Characterization of two cDNAs encoding cysteine proteinases from the soybean cyst nematode *Heterodera glycines*

P. E. URWIN, C. J. LILLEY, M. J. McPHERSON *and* H. J. ATKINSON*

Centre for Plant Biochemistry and Biotechnology, The University of Leeds, Leeds LS2 9JT, UK

(*Received 15 November 1996; revised 13 December 1996; accepted 13 December 1996*)

SUMMARY

Two cDNAs encoding cysteine proteinases were isolated from a cDNA library constructed from feeding females of *Heterodera glycines*. The library was screened with a cysteine proteinase gene fragment originally amplified from cDNA of *H. glycines*. Database searches predict that 1 cDNA (*hgcp-I*) encodes a cathepsin L-like proteinase, while the second (*hgcp-II*) encodes a cathepsin S-like enzyme. Both predicted proteins contain a short secretion signal sequence, a long propeptide and a mature protein of 219 amino acids. Southern blot analysis suggests that the cathepsin S-like enzyme, HGCP-II, is encoded by a single-copy gene in contrast to the cathepsin L-like proteinase, HGCP-I which may have 2 homologues. The regions encoding the mature proteinases were cloned into an expression vector and recombinant protein produced in *E. coli*. HGCP-I was shown, after refolding, to cleave the synthetic peptide Z-Phe-Arg-AMC, and this activity could be inhibited by the engineered rice cystatin Oc-IΔD86. HGCP-II showed no activity against the synthetic substrates tested. The knowledge gained from these studies will improve our understanding of plant nematode proteinases and aid the development of a rational proteinase inhibitor-based approach to plant nematode resistance.

Key words: nematode, cysteine proteinase, cDNA, DNA sequence, protein folding.

INTRODUCTION

Proteolytic enzymes can be divided into 4 main groups, namely, serine, cysteine, aspartyl and metallo-proteinases (Rawlings & Barrett, 1993). They are involved in a wide range of cellular processes in eukaryotes such as intra- and extra-cellular protein metabolism and processing of precursor proteins. Particular attention has focused on understanding the roles of proteinases in host–parasite interactions which may include invasion of host tissues, parasite nutrition and evasion of host immune responses (McKerrow, 1989). Whilst proteinases of all 4 classes have been reported from parasitic helminths (Sakanari, 1990) it is the cysteine proteinases that have been studied most extensively. Genes encoding cysteine proteinases with cathepsin B- and L-like activities have been identified from *Haemonchus contortus* (Pratt *et al.* 1990, 1992), *Schistosoma mansoni* (Klinkert *et al.* 1989; Michel *et al.* 1995) and *Fasciola hepatica* (Heussler & Dobbelaere, 1994). A role for cysteine proteinases in digestive processes of parasites is supported by their high level of expression in actively feeding stages of the life-cycle of *H. contortus* (Pratt *et al.* 1990) and localization to the intestine of a number of animal parasites (Chappell & Dresden, 1986; Maki & Yanagisawa, 1986; Smith *et al.* 1993). Cathepsin B-like cysteine proteinase genes, which show distinct patterns of developmental regulation, have also been isolated from the microbivorous nematode *Caenorhabditis elegans* (Larminie & Johnstone, 1996; Ray & McKerrow, 1992) and the expression of at least 1 of these genes is restricted to the intestine (Ray & McKerrow, 1992).

Little is known about the proteinases of plant-parasitic nematodes or their roles in the host–parasite interaction. Cysteine proteinase activity has been identified in homogenates of feeding females of the potato cyst-nematode, *Globodera pallida* (Koritsas & Atkinson, 1994). We have also achieved reduced growth and fecundity of this nematode by expressing an engineered variant of a rice cysteine proteinase inhibitor in transgenic hairy roots (Urwin *et al.* 1995). More recently we have shown the soybean cyst-nematode, *Heterodera glycines*, contains cathepsin L-like cysteine proteinase activity in the intestine of feeding females (Lilley *et al.* 1996). Here we report the isolation and characterization of cDNA clones encoding 2 cysteine proteinases expressed in female *H. glycines*. These are the first proteinase genes to be characterized from any plant-parasitic nematode.

MATERIALS AND METHODS

Collection of nematodes

Populations of *Heterodera glycines* were maintained on soybean plants and harvested at 18–20 days post-sowing as described previously (Lilley *et al.* 1996).

* Corresponding author. Tel: +0113 233 2863. Fax: +0113 233 3144. E-mail: pabhja@leeds.ac.uk

Copyright © 1997 Cambridge University Press
https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

Young female nematodes were handpicked from males and plant debris under a stereo-binocular microscope. Any female nematodes that were observed as gravid were discarded. Collected nematodes were stored at −70 °C.

cDNA library construction and screening

Female *H. glycines* (500 mg) immersed in liquid N_2 were ground to a fine powder with a mortar and pestle. Poly(A)$^+$ mRNA was isolated using a QuickPrep mRNA Purification Kit (Pharmacia, Uppsala, Sweden) according to the instructions supplied by the manufacturer. Double-stranded oligo(T)-primed cDNA was synthesized from 5 μg of purified mRNA using a ZAP-cDNA synthesis kit (Stratagene, Cambridge, UK) and ligated into the *Eco* RI/*Xho* I sites of the λ Uni-ZAP XR vector (Stratagene, Cambridge, UK). The phage were packaged using Gigapack Gold packaging extracts (Stratagene, Cambridge, UK) to give a cDNA library containing $1{\cdot}4 \times 10^6$ primary recombinants.

Plaques were screened by hybridization with a cloned 153 bp fragment of a *H. glycines* cysteine proteinase gene originally obtained from female *H. glycines* cDNA by PCR amplification with consensus oligonucleotide primers (Lilley *et al*. 1996). The probe fragment was labelled with α^{32}P[dCTP] using a Prime-It II kit (Stratagene, Cambridge, UK) and unincorporated nucleotides were removed using a MicroSpin S-300 HR column (Pharmacia, Uppsala, Sweden). The resulting probe was hybridized to plaques on Hybond N$^+$ nylon membranes (Amersham, UK) at 65 °C overnight in 6 × SSC, 5 × Denhardt's solution, 1 mM EDTA, 0·1 % SDS, 0·05 % sodium pyrophosphate. 1 × SSC is 0·15 M NaCl, 0·015 M Na citrate, pH 7·0. The membranes were washed to a final stringency of 0·5 × SSC, 0·1 % SDS at 65 °C.

Positive phage were identified and the recombinant pBluescript phagemids excised using ExAssist^{TM} helper phage and *E. coli* SOLR strain according to the instructions of the manufacturer (Stratagene, Cambridge, UK).

DNA sequencing and analysis

The nucleotide sequence of the plasmid inserts was determined using M13 forward and reverse primers plus gene-specific primers where appropriate and the Taq DyeDeoxy Terminator Cycle Sequencing System (Applied Biosystems) and an automated sequencer (Applied Biosystems 373A). Fragments of the *hgcp-II* gene were subcloned when necessary into the pBluescript vector to enable complete sequence determination. Nucleotide and amino acid sequence analysis was performed using GCG Package software (GCG, Madison, WI, USA), specifically the FASTA program for comparisons with sequences in the GenBank and SwissProt databases.

Southern blot analysis

Genomic DNA was prepared from female *H. glycines* using a Nucleon I DNA extraction kit (Scotlab, Coatbridge, UK) essentially as described by the manufacturer. Aliquots (50 mg) of frozen nematodes were homogenized in microcentrifuge tubes in extraction buffer containing 10 mM β-mercaptoethanol. DNA (4 μg) was digested with either *Bam* HI, *Hin*d III or *Xho* I, separated by electrophoresis through a 0·8 % agarose gel and transferred to Hybond N$^+$ membrane (Amersham, UK) according to Sambrook, Fritsch & Maniatis (1989). Hybridization with ^{32}P-labelled *hgcp-I* and *hgcp-II* cDNA inserts as probes was carried out overnight at 60 °C as described for library screening. Membranes were washed at 60 °C to a stringency of 1 × SSC/0·1 % SDS.

Cloning and expression of H. glycines *cysteine proteinases*

The sequences encoding the predicted mature proteins of HGCP-I and HGCP-II were amplified from the pBluescript clones by the polymerase chain reaction (PCR) using oligonucleotide primers with *Bam* HI and *Hin*d III restriction enzyme sites added to assist cloning into the expression vector, pQE30 (QIAGEN, Dorking, UK). The primers for *hgcp-I* were 5′ATAGGATCCTTGCCGGAATCGGTG-GACTGG3′ and 5′ATAAAGCTTCACTCCGA-TCAGACCAATGGG3′. Primers for *hgcp-II* were 5′ACAGGATCCTTGCCGGAAAAGTTGGAC-TGG3′ and 5′ACAAAGCTTTCTCAGACCACG-GGGTACG3′. The PCR products were digested with *Bam* HI and *Hin*d III and cloned into pQE30 and the 2 mature proteinases were expressed in *E. coli* as fusion proteins containing a 6 × His N-terminus. Purification of native protein was attempted as described previously (Urwin *et al*. 1995) but proved to be unsuccessful.

Purification of the expressed proteinases under denaturing conditions was achieved as follows. A 500 ml culture of *E. coli* M15 harbouring the expression construct was grown at 37 °C with vigorous shaking to $OD_{600} = 0{\cdot}7$–$0{\cdot}9$, induced with 2 mM IPTG and incubated for a further 2 h. Cells were harvested by centrifugation, resuspended in 8 ml of lysis buffer (8 M urea, 0·1 M Na-phosphate, 10 mM Tris-HCl, pH 8·0) and disrupted by freeze/thawing (−70 °C/37 °C). The lysate was centrifuged at 10000 ***g*** for 10 min and the supernatant mixed with 0·5 ml of Ni-NTA resin (QIAGEN, Dorking, UK) for 30 min at room temperature. Unbound proteins were removed with 5 washes in 8 M urea, 0·1 M Na-phosphate, 10 mM Tris-HCl, pH 6·3 and recombinant protein was eluted from the resin using wash buffer supplemented with 100 mM EDTA. Purity of the expressed, recombinant proteinases

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

```
                                                   caaatatttcaattctcccaattttataaagcaagtaaa 39

atgtttcttcttttcttattatcaatgcttctacttcagacaaatggttggcgtgcccgcgagcgtgcaattgaattggccgactcggac 129
M  F  L  L  F  L  L  S  M  L  L  L  Q  T  N  G  W  R  A  R  E  R  A  I  E  L  A  D  S  D   30
                                              ▲

gaatcgatcgaattgcagaacattggccaacggaaaacggacattcgaacaccgacagaacgaatgtctgctcttcgtcaaatgatcgaa 219
E  S  I  E  L  Q  N  I  G  Q  R  K  T  D  I  R  T  P  T  E  R  M  S  A  L  R  Q  M  I  E   60

cgcggcttttccgattggaatgcttacaaacagaagcatgggaaagcatacgcggaccaagaagtggagaacgaacggatgctgacttat 309
R  G  F  S  D  W  N  A  Y  K  Q  K  H  G  K  A  Y  A  D  Q  E  V  E  N  E  R  M  L  T  Y   90
                                                            ●              ●           ●

ttgagcgccaaacagttcattgacaagcacaacgaggcgtacaaagagggcaaagtgtccttccgagtgggagagactcatattgccgac 399
L  S  A  K  Q  F  I  D  K  H  N  E  A  Y  K  E  G  K  V  S  F  R  V  G  E  T  H  I  A  D   120
      ●           ●           ●

ctgcccttttccgaataccaaaagctgaacggattccgtcgtttgatgggcgacagtttgcgccgcaatgcgtccacttttctggcgcca 489
L  P  F  S  E  Y  Q  K  L  N  G  F  R  R  L  M  G  D  S  L  R  R  N  A  S  T  F  L  A  P   150
                                                                  _  _  _

atgaatgtgggcgatttgccggaatcggtggactggcgggacaaaggatgggtgaccgaagtgaaaaaccagggaatgtgcggctcgtgc 579
M  N  V  G  D  L  P  E  S  V  D  W  R  D  K  G  W  V  T  E  V  K  N  Q  G  M  C  G  S  C   180
             ▲                                                                         ★

tgggcattcagtgccaccggcgcattggagggacaacacgtgcgcgacaagggacatcttgtttcactgtcggaacaaaatctgatcgac 669
W  A  F  S  A  T  G  A  L  E  G  Q  H  V  R  D  K  G  H  L  V  S  L  S  E  Q  N  L  I  D   210

tgctcgaagaagtacggaaacatgggctgcaacggaggcatcatggacaacgccttccaatacattaaggacaacaaaggcatcgacaaa 759
C  S  K  K  Y  G  N  M  G  C  N  G [G] I  M  D  N  A  F  Q  Y  I  K  D  N  K  G  I  D  K   240

gagacggcctacccctacaaggccaagaccggcaaaaagtgtttgttcaagcgcaacgacgtgggggcaaccgactcgggttataacgac 849
E  T  A  Y  P  Y  K  A  K  T  G  K  K  C  L  F  K  R  N  D  V  G  A  T  D  S  G  Y  N  D   270

atagccgaaggggacgaggaggacctgaagatggctgttgcaacgcaagggcccgtctcagttgccattgatgctggtcaccgttccttc 939
I  A  E  G  D  E  E  D  L  K  M  A  V  A  T  Q  G  P  V  S  V  A  I  D  A  G  H  R  S  F   300

caattgtacaccaacggcgtttactttgagaaggaatgcgacccggaaaatttggaccatggtgtgctcgtggtgggctacggcaccgac 1029
Q  L  Y  T  N  G  V  Y  F  E  K  E  C  D  P  E  N  L  D  H  G  V  L  V  V  G  Y  G  T  D   330
                                                         ★

ccaacccaaggcgactattggattgtgaagaacagctggggcacccgctggggcgagcagggatacattcgcatggcacgcaatcgcaac 1119
P  T  Q  G  D  Y  W  I  V  K  N  S  W  G  T  R  W  G  E  Q  G  Y  I  R  M  A  R  N  R  N   360
                              ★

aacaattgcggcatcgcttcccacgcctctttcccattggtctgatcggagtgaatttgttgcccttgcgctgattcagagacatttcat 1209
N  N  C  G  I  A  S  H  A  S  F  P  L  V  *

ttgattaatcgtgcaaaatgataagataattgataatccatcagtcaatcggtcgatttccattttttatgttcgcaattttattcacat 1299

ataaataaattacttattttaa(47)
   ______
```

Fig. 1. Nucleotide and deduced amino acid sequence of the *hgcp-I* cDNA clone. The predicted sites of cleavage of the signal peptide and pro-sequence are indicated with arrows. The 3 residues involved in the active site (Cys, His, Asn) are indicated with asterisks and the conserved glycine residue discussed in the text is boxed. The residues of the interspersed ERFNIN motif are marked with ●. The putative polyadenylation signal is underlined. A dotted line highlights a potential N-glycosylation site. Sequence data have been submitted to GenBank under accession number Y09498.

was assessed by SDS–PAGE using a mini-gel system (BioRad).

Protein renaturation and characterization of activity

Purified, denatured HGCP-I and HGCP-II were exposed to refolding conditions by adding 100 μl of the protein sample as an aerosol from a syringe fitted with a 23 gauge needle into 10 ml of each of 3 buffers, stirring rapidly. The buffers were (i) 50 mM 2-[*N*-morpholino]ethanesulfonic acid (MES), 2 mM DTT, 2 mM EDTA, pH 6·0; (ii) 100 mM Na phosphate, 3 mM DTT, 2 mM EDTA, pH 6·0; (iii) 50 mM Na acetate, 200 mM NaCl, 1 mM EDTA, pH 5·0. The activity of the diluted protein was assayed immediately using the synthetic peptide substrates N-CBZ-Phe-Arg-7-amido-4-methylcoumarin (Z-Phe-Arg-AMC) and Nα-benzoyl-Arg-β-naphthylamide (BANA) (Sigma Chemical Co., Poole, UK). Z-Phe-Arg-AMC was prepared as a 20 mg/ml stock solution in methanol and diluted 500-fold in buffer (i) immediately prior to use. In a microtitre plate, 50 μl of 'refolded' protein solution or papain (100 ng/ml) was pre-warmed at 37 °C for 5 min prior to the addition of 50 μl of substrate solution. Incubation was continued for a further 30 min and fluorescence of the released aminomethylcoumarin was visualized using a UV transilluminator. Where appropriate, 1 μg of the modified rice cystatin Oc-IΔD86 (Urwin *et al.* 1995) or the cowpea trypsin inhibitor, CpTI, was included in the pre-incubation step. Assays using BANA as the substrate were carried out as described by Barrett (1972).

RESULTS

Isolation of H. glycines *cysteine proteinase cDNA clones*

Approximately 500000 recombinant plaques were

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

```
                                                       catttgtcagcccatcccaaaaactgataaaaaaa 35

atggcgttcctctcccgtctctccatccttccaaattcgccaatttcgctgcttgcagtttctttggccgttttggctttcgtcgctttg 125
M  A  F  L  S  R  L  S  I  L  P  N  S  P  I  S  L  L  A  V  S  L  A  V  L  A  F  V  A  L    30

gcatcggccaatccgccgacggcacgcgaaacggcgccaaatgcacagcaaaacaatgccaattcagtggcaactggggaaattgcgaaa 215
A  S  A  N  P  P  T  A  R  E  T  A  P  N  A  Q  Q  N  N  A  N  S  V  A  T  G  E  I  A  K    60
       ▲

aatattgcggaaaagatggagcggatgaatgagttcattaaggcgaagaagttcatcgatgcacataatttggcatttgagaagggcgaa 305
N  I  A  E  K  M  E  R  M  N  E  F  I  K  A  K  K  F  I  D  A  H  N  L  A  F  E  K  G  E    90
         ●           ●           ●        ●           ●           ●

gtgtcgttcaaagttgcgccaaaccatctgatgcattttacacctgcccaatataatcgaattcgcggcttgcaaatgcgcagcaaccga 395
V  S  F  K  V  A  P  N  H  L  M  H  F  T  P  A  Q  Y  N  R  I  R  G  L  Q  M  R  S  N  R   120

caacggcacaacatggcaactctggcgggaaacagcagtactttgccggaaaagttggactggcgcgagaaaggggcggtgaccgaggtc 485
Q  R  H  N  M  A  T  L  A  G  N  S  S  T  L  P  E  K  L  D  W  R  E  K  G  A  V  T  E  V   150
                              -  -  -    ▲

aaagatcaggggactgcggctcgtgttgggcattcagtgccaccggtgccattgagggagcattggcacagaaaaaagcgtcgaaaatt 575
K  D  Q  G  D  C  G  S  C  W  A  F  S  A  T  G  A  I  E  G  A  L  A  Q  K  K  A  S  K  I   180
                        ★

atttcattgtccgaacaaaacctggtcgactgttcgtccaagtacggtaacgagggctgtgacggtggactgatggacagcgcatttgaa 665
I  S  L  S  E  Q  N  L  V  D  C  S  S  K  Y  G  N  E  G  C  D  G  G  L  M  D  S  A  F  E   210

tatgtgcgagacaacaacgggttggacacggaggagtcgtacccgtacgaggccgtaacgggcaaatgccaattcaaaaatgagaccgtg 755
Y  V  R  D  N  N  G  L  D  T  E  E  S  Y  P  Y  E  A  V  T  G  K  C  Q  F  K  N  E  T  V   240
                                                                               -  -  -

ggcggcactgtcgttagcttcaaagacttgaagaaaggcgacgaagagcagctgaagattgccgtcgccacaattgggcccatttccgtt 845
G  G  T  V  V  S  F  K  D  L  K  K  G  D  E  E  Q  L  K  I  A  V  A  T  I  G  P  I  S  V   270

gcgctcgatgccagcaatttgtccttccaattttacaaaaccggcgtttattacgagcggtggtgcagcaaccgatacttggaccacggc 935
A  L  D  A  S  N  L  S  F  Q  F  Y  K  T  G  V  Y  Y  E  R  W  C  S  N  R  Y  L  D  H  G   300
                  -  -  -                                                             ★

gttctcctcgtcggctacggtaccgacgaaacgcacggtgactattggctggtgaagaacagttggggcccgcattggggagagaacggt 1025
V  L  L  V  G  Y  G  T  D  E  T  H  G  D  Y  W  L  V  K  N  S  W  G  P  H  W  G  E  N  G   330
                                                         ★

tacattcgaattgcgcgcaacaaacaaaaccattgtggcattgcgacgatggcatcgtaccccgtggtctgagaaagcgtgggaatgaat 1115
Y  I  R  I  A  R  N  K  Q  N  H  C  G  I  A  T  M  A  S  Y  P  V  V  *

gggacgagaagggatcagaagaagaagcaggcagaccaaatagaagcaattcacaatcattatcattgttatgctttttggtaataaata 1205
                                                                                  ------
                                                                                      -----

aaattgctttgtaa(18)
-
```

Fig. 2. Nucleotide and deduced amino acid sequence of the *hgcp-II* cDNA clone. The most likely signal peptide cleavage site and the N-terminus of the mature enzyme are indicated with arrows. The 3 residues involved in the active site (Cys, His, Asn) are indicated with asterisks and the conserved glycine residue discussed in the text is boxed. The residues of the interspersed ERFNIN motif are marked with ●. Two putative polyadenylation signals are underlined. A dotted line highlights potential N-glycosylation sites. Sequence data have been submitted to GenBank under accession number Y09499.

screened with the cloned 153 bp *H. glycines* cysteine proteinase PCR fragment (Lilley *et al.* 1996), resulting in > 200 positive clones representing > 0·04 % of the library. Ten clones were selected for further characterization and the corresponding pBluescript plasmids recovered by *in vivo* excision. cDNA inserts ranging in size from 1 to 1·4 kbp were completely sequenced. The 10 clones represented two distinct cysteine proteinase genes which we have named *hgcp-I* and *hgcp-II*. Nine clones encoded HGCP-I and 1, more weakly hybridizing clone, encoded HGCP-II.

The full-length *hgcp-I* clone (1368 bp) has an open reading frame of 374 amino acids (aa) (Fig. 1). The *hgcp-II* clone is 1228 nucleotides in length and encodes a polypeptide of 353 aa (Fig. 2). Potential polyadenylation signals (AATAAA) were identified in both transcripts. *Hgcp-I* and *hgcp-II* were compared with sequences in the GenBank database and the highest homologies were to cysteine proteinases. As with most cathepsins, the *H. glycines* cysteine proteinases appear to be synthesized as precursor molecules in a prepro-format (Erickson, 1989). Weight matrix calculations based on the algorithm determined by von Heijne (1986) allow prediction of the most likely cleavage site for the N-terminal secretion signal sequence. In the case of HGCP-I there is a single strongly predicted signal sequence cleavage site after Gly16. Based on comparison with other cathepsins the most likely pro-sequence cleavage site is after Asp155 which indicates a pro-sequence of 139 aa followed by a mature protein of 219 aa of molecular mass 28130 kDa. For HGCP-II the prediction of signal sequence cleavage site is not so certain with 3 very strong predictions for residue -1 as Ala26, Ala31 and Ala33. The latter of these is predicted most strongly and sequence comparisons with other cathepsins also indicate Ala33 as the preferred choice. The pro-sequence may therefore extend from either Phe27, Ser32 or Asn34 to Thr134 resulting in signal/pro-sequence lengths of either 26/108, 31/105 or 33/103 respectively. The mature

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

```
HGCP-I     MFLLFLLSML LLQTNG.... ...WRARERA IELADSDESI ELQNIGQRKT DIRTPTERMS ALRQMIERGF  63
HGCP-II    MAFLSRLSIL PNSPI.SLLA VSLAVLAFVA LASA...... .......... ........NP PTARETAPNA  45
LOB2       MKVAVLFLCG VALAAAS... .......... .......... .......... .......... ..........  17
HUMCATL    MNPTLILAAF CLGIA.SA.. .......... .......... .......... .......... ..TLTFDHSL  25
HUMCATS    MKRLVCVLLV CSS....... .......... .......... .......... .......... VAQLHKDPTL  23
FHEPL      MRFFVLAVLT VG VFA.... .......... .......... .......... .......... ........SN  17
SMANSL     MKVFLLLFSI IISV ..... .......... .......... .......... .......... AIAQHLSLQY  24

HGCP-I     S.DWNAYKQK HGKAYADQEV ENERMLTYLS AKQFIDKHNE AYKEGKVSFR VGETHIADLP FSEYKLNGFR 132
HGCP-II    Q.QNNANSVA TGEIAKNIAE KMERMNEFIK AKKFIDAHNL AFEKGEVSFK VAPNHLMHFT PAQYNRIRGL 114
LOB2       P.SWEHFKGK YGRQYVDAEE DSYRRVIFEQ NQKYIEEFNK KYENGEVTFN LAMNKFGDMT LEEFNA....  82
HUMCATL    EAQWTKWKAM HNRLY.GMNE EGWRRAVWEK NMKMIELHNQ EYREGKHSFT MAMNAFGDMT SEEFRQ....  90
HUMCATS    DHHWHLWKKT YGKQYKEKNE EAVRRLIWEK NLKFVMLHNL EHSMGMHSYD LGMNHLGDMT SEEVMS....  89
FHEPL      DDLWHQWKRI YNKEYNGADD E.HRRNIWGK NVKHIQEHNL RHGLGLVTYK LGLNQFTDLT FEEFKAKYLI  86
SMANSL     DDIWKQWKLK YNKTYSDSNE I.RRKAIFMR YVEKIQQHNL RHDLGLEGYT MGLNQFCDMD WEEIKTIMLS  93

HGCP-I     RLMGDSLRRN A..STFLAPM NVGDLPESVD WRDKGWVTEV KNQGMCGSCW AFSATGALEG QHVRDKGH.L 199
HGCP-II    QMRSNRQRHN MAT....LAG NSSTLPEKLD WREKGAVTEV KDQGDCGSCW AFSATGAIEG ALAQKKASKI 180
LOB2       VMKGNIPRRS APVSVFYPKK ETGPQATEVD WRTKGAVTPV KDQGQCGSCW AFSTTGSLEG QHFLKYGS.L 151
HUMCATL    VMNGFQNRKP RKGKV.FQEP LFYEAPRSVD WREKGYVTPV KNQGQCGSCW AFSATGALEG QMFRKTG.RL 158
HUMCATS    LMSSLRVPSQ WQRNITYKSN PNRILPDSVD WREKGCVTEV KYQGSCGACW AFSAVGALEA QLKLKTG.KL 158
FHEPL      EIPRSSELL. SRGIPY..KA NKLAVPESID WRDYYYVTEV KDQGQCGSCW AFSTTGAVEG QF.RKNERAS 152
SMANSL     KVFGNSPLWD DKKEEL..EL SNDPLPSKWD WRDHGAVTPV KNQGLCGSCW AFSAAGAVEG QL.VKKHKKL 160

HGCP-I     VSLSEQNLID CS.KKYGNMG CNGGIMDNAF QYIKDNKGID KETAYPYKAK TGKKCLFKRN DVGATDSGYN 268
HGCP-II    ISLSEQNLVD CS.SKYGNEG CDGGLMDSAF EYVRDNNGLD TEESYPYEAV TG.KCQFKNE TVGGTVVSFK 248
LOB2       ISLAEQQLVD CS.RPYGPQG CNGGWMNDAF DYIKANNGID TEAAYPYEAR DG.SCRFDSN SVAATCSGHT 219
HUMCATL    ISLSEQNLVD CSGPQ.GNEG CNGGLMDYAF QYVQDNGGLD SEESYPYEA. TEESCKYNPK YSVANDTGFV 226
HUMCATS    VSLSAQNLVD CSTEKYGNKG CNGGFMTTAF QYIIDNKGID SDASYPYKA. MDQKCQYDSK YRAATCSKYT 227
FHEPL      ASFSEQQLVD CT.RDFGNYG CGGGYMENAY EYLKHN.GLE TESYYPYQAV EG.PCQYDGR LAYAKVTGYY 219
SMANSL     ISLSEQQLVD CS.YKYGNDG CQGGTMDQSF AYLEKY.PIE SEKDYKYIGH DS.SCHFRKS KGVVKVKKFV 227

HGCP-I     DIAEGDEEDL KMAVATQGPV SVAIDAGHRS FQLYTNGVYF EKECDPENLD HGVLVVGYG. ..TDPTQGDY 335
HGCP-II    DLKKGDEEQL KIAVATIGPI SVALDASNLS FQFYKTGVYY ERWCSNRYLD HGVVVVGYG. ..TDETHGDY 315
LOB2       NIASGSETGL QQAVRDIGPI SVTIDAAHSS FQFYSSGVYY EPSCSPSYLD HAVLAVGYG. ..SEGGQ.DF 286
HUMCATL    DIP.KQEKAL MKAVATVGPI SVAIDAGHES FLFYKEGIYF EPDCSSEDMD HGVLVVGYGF ESTESDNNKY 295
HUMCATS    ELPYGREDVL KEAVANKGPV SVGVDARHPS FFLYRSGVYY EPSC.TQNVN HGVLVVGYG. ...DLNGKEY 292
FHEPL      TVHSGDEIEL KNLVGTEDLP AVALDA.DSD FMMYQSGIYQ SQTCLPDRLT HAVLAVGYC. ...SQDGTDY 284
SMANSL     DLPARDEEKL QKALYHYGPI SVAIDA.LDD LILYKSGIYE SKQCSSFLLN HGVLAVGYG. ...RENRKDY 292

HGCP-I     WIVKNSWGTR WGEQGYIRMA RNRNNNCGIA SHASFPLV.. .. 373
HGCP-II    WLVKNSWGPH WGENGYIRIA RNKQNHCGIA TMGSYPVV.. .. 353
LOB2       WLVKNSWATS WGDAGYIKMS RNRNNNCGIA TVASYPLV.. .. 324
HUMCATL    WLVKNSWGEE WGMGGYVKMA KDRRNHCGIA SAASYPTV.. .. 333
HUMCATS    WLVKNSWGHN FGEEGYIRMA RNKGNHCGIA SFPSYPEI.. .. 330
FHEPL      WIVKNSWGTW WGEDGYIRFA RNRGNMCGIA SLASVPMVAR FP 326
SMANSL     WLIKNSWGTT WGMNGYFKLR RNKHNMCGIA TNASFPLL.. .. 330
```

Fig. 3. Alignment of the predicted HGCP-I and HGCP-II polypeptides with amino acid sequences of lobster digestive proteinase 2 (LOB2), human cathepsin L (HUMCATL), human cathepsin S (HUMCATS), *Fasciola hepatica* cathepsin L (FHEPL) and *Schistosoma mansoni* cathepsin L (SMANSL). Residues conserved in all 7 sequences are shaded.

proteinase sequence is 219 aa with a predicted molecular mass of 27981 kDa.

The divergent pro-regions both contain homologues of the interspersed ERFNIN motif present in cathepsin L, H and S classes but absent from cathepsin B-like enzymes (Karrer, Peiffer & DiTomas, 1993). The complete amino acid sequences of HGCP-I and HGCP-II have 49% identity and 67% similarity. This rises to 63% identity and 81% similarity if comparison is restricted to the more highly conserved mature enzyme regions. Over the mature proteinase region HGCP-I displays 63·8% identity to chicken cathepsin L, 60·6% identity to human cathepsin L and 60·6% identity to a digestive cysteine proteinase from the American lobster (*Homarus americanus*). HGCP-II is most similar to bovine cathepsin S (58·6% identity) and human cathepsin S (55·9% identity). Fig. 3 shows an alignment between HGCP-I, HGCP-II and other cathepsin L and S-like cysteine proteinases. Amino acid homology within the mature proteinases is centered around the definitive catalytic triad residues of Cys, His and Asn (denoted by asterisks in Figs 1 and 2), characteristic of all cysteine proteinases. In addition to the active site cysteine, both the cysteine proteinases of *H. glycines* contain 6 further conserved cysteine residues which could form disulphide bridges between positions 177 and 220, 211 and 254 and between cysteine residues 313 and 363 (numbering for HGCP-I in Fig. 1). The

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

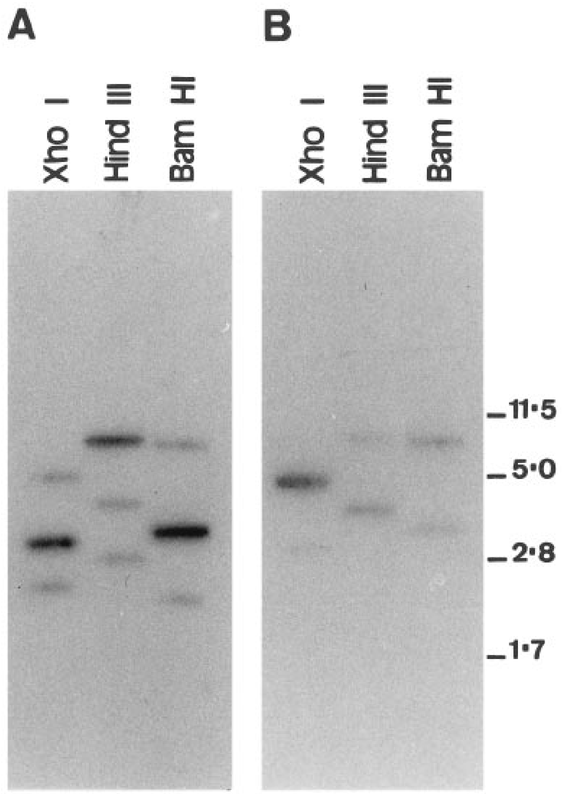

Fig. 4. Southern blot analysis of *Heterodera glycines* genomic DNA (4 μg in each lane) digested with *Xho* I, *Hin*d III or *Bam* HI. The filter in (A) was hybridized with the ^{32}P-labelled *hgcp-I* cDNA then stripped and rehybridized with the ^{32}P-labelled *hgcp-II* cDNA (B). The positions of DNA size markers (kbp) are indicated.

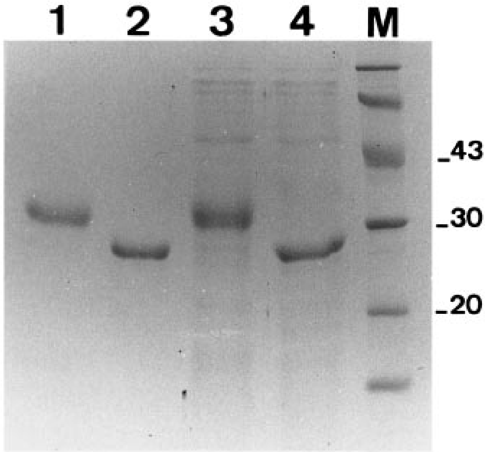

Fig. 5. SDS–PAGE of recombinant HGCP-I and HGCP-II. Lane 1, purified HGCP-II; Lane 2, purified HGCP-I; Lane 3, total extract of *E. coli* expressing HGCP-II; Lane 4, total extract of *E. coli* expressing HGCP-I. M = molecular weight markers, sizes indicated are in kDa.

assignment of the 3 disulfide bridges is based on structural data for papain (Drenth *et al*. 1970) and is characteristic of cathepsin L and S-like enzymes. By contrast cathepsin B enzymes have the potential to form at least 6 disulfide connections (Musil *et al*. 1991). The glycine residue at position 223 of HGCP-I and position 203 of HGCP-II is another conserved amino acid. This is involved in substrate binding in human cathepsin L (Joseph *et al*. 1988). Both proteinases have a potential N-glycosylation site within the pro-peptide region and there are a further 2 within the mature protein of HGCP-II only.

Southern blot analysis

Probes made from the full-length *hgcp-I* and *hgcp-II* cDNA inserts exhibited similar patterns of hybridization to genomic DNA of *H. glycines* on Southern blots but showed differential intensities of hybridization. It was expected that the *hgcp-I* probe would hybridize to *hgcp-II* DNA under the conditions used for blot hybridization and washing since both clones were isolated from a single screening of the cDNA library. Similarly, the *hgcp-II* probe should hybridize to *hgcp-I* DNA. Thus the more intense bands in Fig. 4A probably represent *hgcp-I* sequences, whilst the more strongly hybridizing bands in Fig. 4B represent *hgcp-II* gene fragments. While the general level of hybridization with the *hgcp-II* probe is weaker than with *hgcp-I* due to reprobing of the filter, this does not account for the absence of the third set of hybridization signals observed with the *hgcp-I* probe.

Expression of recombinant H. glycines *cysteine proteinases*

The expressed cysteine proteinases remained in the insoluble cellular fraction and were resistant to solubilization with mild detergent (0·25% Tween-20, 0·1 mM EGTA). Performing cell lysis and protein purification steps in the presence of 8 M urea allowed purification of both HGCP-I and HGCP-II mature enzymes in a denatured form (Fig. 5). Interestingly, the apparent molecular weight of HGCP-II is somewhat larger than the calculated molecular weight and the protein appears as a diffuse band following SDS–PAGE although the reason for this anomaly is unclear.

Characterization of enzyme activity

In an attempt to characterize the activity of the *H. glycines* cysteine proteinases against synthetic substrates it was necessary to refold the denatured, purified proteins into a native, active form. This was achieved for HGCP-I by rapid dilution of the protein from the 8 M urea solubilization buffer into an 'activation buffer' providing optimal conditions for cysteine proteinase activity. This refolding step is only likely to lead to correct folding of a low proportion of HGCP-I molecules but this should be sufficient to allow assays to define the substrate spectrum. The resulting enzyme preparation was capable of cleaving the synthetic peptide Z-Phe-Arg-AMC which is a substrate for both cathepsin B and L-like cysteine proteinases and for certain trypsin-like enzymes (Barrett & Kirschke, 1981). This proteolytic activity was only detectable when the MES and sodium phosphate buffers were used for rapid dilution of the denatured protein. The activity of HGCP-I could be inhibited by the engineered rice cysteine proteinase inhibitor, Oc-IΔD86, but not by the serine proteinase inhibitor, CpTI. In contrast to papain, that was used as the control enzyme, HGCP-I was unable to cleave BANA which acts as a substrate for cathepsin B but not L-like cysteine proteinases. This result agrees with the interpretation of the sequence homology data and confirms that HGCP-I is a cathepsin L-type proteinase. Using identical conditions for the refolding of HGCP-II we were unable to detect any activity against either Z-Phe-Arg-AMC or BANA.

DISCUSSION

In the past few years cysteine proteinase genes have been cloned from several parasitic helminths and protozoa and a number of roles have been proposed for them in host–parasite interactions. Here we report the first cysteine proteinase cDNA sequences from a plant-parasitic nematode, the soybean cyst-nematode, *H. glycines*.

Determination of the first parasite cysteine proteinase sequences suggested that protozoan parasites contained cathepsin L-like enzymes whilst those of helminths were cathepsin B-like (Michel *et al*. 1995). However, a cathepsin B-related proteinase has now been reported from *Leishmania mexicana* (Robertson & Coombs, 1993) and cathepsin L-like genes have been identified from the trematodes *Fasciola hepatica* (Heussler & Dobbelaere, 1994) and *Schistosoma mansoni* (Michel *et al*. 1995; Smith *et al*. 1994). We have now identified cDNAs encoding both a cathepsin L-like proteinase (*hgcp-I*) and most probably a cathepsin S-like enzyme (*hgcp-II*) from *H. glycines*. To our knowledge this is the first report of nematode genes encoding these classes of cysteine proteinases.

The cathepsins B, L, S and H together with papain, cruzipain and a number of other cysteine proteinases have all been grouped into the same molecular family; the members of which share a close evolutionary and structural relationship (Rawlings & Barrett, 1993). In addition, enzymes of this group have been further classified into ERFNIN and cathepsin B-like cysteine proteinases based on the presence of a highly conserved interspersed amino acid motif in the pro-peptide region of all except the cathepsin B-like enzymes. Although neither enzyme contains all 6 of the conserved amino acids, both HGCP-I and HGCP-II can be classified as ERFNIN proteinases. In HGCP-I the phenylalanine is replaced by another amino acid with an aromatic ring, tyrosine, and in both HGCP-I and HGCP-II the first asparagine is replaced by an alanine. Whilst this is not a conservative change, an Asn to Ala substitution also occurs in the ERFNIN motif of the *Trypanosoma brucei* cysteine proteinase (Mottram *et al*. 1989).

Sequence analysis of HGCP-I suggests it is most similar to cathepsin L-like cysteine proteinases and this assignment is confirmed by the substrate specificity of the expressed, recombinant enzyme. Activity is only observed against Z-Phe-Arg-AMC, a substrate for cathepsins B, L and S and not against BANA which is only cleaved by cathepsin B-like cysteine proteinases. HGCP-II is most homologous to human and bovine cathepsin S cysteine proteinases. Within the papain family, cathepsin S enzymes appear to be most closely related to cathepsin L. The classification of HGCP-II as a cathepsin S-like cysteine proteinase can only be provisional without biochemical evidence from assays involving refolded enzyme. No cathepsin S-like activity has previously been reported from a parasite but this may be due to difficulty in distinguishing it from cathepsin L. Both enzymes have high endopeptidase activity against native protein substrates and similar specificities for synthetic substrates. The main distinguishing characteristic is the stability of cathepsin S at pH 7·5, whereas incubation above pH 6·5 will destroy cathepsin L activity (Kirschke & Wiederanders, 1994).

Southern blot analysis suggests that *H. glycines* may have more than 2 cysteine proteinase genes. The most likely explanation for the additional faint bands identified by the *hgcp-I* probe in Fig. 4A is the existence of a third *H. glycines* cysteine proteinase gene with homology to *hgcp-I* but not *hgcp-II*. However, the presence of *Bam* HI, *Hin*d III and *Xho* I restriction sites within the genomic DNA region at the *hgcp-I* locus cannot be discounted. Sequencing of an amplified fragment of *hgcp-I* genomic DNA suggests there are at least 7 introns within the *hgcp-I* gene (data not shown) so these may contain restriction sites not present within the cDNA clone.

The roles of the 2 proteinases *in vivo* have yet to be established. Some animal parasites produce several cysteine proteinases and each may have specialized roles at different stages in the life-cycle. The cathepsin L-like enzymes of *F. hepatica* have been localized to epithelial cells of the intestine (Smith *et al*. 1993) and may have a role in nutrition. This may also be true for the cathepsin B of *S. mansoni* which is secreted into the gut lumen (Chappell & Dresden, 1986) whereas its cathepsin L-like proteinase occurs in the reproductive system of both sexes (Michel *et al*. 1995). The high number of hybridizing plaques suggests that *hgcp-I* is an abundant transcript and may provide the cathepsin L-like activity of the

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

intestine (Lilley *et al.* 1996). If so, it is a good target for an anti-feedant approach to nematode control. The low abundance of *hgcp-II* clones may imply it has a distinct role from that of HGCP-I.

This work indicates that the abundant HGCP-I should provide the focus for future work aimed at disrupting feeding of *H. glycines*. We have previously demonstrated that engineering the rice cysteine proteinase inhibitor, Oc-I, improves both its inhibitory activity and its efficacy as an anti-nematode protein expressed in transgenic plants (Urwin *et al.* 1995). Engineered variants of Oc-I were initially selected by their improved inhibition of papain (Urwin *et al.* 1995). The use of HGCP-I in future assays could help identify both natural and engineered cystatins with enhanced potential against *H. glycines*. Following refolding of HGCP-I we detected proteinase activity which could be inhibited by the engineered rice cystatin Oc-I*Δ*D86 indicating that this inhibitor, previously shown to reduce growth and fecundity of *G. pallida* on transgenic hairy roots, also has potential for control of *H. glycines*. In the present study only the regions encoding the mature proteinases were expressed, since the N-termini of the mature proteins could be predicted with more certainty from early sequence data than the N-termini of the pro-enzymes. In addition, the principal application of the expressed protein is to be in antibody production to allow localization of the enzymes within the nematode. Both HGCP-I and HGCP-II aggregated in the cytoplasm of *E. coli* and could only be solubilized with strong denaturants. It is likely that the activity we observed with recombinant HGCP-I was due to a very small proportion of the molecules adopting the correct conformation. Loss of activity and precipitation of HGCP-I was observed following overnight storage, reflecting the instability of the protein in these conditions. Previous work has demonstrated that the pro-sequence of human cathepsin L is essential for correct folding and/or processing of the molecule (Smith & Gottesman, 1989). It is likely that the lack of activity observed with recombinant HGCP-II was due to its failure to refold correctly under the experimental conditions used since Z-Phe-Arg-AMC acts as a substrate for other cathepsin S enzymes (Kirschke & Wiederanders, 1994). Experiments involving the expression of pro- and prepro-forms of HGCP-I and HGCP-II to produce stable, active proteinase for further studies, are underway.

We are grateful to Mrs Jackie Goodall for technical assistance and Dr Sheena Radford for advice on protein folding strategies. This work was funded by Hilleshög NK and the Scottish Office Agriculture, Environment and Fisheries Department.

REFERENCES

BARRETT, A. J. (1972). A new assay for cathepsin B_1 and other thiol proteinases. *Analytical Biochemistry* **47**, 280–293.

BARRETT, A. J. & KIRSCHKE, H. (1981). Cathepsin B, cathepsin H and cathepsin L. In *Methods in Enzymology*, Vol. 80 (ed. Lorand, L.), pp. 535–61. Academic Press, Orlando.

CHAPPELL, C. L. & DRESDEN, M. H. (1986). *Schistosoma mansoni*: proteinase activity of 'hemoglobinase' from the digestive tract of adult worms. *Experimental Parasitology* **61**, 160–167.

DRENTH, J., JANSONIUS, J. N., KOEKOEK, R., SLUYTERMAN, L. A. A. & WOLTHERS, B. G. (1970). The structure of the papain molecule. *Philosophical Transactions of the Royal Society* **257**, 231–236.

ERICKSON, A. H. (1989). Biosynthesis of lysosomal endopeptidases. *Journal of Cellular Biochemistry* **40**, 31–41.

HEUSSLER, V. T. & DOBBELAERE, D. A. E. (1994). Cloning of a protease gene family of *Fasciola hepatica* by the polymerase chain reaction. *Molecular and Biochemical Parasitology* **64**, 11–23.

JOSEPH, L. J., CHANG, L. C., STAMENKOVICH, D. & SUKHATME, V. P. (1988). Complete nucleotide and deduced amino acid sequences of human and murine preprocathepsin L. *Journal of Clinical Investigation* **81**, 1621–1629.

KARRER, K. M., PEIFFER, S. L. & DiTOMAS, M. F. (1993). Two distinct subfamilies within the family of cysteine proteinase genes. *Proceedings of the National Academy of Sciences, USA* **90**, 3063–3067.

KIRSCHKE, H. & WIEDERANDERS, B. (1994). Cathepsin S and related lysosomal endopeptidases. In *Methods in Enzymology*, Vol. 244 (ed. Barrett, A. J.), pp. 500–511. Academic Press, Orlando.

KLINKERT, M-Q., FELLEISEN, R., LINK, G., RUPPEL, A. & BECK, E. (1989). Primary structures of Sm31/32 diagnostic proteins of *Schistosoma mansoni* and their identification as proteases. *Molecular and Biochemical Parasitology* **33**, 113–122.

KORITSAS, V. M. & ATKINSON, H. J. (1994). Proteinases of females of the phytoparasite *Globodera pallida* (potato cyst nematode). *Parasitology* **109**, 357–365.

LARMINIE, C. G. C. & JOHNSTONE, I. L. (1996). Isolation and characterization of four developmentally regulated cathepsin B-like cysteine protease genes from the nematode *Caenorhabditis elegans*. *DNA and Cell Biology* **15**, 75–82.

LILLEY, C. J., URWIN, P. E., McPHERSON, M. J. & ATKINSON, H. J. (1996). Characterization of intestinally active proteinases of cyst-nematodes. *Parasitology* **113**, 415–424.

MAKI, J. & YANAGISAWA, T. (1986). Demonstration of carboxyl and thiol protease activities in adult *Schistosoma mansoni*, *Dirofilaria immitis*, *Angiostrongylus cantonensis* and *Ascaris suum*. *Journal of Helminthology* **60**, 31–37.

McKERROW, J. H. (1989). Parasite proteases. *Experimental Parasitology* **68**, 111–115.

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press

MICHEL, A., GHONEIM, H., RESTO, M., KLINKERT, M.-Q. & KUNZ, W. (1995). Sequence, characterization and localization of a cysteine proteinase cathepsin L in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* **73**, 7–18.

MOTTRAM, J. C., NORTH, M. J., BARRY, J. D. & COOMBS, G. H. (1989). A cysteine proteinase cDNA from *Trypanosoma brucei* predicts an enzyme with an unusual C-terminal extension. *FEBS Letters* **258**, 211–215.

MUSIL, D., ZUCIC, D., TURK, D., ENGH, R. A., MAYR, I., HUBER, R., POPOVIC, T., TURK, V., TOWATARI, T., KATUNUMA, N. & BODE, W. (1991). The refined 2·15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO Journal* **10**, 2321–2330.

PRATT, D., ARMES, L. G., HAGEMAN, R., REYNOLDS, V., BOISVENUE, R. J. & COX, G. N. (1992). Cloning and sequence comparisons of four distinct cysteine proteases expressed in *Haemonchus contortus* adult worms. *Molecular and Biochemical Parasitology* **51**, 209–218.

PRATT, D., COX, G. N., MILHAUSEN, M. J. & BOISVENUE, R. J. (1990). A developmentally regulated cysteine protease gene family in *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **43**, 181–192.

RAWLINGS, N. D. & BARRETT, A. J. (1993). Evolutionary families of peptidases. *The Biochemical Journal* **290**, 205–218.

RAY, C. & MCKERROW, J. H. (1992). Gut-specific and developmental expression of a *Caenorhabditis elegans* cysteine protease gene. *Molecular and Biochemical Parasitology* **51**, 239–250.

ROBERTSON, C. D. & COOMBS, G. H. (1993). Cathepsin B-like cysteine proteinases of *Leishmania mexicana*. *Molecular and Biochemical Parasitology* **62**, 271–279.

SAKANARI, J. A. (1990). Anisakis – from the platter to the microfuge. *Parasitology Today* **6**, 323–327.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

SMITH, A. M., DOWD, A. J., MCGONIGLE, S., KEEGAN, P. S., BRENNAN, G., TRUDGETT, A. & DALTON, J. P. (1993). Purification of a cathepsin L-like proteinase secreted by adult *Fasciola hepatica*. *Molecular and Biochemical Parasitology* **62**, 1–8.

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

URWIN, P. E., ATKINSON, H. J., WALLER, D. A. & MCPHERSON, M. J. (1995). Engineered oryzacystatin-I expressed in transgenic hairy roots confers resistance to *Globodera pallida*. *The Plant Journal* **8**, 121–131.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* **14**, 4683–4690.

https://doi.org/10.1017/S0031182097008858 Published online by Cambridge University Press