsen, M. S., Kapur, V., Tzipori, S., & Buck, G. A. (2004). The genome of *Cryptosporidium hominis*. *Nature* **431:** 1107-1112.

Y

Yamauchi, K. (1991). The sequence flanking translational initiation site in protozoa. *Nucleic Acids Res.* **19:** 2715-2720.

Yao, C., Donelson, J. E., & Wilson, M. E. (2003). The major surface protease (MSP or GP63) of *Leishmania sp*. Biosynthesis, regulation of expression, and function. *Mol. Biochem. Parasitol.* **132:** 1-16.

Yoshimoto, T., Shimoda, T., Kitazono, A., Kabashima, T., Ito, K., & Tsuru, D. (1993). Pyroglutamyl peptidase gene from *Bacillus amyloliquefaciens*: cloning, sequencing, expression, and crystallization of the expressed enzyme. *J. Biochem. (Tokyo)* **113:** 67-73.

Z

Zaiou, M., Nizet, V., & Gallo, R. L. (2003). Antimicrobial and protease inhibitory functions of the human cathelicidin (hCAP18/LL-37) prosequence. *J. Invest. Dermatol.* **120:** 810-816.

Zangger, H., Mottram, J. C., & Fasel, N. (2002). Cell death in *Leishmania* induced by stress and differentiation: programmed cell death or necrosis? *Cell Death. Differ.* **9:** 1126-1139.

Zhou, S., Kile, A., Kvikstad, E., Bechner, M., Severin, J., Forrest, D., Runnheim, R., Churas, C., Anantharaman, T. S., Myler, P., Vogt, C., Ivens, A., Stuart, K., & Schwartz, D. C. (2004). Shotgun optical mapping of the entire *Leishmania major* Friedlin genome. *Mol. Biochem. Parasitol.* **138:** 97-106.

Ziegler, R., Cushing, A. S., Walpole, P., Jasensky, R. D., & Morimoto, H. (1998). Analogs of Manduca adipokinetic hormone tested in a bioassay and in a receptor-binding assay. *Peptides* **19:** 481-486.

Zilberstein, D., Philosoph, H., & Gepstein, A. (1989). Maintenance of cytoplasmic pH and proton motive force in promastigotes of *Leishmania donovani*. *Mol. Biochem. Parasitol.* **36:** 109-117.

Appendix 1

Alignment of the edited pyroglutamyl peptidase sequences for generation for the phylogenetic tree (Fig. 4.22)

```
                1                                                                    70
AthPPI   (1)    IHITGFKKFHGVAENPTEKMANNLKDVNGCTVLETAGQGALASLYQLLQSAVNTKESGKTIWVHFGVGAT
BamPPI   (1)    VLLTGFDPFGGETVNPSWEAVKRLNAAEPASIVSEQVPTVFYKSLAVLREAIKKHQPDIIICVGQAGGRM
CelPPI   (1)    VVVTGFGPFRGFEENPSSIIIDELTGISGVDLELHKITVAYEDVSKKVPELWNEHKPDLVIHLGAHPVEK
CpePPI   (1)    VLITGFDPFGGESINPALEAVKMIPNIEGAQVIKLEIPTVFRKSLEKIEEKIEEINPDVVISIGQAGGRF
CviPPI   (1)    VLLTGFEPFGGETVNPSWEAARQLDTIAGARVHARLLPCEFGAALDELYRQLDALRPDVAIAVGQAGGRP
DmePPI   (1)    IVVSGFGPFLGHAVNASWEAVKLLPHDGIYDLEKRLVSVEYGAVDEAVAEIWKRQNPYLVIHVGVSGVAK
GgaPPI   (1)    VSRQGFGPFRQYLVNSSWEAVKELSGLGNIDLRIMQLPVVYQKAKEQIVRIWTTLHPILTVHVGLASAKA
 HsPPI   (1)    VVVTGFGPFGEHTVNASWIAVQELEGLGDVDLHVYEIPVEYQTVQRLIPALWEKHSPQLVVHVGVSGMAT
LinPPI   (1)    VFITGYGPFATVKVNPSSDIALRVVMRYTLDVSVTTVAAYFEKVEREIADIIAEHGAILLCHLGVHNTTG
LmjPPI   (1)    VFITGYGPFATVKVNPSSDIALRVAVRYTLDVSVTSVAAYFEKVERDTADIIAEHGAILLCHLGVHNTTG
MtuPPI   (1)    VLVTGFGPYGVTPVNPAQLTAEELDTIAGATVISRIVPNTFFESIAAAQQAIAEIEPALVIMLGEYPGRS
PflPPI   (1)    VLLTGFEPFDQDPVNPSWEAVRQLDQLGSVKIVARRLPCAFATAGECLTRLIDELHPAMVIATGLGPGRS
PfuPPI   (1)    VLVTGFEPFGGEKINPTERIAKDLDKIGDAQVFGRVLPVVFGKAKEVLEKTLEEIKPDIAIHVGLAPGRS
RsoPPI   (1)    VLITGIEPFESDPTNPSWDIARALDRIDGATLVARQLPCVFGRANRELVAAIEATQPSLVLALGLASGRS
SauPPI   (1)    ILVTGFAPFDNQDINPSWEAVTQLEIIGTHTIDKLKLPTSFKKVDTIINKTLASNHYDVVLAIGQAGGRN
ScePPI   (1)    VLITGFAPFGGERVNPSWQAASLVAPPAGLAVTAAELPCVFGESLDALRDAIRADNPDLVLCLGQAGGRP
SpyPPI   (1)    ILVTGFDPFGGEAINPALEAIKKLPTIHGAEIKCIEVPTVFQKSADVLQQHIESFQPDAVLCIGQAGGRT
SsoPPI   (1)    VLLFGFEPFLEYKENPSEQLIVEALNSTLEEVKGVILPVEYEKIEDLIVTKIREMKPILTLGIGVAPGRA
  TbPPI  (1)    LYITGYGPFLEVTENPSATIAQSVAVHHELDVNLEAVSKYFNRLNESVTAHLEATHPVLLVNVGLHSKEK
  TcPPI  (1)    LYMTGFGPFKSITVNPSAMIGESVSVHYELEVGLEAVSAYFDGLEGALAQHMEDNPDVLLLHIGLHSREE
TdePPI   (1)    ILVTGFDPFGGEKINPALETIKLLPEILGAKIIKLEIPTVIGKSVAKIKDMIEKENPDVVLSIGQAGNRA
TliPPI   (1)    VLITGFEPFGGDSKNPTEQIAKYFDQIGNAMVYGRVLPVSVKRATIELKRYLEEIKPEIVINLGLAPTYS
VvuPPI   (1)    VLLTGFEPFGGESINPSLELVKQMAALPQVEIIGCEVPVVRYQAIETVLQAVETHQPDLVLMIGQASGRC
XtrPPI   (1)    VVVTGFGPFGEHTVNASWVAVQELGGLGEVDLHIYEIPVEYQTVQRLIPALRKKHKPKVIVHVGVSGMAT
                71                                                                  140
AthPPI   (71)   KFAIEQQAVNTIEFRCPDELGWKPQNLPIVPSGSTVRKTNLPVEEITKALEKNGFEVISDDAGRFVCNYV
BamPPI   (71)   QITPERVAINLEDARIPDNEGNQPVGEDISQGGPAAYWTGLPIKRIVEEIKKEGINPASYTAGTFVCNHL
CelPPI   (71)   TIKFEQQAFSVNGYCSNDVNGCTPADNKTSCSSHETLVSCINCDFLVTKVTEKCALDGSEDPGRYLCGFS
CpePPI   (71)   GVTPERVAINMDDARIEDNEGNQPIDISIYEDGESAYFSNLPIKAMVKEMVDNGIPASSNTAGTFVCNHV
CviPPI   (71)   DIAVERVAINV-DARFPDNAGRQPIDEPVVANGPAAYFATLPIKAIVAGLRERCGLPASQSAGTFVCNHV
DmePPI   (71)   CVYIEKLAYNYHKFRRADNDKKLANGTCELPNNANVLKTELDVDKIVAVVNENCADCVSKNVGDFLCGYI
GgaPPI   (71)   IIILEQCGKNTKGYEEMDACGFHPEGGCCMLNGPEKIESTINMKTIWKNISVEGIDTISRDAGRYICDYT
 HsPPI   (71)   TVTLEKCGHNTKGYKGLDNCRFCPGSQCCVEDGPESIDSIIDMDAVCKRVTTLGLDVSSQDAGRYLCDFT
LinPPI   (71)   LIRVEVQGYNEASVPDVDGKVLNGHEPIVEDGVTGAVTKNEVTPSDADRKDMAMPAFASRNAGRYLCNCA
LmjPPI   (71)   LICVEVQGCNESSVPDVDGKVLNGHEPIVEDGVTGAVTNNEVTSADTDRKDMAMPTFASRNAGRYLCNCA
MtuPPI   (71)   MITVERLAQNVNRYGLADCAGRVLVGEPTDPAGPVAYHATVPVRAMVLAMRKAGVLPASDAAGTFVCNHL
PflPPI   (71)   DISVERVAININDARIPDNLGEQPIDTAVVADGPAAFFTTLPIKAMVKAVREAGNIAASQTAGTFVCNQV
PfuPPI   (71)   AISIERIAVNAIDARIPDNEGKKIEDEPIVPGAPTAYFSTLPIKKIMKKLHERGIPARSNSAGLYLCNYV
RsoPPI   (71)   ELSVERVAINVIDARIPDNAGNQPVDVPVVADGPAAYFSSLPIKAIVHALRAAGLVPASQSAGTYNCNHL
SauPPI   (71)   AITPERVAINIDDARIPDNDDFQPIDQAIHLDGAPAYFSNLPVKAMTQSIINQGLAPGSNSAGTFVCNHV
ScePPI   (71)   GVTVERVGINVDDARIPDNAGGQPIDEPVVPDGPAAYFSTLPVKACVAAMREAGGVPASNTAGTFVCNHV
SpyPPI   (71)   GLTPERVAINQDDARIPDNEGNQPIDTPIRADGKAAYFSTLPIKAMVAAIHQAGLQPASNTAGTFVCNHL
SsoPPI   (71)   KITPEKIAINYKYSREGDNAGKKYKGEKIDPLGQDGIFTNIPVEDLVDLLNENGIQPASLSAGSYLCNNA
  TbPPI  (71)   VLRLEVRAFNELGNPIDDELPLSTCKDSAFVKGCKLETTTALIEELNAIERNGSDHHISYDAGRYYCNYA
  TcPPI  (71)   KMRLEVCGYNELGSPIDESIPMDVCLESMLVRGCESKATAELIEELNAAISKRGRREVSRDAGRYYCNYT
TdePPI   (71)   DISVERIGINIDDCRIPDNEGNQPIDEPVVKDGPAAYFVTLPIKAIVEKVKAGKIYPASNTAGTFICNHV
TliPPI   (71)   NITVERIAVNIIDARIPDNDGYQPIDEKIEEDAPLAYMATLPVRAITKTLRDNGIPADSYSAGTYLCNYV
VvuPPI   (71)   AITPERVAINLDDYRIEDNAGHQPVDEPIIATGPAAYFSTLPVKAITHALQQAGIQPCSHSAGTFVCNHL
XtrPPI   (71)   AVTLEKCGHNTDGYQGLDNCEFCPGTQCCVEGGPECLHSVIDIDTVCKRAAEACLDVDSTDAGRYLCDFT
```

```
                141                      166
AthPPI  (141) YYHSLRFAEQRSLFVHVPLFVAVDEE
BamPPI  (141) FYGLMDEISRRGGFIHIPYIPEQTLQ
CelPPI  (141) YFLSLHEDCSKSLFIHVPAFKEECTK
CpePPI  (141) MYGVLYLVNKRAGFIHVPYIPAQVVN
CviPPI  (140) MYGLLHRGGPRAGFIHLPFLPEQAIR
DmePPI  (141) YLKSLDMDRKRSLFVHVPPVDRPFSS
GgaPPI  (141) YYTSLYYGSGRAAFIHVPPLSKSVTA
 HsPPI  (141) YYTSLYQSHGRSAFVHVPPLGKPYNA
LinPPI  (141) LYHALRLQEKYGVFIHVVDPIRGKTE
LmjPPI  (141) LYHALRLQEKYGIFVHVVDPIRGKTE
MtuPPI  (141) MYGVLHHLAQRAGWIHLPCLPSVAAL
PflPPI  (141) FYLLQHALAGRSGFIHVPFLPEQVAG
PfuPPI  (141) MYLSLHHSATMSGFIHVPYIPEQIID
RsoPPI  (141) FYGLMHHIATRGGFIHVPATPELAAR
SauPPI  (141) LYHLGYLQDKRFGFIHVPYIPEQVVG
ScePPI  (141) AYGLGHLIATRGGFAHVPWAPEQVPD
SpyPPI  (141) MYQALYLVDKKAGFMHIPFMMEQVVD
SsoPPI  (141) MYIIIREARKLGGFIHVPLHESYAAR
 TbPPI  (141) LYRGVKMQEAFAVFLHIVASTVVCME
 TcPPI  (141) LYRSLKLQKKFAVFVHVANPFTCNNP
TdePPI  (141) CYGVAHIAAAKSGFIHIPFLPEQVIG
TliPPI  (141) MFKTLHFSKIKAGFIHVPYTPDQVVN
VvuPPI  (141) FYGVQHHLHTRSGFIHIPLLPEQASA
XtrPPI  (141) YYTSLYQSHGRSVFIHVPPLGKPYT-
```

Multiple alignment of pyroglutamyl peptidases from various organisms (edited sequences). The Align X (Vector NTI) program was used. Identical residues are on black background; conservative residues are on grey background. Cpe, *Clostridium perfringens* (GenBank Accession No.: NP 562341); Spy, *Streptococcus pyogenes* (NP 268785); Sau, *Staphylococcus aureus* (NP 375809); Tde, *Treponema denticola* (NP 970791); Bam, *Bacillus amyloliquefaciens* (JX0244), Vvu, *Vibrio vulnificus* (NP 936141), Sce, *Streptomyces ceolicor* (NP 624759); Rso, *Ralstonia solanacearum* (NP 520286); Cvi, *Chromobacterium violaceum* (NP 902846); Pfl, *Pseudomonas fluorescens* (A55583); Mtu, *Mycobacterium tuberculosis* (NP 334743); Pfu, *Pyrococcus furiosis* (NP 579028); Tlit, *Thermococcus litoralis* (CAA74299); Sso, *Sulfolobus solfatanicus* (NP 343028); Ath, *Arabidopsis thaliana* (NP 173758); Cel, *Caenorhabditis elegans* (NP 492491); Dme, *Drosophila melanogaster* (NP 730035); Gga, *Gallus gallus* (NP 418244); Hs, *Homo sapiens* (NP 060182); Xtr, *Xenopus tropicalis* (AAH75524); Lin, *Leishmania infantum* (GeneDB CDS: LinJ34.1440); Lmj, *Leishmania major* (GeneDB CDS: LmjF34.2000); Tb, *Trypanosoma brucei* (GeneDB CDS: Tb04.1H19.1350); Tc, *Trypanosoma cruzi* (GeneDB CDS: Tc00.1047053506635.60).