Cell, Vol. 89, 437–444, May 2, 1997, Copyright ©1997 by Cell Press

A Primitive Enzyme for a Primitive Cell: The Protease Required for Excystation of Giardia

Wendy Ward,* Lilia Alvarado,*† Neil D. Rawlings,‡ Juan C. Engel,* Christopher Franklin,* and James H. McKerrow*§
*Department of Pathology
University of California
San Francisco, California 94143
†Unidad de Investigacion Medica en Enfermedades Infecciosas y Parasitarias
CMNSXX1, IMSS
Mexico City DF06725
Mexico
‡Peptidase Laboratory
Department of Immunology
Babraham Institute
Cambridgeshire CB2 4AT
United Kingdom
§Department of Veterans Affairs Medical Center
San Francisco, California 94121

Summary

Protozoan parasites of the genus Giardia are one of the earliest lineages of eukaryotic cells. To initiate infection, trophozoites emerge from a cyst in the host. Excystation is blocked by specific cysteine protease inhibitors. Using a biotinylated inhibitor, the target protease was identified and its corresponding gene cloned. The protease was localized to vesicles that release their contents just prior to excystation. The Giardia protease is the earliest known branch of the cathepsin B family. Its phylogeny confirms that the cathepsin B lineage evolved in primitive eukaryotic cells, prior to the divergence of plant and animal kingdoms, and underscores the diversity of cellular functions that this enzyme family facilitates.

Introduction

Giardia parasites are diplomonad protozoa thought to represent ''biological fossils'' (archezoa) of the earliest lineage of eukaryotic cells (Peattie, 1990; Gillin et al., 1991, 1996; Siddall et al., 1992). Giardia lack mitochondria and peroxisome organelles present even in other protists, and sequence analysis of 16S ribosomal RNA supports their ancient lineage (Sogin et al., 1989; Leipe et al., 1993). The environment in which such early eukaryotic cells evolved was probably both hostile and unstable. The Giardia life cycle includes a cyst form that protects the enclosed trophozoite from desiccation and harsh chemical environments. This would also have facilitated adaptation to parasitism when suitable hosts became available. The success of this adaptation is underscored by the fact that Giardia lamblia is now the most common intestinal parasite of humans worldwide (Adam, 1991; Farthing, 1992; Kappus et al., 1994). It is an important cause of waterborne and restaurant associated outbreaks of diarrhea, travelers' diarrhea, and diarrhea in child care facilities (Levine et al., 1990; Sullivan et al., 1990; Adam, 1991; LeChevallier et al., 1991; Vidal et al., 1991; Quick et al., 1992).

The cyst wall is composed of a carbohydrate–peptide complex (Manning et al., 1992). The cyst ruptures and the motile, replicating trophozoite emerges in the proximal small intestine of the host. We now show that this key step in infection is dependent upon a cysteine protease stored in peripheral vesicles of the trophozoite and released into the space between trophozoite and cell wall during excystation. Specific inhibitors of the protease block excystation of Giardia in a dose-dependent manner.

The excystation protease is the product of one of three cysteine protease genes present in the Giardia genome. In biochemical assays of parasite extracts, cysteine proteases were previously identified as the major proteolytic activity of Giardia (Hare et al., 1989; Parenti, 1989; Werries et al., 1991). Sequence analysis places the Giardia cysteine protease gene family as the earliest known branch of the cathepsin B lineage, a major eukaryotic protease family thought to have evolved and diversified in parallel with emergence of the first eukaryotic cells (Berti and Storer, 1995).

Results

Cysteine Protease Inhibitors Block Excystation of Giardia

G. muris was chosen as the species to test the effects of cysteine protease inhibitors on excystation because of the higher level of in vitro excystation and greater reproducibility of excystation versus G. lamblia (Figure 1). Figure 2A shows that the epoxide cysteine protease inhibitor E-64 and three more specific fluoromethyl ketone–derivatized dipeptides arrested trophozoite excystation from 50% to 90% at 100 μM. Figure 2B confirms a dose-dependent inhibition of excystation with Mu-Tyr(Ome)-hPhe-FMK at 7–100 μM.

Localization of Cysteine Protease to Cytoplasmic Vacuoles in Trophozoites

Fluorescent microscopy following exposure to the fluorescent cysteine protease substrate Z-F-R-4-MNA (Meester et al., 1990) showed that cysteine protease activity in G. lamblia is localized in cytoplasmic vacuoles (Figure 3). These vacuoles are distinct organelles of Giardia (Friend, 1966), which had been reported to contain acid phosphatase (Feely and Dyer, 1987) and protease activity by sedimentation studies (Lindmark, 1988).

Cysteine Protease Inhibitors Have No Effect on Trophozoite Motility or Replication

Addition of 25 μM Mu-Tyr(Ome)hPhe-FMK in DMSO to culture medium inhibited trophozoite cysteine protease activity 100% as assessed by the Z-F-R-4-MNA assay (Figure 3). The presence of inhibitor did not decrease the rate of replication or alter morphology as judged by

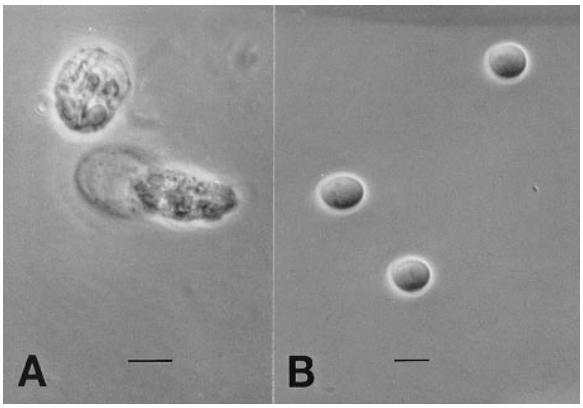

Figure 1. G. muris Cysts with Trophozoites Emerging and Intact

The effect of protease inhibitors on excystation was scored by such a light microscopic assessment. Bar, 16 μm. (A) Trophozoites emerging; magnification, 1000×. (B) Trophozoites intact; magnification, 400×.

light microscopy. These results suggest that G. lamblia trophozoites do not depend on cysteine proteases for growth or cytokinesis. Induction of encystation in the presence of 25 μM Mu-Tyr(Ome)hPhe-FMK produced water-resistant cysts of the same quantity and quality as controls.

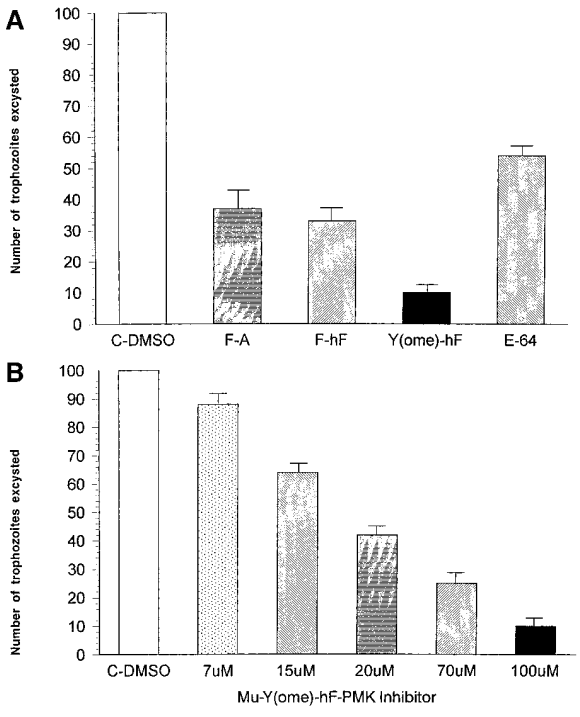

Figure 2. Inhibition of Giardia Excystation

(A) Effect of 100 μM cysteine protease inhibitors on Giardia excystation (n = 4). (C-DMSO), DMSO control; (F-A), Z-F-A-FMK; (F-hF), Mu-F-hF-FMK; (Y(Ome)-hF), Mu-Y(Ome)-hF-FMK; (E-64), E-64c.

(B) Dose dependence of inhibition of Giardia excystation by Mu-Y(Ome)hF-FMK (n = 4).

Identification and Purification of the G. lamblia Excystation Cysteine Protease, CP2, Targeted by Fluoromethyl Ketone Inhibitors

Soluble proteins from a G. lamblia trophozoite extract were centrifuged, desalted, and purified by gel chromatography and anion exchange chromatography. Cysteine protease activity was monitored by cleavage of the fluorescent substrate Z-F-R-AMC in the presence of 5 mM dithiothreitol (DTT). Active fractions following anion exchange were pooled and incubated with biotinylated F-A-FMK (Bt-F-A-FMK). Bt-F-A-FMK, a biotinylated derivative of one of the fluoromethyl ketones that inhibited excystation (Figure 2A), covalently binds to the active site of targeted cysteine proteases (Eakin et al., 1992). The pooled active fractions, including any biotin-tagged cysteine proteases, were resolved by SDS polyacrylamide gel electrophoresis (SDS–PAGE). A single 30 kDa cysteine protease was detected by Western blot after incubation with streptavidin-horseradish peroxidase and substrate development (Figure 4). A corresponding Coomassie-stained band was excised from a parallel blot for N-terminal amino acid sequencing (Figure 4). The deduced N-terminal amino acid sequence (DKDDVPESFDFREE) was most similar to the N-terminus of the 29 kDa cathepsin B of Sarcophaga peregrini

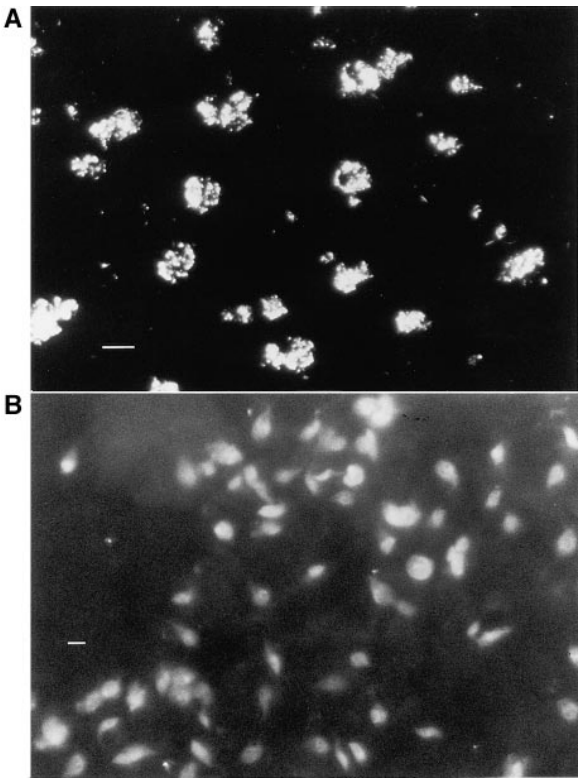

Figure 3. Protease Activity and Inhibition of Cleavage

(A) Localization of cysteine protease activity in trophozoite vacuoles by the yellow fluorescence released following protease cleavage of Z-F-R-4-MNA (25). Bar, 16 μm.

(B) Inhibition of Z-F-R-4MNA cleavage activity by preincubation of trophozoites with 25 μM Mu-Y(Ome)hF-FMK. The trophozoites can only be visualized by their blue autofluorescence after longer automatic exposure of film then in (A). Bar, 16 μm.

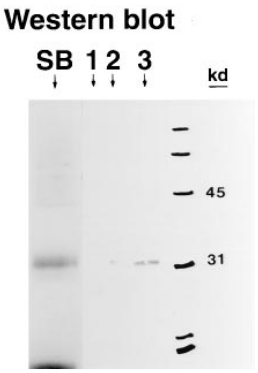

Figure 4. Identification of Protease Target of Fluoromethyl Ketone Inhibitor Following Incubation of Fractionated Giardia Extracts with Biotinylated Inhibitor

Proteins were resolved by SDS–PAGE and blotted to nitrocellulose. The membrane was incubated with streptavidin-HRP and developed with H_2O_2 and DAB peroxidase substrate. Note single species at ~30 kDa reacting with peroxidase-strepavidin. Lanes: 1, control with no inhibitor; 2, crude Giardia extract; 3, purified cysteine protease. SB, Coomassie-stained SDS–PAGE gel of Giardia extract that was incubated with Bt-F-A-FMK and target protease captured on streptavidin beads; beads were then washed and protein released by boiling in sample buffer.

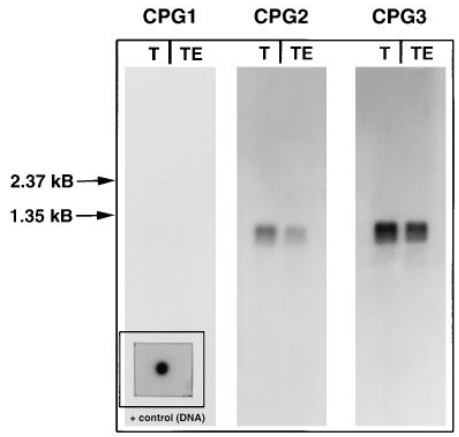

Figure 5. Expression of Cysteine Protease Genes

Northern blot showing expression of cysteine protease genes *CP2* (CPG2) and *CP3* (CPG3) but not *CP1* (CPG1) in mRNA of replicating (T) and encysting (TE) trophozoites in medium without supplemental iron.

(Kurata et al., 1992). It was used as a template to isolate the corresponding gene by PCR.

Identification, Cloning and Sequencing of the Excystation Cysteine Protease Gene of G. lamblia

Using Giardia codon bias, a gene fragment of the excystation cysteine protease (CP2) was amplified from G. lamblia genomic DNA by PCR using a degenerate primer based on the amino-terminal amino acid sequence of the target cysteine protease identified using the biotinylated FMK inhibitor, and a primer homologous to the conserved region around the active site Asn. A 595 bp PCR band was amplified, cloned, and sequenced. The deduced amino acid sequence was 100% homologous to the amino-terminal amino acid sequence of purified CP2 in the overlapping region. Generic cysteine protease primers (Eakin et al., 1990) were also used to screen the G. lamblia genome for papain family cysteine protease genes by PCR. Several bands between 300 and 750 bp were amplified. A 520 bp band contained the sequences of the original primers and sequence homology with cathepsin B. This cysteine protease gene fragment (CP1) was radiolabeled and used to probe a G. lamblia genomic DNA phage library. A full-length cysteine protease gene (*CP1*) and flanking sequences were cloned. The first ATG of the longest open reading frame (908 bp) was presumed to be the start codon. The first stop codon is followed by 42 bases and the conserved Giardia polyadenylation signal AGTPuAAPyr (Adam, 1991). Like most protozoan genes and all known G. lamblia genes, no introns are present in *CP1*. The deduced protein (including a putative proregion) is 303 aa and is 42% identical to human cathepsin B. Radiolabelled *CP2* was also used to reprobe the G. lamblia genomic DNA library. A third full-length cysteine protease gene was isolated (*CP3*). The 897 bp open reading frame when translated showed 80% nucleotide identity to *CP2*, and 83% identity at the predicted amino acid level. *CP1*, by contrast, is only 52% identical. The nucleotide sequences of the three Giardia cysteine protease genes have been deposited in GenBank (U83275, U83276, U83277).

A Southern blot of pulsed-field gel electrophoresis of G. lamblia DNA was probed sequentially with the three G. lamblia cysteine protease gene fragments. *CP1* is located on the small chromosome II, *CP2* and *CP3* are located on chromosome IV. Restriction digest and Southern blot analysis suggest *CP1* is a single gene and *CP2* and *CP3* are single or low copy number (not shown). Northern blot analysis of messenger RNA isolated from trophozoites and encysting trophozoites show that under culture conditions without supplemental iron, *CP1* is not expressed, but both *CP2* and *CP3* are (Figure 5).

Relationship of the Giardia Cysteine Protease Gene Family to Cathepsin B

Sequence analysis, database searches, and alignment (Figure 6) of the three G. lamblia cysteine proteases suggested that they are prototypes of the cathepsin B subgroup of the peptidase family C1 (Rawlings and Barrett, 1994). While lacking the active site ''loop'' found in higher eukaryote cathepsin Bs, the Giardia proteases have more homology to cathepsin B than to papain or cathepsin L. Three cysteines (residues 14, 43, and 100) unique to cathepsin B align well with the Giardia sequences (Figure 6). The S2 substrate binding pocket of the Giardia proteases is also similar to cathepsin B, with glycine at residue 198, and a negatively charged glutamate at residue 245. This is thought to allow for the cleavage of substrates with positively charged arginine in the P2 position (Hasnain et al., 1992), a biochemical property of cathepsin B. To confirm this prediction,

Figure 6. Alignment of Giardia Cysteine Proteases, Cathepsin B, and Papain

Alignment of two Giardia cysteine proteases with cathepsin B (CATB) and papain showing α helix (a) and β structure (b). Single underline identifies regions where CP1 and CP2 are more like cathepsin B than papain, and double underline indicates regions of homology for all four proteases. Note identity of cathepsin B cysteines at residues 14, 43, and 100, but absent active site ''loop'' (104–126) in Giardia proteases. Numbering system is for cathepsin B (Musil et al., 1991). Only residues 1–221 are shown for simplicity. Complete sequences have been deposited in GenBank (accession numbers U83275, U83276, and U83277).

the Giardia protease was incubated with two fluorescently labeled dipeptide substrates, Z-F-R-AMC and Z-R-R-AMC. Both were cleaved as measured by release of the fluorescent AMC group (2.2 fu/min and 0.44 fu/min, respectively). Under the same conditions, papain had no activity versus Z-R-R-AMC.

Phylogenetic analysis (Figure 7) confirms the position of the G. lamblia cysteine protease gene family as the first branch of the cathepsin B subgroup of peptidase family C1. The Giardia proteases are ancestral to both the divergence of the dipeptidyl peptidase I group, and even the divergence of plant and animal cathepsin B.

Discussion

In Giardia, an archezoan ''biological fossil'' of eukaryotic evolution, a prototype cathepsin B protease is required by the organism to rupture and to emerge from the cyst form that provides resistance to desiccation and gastric acid. A simple life cycle, of cyst formation followed later by cyst degradation to release a replicative form, may have been essential to the adaptation and evolution of early eukaryotic cells in the hostile environment in which they evolved. The eventual adaptation of protozoa like Giardia to parasitism would have been aided by the ability of the cyst to survive gastric acid before trophozoites emerge in the nutrient-rich intestine of the host. This evolutionary theme may have been expressed again in more advanced eukaryotic parasites—cysteine proteases have also been implicated in excystation of metacercariae of the parasitic lung fluke Paragonimus westermani (Chung et al., 1995).

The excystation protease is encoded by one of three cysteine protease genes identified in G. lamblia. CP2 was identified as the excystation protease by first using the biotinylated form of the irreversible cysteine protease inhibitor to ''tag'' its targets in Giardia. Only one protein species was targeted and its amino-terminal amino acid sequence determined. CP2 showed 100% identity in the region of overlap. Furthermore, Northern blot analysis indicated that the *CP1* gene was not expressed under the culture conditions used in these studies. All three genes are prototypes of the cathepsin B subfamily of peptidase family C1 (the papain family), confirming speculation from previous phylogenetic studies that this gene subfamily had its origin in very primitive eukaryotic cells (Berti and Storer, 1995). As the first known branch of this cathepsin B subfamily, the Giardia gene sequences are ancestral to both mammalian cathepsin B and dipeptidyl-peptidase I. While dipeptidyl-peptidase I would be expected to be absent from Giardia, it should be present in more advanced protozoa, because the divergence of cathepsin B and DPP1 precedes the divergence of cathepsin B of Leishmania and plants. The Giardia genes branch later than the primordial divergence of the bacterial cytosolic aminopeptidases, papain, and the cathepsin L subgroups (Figure 7).

The predicted amino acid sequence of the Giardia cathepsin B is an interesting ''chimera'' between mammalian cathepsin B and papain or cathepsin L. For example, several of the signature cysteines of mammalian cathepsin B are present (Figure 6), but the active site loop that confers dipeptidyl-peptidase activity is absent. Structural additions like the loop, as well as specific interdomain charged residues identified by crystallographic studies of mammalian cathepsin B (Musil et al., 1991), were in place as early as the evolution of the more advanced kinetoplastid protozoan parasites like Leishmania, and before the divergence of the plant and animal kingdoms.

The histochemical localization of cysteine protease activity (Figure 3) is consistent with its proposed role in excystation. The protease activity is targeted to vacuoles in the cytoplasm that were previously shown to move to the margin of the cells at the time of excystation, and then to release their contents into the space between trophozoite and cyst wall (Coggins and Schaefer, 1986; Feely and Dyer, 1987; Lindmark, 1988). No effects of cysteine protease inhibitors were noted on trophozoite or cyst viability, trophozoite motility, or trophozoite replication. It has been suggested that processing of cyst wall precursors may be dependent on cysteine protease activity (H. Lujan and T. Nash, personal communication). Further studies will be necessary to evaluate this possibility.

Members of the cathepsin B subgroup were long thought to function primarily, if not exclusively, like mammalian lysosomal cathepsin B in intracellular protein degradation. However, the Giardia cathepsin B is released extracellularly for cyst wall digestion; blood fluke cathepsin B is secreted into the worm intestine to aid in hemoglobin degradation (Wasilewski et al., 1996); and even mammalian cathepsin B may be membrane-associated or secreted in invading tumor cells (Elliott and Sloane, 1996). These observations now underscore

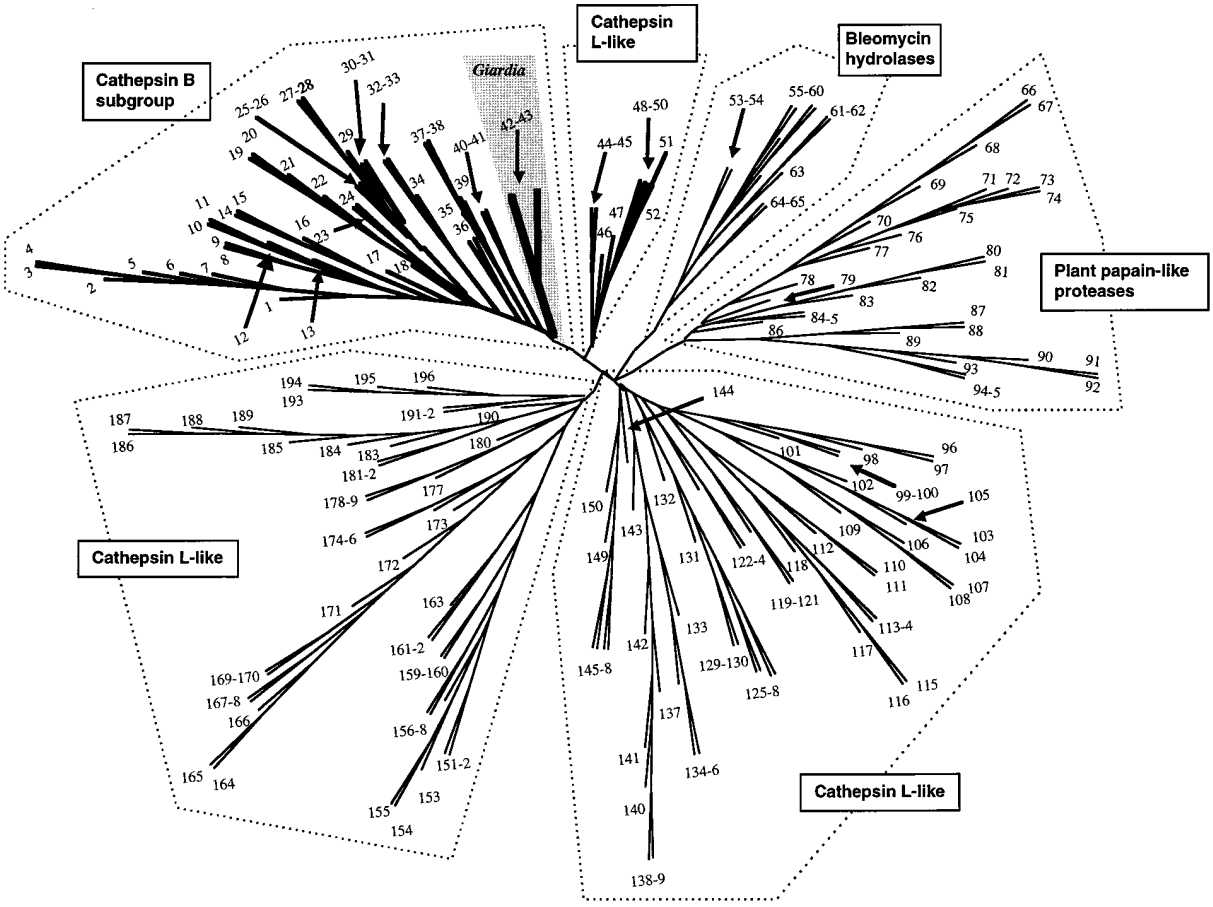

Figure 7. A Consensus Neighbor-Joining Phylogenetic Tree for Peptidase Sequences from the Entire C1 Peptidase Family

The cathepsin B subgroup is highlighted and the Giardia branch labeled. The sequences are numbered 1 to 196. The following abbreviations are used: cysteine endopeptidase (CE), cathepsin B (CB), cathepsin L (CL), cathepsin H (CH), cathepsin S (CS), cathepsin K (CK), dipeptidyl-peptidase I (DPP-I), and aminopeptidase (AP). Key to sequences: 1, Sarcophaga peregrina CB; 2, mouse CB; 3–4, chicken CB; 5, rat CB; 6, cattle CB; 7, human CB; 8–13, Caenorhabditis elegans CB; 14–15, Schistosoma japonicum CB; 16, Schistosoma mansoni CB; 17, Leishmania mexicana CB; (other cathepsin B-like sequences) 18, Aedes aegypti CE; 19–22, Ancyclostoma caninum CE; 23, Strongyloides ratti CE; 24, Haemonchus contortus CE; 25–26, Ostertagia ostertagi CE; 27–31, Haemonchus contortus CE; 32–33, wheat CE; 34, tobacco CE; 35, Fasciola hepatica CE; 36, rabbit tubulointerstitial nephritis antigen; 37, human DPP-I; 38, rat DPP-I; 39, Schistosoma mansoni DPP-I; 40, Cattle cathepsin X; 41, Urechis caupo CE; 42, Giardia lamblia CE1; 43, Giardia lamblia CE2; 44, Toxocara canis CE; 45, Caenorhabditis elegans CE; 46, yeast AP; 47, Tritrichomonas foetus CE; 48, Streptococcus thermophilus AP; 49, Lactococcus lactis AP; 50, Lactobacillus delbrueckii AP; 51, Lactobacillus helveticus AP; 52, rabbit bleomycin hydrolase; 53, alder CE; 54, Arabidopsis thaliana CE; 55, Phalaenopsis sp. CE; 56–57, day lily CE; 58–59, barley CE; 60, rice CE; 61, French bean CE; 62, mung bean CE; 63, spring vetch CE; 64–65, Mesembryanthemum crystallinum CE; 66, chickpea CE; 67, spring vetch CE; 68, Phaseolus vulgaris CE; 69, Brassica napus CE; 70, pea CE; 71, tomato CE; 72, pea CE; 73, Arabidopsis thaliana CE; 74, oryzain a; 75, carnation CE; 76, oryzain β; 77, pseudotzain; 78, Zinnia elegans CE; 79, actinidian; 80, caricain; 81, glycyl endopeptidase; 82, papain; 83, chymopapain; 84, stem bromelain; 85, Tetrahymena thermophila CE; 86, soya bean oil bodies-associated protein P34; 87, mouse CH; 88, rat CH; 89, human CH; 90, tomato CE; 91–92, petunia CE; 93, pea CE; 94, oryzain g; 95, aleurain; 96, mouse CL; 97, rat CL; 98, human CL; 99, rat testin; 100, rat placental cathepsin; 101, chicken CL; 102, carp CE; 103, rat CS; 104, cattle CS; 105, human CS; 106, mouse CK; 107, rabbit CK; 108, human CK; 109, chicken CK; 110, Drosophila melanogaster CE; 111, Sarcophaga peregrina CE; 112, silk moth CE; 113, Homarus americanus CE3; 114, Nephrops norvegicus stomach CE; 115, Homarus americanus CE1; 116, Nephrops norvegicus eye-stalk CE; 117, Homarus americanus CE2; 118, Penaeus vannamei CE; 119–121, Dictyostelium discoideum CE2; 122, Schistosoma mansoni CE; 123, Schistosoma japonicum CE; 124, Spirometra mansoniodes CE; 125–132, Fasciola hepatica CE's; 133–134, Entamoeba histolytica CE's; 135, Entamoeba dispar CE; 136–137, Entamoeba histolytica CE's; 138–139, Entamoeba invadens CE's; 140–143, Entamoeba histolytica CE's; 144, Tritrichomonas foetus CE3; 145–148, Trichomonas vaginalis CE's; 149–150, Tritrichomonas foetus CE's; 151–152, Trypanosoma congolense CE's; 153–155, Trypanosoma brucei CE's; 156, Trypanosoma rangeli CE; 157–158, Trypanosoma cruzi CE's; 159, Leishmania pifanoi CE1; 160, Leishmania mexicana CE1; 161, Leishmania pifanoi CE2; 162–163, Leishmania mexicana CE's; 164–5, tobacco CE7; 166, tomato CE2; 167–168, Arabidopsis thaliana CE's; 169, spring vetch CE; 170; pea CE; 171, maize CE; 172, soya bean CE; 173, Dictyostelium discoideum CE1; 174, Naegleria fowleri CE; 175, Schistosoma japonicum CE; 176, Schistosoma mansoni CE; 177, Choristoneura fumiferana nuclear polyhedrosis virus CE; 178, Bombyx mori nuclear polyhedrosis virus CE; 179, Autographa californica nuclear polyhedrosis virus CE; 180, Paragonimus westermani CE; 181, Plasmodium vinckei CE; 182, Plasmodium berghei CE; 183, Plasmodium gallinaceum CE; 184, Plasmodium falciparum CE; 185, Plasmodium fragile CE; 186, Plasmodium ovale; 187, Plasmodium malariae CE; 188, Plasmodium vivax; 189, Plasmodium cynomolgi; 190, human cathepsin O; 191, Theileria annulata CE; 192, Theileria parva CE; 193–195, house dust mite faecal antigens; 196, Euroglyphus maynei faecal antigen. (From Rawlings and Barrett, 1994.)

the versatility of this enzyme family and suggest why there was an ''explosive'' divergence and dispersion of cysteine protease genes coincident with the evolution of the first eukaryotic cells (Berti and Storer, 1995).

Experimental Procedures

Reagents

All buffer reagents were purchased from Sigma Chemical Co. (St. Louis, MO) unless specified otherwise. Fluorescently labeled peptide substrates were from Enzyme Systems Products (Dublin, CA). All fluoromethyl ketone inhibitors were gifts of Mary Zimmerman and Robert Smith of Prototek, Inc. (Dublin, CA).

Giardia Cell Culture

G. lamblia WB trophozoites (Smith et al., 1982) were a gift of Alice Wang of the Department of Pharmaceutical Chemistry, University of California, San Francisco, and were subsequently maintained in modified TYI-S-33 medium (omitting ferric ammonium citrate) as previously described by Keister (Keister, 1983).

Encystation was induced by a modification of the procedure described by Kane et al. (1991): trophozoites were grown to near confluency; expended medium and unattached trophozoites were poured off; encystation medium (TYI-S-33 [pH 7.8] with 5 mg/ml bovine bile) was added and flasks were incubated at 37°C for 48–72 hr. Remaining trophozoites were lysed by incubation in distilled water for 45 min. Water-resistant cysts were collected by centrifugation (300 g) and stored at 4°C no longer than 7 days.

G. muris cysts were collected from feces of chronically infected nude mice. Cysts were isolated by pelleting and washing fecal sediment ten times with distilled water followed by centrifugation at 700 × g for 20 min over a 1 M sucrose cushion. Cysts recovered from the middle opalescent layer were stored at 4°C in distilled water prior to induction of excystation.

Excystation of G. muris fecal cysts was induced by incubation in 0.1 M potassium phosphate (pH 7.0), 0.3 M sodium bicarbonate for 30 min at 37° (Feely et al., 1991). Effects of cysteine protease inhibition on G. muris excystation were evaluated by incubation of fecal cysts with 7–100 μM of the peptidyl fluoromethyl ketone morpholine-tyrosine (O-methyl)-homophenylalanine-fluoromethyl ketone (Mu-Tyr(Ome)-hPhe-FMK) (24) or 10–150 μM L-trans-epoxysuccinyl-L-leucylamide-(4-guanidino)-butane (E-64). Inhibitor was added for 30 min at 37°C during the induction of excystation in bicarbonate and in media washes, and for 1.5 hr at 37°C in culture medium (TYI-S-33). Assessment of total excystation, partial excystation, or failure to excyst was determined with a Zeiss microscope adapted with Hoffman optics and objectives.

Cysteine protease activity was inhibited in cultured trophozoites by 25 μM of Mu-Tyr(Ome)hPhe-FMK (Rasnick, 1985) dissolved in DMSO. The final concentration of DMSO in culture medium was .025%. Adequacy of cysteine protease inhibition was assessed by Z-F-R-MNA fluorescence assay (described below). Parallel assays of untreated controls and DMSO controls were performed.

Cysteine protease activity was inhibited during encystation by maintaining 25 μM Mu-Tyr(Ome)hPhe-FMK until cysts were harvested.

Assay of Cysteine Protease Activity in Trophozoite Extracts

Trophozoites grown to confluency were pelleted at 700 × g and washed three times in PBS. They were then lysed in the presence of 0.1% Triton X-100, in a Fisher Model 300 ultrasonic homogenizer set at 60% and equipped with a microtip, for 7 min or until no intact trophozoites were visualized by light microscopy. The homogenate was centrifuged at 40,000 g for 60 min and the supernatant fraction was stored at −70°C prior to use. Cysteine protease activity was assayed by the liberation of the fluorescent leaving group 7-amino-4-methyl coumarin (AMC, Enzyme Systems Products, Livermore, CA) from the peptide substrates Z-Phe-Arg-AMC or Z-Arg-Arg-AMC (where Z is benzyloxycarbonyl). Substrate (5 μM) was dissolved in 100 mM sodium acetate, 5 mM DTT (pH 7.5). Fluorescence was measured upon addition of enzyme sample with a Perkin-Elmer LS 30 spectrofluorometer, excitation = 380 nm; emission = 460 nm. Cysteine protease activity was assayed against seven specific, irreversible peptidyl fluoromethyl ketone (FMK) cysteine protease inhibitors (a gift of Mary Zimmerman and Robert Smith, Prototek, Inc., Dublin, CA) by preincubating enzyme with inhibitor for 10 min at 22°C before adding to substrate.

Localization of Cysteine Protease Activity in Giardia

The dipeptide substrate Z-F-R-4-methoxy-2-napthylamine (MNA, Enzyme Systems Products, Livermore, CA) was used to identify cysteine protease activity in intact trophozoites and cysts as previously described by Meester et al. (1990). Trophozoites were observed and photographed with a Zeiss fluorescence microscope set at excitation = 340 nm; emission = 425 nm.

Isolation of G. lamblia Genomic DNA

Genomic DNA was prepared from approximately 5×10^8 G. lamblia trophozoites grown to confluency. Trophozoites were pelleted at 700 × g for 7 min, washed three times in PBS, lysed with .5% Sarkosyl, .5 M EDTA (pH 8.0) with 200 μg/ml proteinase K at 50°C for 2 hr, extracted three times with phenol/chloroform, dialyzed against 500 volumes of TE at 4°C, and precipitated with 5% C-TAB in .5M NaCl (Wang and Wang, 1986).

PCR Amplification, Cloning, and Sequencing of G. lamblia Cysteine Protease Genes

To identify all Giardia cysteine protease genes, degenerate primers designed around conserved structural motifs flanking active site Cys and Asn residues of papain family cysteine proteases (Eakin et al., 1990) were used to amplify DNA fragments from G. lamblia genomic DNA. To specifically identify the gene coding for the excystation protease, a degenerate primer was designed based on the N-terminal amino acid sequence of the purified cysteine protease identified as the inhibitor target. It was coupled with a primer homologous to the active site Asn to amplify a DNA fragment of the excystation protease gene (*CP2*) from genomic DNA by PCR. In reactions with one degenerate primer and one homologous primer, the annealing temperature was 65°C. In reactions with two degenerate primers, the annealing temperature was 45°C. Ampli-Taq polymerase (Perkin-Elmer Cetus, Foster, CA) or Taq polymerase from Boehringer Mannheim (Indianapolis, IN) was used in all PCR reactions. The DNA thermocycler was obtained from Perkin Elmer Cetus Instruments (Foster City, CA).

PCR fragments were gel-purified with Sephaglass (Pharmacia, Uppsala, Sweden) and subcloned into Bluescript KS^- plasmid (pBS) from Strategene (La Jolla, CA). Double-stranded DNA was sequenced directly using pBS priming sites and Sequenase enzyme and reagents from U.S. Biochemical (Cleveland, OH) or Amplitaq polymerase and cycle sequencing reagents from Perkin Elmer Cetus (Foster, CA) according to manufacturers' instructions.

Genomic Library Construction, Cloning, and Sequencing

A G. lamblia genomic library of 15–20 kb fragments was constructed in Lambda EMBL phage using the Lambda GEM-11 system (Promega Corporation, Madison, WI) according to the manufacturer's instructions. The library was screened consecutively by plaque hybridization using the PCR-generated gene fragments labeled with [α-^{32}P]dCTP (Amersham, Arlington Heights, IL) by random priming. Phage DNA was transferred to NYTRAN circles (Schleicher and Schuell, Keene, NH) and hybridized in 6 × SSC, 5% BLOTTO at 68°C, then washed stringently with 0.1 × SSC, 0.1% SDS at 68°C according to Maniatis et al. (1987). For ease of handling, large inserts from positive clones were digested with various restriction endonucleases, electrophoresed on a 1% agarose, Tris acetate–buffered gel, and transferred to NYTRAN (Schleicher and Schuell, Keene, NH) by Southern blot. Bands containing cysteine protease genes were identified by probing the blots with the corresponding labeled PCR fragment. Double-stranded DNA was sequenced as described above, designing sequencing primers homologous to internal sequence previously determined from PCR fragments. The sequences were confirmed for both stands of DNA in each instance.

Southern Blots and Pulsed-Field Gel Electrophoresis of Giardia

Genomic DNA (1 μg/reaction) from G. lamblia was digested with excess Bgl II and Pst I (Boehringer Mannheim, Indianapolis, IN). The DNA was transferred to NYTRAN (Schleicher and Schuell, Keene, NH) in neutral transfer buffer with the Schleicher and Schuell (Keene, NH) turboblotter downward transfer system according to manufacturer's instructions. The blot was probed sequentially with PCR-amplified gene fragments of *CP1*, *CP2*, and *CP3*. Hybridization was carried out in 50% formamide, 5 × Denhardt's, 1% SDS, 1 μg/ml tRNA at 45°C O/N and washed stringently with .1 × SSC, .1% SDS at 65°C. Between probes, the blot was stripped with .4N NaOH at 45°C for 30 min and neutralized with two changes of 200 mM Tris-Cl (pH 7.0), .1 × SSC, .1% SDS (Maniatis et al., 1987). Chromosomes were separated by pulsed-field gel electrophoresis on .9% chromosomal grade agarose (Bio Rad, Hercules, CA), Tris-acetate-buffered gel at 75 V for 82 hr using Hula Gel (Hoeffer, San Francisco, CA) with ramped pulse time of 1000–2000 s, and 106° reorientation angle with S. pombe and S. cerevisiae size markers (FMC, Rockland, ME). Chromosomes were transferred to Hybond-N (Amersham, Arlington Heights, IL) with a Posiblot (Hoefer, San Francisco, CA) and probed as described for Southern blots.

Northern Blots

Total RNA was extracted from 5×10^8 trophozoites and 5×10^8 trophozoites exposed for 24 hr to encystation stimuli using RNAzol B (BIOTECX, Houston, TX) according to manufacturer's instructions. Ten micrograms of total RNA from trophozoites was electrophoresed on a formaldehyde 1% agarose gel in formamide running buffer (Ausubel et al., 1994) and transferred to NYTRAN (Schleicher and Schuell, Keene, NH) with a Turboblotter downward transfer system (Schleicher and Schuell, Keene, NH) according to manufacturer's instructions. Messenger RNA from encysting trophozoites was isolated from total RNA with DYNAbeads (DYNAL, Oslo, Norway), electrophoresed, and transferred as described for total RNA. Blots were hybridized sequentially with labeled PCR fragments from *CP1*, *CP2*, and *CP3* in 50% formamide, 5 × Denhardt's, 1% SDS, 1 μg/ml tRNA at 45°C O/N and washed stringently with .1 × SSC, .1% SDS at 65°C.

Purification and Amino-Terminal Sequencing of the Excystation Protease Targeted by Fluoromethyl Ketone Inhibitors

Supernatant from sonicated trophozoites was prepared as described above from 100 flasks of confluent trophozoites (approximately 1×10^{10}). The supernatant was applied to a 1 m ACA 54 column in 200 mM sodium acetate, 2 mM DTT (pH 7.8). Fractions were collected at .2 ml/min over 18 hr and assayed for cysteine protease activity with the fluorescent substrate Z-F-R-AMC. Ten microliters of elute was added to 100 μl 50 mM HEPES (pH 7.4), 5 mM DTT, 5 μM Z-F-R-AMC and assessed with a Labsystems Fluoroscan II spectrofluorometric plate reader governed by Delta Soft II software (excitation = 355 nm and emission = 460 nm). All active fractions were pooled and concentrated with Centriprep 10 size selective filters (Amicon, Beverly, MA). Concentrate was applied to a MonoQ HR 5/5 anion exchange column (Pharmacia, Uppsala, Sweden) in 20 mM Tris (pH 7.4) with NaCl gradient from 250 to 750 mM, flowing at 1 ml/min. All active fractions were again pooled, concentrated, and the target of the fluoromethyl ketone cysteine protease inhibitors identified by incubation with 20 μM biotinylated Z-F-A-FMK (Prototek, Inc., Dublin, CA) in 50 mM HEPES (pH 7.4) 5 mM DTT for 3 hr at room temperature. The sample was suspended in loading buffer with 100 mM DTT, heated to 100°C for 5 min, and resolved by SDS–PAGE (Laemmli, 1970). The proteins were electroblotted to Pro-Blot (Applied Biosystems, Foster City, CA) using a TE 70 Semiphor Transfer Unit (Hoefer Scientific Instruments, San Francisco, CA) with 10 mM CAPS buffer according to manufacturer's instructions. Biotinylated cysteine proteases were localized by incubating the blot with streptavidin-HRP (PIERCE, Rockford, IL) and developing with diaminobenzidine. A corresponding Coomassie-stained band from a parallel SDS–PAGE was excised from Pro-Blot (Applied Biosystems, Foster City, CA). Peptide sequencing of the amino terminus was performed at the Biomolecular Resource Center, University of California at San Francisco, using a gas phase sequencer (Applied Biosystems Inc., Foster City, CA).

Sequence Alignment and Protease Phylogeny

An alignment of 196 members of the papain family was prepared by use of the PILEUP program from the GCG package (Genetics Computer Group, 1994). Only that part of the alignment corresponding to the catalytic domain was used. The alignment was further modified to reflect the structural alignments derived from comparisons of tertiary structures (Musil et al., 1991; Joshua-Tor et al., 1995), by removal of inserts present in only a few sequences, trimming of the C-termini, and manually aligning all conserved cysteine residues thought to be involved in disulfide bonds. The phylogenetic analysis was performed using the Neighbor-Joining algorithm from the Phylip package (Saitou and Nei, 1987; Felsenstein, 1989), and the final tree was a consensus of 30 trees generated by bootstrap analysis.

Acknowledgments

Wendy Ward was supported by a Howard Hughes Postdoctoral Research Fellowhip for Physicians. James H. McKerrow is a recipient of the Burroughs Wellcome Molecular Parasitology Scholar Award. This work was supported by a TDRU grant (#AI35707) from the NIAID. The authors thank Francis Gillin for advice and training in the excystation assays, Alan Barrett for advice on phylogenetic analysis, Wolfram Bode for helpful discussions of cathepsin B structural motifs, and Xiaowu Chen for help with structural alignments.

Received February 11, 1997; revised March 21, 1997.

References

Adam, R.D. (1991). The biology of *Giardia spp*. Microbiol. Rev. *55*, 706–732.

Ausubel, F.M., Brent, R., Kingston, R.L., Moore, D.D., Seidman, J. G., Smith, J.A., and Struhl, K., eds. (1994). Current Protocols in Mol. Biol. (New York: John Wiley & Sons, Inc.).

Berti, P.J., and Storer, A.C. (1995). Alignment/phylogeny of the papain superfamily of cysteine proteases. J. Mol. Biol. *246*, 273–283.

Chung, Y.-B., Kong, Y., Joo, I.-J., Cho, S.-Y., and Kang, S.-Y. (1995). Excystment of *Paragonimus westermani* metacercariae by endogenous cysteine protease. J. Parasitol. *81*, 137–142.

Coggins, J.R., and Schaefer, F.W., III (1986). *Giardia muris*: ultrastructural analysis of in vitro excystation. Exp. Parasitol. *61*, 219–228.

Eakin, A.E., Bouvier, J., Sakanari, J.A., Craik, C.S., and McKerrow, J.H. (1990). Amplification and sequencing of genomic DNA fragments encoding cysteine proteases from protozoan parasites. Mol. Biochem. Parasitol. *39*, 1–8.

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

Elliott, E., and Sloane, B. (1996). The cysteine protease cathepsin B in cancer. In Cysteine Proteases, J.H. McKerrow and M.N.G. James, eds., Series: Perspectives in Drug Discovery Design, P.S. Anderson, G.L. Kenyon, and G.R. Marshall, eds. (Leiden: ESCOM), pp. 12–32.

Farthing, M.J.G. (1992). New perspectives in giardiasis. J. Med. Microbiol. *37*, 1–2.

Feely, D.E., and Dyer, J.K. (1987). Localization of acid phosphatase activity in *Giardia lamblia* and *Giardia muris* trophozoites. J. Protozool. *34*, 80–83.

Feely, D.E., Gardner, M.D., and Hardin, E.L. (1991). Excystation of *Giardia muris* induced by a phosphate-bicarbonate medium: localization of acid phosphatase. J. Parasitol. *77*, 441–448.

Felsenstein, J. (1989). PHYLIP—phylogeny inference package. Cladistics *5*, 164–166.

Friend, D.S. (1966). The fine structure of *Giardia muris*. J. Cell Biol. *29*, 317–332.

Gillin, F.D., Reiner, D.S., and McCaffrey, M. (1991). Organelles of protein transport in *G. lamblia*. Parasitol. Today *7*, 113–116.

Gillin, F.D., Reiner, D.S., and McCaffery, J.M. (1996). Cell biology of the primitive eukaryote *Giardia lamblia*. Annu. Rev. Microbiol. *50*, 679–705.

Hare, D.F., Jarroll, E.L., and Lindmark, D.G. (1989). *Giardia lamblia*: characterization of proteinase activity in trophozoites. Exp. Parasitol. *68*, 168–175.

Hasnain, S., Huber, C.P., Muir, A., Rowan, A.D., and Mort, J.S. (1992). Investigation of structure function relationships in cathepsin B. Biol. Chem. Hoppe-Seyler *373*, 413–418.

Joshua-Tor, L., Xu, H.E., Johnston, S.A., and Rees, D.C. (1995). Crystal structure of a conserved protease that binds DNA: the bleomycin hydrolase, Gal6. Science *269*, 945–950.

Kane, A.V., Ward, H.D., Keusch, G.T., and Pereira, M.E. (1991). *In vitro* encystation of *Giardia lamblia*: large-scale production of *in vitro* cysts and strain and clone differences in encystation efficiency. J. Parasitol. *77*, 974–981.

Kappus, K.D., Lundgren, R.G., Jr., Juranek, D.D., Roberts, J.M., and Spencer, H.C. (1994). Intestinal parasitism in the United States: update on a continuing problem. Am. J. Trop. Med. Hyg. *50*, 705–713.

Keister, D.B. (1983). Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. Trans. R. Soc. Trop. Med. Hyg. *77*, 487–488.

Kurata, S., Saito, H., and Natori, S. (1992). Purification of a 29-kDa hemocyte proteinase of *Sarcophaga peregrina*. Eur. J. Biochem. *204*, 911–914.

Laemmili, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *277*, 680–685.

LeChevallier, M.W., Norton, W.D., and Lee, R.G. (1991). Giardia and *Cryptosporidium spp.* in filtered drinking water supplies. Appl. Environ. Microbiol. *57*, 2617–2621.

Leipe, D.D., Gunderson, J.H., Nerad, T.A., and Sogin, M.L. (1993). Small subunit ribosomal RNA+ of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Mol. Biochem. Parasitol. *59*, 41–48.

Levine, W.C., Stephenson, W.T., and Craun, G.F. (1990). Waterborne disease outbreaks, 1986–1988. Morbid. Mortal. Weekly Report CDC Surveillance Summaries *89*, 1–13.

Lindmark, D.G. (1988). *Giardia lamblia*: localization of hydrolase activities in lysosomelike organelles of trophozoites. Exp. Parasitol. *65*, 141–147.

Maniatis, T., Fritsch, E.F., and Sambrook, J., eds. (1987). Molecular Cloning, A Laboratory Manual, (New York: Cold Spring Harbor Laboratories).

Manning, P., Erlandsen, S.L., and Jarroll, E.L. (1992). Carbohydrate and amino acid analyses of *Giardia muris* cysts. J. Protozool. *89*, 290–296.

Meester, F.D., Shaw, E., Scholze, H., Stolarsky, T., and Mirelman, D. (1990). Specific labeling of cysteine proteinases in pathogenic and nonpathogenic *Entamoeba histolytica*. Infect. Immun. *58*, 1396–1401.

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

Parenti, D.M. (1989). Characterization of a thiol proteinase in *Giardia lamblia*. J. Infect. Dis. *160*, 1076–1080.

Peattie, D.A. (1990). The giardins of *Giardia lamblia*: genes and proteins with promise. Parasitol. Today *6*, 52–56.

Genetics Computer Group. (1994). Program Manual, Wisconsin Sequence Analysis Package, Vers. 8. (Madison, Wisconsin: GCG).

Quick, R., Paugh, K., Addiss, D., Kobayashi, J., and Baron, R. (1992). Restaurant-associated outbreak of giardiasis. J. Infect. Dis. *166*, 673–676.

Rasnick, D. (1985). Synthesis of peptide fluoromethyl ketones and the inhibition of cathepsin B. Anal. Biochem. *149*, 461–465.

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

Saitou, N., and Nei, M. (1987). The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Siddall, M.E., Hong, H., and Desser, S.S. (1992). Phylogenetic analysis of the diplomonadida (Wenyon, 1926) Brugerolle, 1975: evidence for heterochrony in protozoa and against *Giardia lamblia* as a "issing link." J. Protozool. *39*, 361–367.

Smith, P.D., Gillin, F.D., Kaushal, N.A., and Nash, T.E. (1982). Antigenic analysis of *Giardia lamblia* from Afghanistan, Puerto Rico, Ecuador, and Oregon. Infect. Immun. *36*, 714–719.

Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A., and Peattie, D.A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science *243*, 75–77.

Sullivan, P.B., Marsh, M.N., Phillips, M.B., Dewit, O., Neale, G., Cevllos, A.M., Yamson, P., and Farthing, M.J.G. (1990). Prevalence and treatment of giardiasis in chronic diarrhoea and malnutrition. Arch. Dis. Childhood *65*, 304–306.

Vidal, M.F.C., Gilman, R.H., Ungar, B.L.P., Verastegui, M.R., Benel, A.D., Marquis, G., Penny, M., Lanata, C., and Mirana, E. (1991). Detection of *Giardia lamblia* antigen in children living in a Peruvian periurban shantytown (Pueblo Joven). J. Clin. Microbiol. *29*, 636–637.

Wang, A.L., and Wang, C.C. (1986). Discovery of a specific double-stranded RNA virus in *Giardia lamblia*. Mol. Biochem. Parasitol. *12*, 269–276.

Wasilewski, M.M., Lim, K.C., Phillips, J., and McKerrow, J.H. (1996). Cysteine protease inhibitors block schistosome hemoglobin degradation *in vitro* and decrease worm burden and egg production in vivo. Mol. Biochemical Parasitol. *61*, 179–189.

Werries, E., Franz, A., Hippe, H., and Acil, Y. (1991). Purification and substrate specificity of two cysteine proteinases of *Giardia lamblia*. J. Protozool. *38*, 378–383.

GenBank Accession Numbers

The nucleotide sequences of the three Giardia cysteine protease genes, *CP1*, *CP2*, and *CP3*, have been deposited in GenBank with accession numbers U83275, U83276, and U83277, respectively.