The secreted cathepsin L-like proteinases of the trematode, *Fasciola hepatica*, contain 3-hydroxyproline residues

Gene L. WIJFFELS,*‡ Michael PANACCIO,* Liliana SALVATORE,* Lachlan WILSON,* Ian D. WALKER† and Terry W. SPITHILL*

*Victorian Institute of Animal Science, Department of Agriculture, Attwood, 3049, Australia,
and †School of Veterinary Science, University of Melbourne, Parkville, 3052, Australia

The cysteine proteinases synthesized by the adult stage of the trematode *Fasciola hepatica* were found to be a very heterogeneous group of proteins as demonstrated by one- and two-dimensional gel analyses. N-terminal amino acid sequencing indicated the presence of at least two distinct gene products among the secreted cysteine proteinases. Enzymic studies and peptide sequence analysis of the excreted/secreted cysteine proteinases suggested a close relationship to the plant thiol cathepsins and the mammalian cathepsin L subfamily. The cloning of a representative cDNA for a putative *Fasciola* cathepsin confirmed similarities to the cathepsin L subfamily but revealed low identity with the cathepsin-like proteinases of the related trematode, *Schistosoma*, nematode cathepsins and the mammalian cathepsin B subfamily. Furthermore, peptide and protein sequencing revealed the modification of certain highly conserved prolines to unusual 3-hydroxyproline derivatives. This is the first report of modified prolines in any proteinase. This finding, as well as the high activities of these cathepsins at neutral to alkaline pH values, raises a number of questions as to the physiological function of these thiol cathepsins and their interaction with host tissues.

INTRODUCTION

The thiol cathepsins are a highly conserved superfamily of cysteine proteinases with representative members expressed in numerous species of plants and in animals from protozoa through to mammals. The accepted functions of these proteinases are already broad, and many more speculative functions have been inferred (Erickson, 1989). Within a species, the regulated expression of the thiol cathepsin families is observed in a wide variety of tissues. The processing of these enzymes is complex. The mammalian thiol cathepsins are synthesized as preproenzymes, with the signal sequence ensuring their initial localization to the endoplasmic reticulum where the proenzyme is glycosylated. In the Golgi apparatus, certain mannose residues on most cathepsin polypeptides are phosphorylated, directing the enzymes to the lysosomes. In some situations, certain cathepsin polypeptides are further glycosylated in the Golgi apparatus to form complex carbohydrates before the secretion of mature molecules from the cell. These glycosylated proenzyme forms are active near the cell surface. In the lysosomal compartment, the proregion is generally cleaved off, and there may be further cleavage of the mature polypeptide into a light chain/heavy chain heterodimer linked together by disulphide bonds [reviewed in Erickson (1989)]. Owing to their localization to the lysosomes, most of the thiol cathepsins are active at low pH.

In recent years, many studies have been reported characterizing thiol cathepsin-related proteinases and cDNAs encoding these proteinases from several invasive parasites. Parasites tend to express several cysteine proteinases and the more specialist functions of these proteinases in parasites may overlap with the 'housekeeping' functions of the lysosomal cathepsins general in diverse mammalian cells and tissues. A function of thiol cathepsins peculiar to parasites might be the supply of the nutritional requirements of the parasite (Zussman et al., 1970; Maki and Yanagisawa, 1986; McKerrow and Doenhoff, 1988). Alternatively, by analogy with metastasizing tumours, parasites may elaborate cysteine proteinases to enhance invasiveness into host tissues, to promote mobility and to disrupt immune function (Poole et al., 1978). In studies thus far, protozoa appear to synthesize cathepsin L-like cysteine proteinases (Tannich et al., 1991; Baylis et al., 1992; Rosenthal and Nelson, 1992; Eakin et al., 1992), whereas the nematodes, *Caenorhabditis elegans* and *Haemonchus contortus*, and two species of the trematode, *Schistosoma*, elaborate cathepsin B-like proteinases (Klinkert et al., 1989; Cox et al., 1990; Pratt et al., 1992; Ray and McKerrow, 1992; Rege et al., 1992). Generally, these cathepsins, like their mammalian counterparts, are 25–35 kDa in size and are optimally active at acidic pH values.

Many studies have attributed cysteine proteinase-like activities to the secretions and tissues of the trematode *Fasciola hepatica* (Halton, 1967; Aoki, 1980; Chapman and Mitchell, 1982; Rege et al., 1989; Dalton and Heffernan, 1989; Yamasaki et al., 1989). In this report, we have isolated and characterized the excreted/secreted cysteine proteinases from *F. hepatica*. Although several features of this heterogeneous group of enzymes reveal a clear relationship to the cathepsin L family and several plant cathepsins, this study reveals two unusual features of the *Fasciola* cathepsins which critically distinguish them from other thiol cathepsins: first, the more alkaline pH optima of these enzymes in *F. hepatica*, and secondly, the unexpected modification of two proline residues to 3-hydroxyproline. The sites containing these modified proline residues do not conform to known prolyl hydroxylase recognition sites and we speculate on the possible role of a novel prolyl 3-hydroxylase in *F. hepatica*.

Abbreviations used: *F. hepatica*: *Fasciola hepatica*; N-Cbz, *N*-carbobenzoxy-; NHMec, 7-amido-4-methylcoumarin; PMSF, phenylmethanesulphonyl fluoride; IAA, iodoacetamide; E64, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; DTT, dithiothreitol; TFA, trifluoroacetic acid; endoGluC, endoproteinase-Glu-C; PTH, phenylthiohydantoin.

‡ To whom correspondence should be addressed at: CSIRO, Division of Tropical Animal Production, Private Bag 3, PO Indooroopilly, 4068, Australia.

MATERIALS AND METHODS

Gel analyses

SDS/PAGE analyses were carried out on 15% (w/v) polyacrylamide gels (Laemmli, 1970) which were silver stained (Morrissey, 1981). For two-dimensional gel analysis, 30 μg of the purified cysteine proteinases was applied to an isoelectric focusing gel (O'Farrell, 1975) in the first dimension (incorporating Pharmacia ampholytes pH 2.5–5 and pH 5–7 mixed at a ratio of 1:2). SDS/PAGE was conducted using 15% (w/v) acrylamide gels in the second dimension under reducing conditions and gels were silver stained (Morrissey, 1981). Western-blot analysis using an ovine cysteine proteinase antiserum was conducted as previously described (Wijffels et al., 1994). Whole fluke lysates were prepared as described (Wijffels et al., 1992).

Purification of *F. hepatica* cysteine proteinases

Essentially, the cysteine proteinases were purified from adult liver fluke regurgitant by a procedure adapted from Wijffels et al. (1994). Samples of *F. hepatica* regurgitant concentrated 50-fold by ultrafiltration were subjected to a single molecular-sieving step in 50 mM Tris/HCl, pH 6.0, containing 0.6 M NaCl. Fractions were collected from the known cysteine proteinase peaks and assessed for purity by non-reducing SDS/PAGE followed by Western-blot analyses. Fractions containing only the cysteine proteinase complex were pooled, dialysed overnight against distilled water and further concentrated by vacuum centrifugation (Savant Instruments, Hicksville, NY, U.S.A.). This final preparation was stored at −20 °C. Estimations of protein concentration were performed with a Bio-Rad protein assay kit.

Construction of cDNA libraries in phage vector λZAP

Metacercariae of *F. hepatica* were purchased from Compton Paddock Laboratories (U.K.) and used to infect sheep from which adult parasites were obtained after 35 weeks. Total RNA was extracted from adult *F. hepatica* by the method of Chirgwin et al. (1979). Polyadenylated RNA was selected by oligo(dT) chromatography (Aviv and Leder, 1972). The cDNA libraries were constructed in phage vectors λZAPII (Clontech, Palo Alto, CA, U.S.A.).

Immunoscreening of cDNA libraries

The λZAP cDNA library was used to infect *Escherichia coli* BB4 cells which were then plated at a density of 50000 plaque-forming units per 150-mm-diam. Luria Broth/agar plate. Approx. 5×10^5 plaque-forming units were screened for expression of cysteine proteinases of *F. hepatica* using the Protoblot method as described in the Protoblot Technical Manual purchased from Promega (Madison, WI, U.S.A.). The library was screened with a pooled sheep antiserum raised to the purified cysteine proteinases of adult *F. hepatica* which was used at a dilution of 1:600. Filters were blocked in a buffer containing 10 mM Tris/HCl, pH 8.0, 150 mM NaCl, 0.05% Tween 20 and 1% (w/v) gelatin. Positive plaques identified in a primary screen were picked, replated at a lower density and rescreened with the ovine antiserum until individual positive plaques were identified.

Isolation and sequencing of cDNA inserts

Phagemid DNA containing cDNA inserts from positive λZAP phage clones was isolated by excision *in vivo* of the pBluescript phagemid under the conditions recommended by Stratagene (La Jolla, CA, U.S.A.). Phagemid DNA was extracted by the method of Birnboim and Doly (1979). DNA sequencing of cDNA inserts was performed by the chain-termination method (Sanger et al., 1977) after the plasmid DNA was denatured by treatment with NaOH.

Production of peptides

Peptides were generated from several digests of purified *F. hepatica* cysteine proteinases. Several peptides were produced from a digest using endoproteinase-Glu-C (endoGluC; Boehringer-Mannheim, Mannheim, Germany). Approx. 20 μg of the cysteine proteinase preparation was *S*-pyridylethylated and co-precipitated (in acetone) with 2% (w/v) endoGluC. Digestion was performed at 37 °C for 9 h in 100 μl of 0.1 M NH_4HCO_3, pH 8.0. On completion of digestion, the mixture was dried down by vacuum centrifugation and resolubilized in 6 M guanidinium chloride in 10 mM Tris/HCl, pH 8.0, and injected on to a C18 Nova-Pak reversed-phase column (Waters–Millipore, Milford, MA, U.S.A.). Peptides were resolved with a 5–60% acetonitrile gradient [in 0.1% trifluoroacetic acid (TFA)] delivered by h.p.l.c. at 0.5 ml/min (Waters 625 LC system). Elution was monitored at 214 nm, and high-absorbance peaks were collected by a timed loop. The contents of selected peaks were repurified on the same system using a 0–40% acetonitrile (in 0.1% TFA) gradient.

In an attempt to obtain larger peptides, a chymotryptic digest was performed on 100 μg of purified protein that had not been reduced and alkylated. Digestion was performed over 4 h but otherwise was conducted as described by Wijffels et al. (1992). Ensuing peptides were purified and refractionated on a C8 Nova-Pak (Millipore–Waters) reversed-phase column using a 5–60% acetonitrile (in 0.1% TFA) gradient delivered by h.p.l.c. at 0.5 ml/min. This digest yielded the chymotryptic peptides CT13.2 and CT11.3. Two other chymotryptic peptides, CT21.2 and CT13.3, were obtained from a digest of the same preparation of cysteine proteinases described in Wijffels et al. (1994). The chymotryptic digest was performed as previously described (Wijffels et al., 1992). Peptides were purified as for the endoGluC digest.

Purified peptides were dried by vacuum centrifugation before N-terminal sequencing conducted by a model 471A protein sequencer (Applied Biosystems, Foster City, CA, U.S.A.). Phenylthiohydantoin (PTH)-amino acids were identified by h.p.l.c. using a Brownlee Laboratories microgradient system to deliver a linear acetonitrile gradient.

Preparation of collagen peptides for amino acid sequencing

CNBr digests of the α1 and α3 chains of human type-I collagen were a gift from Mr. Danny Chan and Dr. John Bateman of the Paediatrics Department, Melbourne University, and were prepared as previously described (Cole and Chan, 1981). Digests were solubilized at 1 mg of protein/ml in reducing SDS/PAGE sample buffer, and loaded at 50 μg of protein/lane on to a 12% SDS/PAGE minigel system. On completion of the run the gel was electrotransferred on to a Problot membrane (Applied Biosystems Inc.) at 100 V for 45 min in a buffer containing 10% (v/v) ethanolamine and 20% (v/v) methanol. The blotted membrane was stained by 1% (w/v) Coomassie Blue for 5–10 min, followed by a brief rinse in methanol, and the peptide bands were slowly visualized in washes of 7% (v/v) acetic acid. The membrane was air-dried and the CB6 peptides (migrating at approx. 25 kDa in this system) were cut from the stained tracks,

shredded and applied to the model 471A protein sequencer (Applied Biosystems).

Enzyme assays

Solutions of purified cysteine proteinases were assessed for proteolytic activity by a fluorogenic assay based on that of Barrett (1980). The mixed buffer solution consisted of 100 mM trisodium citrate, 100 mM Hepes, 100 mM NaH_2PO_4 and 100 mM boric acid, and the pH was adjusted as indicated. Assays were conducted in 1 ml volumes containing the buffer to which dithiothreitol (DTT) (2 mM), Brij 35 (0.05%, v/v) and proteinase substrate (5 μM) had been added. Either papain (150 ng) or purified *F. hepatica* cysteine proteinase (40 ng) was then added to assay tubes which were then incubated for 17 min at 37 °C. Brij 35 (0.1%) was then added and the fluorescence determined by a fluorescence spectrophotometer (Perkin–Elmer model MPF-4) set with excitation and emission wavelengths of 349 and 437 nm respectively. To determine the inhibitory effect of various agents, the enzyme was preincubated with buffer, DTT and inhibitor for 10 min at ambient temperature before addition of substrate. Substrates [*N*-carbobenzoxy-Phe-Arg-7-amido-4-methylcoumarin (N-Cbz-Phe-Arg-NHMec) and N-Cbz-Arg-NHMec] and inhibitory reagents {phenylmethanesulphonyl fluoride (PMSF), iodoacetamide (IAA), leupeptin, antipain, aprotinin and E64 [trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane]} were all purchased from Sigma. EDTA was obtained from BDH.

RESULTS

Heterogeneity of cysteine proteinases of *F. hepatica*

Gel filtration was conducted on a 50-fold concentrate of the *F. hepatica* excretory/secretory material followed by assay of individual fractions for the presence of cysteine proteinases. The absorbance profile of the chromatographed material is depicted in Figure 1(a) and the proteinase-rich zone indicated was pooled for further analysis. The occurrence of proteinase activity associated with both peaks A and B of Figure 1(a) suggested apparent size heterogeneity within the proteinases. Only fractions of the later half of the major cysteine proteinase peak A and most of the minor peak B were collected in an attempt to obtain cysteine proteinases free of other proteins. These fractions were analysed by non-reducing SDS/PAGE which revealed a dominant protein species at 26 kDa in all fractions (Figure 1b). Fraction 1 contained other low-molecular-mass species.

When those fractions containing only the 26 kDa species were pooled and further characterized by Western blot and SDS/PAGE under reducing and non-reducing conditions, heterogeneity was apparent. Within the reduced preparation a predominant zone of 28 kDa was observed together with less intensely staining species of 14, 15, 17 and 20 kDa (Figure 2, lane A) whereas the non-reduced material was relatively homogeneous with a major complex at 26 kDa (Figure 2, lane B). Western-blot analyses employing an ovine antiserum raised to the purified *F. hepatica* cysteine proteinases revealed a staining pattern comparable with the silver-stained SDS/PAGE (Figure 2, lanes C and D). In whole fluke lysate, treated with reducing SDS/PAGE sample buffer, numerous species were reactive to the ovine cysteine proteinase antiserum (Figure 2, lane E). Although the 28 kDa complex was prominent, a ladder of higher-molecular-mass species at 30–43 kDa was also detected. In the non-reduced preparation a simpler pattern was observed: a complex of three bands at 26–28 kDa and a second complex of three bands at 35–42 kDa (Figure 2, lane F).

Two-dimensional SDS/PAGE analysis of the purified excretory/secretory cysteine proteinases resolved several components at 28 kDa, with at least three closely migrating dominant species of slightly differing pI values and four minor species of higher pI (Figure 3). Lower-molecular-mass species (at approx. 20 and 15 kDa) were also detected but at the same pI as the three most prominent components, suggesting that they are either fragments of forms of the cysteine proteinases which are

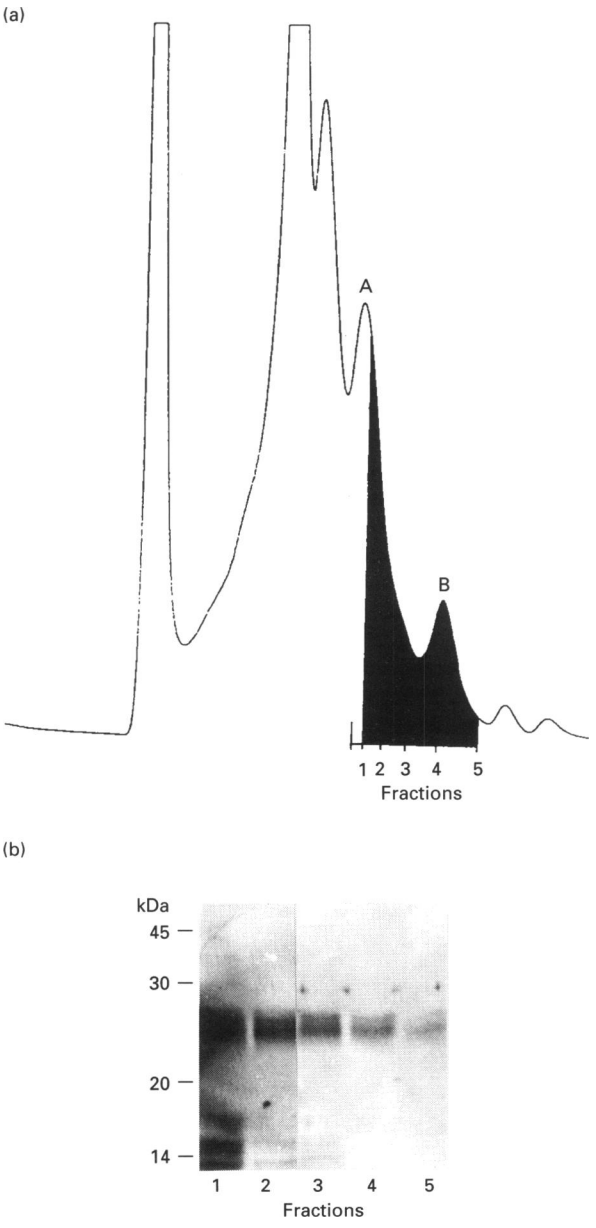

Figure 1 Gel filtration of concentrated regurgitant of adult *F. hepatica*

Approx. 1 mg of concentrated regurgitant was loaded on to a Superose 6 f.p.l.c. column (see the Materials and methods section). (**a**) Gel-filtration profile of the regurgitant concentrate with elution monitored at 280 nm (1.0 absorbance unit, full-scale). The cysteine proteinases are eluted under peaks A and B. Fractions were collected where indicated. The shaded region depicts the fractions later pooled for further analysis. (**b**) SDS/PAGE analysis (followed by silver staining) of gel-permeation fractions 1–5. Approx. 1% of each fraction was loaded in non-reducing sample buffer.

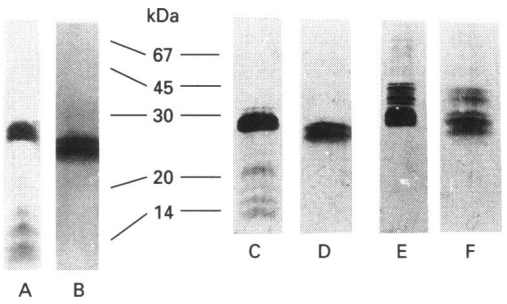

Figure 2 Comparison of cysteine proteinases from the regurgitant and whole fluke lysate of adult *F. hepatica*

In lanes A and B, cysteine proteinases purified from adult regurgitant were resolved by SDS/PAGE in reducing and non-reducing conditions respectively. The proteins were visualized by silver staining. Western-blot analysis was conducted on the same preparation (lanes C and D) and on whole fluke lysate (0.5 μl; lanes E and F) using the ovine *Fasciola* cysteine proteinase antiserum. Lanes D and F were obtained under non-reducing conditions.

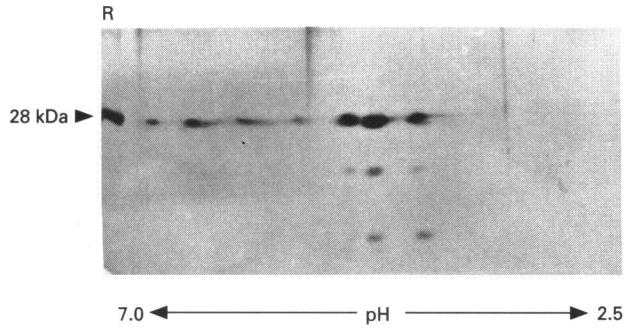

Figure 3 Two-dimensional SDS/PAGE analysis of purified excretory/secretory cysteine proteinases of adult *F. hepatica*

Isoelectric focusing in the first dimension was conducted over a pH range of 2.5–7.0 with 30 μg of purified proteinases. Proteins were detected in the second dimension (SDS/PAGE, 15% gels) by silver staining. The lane (R) is a single-dimensional reducing SDS/PAGE run of the cysteine proteinase preparation.

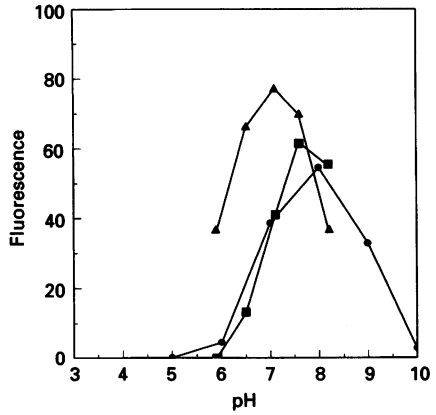

Figure 4 pH optima of the excretory/secretory cysteine proteinases of adult *F. hepatica*

The *F. hepatica* cysteine proteinase pool (40 ng) was assayed for activity at various pH values in 100 mM phosphate buffer (■) and a mixed buffer solution (●; see the Materials and methods section). Papain (150 ng) was also assessed in the phosphate buffer (▲). The reactions were performed over the pH range indicated using the fluorogenic substrate, N-Cbz-Phe-Arg-NHMec, as described in the Materials and methods section.

Table 1 Inhibition of *F. hepatica* cysteine proteinases by various proteinase inhibitors

The purified *F. hepatica* cysteine proteinase pool (50 ng) was preincubated with the indicated inhibitors and assessed for activity using N-Cbz-Phe-Arg-NHMec as substrate as described in the Materials and methods section. These assays were conducted twice with all inhibitors.

Inhibitory reagent	Proteinase class specificity	Concentration (μM)	Inhibition (%)
PMSF	All serine and some cysteine proteinases	100	4
E64	All cysteine proteinases	0.5	92
Leupeptin	Most cysteine and trypsin-like serine proteinases	1	93
IAA	Cysteine proteinases	100	74
		1	25
Aprotinin	Serine proteinases	100 μg/ml	0
Antipain	Trypsin-like serine and some cysteine proteinases	1	91
EDTA	Metalloproteinases	1000	0
No DTT	–	0	93

disulphide-linked oligomers or, alternatively, products of autoproteolysis of the intact polypeptide.

Enzymic characterization of *F. hepatica* cysteine proteinases

Of the two fluorogenic substrates (N-Cbz-Arg-NHMec and N-Cbz-Phe-Arg-NHMec) tested for sensitivity to the proteolytic action of the purified proteinases, only N-Cbz-Phe-Arg-NHMec proved susceptible and was used in all further enzymic characterization. There was no detectable activity with N-Cbz-Arg-NHMec even at high concentrations of the enzyme (results not shown). The *F. hepatica* cysteine proteinase pool achieved maximum activity at pH 7.6 in 100 mM phosphate buffer (Figure 4), and an average K_m of 46 μM was derived by a Michaelis–Menten plot of data obtained using this buffer at pH 7.45 (results not shown). Under the same conditions, papain produced a maximum activity at pH 7.0 (Figure 4), as expected (Bond, 1989). In a mixed-buffer system, activity over a broad pH range (pH 7–9) was detected, with maximum activity at pH 8.0 (Figure 4). In a Tris buffer, high levels of activity were detected at very alkaline pH values (results not shown).

The inhibition assays revealed that the cysteine proteinases were highly sensitive to very low concentrations of the classic inhibitors of this class of proteinases (Table 1), although IAA was less effective than leupeptin or E64, producing only 74% inhibition at a 100 μM concentration. PMSF had a very slight inhibitory effect whereas aprotinin and EDTA caused no decrease in activity. DTT was found to be an absolute requirement for activity (Table 1). These data were reproducible over two assays.

Sequence similarity to the thiol cathepsins

An adult *F. hepatica* cDNA expression library was screened with the ovine anti-(cysteine proteinase) serum. A high frequency (0.1–1.0%) of clones was found to react with this antiserum (which had been extensively adsorbed to remove antibodies directed to *E. coli*). Nine randomly selected positive clones were purified and their cDNA inserts were partially sequenced. They were all found to be identical in sequence at their 3′ end. The

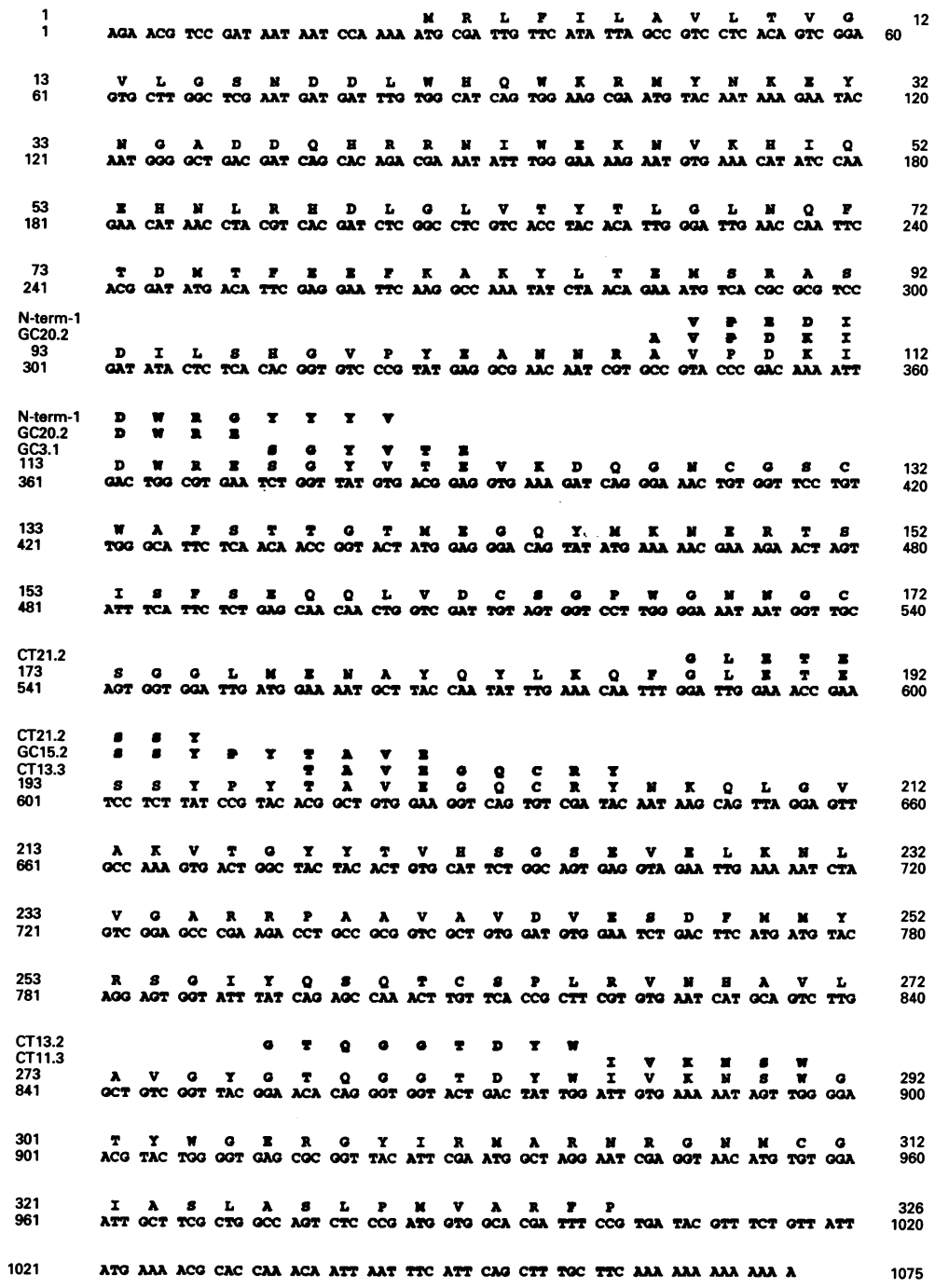

Figure 5 DNA sequence of the Fhcat-1 cDNA with its predicted amino acid sequence and homologous peptides

The peptides were derived from chymotryptic digests (CT) and endoGluC digests (GC) of the purified *F. hepatica* cysteine proteinases (see the Materials and methods section), and are aligned to the Fhcat-1 predicted amino acid sequence. The N-terminal sequence (N-term-1) of intact material is also mapped to the Fhcat-1 primary sequence.

insert of the largest clone, Fhcat-1, was sequenced on both strands, and on analysis revealed an open reading frame that extended from nucleotides 25 to 1002 (Figure 5). The predicted open reading frame encodes a putative protein of 326 amino acids in length which has homology to members of the thiol cathepsin family of cysteine proteinases (EC 3.4.22.-) (Figure 6). The putative Fhcat-1 preproprotein has 44% identity with human preprocathepsin L and 39% identity with preprocathepsin H. There was limited identity with human preprocathepsin B (23%) and cysteine proteinases from *Schistosoma* (20%, Figure 6) and *Haemonchus* (20%, results not shown).

By comparison with the mammalian thiol cathepsins, amino acids 1–17 encode the pre region and amino acids 18–106 encode the pro region. The pre and pro regions would be cleaved off to form a mature proteinase of 219 amino acids in length. Residues 126–136 of the predicted Fhcat-1 mature protein contain similarities to the thiol cathepsin consensus pattern Gln-Xaa(3)-[Gly-Glu]-Xaa-Cys-Trp-Xaa(2)-[Ser-Thr-Ala-Gly]. The best overall

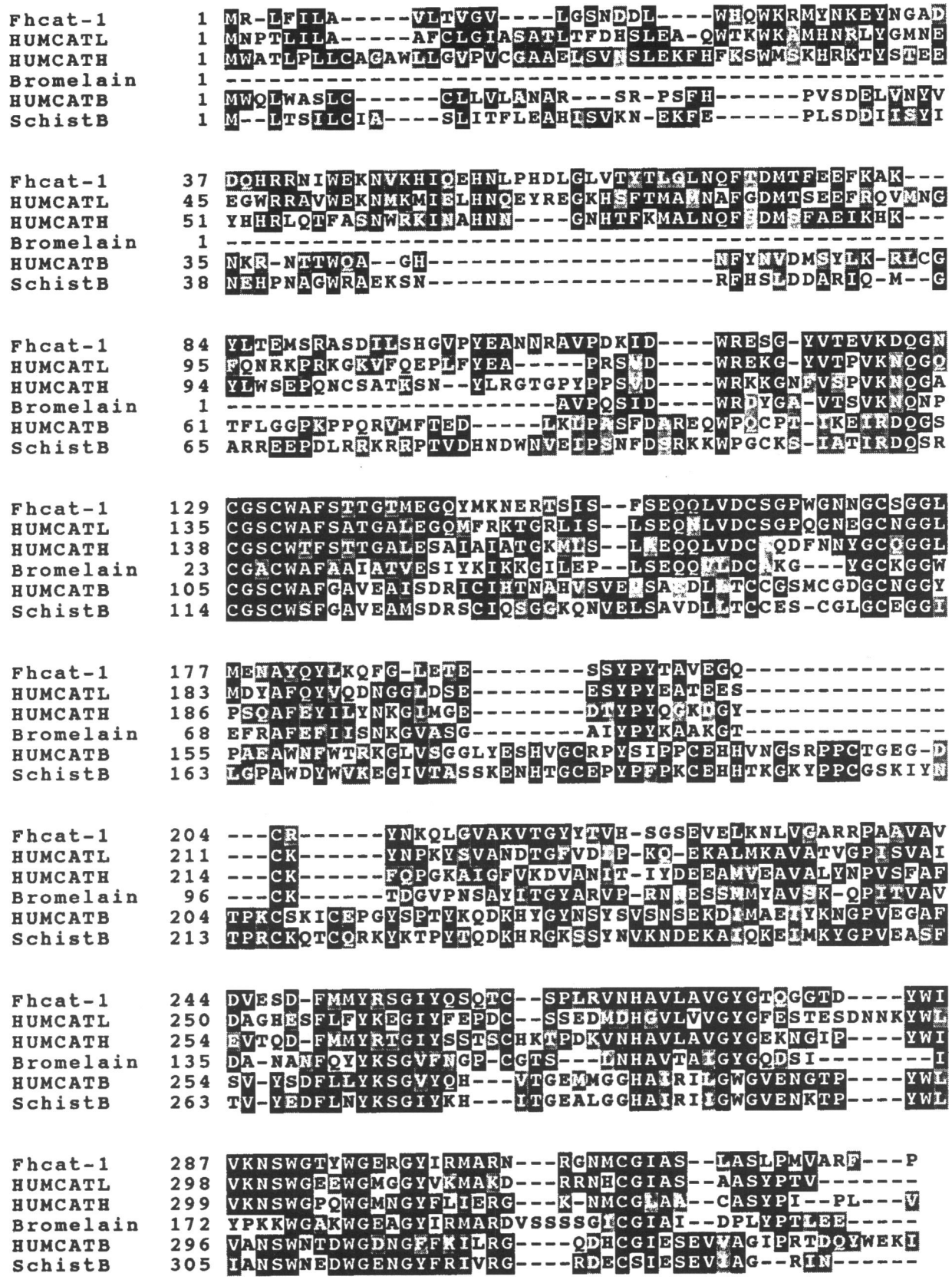

Figure 6 Comparison of the predicted Fhcat-1 polypeptide sequence with other thiol cathepsins

Alignment of the predicted amino acid sequences of the Fhcat-1 preproprotein with the sequences of human preprocathepsin L (HUMCATL: P07711), human preprocathepsin H (HUMCATH: P09668), bromelain (P14518), human preprocathepsin B (HUMCATB: P07858), and a *Schistosoma* thiol cathepsin (SchistB: N21309) (GenBank accession numbers indicated in parentheses). Shaded areas represent conservative changes.

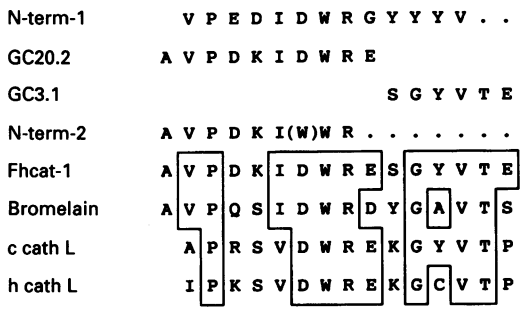

Figure 7 N-terminal sequences of the excretory/secretory cysteine proteinases of adult *F. hepatica*

N-terminal amino acid sequences (N-term-1 and N-term-2) of intact purified *F. hepatica* cysteine proteinase derived from two different collections of regurgitant are shown. Peptides (GC3.1 and GC20.2) were isolated from an endoGluC digest of the same pool of purified material that produced the N-term-1 sequence, and can be aligned with the Fhcat-1 N-terminus which is predicted by its identity with the bromelain N-terminus. Alignments were found to the N-termini of bromelain (Ritonja et al., 1989) and two thiol cathepsins: chicken cathepsin L (c cath L) (Wada et al., 1987) and human cathepsin L (h cath L) (Mason et al., 1986).

homologies between the mature Fhcat-1 polypeptide and the thiol cathepsins are found to cathepsins L and H (54 and 44% identity respectively). Good sequence similarity was also obtained to bromelain (39%, Figure 6), as well as various other plant thiol cathepsin species such as barley aleurain (47%), actinidin (42%) and papain (35%) (results not shown).

To ensure that the cloned cDNA was representative of the secreted/excreted cysteine proteinases of *F. hepatica*, purified peptides were generated from chymotryptic and endoGluC digests of the purified cysteine proteinases. One series of peptides (CT21.2, GC15.2 and CT13.3) overlapped with each other, and their contiguous sequence corresponds exactly to positions 188–206 of the putative Fhcat-1 polypeptide (Figure 5). Peptide CT13.3 displayed one difference at position 204 where both Tyr and Cys were detected. A second series of peptides (CT13.2 and CT11.3) were also contiguous when mapped to positions 277–291 of the predicted amino acid sequence of the Fhcat-1 cDNA (Figure 5). CT11.3 in particular has a highly conserved region containing the Asn residue which makes up part of the catalytic triad of the thiol cathepsins (Figure 6 and Musil et al., 1991).

N-terminal sequences of *F. hepatica* cysteine proteinases

N-terminal sequencing was conducted on two preparations of the purified cysteine proteinases that had been obtained from different batches of source material. Both runs revealed a major sequence which was representative of N-termini of other members of the thiol cathepsin family, but these two sequences were not identical (Figure 7). In the first instance, the N-terminal sequencing run yielded a sequence (N-term-1: Val-Pro-Glu-Asp-Ile-Asp-Trp-Arg-Gly-Tyr-Tyr-Tyr-Val-...) which was not predicted by the Fhcat-1 cDNA but which had obvious homology to the N-termini of several thiol cathepsins (Figures 6 and 7). PTH-Ala was also detected in the first sequencing cycle of this run, but no clear minor sequence was discernible. The N-term-1 sequence has only the one residue (valine) before the proline in position 2, a feature shared with most thiol cathepsins. This sequence then deviates from most of the thiol cathepsin N-terminal sequences after the Asp-Trp-Arg region, resulting in a lower level of sequence similarity to bromelain (61%) and the cathepsin L subfamily (Figure 7). However, an EndoGluC digest of a portion of this same material yielded a peptide (GC20.2) that could be aligned to the predicted N-terminus of the Fhcat-1 polypeptide (Figure 5). The peptide, GC20.2, terminated with a Glu at position 9, but it is likely to be contiguous with the peptide GC3.1 which aligns at this exact region in the predicted Fhcat-1 polypeptide (Figure 5). This extended sequence has 70% identity with the bromelain N-terminus but also has similar levels of identity with the N-termini of cathepsins L and H (Figures 6 and 7).

In order to confirm the N-terminus of the *Fasciola* cysteine proteinases and to determine the relative proportions of 3-hydroxyproline to proline (see below), a second N-terminal sequencing run was conducted on intact material derived from a different collection of regurgitant. On this occasion, a minor sequence similar to the N-term-1 sequence was detected (results not shown). Accompanying this N-term-1-like sequence in twice the molar abundance was the dominant sequence [N-term-2: Ala-Val-Pro-Asp-Ile-(Trp)-Trp-Arg-...]. The N-term-2 sequence is clearly similar to the predicted N-terminus of the Fhcat-1 polypeptide and the contiguous sequence of the peptides GC20.2 and GC3.1 (Figure 7).

Hydroxyprolines in the excreted/secreted cysteine proteinases of *F. hepatica*

Amino acid sequencing of peptides GC15.2 and GC20.2 yielded an unusual product at cycles 4 and 3 respectively in the absence of an increase in any other PTH-amino acid derivatives (Figure 8c). Small amounts of a product with an identical chromatographic retention were also detected in N-terminal sequencing runs of the purified intact preparations of the *Fasciola* cathepsins, along with significant amounts of PTH-proline (results not shown). As the Fhcat-1 polypeptide sequence predicted from its cDNA indicated that a proline should occur at these sites, and, as invariant proline residues occurred at the corresponding positions in other thiol cathepsins (Figures 6 and 7), we investigated the possibility of a modified proline within these peptides.

The most common simple modification of proline occurs in the collagen chains, where proline is hydroxylated at either C-3 or C-4 generating 3-hydroxyproline (3-Hyp) or 4-hydroxyproline (4-Hyp) respectively. To determine if the unusual PTH-amino acids derived from the *Fasciola* cysteine proteinases were indeed hydroxylated prolines, we chromatographically compared both 3-Hyp-PTH and 4-Hyp-PTH derived from the CB6 peptides of the α1 and α3 chains of human type-I collagen. Partial hydroxylation of proline to 3-Hyp and 4-Hyp has been demonstrated in the α1 collagen chain CB6 peptide which has the sequence: Gly-Pro/3-Hyp-Pro/4-Hyp-Gly-Leu-Ala-Gly-Pro-Pro/4-Hyp-.... On Edman degradation of this peptide a small amount of 3-Hyp-PTH was detected at cycle 2 accompanied by large amounts of proline-PTH (Figure 8a), whereas the two major adducts of 4-Hyp-PTH are distinctive in cycles 3 and 9 (Figure 8d). Similarly, we detected partial but more significant modification of the proline to 3-Hyp in the second cycle of sequencing the α3 CB6 peptide (Gly-3-Hyp/Pro-Arg-Gly-Ala-Pro-Gly-...) (Figure 8b).

When the chromatograms of the collagen-derived 3-Hyp-PTH are aligned with that of the unidentified Edman degradation product derived from the cysteine proteinase peptide GC20.2, co-elution was obvious (Figures 8b and 8c). Similarly, the Edman degradation product at cycle 4 of the GC15.2 peptide was chromatographically identical with the collagen-derived 3-Hyp-PTH (results not shown). In both these purified *Fasciola* peptides, proline-PTH was also detected as a minor product, 16% and

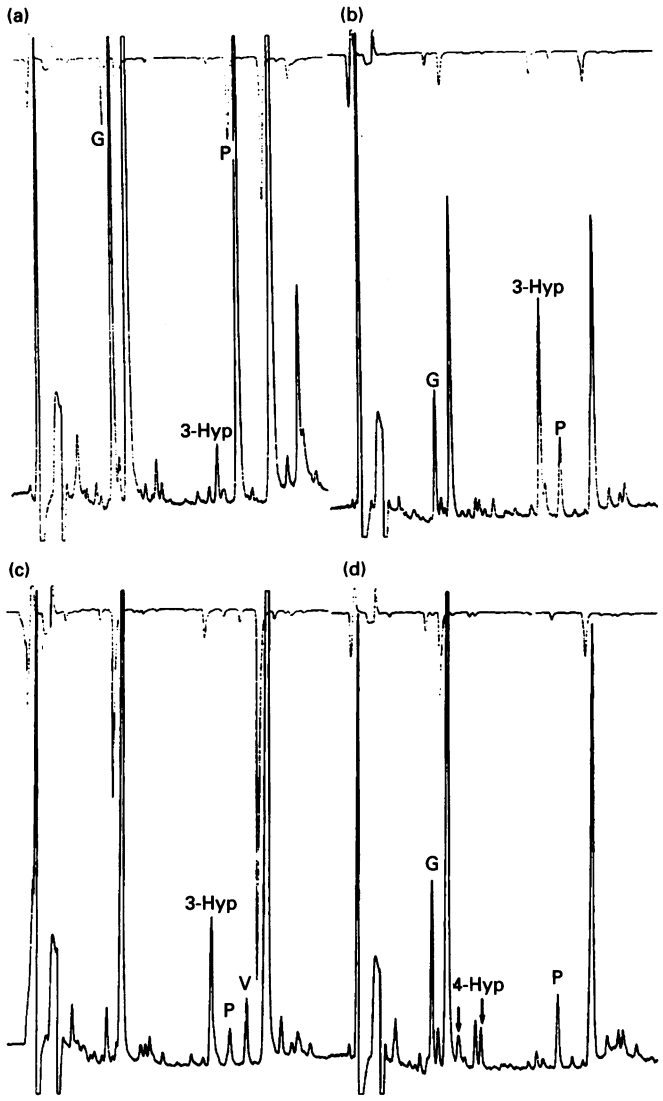

Figure 8 Amino acid sequencing chromatograms of hydroxyprolines

(**a**) Residue 2 (3-hydroxyproline) of the CB6 peptide of α1 human collagen; (**b**) residue 2 (3-hydroxyproline) of the CB6 peptide of α3 human collagen; (**c**) residue 3 (3-hydroxyproline) of the GC20.2 peptide of the *Fasciola* cysteine proteinases; (**d**) residue 9 (4-hydroxyproline) of the CB6 peptide of α1 human collagen. Amino acid sequencing and identification of the PTH-amino acids were conducted as described in the Materials and methods section.

12% of total proline for GC20.2 and GC15.2 respectively (Figure 8c). N-terminal sequencing of the purified *Fasciola* cysteine proteinases, however, indicated that hydroxylation near the N-terminus was a minor event overall, with only 12 and 21% of prolines found in the hydroxylated form in two different sequencing runs (results not shown). Significantly, no 4-Hyp-PTH was detected in any amino acid sequencing of the *Fasciola* products.

DISCUSSION

Cysteine proteinase material of *F. hepatica* was purified from concentrated adult excretory/secretory material in a single gel-permeation step. This material was judged to be sufficiently pure to permit direct analysis but probably corresponds to a subset of the *F. hepatica* cysteine proteinase repertoire. Although a purification regime for the entire excreted/secreted cysteine proteinase population of *F. hepatica* has been developed, it involves steps that may inactivate the enzymes or reduce their solubility and possibly render them unsuitable for functional studies (Wijffels et al., 1994).

The cysteine proteinases that are characterized here were resolved as a discrete 26 kDa zone in non-reducing SDS/PAGE and Western blots. When subjected to reducing SDS/PAGE, a shift in migration to a slightly less mobile species (28 kDa) was apparent, suggesting internal disulphide bonding. Furthermore, several lower-molecular-mass but minor species at 14, 15, 17 and 20 kDa were also produced as a result of reduction. Clearly the majority of the excreted/secreted cysteine proteinases of *F. hepatica* are single polypeptides. Two-dimensional SDS/PAGE analysis focused the cysteine proteinases into seven components and exemplified the heterogeneity of this group of enzymes. The majority of the cysteine proteinases resolved into three similarly charged species of acidic pI and four species of divergent charge and higher pI. Lower-molecular-mass species of 17 and 15 kDa are also present but at the same pI as the major components, suggesting that these are products of *in situ* autodegradation.

The excreted/secreted cysteine proteinases were compared with the cysteine proteinase of whole adult fluke extract by Western-blot analysis. The lysate revealed predominant forms at 26 kDa (28 kDa when reduced) but no low-molecular-mass species were apparent in the reduced lysate. The numerous higher-molecular-mass species (30–43 kDa) detected in the reduced preparations may indicate the presence of intracellular precursor forms common to several members of the thiol cathepsin family (Chan et al., 1986; Portnoy et al., 1986; Klinkert et al., 1989). This would suggest that only fully processed cysteine proteinases are released into the gut of *F. hepatica*. The predicted molecular mass of the Fhcat-1 preproprotein and proenzyme are 39 and 36 kDa respectively. The likely presence of these species does not fully account for the ladder of bands detected in the whole fluke. There are a number of possibilities. For example, the proproteins of many thiol cathepsins are known to undergo multistep cleavage before the final N-terminus is exposed (Erickson, 1989). Alternatively, the 30–45 kDa species detected in the adult lysate may represent tissue cysteine proteinases that are antigenically related to the cysteine proteinases in the gut contents.

The *Fasciola* cathepsins exhibit marked differences from their plant and mammalian counterparts, most notably their preference for neutral to alkaline pH for optimal activity. As with the B and L cathepsin subfamilies, the *F. hepatica* gut cathepsins exhibit strict substrate preference for the synthetic fluorogenic substrate, N-Cbz-Phe-Arg-NHMec, over a similar synthetic substance, N-Cbz-Arg-NHMec which is a highly sensitive substrate for the cathepsin H subfamily (Barrett and Kirschke, 1981). Absolute dependence on the presence of free thiol groups for activity and great sensitivity to the inhibitors leupeptin and E64 are further evidence of cathepsin B- and L-like properties (Barrett and Kirschke, 1981). However, in contrast with these cathepsins, the liver fluke-excreted/secreted thiol cathepsins were found to be active in neutral to alkaline conditions. Cathepsins B and L function optimally at pH 6.0 and pH 5.5 respectively (Barrett and Kirschke, 1981), and are irreversibly inactivated at neutral pH (Maciewicz and Etherington, 1988).

The predicted amino acid sequence of a single cDNA molecule isolated from an *F. hepatica* cDNA library revealed the typical cathepsin-like structure of a preproenzyme containing a 17-amino acid leader sequence, a 90-amino acid proregion and a

219-residue mature protein. The primary structure of the mature segment maintained reasonable identity with plant thiol cathepsins (35–47 %) and the mammalian cathepsin L subfamily (54 %). Regions of highest conservation are found in the proximity of residues involved in the catalytic site (Cys-132, His-269, Asn-289) and disulphide bonds (Musil et al., 1991). By comparison with other thiol cathepsins, three conserved thiol linkages are likely to be retained in the Fhcat-1 structure (Cys-129–Cys-172; Cys-163–Cys-204; Cys-262–Cys-311).

Peptide sequences generated from digests of the purified *F. hepatica* cathepsins and N-terminal analysis of the intact material confirmed that the putative Fhcat-1 protein, or closely related proteins, were predominant in the excreted/secreted cathepsin pool. Two EndoGluC peptides and one direct N-terminal sequencing run indicated that the N-terminus was bromelain-like (Ritonja et al., 1989). However, the presence of a second distinct polypeptide was revealed by the N-terminal sequencing of the intact material derived from a different collection of adult regurgitant. This sequence was more typical of the cathepsin L subfamily. It is clear that there are two distinct but related gene products giving rise to two different mature proteinases and it is possible that their relative proportions may vary in different collections of the secreted material.

N-terminal sequencing and cDNA cloning has accounted for two polypeptide variants, but modified amino acids provide a further possible source of heterogeneity. It is apparent that some prolines of the cysteine proteinases of *F. hepatica* undergo modification to 3-hydroxyproline (3-Hyp). The conserved prolines in two peptides were almost fully hydroxylated to the 3-Hyp form, whereas a smaller proportion (10–20 %) of the conserved proline at the N-terminus was hydroxylated in both N-terminal variants. Thus far there has been no report of modified prolines occurring in any enzyme, other than chitinase from tobacco which was demonstrated to have fully and partially hydroxylated prolines. However, the 4-Hyp residues were restricted to the spacer arm between the enzymically active region and the carbohydrate-binding domain of the protein (Sticher et al., 1992). The most common sources of hydroxyprolines are structural proteins such as collagens in animals and extensins in plants, in which they are often further modified with extensive carbohydrate attachments (Showalter and Rumeau, 1990). In these proteins the proline is hydroxylated at C-4, with the 3-Hyp isomer being relatively rare even in collagen (Fietzek et al., 1972) and not detected in *Fasciola* collagens (Nordwig et al., 1970). It is most striking that the modification of the prolines in the *Fasciola* thiol cathepsins is restricted to the 3-Hyp form. It would appear that the responsible prolyl 3-hydroxylase of *F. hepatica* is quite different from the known mammalian and plant enzymes. The sites of hydroxylation in the *F. hepatica* cathepsins have no features in common with sites of known prolyl hydroxylase activity (Sticher et al., 1992). Considering that *Fasciola* may lack a collagen prolyl 3-hydroxylase, it is likely that this trematode has a novel prolyl 3-hydroxylase which is not confined to the substrates of the other known proline hydroxylases.

It is formally possible that modification of the prolines in these cathepsins may be due to the promiscuous activity of a prolyl hydroxylase present during purification. Clearly, this activity is not due to collagen prolyl hydroxylases. Moreover, hydroxylation of prolines is not common to proteins derived from *F. hepatica*, as purification of the glutathione S-transferases from whole homogenates of fluke did not reveal any altered prolines on Edman degradation (Wijffels et al., 1992).

Unlike other trematodes and the related nematodes, *F. hepatica* has opted for the cathepsin L-like proteinases in the pool of excreted/secreted cathepsins rather than the cathepsin B-like enzymes (Klinkert et al., 1989; Cox et al., 1990; Ray and McKerrow, 1992). The cathepsin L subfamily is renowned for its ability to digest collagens (Delaisse et al., 1991; Maciewicz and Etherington, 1988). It is curious that *Fasciola* should secrete a cathepsin L-like proteinase that shares a modification that is almost exclusive to the collagens. Hydroxylation of prolines in a thiol cathepsin is an heretofore unknown event in this family of enzymes and its functional significance is not clear.

We thank Mr. Danny Chan and Dr. John Bateman of the Orthopaedic Unit, Department of Paediatrics, University of Melbourne, Royal Children's Hospital for their advice and gifts of purified collagen digests.

REFERENCES

Aoki, T. (1980) Jpn. J. Parasitol. **29**, 325–332
Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 1408–1412
Barrett, A. J. (1980) Biochem. J. **187**, 909–912
Barrett, A. J. and Kirschke, H. (1981) Methods Enzymol. **80**, 535–561
Baylis, H. A., Megson, A., Mottram, J. C. and Hall, R. (1992) Mol. Biochem. Parasitol. **54**, 105–108
Birnboim, H. C. and Doly, J. (1979) Nucleic Acids Res. **7**, 1513–1518
Bond, J. S. (1989) in Proteolytic Enzymes: A Practical Approach (Beynon, R. J. and Bond, J. S., eds.), p. 233, IRL Press, Oxford
Chan, S.-J., Segunda, B. S., McCormick, M. B. and Steiner, D. F. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 7721–7725
Chapman, C. B. and Mitchell, G. F. (1982) Vet. Parasitol. **11**, 165–178
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Cole, W. G. and Chan, D. (1981) Biochem. J. **197**, 377–383
Cox, G. N., Pratt, D., Hageman, R. and Boisevenue, R. J. (1990) Mol. Biochem. Parasitol. **41**, 25–34
Dalton, J. P. and Heffernan, M. (1989) Mol. Biochem. Parasitol. **35**, 161–166
Delaisse, J.-M., Ledent, P. and Vaes, G. (1991) Biochem. J. **279**, 167–174
Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H. and Craik, C. S. (1992) J. Biol. Chem. **267**, 7411–7420
Erickson, A. H. (1989) J. Cell. Biochem. **40**, 31–41
Fietzek, P. P., Rexrodt, W., Wendt, P., Stark, M. and Kuhn, K. (1972) Eur. J. Biochem. **30**, 163–168
Halton, D. W. (1967) Parasitology **57**, 639–660
Klinkert, M.-Q., Felleisen, R., Link, G., Ruppel, A. and Beck, E. (1989) Mol. Biochem. Parasitol. **33**, 113–122
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Maciewicz, R. A. and Etherington, D. J. (1988) Biochem. J. **256**, 433–440
Maki, J. and Yanagisawa, T. (1986) J. Helminthol. **60**, 31–37
Mason, R. W., Walker, J. E. and Northrop, F. D. (1986) Biochem. J. **240**, 373–377
McKerrow, J. H. and Doenhoff, M. J. (1988) Parasitol. Today **4**, 334–340
Morrissey, J. H. (1981) Anal. Biochem. **117**, 307–310
Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N. and Bode, W. (1991) EMBO J. **10**, 2321–2330
Nordwig, A., Rogall, E. and Hayduk, U. (1970) in Chemistry and Molecular Biology of the Intercellular Matrix (Balazs, E. A., ed.), vol. 1, pp. 27–41, Academic Press, London
O'Farrell, P. H. (1975) J. Biol. Chem. **250**, 4007–4021
Poole, A. R., Tiltman, K. J., Recklies, A. D. and Stoker, T. A. M. (1978) Nature (London) **273**, 545–547
Portnoy, D. A., Erickson, A. H., Kochan, J., Ravetch, J. V. and Unkeless, J. C. (1986) J. Biol. Chem. **261**, 14697–14703
Pratt, D., Armes, L. G., Hageman, R., Reynolds, V., Boisevenue, R. J. and Cox, G. N. (1992) Mol. Biochem. Parasitol. **51**, 209–218
Ray, C. and McKerrow, J. H. (1992) Mol. Biochem. Parasitol. **51**, 239–250
Rege, A. A., Herrera, P. R., Lopez, M. and Dresden, M. H. (1989) Mol. Biochem. Parasitol. **35**, 89–96
Rege, A., Wang, W. and Dresden, M. H. (1992) J. Parasitol. **78**, 16–23
Ritonja, A., Rowan, A. D., Buttle, D. J., Rawlings, N. D., Turk, V. and Barrett, A. J. (1989) FEBS Lett. **247**, 419–424
Rosenthal, P. J. and Nelson, R. G. (1992) Mol. Biochem. Parasitol. **51**, 143–152
Sanger, F., Nicklen, S.. and Coulson, A. R. (1977) Proc. Nat. Acad. Sci. U.S.A. **74**, 5463–5469
Showalter, A. M. and Rumeau, D. (1990) in Organization and Assembly of Plant and Animal Extracellular Matrix (Adair, W. S. and Mecham, R. P., eds.), pp. 247–281, Academic Press,, New York
Sticher, L., Hofsteenge, J., Milani, A., Neuhaus, J.-M. and Meins, Jr., F. (1992) Science **257**, 655–657

Tannich, E., Scholze, H., Nickel, R. and Horstmann, R. D. (1991) J. Biol. Chem. **266**, 4798–4803
Wada, K., Takai, T. and Tanabe, T. (1987) Eur. J. Biochem. **167**, 13–17
Wijffels, G. L., Sexton, J. L., Salvatore, L., Pettitt, J. M., Humphries, D. C., Panaccio, M. and Spithill, T. W. (1992) Exp. Parasitol. **74**, 87–99

Received 15 July 1993/12 November 1993; accepted 24 November 1993

Wijffels, G. L., Salvatore, L., Dosen, M., Waddington, J., Thompson, C., Campbell, N., Sexton, J., Wicker, J., Bowen, F., Friedel, T. and Spithill., T. W. (1994) Exp. Parasitol., in the press
Yamasaki, H., Aoki, T. and Hiroshi, O. (1989) Jpn. J. Parasitol.. **38**, 373–384
Zussman, R. A., Bauman P. M. and Petruska, J. C. (1970) J. Parasitol. **56**, 75–79