Experimental Parasitology **94,** 201–207 (2000)
doi:10.1006/expr.2000.4489, available online at http://www.idealibrary.com on IDEAL®

RESEARCH BRIEF

The New Subfamily of Cathepsin-Z-like Protease Genes Includes *Tc-cpz-1*, a Cysteine Protease Gene Expressed in *Toxocara canis* Adults and Infective Stage Larvae[1]

Franco H. Falcone, Kevin K. A. Tetteh,[2] P. Hunt,[3] M. L. Blaxter, Alex Loukas,[4] and Rick M. Maizels[5]

Institute of Cell, Animal, and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, United Kingdom

Falcone, F. H., Tetteh, K. K. A., Hunt, P., Blaxter, M. L., Loukas, A., and Maizels, R.M. 2000. The new subfamily of cathepsin-Z-like protease genes includes *Tc-cpz-1*, a cysteine protease gene expressed in *Toxocara canis* adults and infective stage larvae. *Experimental Parasitology* **94,** 201–207.

Cathepsins are lysosomal proteases, most of which belong to the papain or C1 family of cysteine proteases (Berti and Storer 1995, Barrett and Rawlings 1996). Within the C1 family, the cathepsins can be subdivided into more than 10 subfamilies on the basis of primary sequence and substrate preferences (Berti and Storer 1995, Barrett and Rawlings 1996). Parasitic cathepsin-like proteases are of considerable interest, since the knowledge of their mechanism of action may facilitate development of novel chemotherapeutic agents (Coombs and Mottram 1997, Engel *et al.* 1998; Ward *et al.* 1997). It is known that a wide range of cathepsins is expressed in the primitive metazoan protostome phyla (Tort *et al.* 1999). For example, cathepsin-B-like proteases are produced by the nematodes *Ancylostoma caninum* (Harrop *et al.* 1995) and *Haemonchus contortus* (Pratt *et al.* 1990) and the trematodes *Fasciola hepatica* (Wilson *et al.* 1998) and *Schistosoma japonicum* (Merckelbach *et al.* 1994). Cathepsin-L-like genes have been identified in nematodes (Loukas *et al.* 1998), cestodes (Liu *et al.* 1996), and trematodes (Roche *et al.* 1997). Cathepsin-C-like sequences are known for *S. mansoni* (Michel *et al.* 1995) and *S. japonicum* (Hola-Jamriska *et al.* 1998), and a cathepsin-S-like gene was reported in *Heterodera glycines* (Urwin *et al.* 1997). Among the most recently described subfamilies is one whose members have been variously termed cathepsin P (Acc. No. AF009923), X (Gay and Walker 1985; Nagler and Menard 1998), Y (Acc. No. AB023781), or Z (Santamaría *et al.* 1998), which are characterized by a unique 3 amino acid (aa) insertion close to the active site.

We report here the cloning of *Tc-cpz-1*, a cysteine protease gene expressed in infective stage larvae and adults of the parasitic nematode *Toxocara canis* that has a high level of identity (59%) with human cathepsin Z (Santamaría *et al.* 1998) and shares the features thought to be distinctive in its functions. We prefer the classification cathepsin Z as this was the term given to the first full-length sequence deposited in GenBank (for Acc. No. see Fig. 2 legend). Our analysis of the cathepsin family in the light of this sequence indicates that the cathepsin Z subfamily is certainly expressed among both parasitic (*Onchocerca volvulus* and *T. canis*) and free-living protostomes (*Caenorhabditis elegans* and *Urechis caupo*) and represents a novel monophyletic group in the C1 family of proteases.

A partial cDNA sequence from larval *T. canis* was first obtained

[1]The nucleotide sequence of *Tc-cpz-1* has been deposited in GenBank under Accession Number AF143817.

[2]Present address: London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

[3]Present address: CSIRO Plant Industry, GPO Box 160, Canberra, A.C.T. 2601, Australia.

[4]Present address: Molecular Parasitology Unit, Queensland Institute of Medical Research, Royal Brisbane Hospital, Queensland 4029, Australia.

[5]To whom correspondence should be addressed. Fax: +44-131-650 5450. E-mail: r.maizels@ed.ac.uk.

0014-4894/00 $35.00

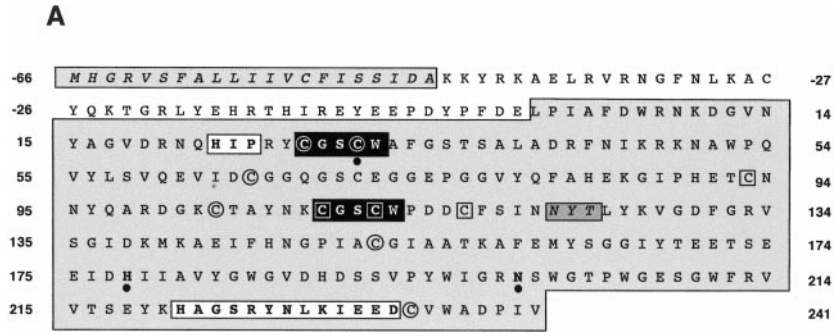
from an expressed sequence tag project (Tetteh *et al.* 1999) (clone *Tc*-EST-205, 5′: AI080921 and 3′: AI080922). The 800-bp insert in *Tc*-EST-205 was truncated at the 5′ end, since there was no start methionine codon. From the available single-pass sequences, an internal antisense primer was designed (5′ CGTATTCCCTTATG TGGGTCCTG-TGCTCG 3′). This was used in combination with a vector-derived primer to amplify the 5′ end from the cDNA library by PCR. The sequences obtained were aligned using AssemblyLign v1.0.5 (Kodak International Biotechnologies, Eastman Kodak Company, New Haven, CT, U.S.A.) and a consensus sequence for the full 5′ end was derived. Two new primers were designed from the 5′ untranslated region (UTR) (5′-ATGCATGGTCGAGTCTCGTTTGCC-3′, 24 nt overlap) and the 3′ UTR (5′-CGCGCCGGATCCCTA TACGATCGGGTCAGC-3′, 18 nt overlap). Full-length cDNA was amplified using the above primers from a λ-phage (Uni-ZAPXR) cDNA library (kindly provided by Drs. C. Tripp and R. Grieve, Heska, Fort Collins, CO, U.S.A.) using *Taq* polymerase (Promega, Madison, WI, U.S.A.) and the following PCR protocol: 45 s at 94°C, 45 s at 55°C, and 2 min at 72°C, 35 cycles, followed by a 10-min final extension at 72°C. Figure 1A shows the predicted aa sequence of *Tc*-CPZ-1. The open reading frame encodes a 307 amino acid protein that consists of a hydrophobic signal sequence (positions -66 to -46, boxed, predicted cleavage site Ala (-46)) (Nakai and Horton 1999) and an unusually short predicted propeptide (positions -45 to -1), in comparison with other cysteine proteases. The propeptide contains neither the ERFNIN motif (Karrer *et al.* 1993) that is characteristic of non-cathepsin-B papain-like cysteine proteases nor the GNFD signature, a motif that plays a role in the regulation of the intramolecular processing and subsequent activation of papain (Vernet *et al.* 1995), and is found in cathepsin B and L proregions. Both motifs are present in *Tc*-CPL-1 (Loukas *et al.* 1998), a previously described cathepsin-L-like enzyme from the same stage of *T. canis*. The absence of the GNFD motif in the cathepsin-Z-like proregion, as well as the notably shorter length of the propeptide (45 amino acids, typical average lengths for other proregions of cathepsins are cathepsins B: 60 aa, cathepsins L: 100 aa, cathepsins C: 200 aa), suggest that it might not be an effective inhibitor of the mature enzymes, although in human cathepsin L the first 20 N-terminal residues are more important in inhibition than the rest of the propeptide (Carmona *et al.* 1996).

The predicted mature protein (Fig. 1A, boxed) consists of a 241-aa polypeptide chain with a single putative *N*-glycosylation site (Asn-122, dark gray box, italics). As is found for most cysteine proteases, proline is the second aa in the mature protein. The protein contains the residues found in papain to form the catalytic dyad (Cys-31 and His-178, or Cys-25 and His-159 papain numbering) and a hydrogen bond with the active site histidine (Asn-200, or Asn-175 papain numbering, highlighted by a dot in Fig. 1A) (Storer and Menard 1994). Interestingly, the protein sequence contains a complete repeat of the conserved CGSCW motif around the active site cysteine (Cys-31) at positions 109–113 (Fig. 1A, black box). Furthermore, at least two specific inserts characteristic of the human cathepsin Z (Santamaría *et al.* 1998) are also found in *Tc*-CPZ-1, namely a HIP motif at position 23–25, shortly before the active site Cys-31 (black background in Fig. 1A), and a second, larger insert that is not found in other cathepsins, from position 221 to 233 (both inserts in white boxes in Fig. 1A). Performing a BLASTP search (Altschul *et al.* 1990) with *Tc*-CPZ-1 against the GenBank nonredundant protein database identified human cathepsin Z and six closely related sequences (Acc. Nos. in Fig. 2 legend) from *O. volvulus* (LOVCP), *U. caupo* (innkeeper worm), *Rattus norvegicus*, *Mus musculus*, and *C. elegans* (F32B5.8 and M04G12.2, The *C. elegans* Genome Sequencing Consortium 1998).

Figure 1B shows an alignment of the protein sequences of mature *Tc*-CPZ-1 and its homologs. The closest homologs of *Tc*-CPZ-1 are LOVCP (Lustigman *et al.* 1996), a larval cysteine protease from *O. volvulus* with 78% identity, followed by a cysteine protease of *U. caupo*, which had previously been described as cathepsin-B-like (Rosenthal 1993) (61% identity), as well as rat cathepsin Y and mouse cathepsin Z (both 61% identity), and by human cathepsin Z (59% identity). All nematode sequences show the complete repeat of the CGSCW motif at positions 31–35 and 109–113. As can be seen from the sequence alignment, the regions around the active site cysteine and histidine in cathepsins (highlighted by black dots in Figs. 1A and 1B) are highly conserved in all homologs of *Tc*-CPZ-1. This is also true for further residues that are thought to be involved in the mechanism of catalysis, such as Gln-22 (Gln-19 papain numbering) and Asn-200 (Asn-158 papain numbering) (Storer and Menard 1994). Interestingly, there is a region of divergence between the nematode and human cathepsin Z between Lys-158 and Ala-181 with only 8/25 identities, indicating a possible target for selective parasite cathepsin Z inhibitors.

The most remarkable feature of all eight proven or putative cathepsin-Z-like cysteine proteases is the perfectly conserved HIP insert between Asn-22 and the active site cysteine Cys-31. The introduction of a three-residue insert that is potentially positively charged in close proximity to the active site and the oxyanion hole is likely to affect the mechanism

FIG. 1. (A): Deduced protein sequence of *Tc-cpz-1*. The amplicons obtained were purified using the QIAquick gel extraction kit (Qiagen, Crawley, UK) and ligated into pGEM-T (pGEM-T East Vector System, Promega). Recombinant plasmids were purified with the QIAprep spin miniprep kit (Qiagen) from cultures of transformed JM109 bacteria and subsequently sequenced in both directions on an ABI Prism 377 DNA sequencer (Applied Biosystems, Foster City, CA, U.S.A.) using the AmpliTaq DyeDeoxy terminator cycle sequencing system (Applied Biosystems) with specific (above) and vector primers (M13 forward, M13 reverse). The sequences were analyzed and translated into a putative protein sequence using MacVector 6.0 software (Oxford Molecular Group, Oxford, UK). The signal sequence and its cleavage site (boxed in gray, italics) were predicted with PSORT II on the PSORT WWW Server (Nakai and Horton 1999). The amino acids are numbered using negative numbers for the preproregion and positive numbers for the mature protein. The active site cysteine (Cys-31), histidine (His-178), and asparagine (Asn-200) are highlighted with a black dot. The single potential N-glycosylation site, is boxed in dark gray (pos. 122–124). The two inserts found only in cathepsin-Z-like proteases are boxed in white (pos. 23–25 and 221–233). The cysteines also present in papain are circled; the four additional cysteines conserved in cathepsin-Z-like enzymes are boxed. The two repeats of the CGSCW motif are highlighted by black boxes. (28–32 and 109–113). (B): Multiple alignment of the deduced mature protein sequences of *Tc*-CPZ-1 and its homologs identified by a PBLAST search on the GenBank server (Altschul *et al.* 1990). Gaps have been introduced to maximize alignment. Boxes designate identical aa. The aa residues essential for activity are marked by a dot.


A
-66 MHGRVSFALLIIVCFISSIDAKKYRKAELRVRNGFNLKAC -27
-26 YQKTGRLYEHRTHIREYEEPDYPFDELPIAFDWRNKDGVN 14
15 YAGVDRNQHIPRYCGSCWAFGSTSALADRFNIKRKNAWPQ 54
55 VYLSVQEVIDCGGQGSCEGGEPGGVYQFAHEKGIPHETCN 94
95 NYQARDGKCTAYNKCGSCWPDDCFSINNYTLYKVGDFGRV 134
135 SGIDKMKAEIFHNGPIACGIAATKAFEMYSGGIYTEETSE 174
175 EIDHIIAVYGWGVDHDSSVPYWIGRNSWGTPWGESGWFRV 214
215 VTSEYKHAGSRYNLKIEEDCVWADPIV 241
B
Tc-CPZ-1
O. volvulus LOVCP
C. elegans F32B5.8
R. norvegicus cathepsin Y
M. musculus cathepsin Z
U. caupo cathepsin B-like
H. sapiens cathepsin Z
C. elegans M04G12.2
R. norvegicus cathepsin Y 290 DTRTRSVGCTPRGDRKG

of catalysis (Storer and Menard 1994). It will be interesting to examine what effect this insertion has on substrate specificity and affinity (e.g., by expressing active deletion mutants without the HIP insertion). Furthermore, whereas mature papain contains six cysteine residues that are known to be involved in three intramolecular disulfide bonds (circled cysteines in Fig. 1A), *Tc*-CPZ-1 and its homologs contain four additional conserved cysteines (boxed cysteines in Fig. 1A), suggesting that cathepsin-Z-like proteases might possess up to two additional intramolecular disulfide bonds.

Since human cathepsin Z, *Tc*-CPZ-1, and the six homologs identified in GenBank shared these novel sequence characteristics, we asked whether these cathepsin-Z-like proteases form a distinct branch separated from other cathepsin subfamilies in a phylogenetic analysis. Figure 2A shows a phylogenetic tree generated using neighbor joining with an alignment of the mature protein sequences of various papain-like cathepsins in which *Tc*-CPZ-1 is grouped in a branch together with the other identified cathepsin-Z-like proteases. This branch had a bootstrap value of 100% and excluded mammalian and parasitic cathepsins B, C, or L. A very similar phylogenetic tree was obtained using maximum parsimony (not shown), with slightly different bootstrap values. The four different cathepsin subfamilies (Z, B, C, and L) each had a bootstrap support of 100%. We generated phylograms including or excluding regions of uncertain alignment, cathepsin-Z-specific indels, or the region corresponding to the occluding loop of cathepsins B obtaining very similar trees (not shown). The internal branch obtained with our selection of representative cathepsins (Fig. 2A) suggests that cathepsins Z are more closely related to cathepsins L than cathepsins B or C, but this is supported at the margin of significance (66%). Therefore, we cannot speculate on the wider relationship of the Z subfamily of C1 peptidases. Our phylogenetic analysis, however, strongly suggests that the cathepsin Z subfamily represents a very old lineage, whose evolution predates the divergence of a common bilaterian ancestor into the protostome and deuterostome branches.

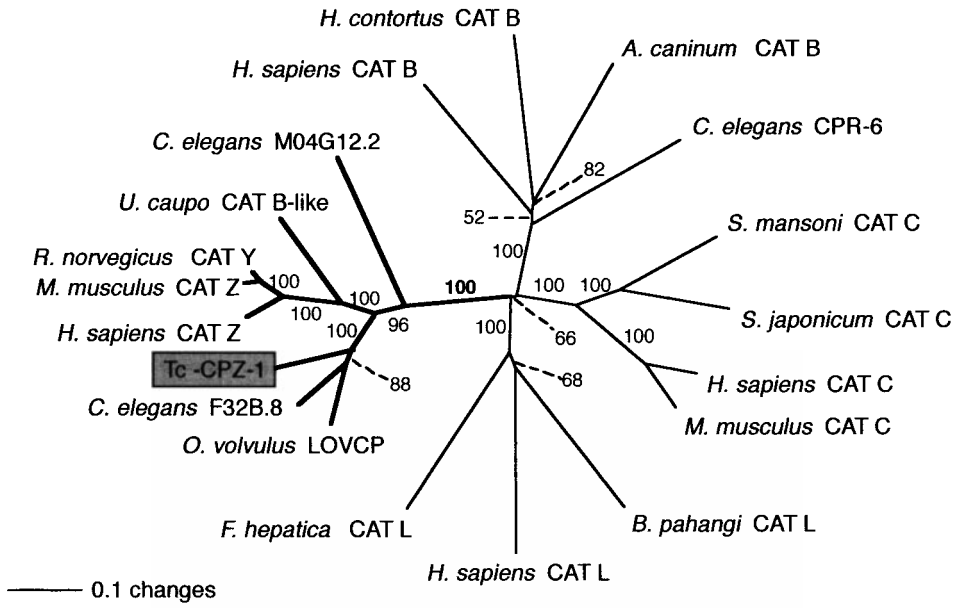
The very high degree of conservation (78% identity) between *Tc*-CPZ-1 and LOVCP implies that *Tc*-CPZ-1 might have an important function in parasite physiology. Lustigman *et al.* (1996) found that LOVCP is involved in the molting process of larvae. *Tc-cpz-1*, however, was isolated from a cDNA library obtained from arrested-stage larvae, which are not in the process of molting, but which *in vivo* are activated by factors such as pregnancy in the canine host. We therefore asked whether *Tc-cpz-1* expression is upregulated by the addition of pregnant canine serum to the culture medium, under conditions known to induce enhanced protein synthesis (A. P. Page, L. X. Liu, and R. M. Maizels, unpublished observation). We performed RT-PCR experiments using RNA derived from larvae cultivated with medium only (Hepes-buffered, serum-free RPMI 1640 supplemented with 1% (w/v) glucose and penicillin/streptomycin) or medium additionally supplemented with either 10% normal or pregnant canine serum. The larvae were incubated at 37°C for 48 h. RNA was also isolated from different adult *T. canis* tissues (total adult male or female parasites, heads, ovaries, or hypodermal/muscle preparations) as described previously (Loukas *et al.* 1999). The results, shown in Fig. 2B, indicate constitutive *Tc-cpz-1* expression in larvae under all conditions, in adults of both sexes and in different tissues, with the notable exception of the head. The lack of upregulation of *Tc-cpz-1* in induced larvae and its constitutive expression in the adult stage preclude an exclusive role of *Tc*-CPZ-1 in larval molting, but rather suggest a more fundamental biochemical function.

We have also generated recombinant mature *Tc*-CPZ-1 with an N-terminal poly-His purification tag and a thrombin cleavage site in bacteria. Despite high levels of expression, recombinant protein has been poorly soluble and enzymatically inactive. Moreover, antibodies to the insoluble fusion protein failed to detect native *Tc*-CPZ-1 in larval or adult soluble or detergent (0.2% Empigen) *T. canis* extracts. It seems likely that eukaryotic expression will be required to advance both enzymatic and immunological analyses.

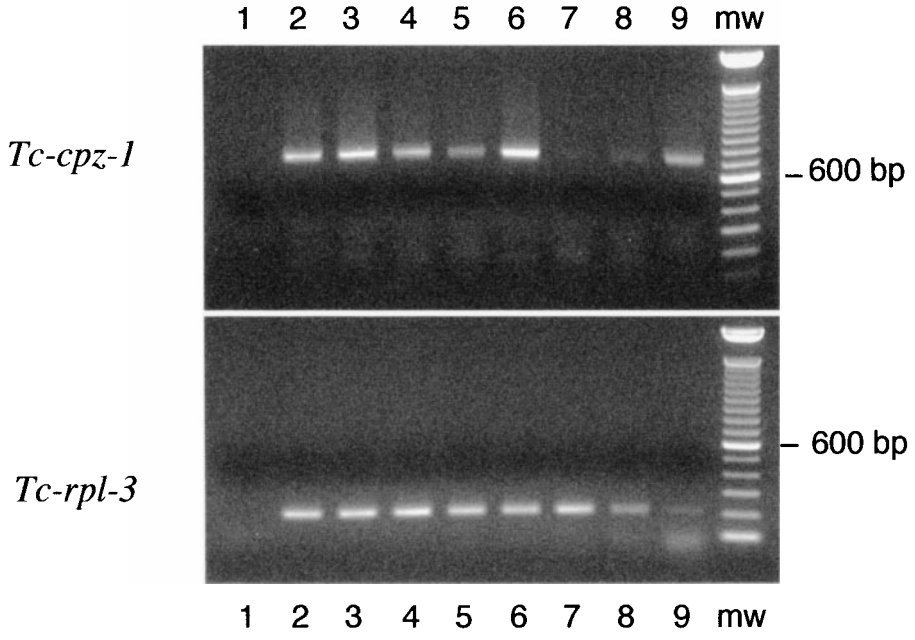
Further characterization of the substrate specificity of *Tc*-CPZ-1 would be highly desirable. Human cathepsin Z has been shown to cleave the synthetic cathepsin L and B substrate CBZ-Phe-Arg-MCA (Santamaría *et al.* 1998), but no cathepsin-Z-specific substrates are available as yet. We have previously observed that the excretory/secretory antigens of *T. canis* (TES) do not cleave CBZ-Phe-Arg-MCA, whereas somatic extracts (TEX) do (Loukas *et al.* 1998). We cannot yet determine whether this cleavage is attributable to the activity of *Tc*-CPL-1 (Loukas *et al.* 1998) or to the combined activities of *Tc*-CPL-1 and *Tc*-CPZ-1. Future studies will be aimed at elucidating the roles of both these cysteine proteases in parasite physiology and the likelihood of their acting as targets of novel chemotherapeutic compounds.

FIG. 2. (A): The neighbor joining phylogram obtained from aligned mature proteases of the cathepsin Z subfamily and other members of the C1 family using PAUP 4.b1 software (Sinauer Associates, Sunderland, MA, U.S.A.). The alignment is available from the authors. Bootstrap values obtained by 1000 replicates are shown next to the branches. GenBank Accession Numbers (clockwise): *A. caninum* CAT B (AcCP-1), U18911; *C. elegans* CAT B (protein cpr-6), U41556; *S. mansoni* CAT C, Q26563; *S. japonicum* CAT C, U77932; *Homo sapiens* CAT C, P53634; *M. musculus* CAT C, P97821; *B. pahangi* CAT L, AF031819; *H. sapiens* CAT L, Y18462; *F. hepatica* CAT L, L33771; *O. volvulus* LOVCP, U71150; *T. canis Tc*-CPZ-1, AF143817; *C. elegans* F32B5.8, P92005; *H. sapiens* CAT Z, AF032906; *M. musculus* CAT Z, AJ242663; *R. norvegicus* CAT Y, AB023781; *U. caupo* cathepsin B-like, Q27125; *C. elegans* M04G12.2, O01850; *H. sapiens* CAT B, L16510; *H. contortus* CAT B, P19092. (B): RT-PCR reaction performed on RNA extracted from adult worms or infective larvae (cultured under different conditions), representative of four experiments with the same results. Total RNA was isolated using RNA Stat-60 (TEL-Test, Inc., Friendswood, TX, U.S.A.) and first-strand cDNA obtained with the GeneAMP RNA PCR kit (Perkin–Elmer) using oligo-dT primer. (Lane 1): Negative control without template; (lane 2): larvae with medium alone; (lane 3): larvae in medium with 10% normal canine serum; (lane 4): medium with 10% pregnant canine serum; (lane 5): adult female; (lane 6): adult male; (lane 7): adult head; (lane 8): adult hypodermis/muscle scrape; (lane 9): adult ovaries; mw: 100-bp ladder (Gibco). The primers and PCR conditions used were as used for the amplification of the full-length sequence from the cDNA library. The lower gel shows a positive control which would also detect potential contamination with genomic DNA; the primers (described in Loukas *et al.* 1999) flank an intron and amplify a 220-bp fragment from the cDNA of *Tc-rpl-3*, a gene encoding a constitutively expressed ribosomal protein from *T. canis* (Acc. No. U17358).

A


H. contortus CAT B
A. caninum CAT B
H. sapiens CAT B
C. elegans CPR-6
C. elegans M04G12.2
82
U. caupo CAT B-like
52
S. mansoni CAT C
100
R. norvegicus CAT Y
M. musculus CAT Z
100
100
100
100
100
100
S. japonicum CAT C
H. sapiens CAT Z
100
96
100
66
100
Tc -CPZ-1
88
68
H. sapiens CAT C
C. elegans F32B.8
M. musculus CAT C
O. volvulus LOVCP
F. hepatica CAT L
B. pahangi CAT L
H. sapiens CAT L
0.1 changes

B


1 2 3 4 5 6 7 8 9 mw
Tc-cpz-1
600 bp
Tc-rpl-3
600 bp
1 2 3 4 5 6 7 8 9 mw

ACKNOWLEDGMENTS

This work was supported by the DFG (Fa 359/1-1), the Medical Research Council, and the Wellcome Trust. The authors express their thanks to David Guiliano, William F. Gregory, and Xing-Xing Zang for their help and expertise.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403–410.

Barrett, A. J., and Rawlings, N. D. 1996. Families and clans of cysteine proteases. *Perspectives in Drug Discovery and Design* **6,** 1–11.

Berti, P. J., and Storer, A. C. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. *Journal of Molecular Biology* **246,** 273–283.

Butler, R., Michel, A., Kunz, W., and Klinkert, M.-Q. 1995. Sequence of *Schistosoma mansoni* cathepsin C and its structural comparison with papain and cathepsins B and L of the parasite. *Protein and Peptide Letters* **2,** 313–320.

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

Coombs, G. H., and Mottram, J. C. 1997. Parasite proteinases and amino acid metabolism: Possibilities for chemotherapeutic exploitation. *Parasitology* **114,** S61–S80.

Engel, J. C., Doyle, P. S., Hsieh, I., and McKerrow, J. H. 1998. Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *Journal of Experimental Medicine* **188,** 725–734.

Gay, N. J., and Walker, J. E. 1985. Molecular cloning of a bovine cathepsin. *Biochemical Journal* **225,** 707–712.

Harrop, S. A., Sawangjaroen, N., Prociv, P., and Brindley, P. J. 1995. Characterization and localization of cathepsin B proteinases expressed by adult *Ancylostoma caninum* hookworms. *Molecular and Biochemical Parasitology* **71,** 163–171.

Hola-Jamriska, L., Tort, J. F., Dalton, J. P., Day, S. R., Fan, J., Aaskov, J., and Brindley, P. J. 1998. Cathepsin C from *Schistosoma japonicum* cDNA encoding the preproenzyme and its phylogenetic relationships. *European Journal of Biochemistry* **255,** 526–534.

Karrer, K. M., Peiffer, S. L., and DiTomas, M. E. 1993. Two distinct gene subfamilies within the family of cysteine protease genes. *Proceedings of the National Academy of Sciences of the USA* **90,** 3063–3067.

Liu, D. W., Kato, H., Nakamura, T., and Sugane, K. 1996. Molecular cloning and expression of the gene encoding a cysteine proteinase of *Spirometra erinacei. Molecular and Biochemical Parasitology* **76,** 11–21.

Loukas, A., Hunt, P., and Maizels, R. M. 1999. Cloning and expression of an aquaporin-like gene from a parasitic nematode. *Molecular and Biochemical Parasitology* **99,** 287–293.

Loukas, A., Selzer, P. M., and Maizels, R. M. 1998. Characterisation of *Tc*-CPL-1, a cathepsin L-like cysteine protease from *Toxocara canis* infective larvae. *Molecular and Biochemical Parasitology* **92,** 275–289.

Lustigman, S., McKerrow, J. H., Shah, K., Lui, J., Huima, T., Hough, M., and Brotman, B. 1996. Cloning of a cysteine protease required for the molting of *Onchocerca volvulus* third stage larvae. *Journal of Biological Chemistry* **271,** 30181–30189.

Merckelbach, A., Hasse, S., Dell, R., Eschlbeck, A., and Ruppel, A. 1994. cDNA sequences of *Schistosoma japonicum* coding for two cathepsin B-like proteins and Sj32. *Tropical Medicine and Parasitology* **45,** 193–198.

Michel, A., Ghoneim, H., Resto, M., Klinkert, M. Q., and Kunz, W. 1995. Sequence, characterization and localization of a cysteine proteinase cathepsin L in *Schistosoma mansoni. Molecular and Biochemical Parasitology* **73,** 7–18.

Nagler, D. K., and Menard, R. 1998. Human cathepsin X: A novel cysteine protease of the papain family with a very short proregion and unique insertions. *FEBS Letters* **434,** 135–139.

Nakai, K., and Horton, P. 1999. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Science* **24,** 34–36.

Pratt, D., Cox, G. N., Milhausen, M. J., and Boisvenue, R. J. 1990. A developmentally regulated cysteine protease gene family in *Haemonchus contortus. Molecular and Biochemical Parasitology* **43,** 181–191.

Roche, L., Dowd, A. J., Tort, J., McGonigle, S., McSweeney, A., Curley, G. P., Ryan, T., and Dalton, J. P. 1997. Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae. European Journal of Biochemistry* **245,** 373–380.

Rosenthal, E. 1993. Sequence analysis of translationally controlled maternal mRNAs from *Urechis caupo. Developmental Genetics* **14,** 485–491.

Santamaría, I., Velasco, G., Pendas, A. M., Fueyo, A., and Lopez-Otin, C. 1998. Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *Journal of Biological Chemistry* **273,** 16816–16823.

Storer, A. C., and Menard, R. 1994. Catalytic mechanism in papain family of cysteine peptides. *Methods of Enzymology* **244,** 486–500.

Tetteh, K. K. A., Loukas, A. C., Tripp, C., and Maizels, R. M. 1999. Cloning of novel and conserved genes from *Toxocara canis* by an EST strategy. *Infection and Immunity* **67,** 4771–4779.

The *C. elegans* Genome Sequencing Consortium. 1998. Genome sequence of *Caenorhabditis elegans*: A platform for investigating biology. *Science* **282,** 2012–2018.

Tort, J., Brindley, P. J., Knox, D., Wolfe, K. H., and Dalton, J. P. 1999. Proteinases and associated genes of parastic helmiths. *Advances in Parasitology* **43,** 161–266.

Urwin, P. E., Lilley, C. J., McPherson, M. J., and Atkinson, H. J. 1997. Characterization of two cDNAs encoding cysteine proteinases from the soybean cyst nematode *Heterodera glycines. Parasitology* **114,** 605–613.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C., and Thomas, D. Y. 1995. Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *Journal of Biological Chemistry* **270,** 10838–10846.

von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* **14,** 4683–4690.

Ward, W., Alvarado, L., Rawlings, N. D., Engel, J. C., Franklin, C., and McKerrow, J. H. 1997. A primitive enzyme for a primitive cell: The protease required for the excystation of *Giardia*. *Cell* **89,** 437–444.

Wilson, L. R., Good, R. T., Panaccio, M., Wijffels, G. L., Sandeman, R. M., and Spithill, T. W. 1998. *Fasciola hepatica*: Characterization and cloning of the major cathepsin B protease secreted by newly excysted juvenile liver fluke. *Experimental Parasitology* **88,** 85–94.

Received 17 September 1999, accepted with revision 31 January 2000