The Intestinal Protozoan Parasite Entamoeba histolytica Contains 20 Cysteine Protease Genes, of Which Only a Small Subset Is Expressed during In Vitro Cultivation

Iris Bruchhaus,1* Brendan J. Loftus,2 Neil Hall,3 and Egbert Tannich1

Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany,1 The Institute for Genomic Research, Rockville, Maryland 20850,2 and The Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom3

Received 17 October 2002/Accepted 13 February 2003

Cysteine proteases are known to be important pathogenicity factors of the protozoan parasite Entamoeba histolytica. So far, a total of eight genes coding for cysteine proteases have been identified in E. histolytica, two of which are absent in the closely related nonpathogenic species Entamoeba dispar. However, present knowledge is restricted to enzymes expressed during in vitro cultivation of the parasite, which might represent only a subset of the entire repertoire. Taking advantage of the current E. histolytica genome-sequencing efforts, we analyzed databases containing more than 99% of all ameba gene sequences for the presence of cysteine protease genes. A total of 20 full-length genes was identified (including all eight genes previously reported), which show 10 to 86% sequence identity. The various genes obviously originated from two separate ancestors since they form two distinct clades. Despite cathepsin B-like substrate specificities, all of the ameba polypeptides are structurally related to cathepsin L-like enzymes. None of the previously described enzymes but 7 of the 12 newly identified proteases are unique compared to cathepsins of higher eukaryotes in that they are predicted to have transmembrane or glycosylphosphatidylinositol anchor attachment domains. Southern blot analysis revealed that orthologous sequences for all of the newly identified proteases are present in E. dispar. Interestingly, the majority of the various cysteine protease genes are not expressed in E. histolytica or E. dispar trophozoites during in vitro cultivation. Therefore, it is likely that at least some of these enzymes are required for infection of the human host and/or for completion of the parasite life cycle.

Cysteine proteases (EC 3.4.22) of the papain superfamily occur in a wide range of organisms including bacteria, plants, invertebrates, and vertebrates (7). In mammals, they are well documented as intracellular enzymes involved in protein turnover within lysosomes. In addition, extracellular cysteine proteases have been implicated in various physiological and pathophysiological processes, including tumor invasion and metastasis (33, 43). Cysteine proteases are important virulence factors of various infectious agents and the main proteolytic enzymes in many protozoan parasites (22, 29). The protozoan Entamoeba histolytica, the causative agent of human amoebiasis, is characterized by its great capacity to destroy host tissue, leading to potentially life-threatening diseases such as ulcerative colitis or liver abscess. Convincing evidence exists that cysteine proteases are essential for E. histolytica-induced pathology. Treatment of the ameba with sublethal doses of a specific cysteine protease inhibitor or the addition of laminin, which blocks the substrate-binding pocket of cysteine proteases, greatly reduces its ability to produce liver abscesses in laboratory animals (18, 37). Likewise, liver abscess formation was totally blocked and significantly less gut inflammation and damage to the intestinal permeability barrier were observed when ameba genetically engineered to produce low levels of cysteine protease activity were used to infect the animals (1, 2, 47).

Until now, a total of eight genes coding for cysteine proteases in E. histolytica have been identified (ehcp1 to ehcp7 and ehcp112) (8, 12, 26, 38, 39) (EhCP7 accession number, AJ409105). Three of the corresponding proteins, namely, EhCP1, EhCP2, and EhCP5, have been purified (15, 30, 32) and were found to account for approximately 90% of the total cysteine protease activity present in lysates of axenically cultured E. histolytica trophozoites (8). Accordingly, Northern blot analysis indicated high-level expression of ehcp1, ehcp2, and ehcp5, low-level expression of ehcp3, and no expression of ehcp4 and ehcp6 (8). So far, the expression pattern of ehcp3 and ehcp112 has not been investigated. Interestingly, in contrast to all other cysteine proteases, functional genes corresponding to ehcp1 and ehcp5 are absent in Entamoeba dispar, a closely related but nonpathogenic Entamoeba species. By comparing the E. histolytica and E. dispar genomic loci containing the gene for CP5, it was found that the position of cp5 within the genomic context is conserved between the two organisms. However, the gene is highly degenerated in E. dispar, since it contains numerous nucleotide exchanges, insertions, and deletions, resulting in multiple stop codons within the cp5 reading frame (46). With respect to E. histolytica pathogenicity, EhCP5 appears to be of special importance, since it is the only cysteine protease known so far that is present on the ameba surface (15). However, the mechanism of surface association remains to be determined, since the molecule does not contain any known surface attachment moiety.

Our current knowledge of the activity of cysteine proteases in E. histolytica is based on analysis of cultured ameba trophozoites. However, culture medium represents an artificial envi-
vironment which is not comparable to the situation found during infection of the human host. The various ehcp genes known so far have been cloned by screening E. histolytica cDNA or genomic libraries by using (i) antibodies or N-terminal sequence information of purified ameba enzymes, (ii) PCR-amplified fragments obtained with degenerate primers from regions highly conserved between cysteine protease genes of various organisms, or (iii) already cloned ameba ehcp genes as cross-hybridizing probes. The last approach identified only two additional E. histolytica sequences, since the various ehcp genes have only 40 to 80\% sequence identity (8). All of the ameba cp genes encode mature cysteine proteases with calculated molecular masses between 24 and 35 kDa. However, earlier studies using substrate gel electrophoresis suggested that E. histolytica contains a considerable number of cysteine proteases ranging from 16 to more than 100 kDa, some of which were considered to be associated with the ameba membrane (3, 19, 24, 27, 30, 35). This may imply that only a subset of E. histolytica cysteine proteases has been identified so far.

The recent progress in sequencing the E. histolytica genome already allows the identification of nearly all ameba genes, since the amount of sequence information deposited in the two available databases is considered to cover more than 99\% of the entire E. histolytica genome.

Here we report the identification and analysis of 12 novel genomics sequences encoding cysteine protease in E. histolytica, of which only 4 are expressed in cultured trophozoites.

**MATERIALS AND METHODS**

**Sequence identification.** Homology searches were performed using the E. histolytica databases of The Institute for Genomic Research (http://www.tigr.org) and The Wellcome Trust Sanger Institute (http://www.sanger.ac.uk). The eight known members of the E. histolytica cysteine protease family were used to query the databases, and additionally identified cysteine protease genes were then used for further analysis of the databases.

The sequencing effort of The Institute for Genomic Research is part of the International Entamoeba Genome Sequencing Project, and the funding for this project is being provided by the National Institute of Allergy and Infectious Diseases. The Wellcome Trust has funded the Sanger Institute Pathogen Sequencing Unit, in collaboration with Graham Clark at the London School of Hygiene and Tropical Medicine, to undertake whole-genome shotgun sequencing of E. histolytica.

Signal sequences and their cleavage sites were identified using the SignalP program (http://www.cbs.dtu.dk/services/SignalP). The prediction of the glycosylphosphatidylinositol (GPI) anchor and cleavage site was made by using DGPI (http://dgpi1.pathbot.com). The prediction of transmembrane helices in proteins was made by using TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0).

![FIG. 1. Structural organization of 20 E. histolytica cysteine proteases. Numbers indicate the number of amino acid residues forming the predomains, prodomains, or catalytic domains.](image_url)
For the prediction of protein-sorting signals and protein localization sites in cells, PSORTII software was used (http://psort.nibb.ac.jp/). The NetNGly WWW server was used to predict the N-glycosylation sites within the proteases (http://www.cbs.dtu.dk/services/NetNGlyc/) (R. Gupta, E. Jung, and S. Brunak, unpublished data). Homology searches were done using BLAST (http://www.ncbi.nlm.nih.gov/blast/). Multiple alignments were performed by CLUSTAL W (http://www2.ebi.ac.uk/clustalw/) (40). The phylogenetic tree was generated using the alignment performed by CLUSTAL W and the Phylip program package with the neighborhood-joining approach (http://evolution.genetics.washington.edu/phylip.html) (11).

**FIG. 2.** Phylogenetic tree of all 20 *E. histolytica* cysteine proteases and representative family members of other organisms. The tree was generated using sequence alignments of the region between the active-site Cys and the active-site His as described in Materials and Methods. Human cathepsin B was used as an outgroup. Two distinct groups are clustered within the *E. histolytica* protease family (EhCP-A and EhCP-B). The accession numbers for the various proteases are as follows: EhCP1, Q01957; EhCP2, Q01958; EhCP3, CAA60673; EhCP4, CAA62833; EhCP5, CAA62835; EhCP6, CAA62835; EhCP7, CA42109; EhCP8, AY156066; EhCP9, AK156067; EhCP10, AY156068; EhCP11, AY156096; EhCP12, AY156078; EhCP13, AY156079; EhCP14, AY156080; EhCP15, AY156075; EhCP16, AY156075; EhCP17, AY156076; EhCP18, AY156076; EhCP19, AY156072; EhCP20, AY156073; B. malayi, AAK16513; human cathepsin C, Q95029; Giardia lamblia cathepsin C (CatC), AAK97078; human cathepsin L, KHHUL; Leishmania mexicana, CAA44094; Litomosoides sigmodontis, AAK16515; Mus musculus, NP034114; Naegleria fowleri, AAB01769; Onchocerca volvulus, AAK16514; papain, P00784, Plasmodium falciparum, A45624; Schistosoma mansoni, CAA83538; Trypanosoma brucei, S07051; Trypanosoma congolense, S37048; Trypanosoma cruzi, P25779; and Trichomonas vaginalis, S41427.
about 10 cm, were used. Each slot was loaded with 200 μg of total *E. histolytica* (gel 1) or *E. dispar* (gel 2) RNA. After RNA separation, the gels were blotted onto nylon membranes, and subsequently the membranes were cut into 20 strips. One *E. histolytica* RNA-containing strip and one *E. dispar* RNA-containing strip were hybridized in parallel. Each pair of strips was sequentially hybridized with the respective *ehcp* probe as well as with *ehcp9* and an *E. histolytica* actin gene probe. Hybridization was performed in 0.5 M Na₂HPO₄–7% sodium dodecyl sulfate–1 mM EDTA (pH 7.2) at 55°C. The blots were washed in 40 mM Na₂HPO₄–1% sodium dodecyl sulfate (pH 7.2) at 55°C. The following oligonucleotides were used for PCR amplification to generate the various hybridization probes: EhCP8S (5′-GAGA GGTACC ATG TT T GCA GT T AT T CT G TTA AAC, EhCP10S (5′-GAGA GGTACC ATG TAT CGA AAT AAC CTT TTT TTT CTT), EhCP10AS (5′-GAGA GGATCC TTA CAA GAT AAC CTT CTT TTT TTT CTT), EhCP11S (5′-GAGA GGTACC ATG TATA GGA GTT TTG TTG), EhCP11AS (5′-GAGA GGTACC ATG TATA GGA GTT TTG TTG), EhCP12S (5′-GAGA GGTACC ATG AGC CAT TTG ATT ATT ATT GTT), EhCP12AS (5′-GAGA GGATCC TCA ATT AAA AAT TAT TGA GCG ATG), EhCP13S (5′-GAGA GGTACC ATG TCA ATT TTG CTT ATC ATC TTG, EhCP13AS (5′-GAGA GGTACC ATG TCA ATT TTG CTT ATC ATC TTG, EhCP14S (5′-GAGA GGTACC ATG TTT TTG TTC CTC GTA TTT C), EhCP14AS (5′-GAGA GGATCC ATG TTT TTG TTC CTC GTA TTT C) and EhCP15S (5′-GAGA GGTACC ATG TTT TTG TTC CTC GTA TTT C). Southern blot analysis. DNA was isolated from *E. histolytica* (HM-1:IMSS) and *E. dispar* (SAW 760) trophozoites by using DNAEasy (InVitrogen). For Southern blot analysis, agarose gels were loaded with 20 μg of restriction enzyme-digested DNA. The transfer was performed as described for the Northern blot analysis.

**RESULTS AND DISCUSSION**

Comprehensive BLAST search analysis of the two currently available *E. histolytica* genome databases identified a total of

| EhCP1 | APKAVDWRK | GK | VTPIRDQGCGS |
| EhCP2 | APESVDWRKE | GK | VTPIRDQGCGS |
| EhCP3 | APESVDWRSE | MN | PAKDQGCGS |
| EhCP4 | TATTKDRWN | GK | VTPVRDQGCGS |
| EhCP5 | DVPEVDWRN | GK | VPAIRDQGCGS |
| EhCP6 | TPIADWRNE | GK | LTPIRDHTQCGS |
| EhCP7 | VPANYLTCTSEAEN | YC5TN1DQGNGC |
| EhCP11 | VPQYLTCTSEAEN | YC5TN1DQGNGC |
| EhCP12 | VPINSACNQKTLFGKLNPEGIDECNGIEFQDGGS |
| EhCP13 | VPTQYASCLQNKLLGQNSNINDGTVQESEQGGS |
| EhCP14 | VPTMYSLCNGKINYSNADGKVDRCSCGLQKCLRCC |
| EhCP15 | IPSLSYSCGYYRKNPHKTE | DYC1KTMQNCSECCGGS |
| EhCP16 | PRRSSFRACSKYTLNNSNMRKPDSTD | YWSSSCCGGS |
| EhCP18 | LPEELSCDYGVVNTDTDPK WNLCLTP YQDGSCGGS |
| EhCP112 | LPQNYAPFCEYVSNKTDRPK WDLCGE VQFQCGGGS |

**FIG. 4.** Alignment of the ERFNIN motifs located within the prodomains of the various *E. histolytica* cysteine proteases. Conserved amino acid residues are printed in bold type.
20 different genes with conceptual open reading frames for cysteine proteases of the papain superfamily. These comprise all 8 ehcp genes previously identified (ehcp1 to ehcp7 and ehcp12) as well as 12 so far undescribed ehcp genes, designated ehcp8 to ehcp19. Since the two databases are considered to contain >99% of all E. histolytica genes, it is likely that these 20 ehcp sequences represent the entire repertoire of cysteine protease genes from this protozoan parasite. The various DNA-derived amino acid sequences could be attributed to cysteine proteases since they revealed significant similarity to cathepsin B- or cathepsin L-like enzymes. Sequence alignments revealed the highest similarity of all the ameba polypeptides to members of the cathepsin L-like subclass, irrespective of whether they belong to subfamily EhCP-A or EhCP-B. This was further supported by the finding that 19 of the 20 ameba molecules contain an ERFNIN motif positionally conserved within the pro-region (Fig. 4). This motif, Glu-(X)3-Arg-(X)2-Ile/Val-Phe-(X)2-Asn-(X)3-Ile-(X)3-Asn, and its position within the pro-region are known to be specific for cathepsin L-like proteases (16). However, despite their sequence similarity to cathepsin L-like enzymes, ameba cysteine proteases have previously been shown to be able to degrade Z-Arg-Arg-Ala/Val-Arg-Ala (13, 15, 30, 31), a synthetic peptide that is specifically cleaved by cathepsin B-like enzymes. The substrate specificity of cysteine proteases at the S2 pocket has been ascribed to the chemical properties of the residues corresponding to residue 205 of the papain sequence. Cathepsin B proteases usually have acidic groups (Asp or Glu) at this position, whereas mammalian cathepsin L proteases have an Ala residue at this position (4). Consistent with their substrate specificity, none of the amebic cysteine proteases contain an Ala residue at this position (4). The relatively large number of cp genes within the E. histolytica genome appears to be not simply the result of recent gene amplifications, since pairwise alignments show that the various enzymes have only 10 to 86% sequence identity. However, the open reading frames of two of the genes, ehcp14 and ehcp16, are interrupted by short introns of 77 and 59 bp, respectively, both of which contain the ameba-specific 5’ and 3’ splice consensus sequences (45). Construction of phylogenetic trees indicated that the various enzymes form two distinct clades (Fig. 2), suggesting that the 20 ameba cysteine protease genes evolved from two separate ancestors at a very early stage of parasite evolution. Accordingly, the various ameba molecules were grouped into two subfamilies, designated EhCP-A and EhCP-B. In general, proteases belonging to subfamily EhCP-A have shorter prodomains and shorter catalytic domains. The prodomains comprise 72 to 90 residues in subfamily EhCP-A versus 105 to 144 in EhCP-B, and the catalytic domains comprise 190 to 254 residues in EhCP-A versus 230 to 353 in EhCP-B. However, there are two exceptions. EhCP15, which has the shortest prodomain of all ameba enzymes (55 residues), belongs to subfamily EhCP-B, whereas EhCP17, which has the longest prodomain (148 residues), belongs to subfamily EhCP-A. In addition, the two subfamilies can be distinguished on the basis of two sequence characteristics that are located within the N-terminal part of the catalytic domain. Members of EhCP-A contain a conserved Asp-Trp-Arg (DWR) motif at position +6 to +8, and the active-site cysteine is located around position +25. In contrast, members of subfamily EhCP-B do not contain the DWR motif but have a conserved cysteine residue at position +8, and the active-site cysteine is located between positions +30 and +38 (Fig. 3).

Based on primary sequence and substrate specificity, cysteine proteases can be divided into distinct subclasses such as cathepsin B- or cathepsin L-like enzymes. Sequence alignments revealed the highest similarity of all the ameba polypeptides to members of the cathepsin L-like subclass, irrespective of whether they belong to subfamily EhCP-A or EhCP-B. This was further supported by the finding that 19 of the 20 ameba molecules contain an ERFNIN motif positionally conserved within the pro-region (Fig. 4). This motif, Glu-(X)3-Arg-(X)2-Ile/Val-Phe-(X)2-Asn-(X)3-Ile-(X)3-Asn, and its position within the pro-region are known to be specific for cathepsin L-like proteases (16). However, despite their sequence similarity to cathepsin L-like enzymes, ameba cysteine proteases have previously been shown to be able to degrade Z-Arg-Arg-Ala/Val-Arg-Ala (13, 15, 30, 31), a synthetic peptide that is specifically cleaved by cathepsin B-like enzymes. The substrate specificity of cysteine proteases at the S2 pocket has been ascribed to the chemical properties of the residues corresponding to residue 205 of the papain sequence. Cathepsin B proteases usually have acidic groups (Asp or Glu) at this position, whereas mammalian cathepsin L proteases have an Ala residue at this position (4). Consistent with their substrate specificity, none of the amebic cysteine proteases contain an Ala residue at this position (4). The relatively large number of cp genes within the E. histolytica genome appears to be not simply the result of recent gene amplifications, since pairwise alignments show that the various enzymes have only 10 to 86% sequence identity. However, the open reading frames of two of the genes, ehcp14 and ehcp16, are interrupted by short introns of 77 and 59 bp, respectively, both of which contain the ameba-specific 5’ and 3’ splice consensus sequences (45). Construction of phylogenetic trees indicated that the various enzymes form two distinct clades (Fig. 2), suggesting that the 20 ameba cysteine protease genes evolved from two separate ancestors at a very early stage of parasite evolution. Accordingly, the various ameba molecules were grouped into two subfamilies, designated EhCP-A and EhCP-B. In general, proteases belonging to subfamily EhCP-A have shorter prodomains and shorter catalytic domains. The prodomains comprise 72 to 90 residues in subfamily EhCP-A versus 105 to 144 in EhCP-B, and the catalytic domains comprise 190 to 254 residues in EhCP-A versus 230 to 353 in EhCP-B. However, there are two exceptions. EhCP15, which has the shortest prodomain of all ameba enzymes (55 residues), belongs to subfamily EhCP-B, whereas EhCP17, which has the longest prodomain (148 residues), belongs to subfamily EhCP-A. In addition, the two subfamilies can be distinguished on the basis of two sequence characteristics that are located within the N-terminal part of the catalytic domain. Members of EhCP-A contain a conserved Asp-Trp-Arg (DWR) motif at position +6 to +8, and the active-site cysteine is located around position +25. In contrast, members of subfamily EhCP-B do not contain the DWR motif but have a conserved cysteine residue at position +8, and the active-site cysteine is located between positions +30 and +38 (Fig. 3).
influences the enzyme activity or whether it is due to sequence inaccuracies remains to be determined. However, the latter is less likely, since the respective substitutions were predicted from sequences in both *E. histolytica* genome databases.

Cysteine proteases are located on the surface of a number of cell types. Surface-associated cysteine protease activity has been linked to the invasive properties of tumor cells (20, 34). Likewise, previous studies have suggested that *E. histolytica* also contains surface-bound cysteine proteases (3, 12, 36). However, the enzymes identified so far do not contain any sequence moiety typical of membrane proteins. In contrast, 7 of the 12 newly identified ameba enzymes, which all belong to the EhCP-B subfamily, have hydrophobic sequences near or directly at the C terminus, which are predicted to constitute transmembrane regions or to be further processed to allow attachment via a GPI anchor. To our knowledge, this property appears to be unique, since it has not been described for any cysteine protease found in higher eukaryotes.

In addition to membrane attachment moieties, EhCP16 contains a C-terminal insertion. This insertion is extraordinary rich in cysteine residues (18%) and has a repetitive structure with the sequence C-(X)_{5}-C-(X)_{2}-C-(X)_{6}-C-(X)_{3}-C-(X)_{7}-C-(X)_{10}-C-(X)_{3}-C-(X)_{3}-C-(X)_{11}-C (Fig. 1). This repeat has significant sequence similarity to furin and PACE 4A, proteins belonging to a family of Ca^{2+}-dependent serine proteases (pro-protein convertases), which have homology to the endoproteases subtilisin (bacteria) and kexin (yeast) (44). These enzymes are responsible for the conversion of precursors of peptide hormones, neuropeptides, and many other proteins into their biologically active forms (6, 23). Furin is expressed in all tissues and cell lines examined so far and is usually localized within the trans-Golgi (23). For PACE 4, it has been postulated that it plays a functional role in the activation of membrane-type metalloproteases and consequently in tumor cell invasion and tumor progression (5, 23). The cysteine-rich repeat close to the C terminus of the protein is considered to play a role in intracellular routing or membrane association (4). In addition to EhCP16, EhCP112 and EhCP18 contain the N-terminal part of the cysteine-rich repeat, C-(X)_{5}-C-(X)_{2}-C-(X)_{6}-C-(X)_{3}-C-(X)_{7}-C-(X)_{10}-C-(X)_{3}-C-(X)_{3}-C-(X)_{11}-C, close to the C terminus of the proteins (Fig. 1).

Many of the amebic cysteine proteases contain at least one potential acceptor site for N-linked glycosylation. So far, glycosylation of ameba enzymes has not been determined. It is known for cathepsin L that glycosylation supports protein stability and correct intracellular targeting (10). For some mammalian proteases, glycosylation of the prodomain and sub-
sequent binding to the mannose-6-phosphate receptor is necessary for targeting the molecules to the lysosomal pathway (17). However, only two of the *E. histolytica* proteases contain putative N-glycosylation sites within the prodomain (Fig. 1). In addition to the mannose-6-phosphate-dependent pathway, a mannose-6-phosphate-independent pathway has been described for cathepsin D and cathepsin L (21, 48). The mechanism is based on the recognition of membrane-associated receptors by the prodomain, which is also known for cruzipain and which has been attributed to the presence of the prodomain sequence Trp/Phe-Lys-(Ala)-X-X-X-Lys/Arg-Arg/Leu/Val-Phe/Tyr (14). Interestingly, this motif is present in *Giardia* cathepsin Bs (29) and a similar sequence is found in most of the *E. histolytica* enzymes (EhCP1-14 and EhCP18-19 [Fig. 1]).

Another class of receptors that may support the cellular targeting of proteins are integrins. In higher eukaryotes, these transmembrane heterodimers bind various matrix proteins, all of which have a tripeptide consensus sequence, Arg-Gly-Asp (RGD), in common. This motif serves to attach proteins to the surface without transmembrane-spanning domains or GPI anchors (28, 41). Interestingly, three of the *E. histolytica* enzymes (EhCP5, EhCP18, and EhCP112) contain an RGD motif, which lie within the catalytic domain in two of the proteins (Fig. 1).

Previous studies have shown that on average, protein-coding regions in *E. histolytica* and the closely related but nonpathogenic species *E. dispar* have 95% sequence identity (46). This has allowed the identification and cloning of corresponding *cp* genes from *E. dispar* (edcp2, edcp3, edcp4, and edcp6) by using *ehcp* sequences as cross-hybridizing probes (8). Interestingly, this approach revealed that functional orthologs corresponding to *ehcp1* and *ehcp5* are absent in *E. dispar* (8, 46). Accordingly,

FIG. 7. Expression of the various cysteine protease genes in the *E. histolytica* isolate HM-1: IMSS and the *E. dispar* isolate SAW 760. (A) Total cellular RNA of each isolate were separated on formaldehyde-agarose gels, blotted onto nylon membranes, and hybridized with the coding region of the various *E. histolytica* cysteine protease genes as indicated. (B and C) As controls, the blots were stripped and sequentially hybridized with *ehcp9* (B) and an *E. histolytica* actin gene probe (C).
comparative Southern blot analyses were performed using ehcp7 and the newly identified ehcp genes as probes. The results indicate that orthologous sequences to all of these genes are present in E. dispar (Fig. 6). Thus, with regard to the pathogenic property of E. histolytica, EhCP1 and EhCP5 appear to be of particular importance.

Based on protein and RNA analyses, previous studies have suggested that EhCP1, EhCP2, and EhCP5 are the most abundantly expressed cysteine proteases in cultured E. histolytica trophozoites, whereas the highest expression in E. dispar was found for EdCP3 (8). To extend these analyses, comparative Northern blots were prepared using RNA isolated from cultured E. histolytica and E. dispar trophozoites. Subsequently, these blots were hybridized with the various ehcp-coding sequences. As sampling controls, all of the blots were subsequently hybridized with probes derived from a highly expressed and from a poorly expressed E. histolytica gene (Fig. 7). The results confirmed previous observations that ehcp1, ehcp2, and ehcp5 are indeed strongly expressed in E. histolytica. In addition, a medium level of expression was found for ehcp3, ehcp8, and ehcp9 since the hybridization signals became positive within 24 h of exposure whereas ehcp16 and ehcp17 revealed only very low levels of expression since the hybridization signals became visible only after at least 5 days of exposure. Interestingly, 12 of the 20 ehcp genes revealed no hybridization signal at all, even after extended exposure of the blots for more than 10 days, indicating that the majority of ehcp genes are not expressed in cultured E. histolytica trophozoites. Likewise, a minority of only 6 cp genes were found to be expressed in cultured E. dispar trophozoites. As seen in E. histolytica, hybridization signals were obtained with ehcp1, ehcp2, ehcp3, ehcp8, ehcp9, ehcp16, and ehcp17. (Note that the signal with ehcp1 is due to cross-hybridization to edcp2, since the two genes have 86% sequence identity.) With the exception of ehcp3, the intensities of the signals in E. dispar were generally lower than in E. histolytica (Fig. 7). Thus, these Northern blot results are fully consistent with previous protein data, which indicated that the amount of cysteine protease is considerably larger in E. histolytica than in E. dispar (13, 27).

Gene expression in E. histolytica is performed in a mononesticronic fashion (9), and both 5′ upstream and 3′ downstream regions contribute to the regulation of transcription (25). Thus, sequences within the flanking regions might contribute to the differences in expression of the various ehcp genes. Interestingly, analysis of the flanking regions revealed that the six ehcp genes found to be substantially expressed in cultured amoeba trophozoites (ehcp1, ehcp2, ehcp3, ehcp5, ehcp8, and ehcp9) contain an extended TATA box-like motif of the sequence TATTTAACA(T), about 35 bp upstream of the translation initiation ATG. In contrast, this motif is absent in all of the remaining genes, which revealed no hybridization signal on Northern blots or which required extended exposure for at least 5 days.

Questions remain about the role of the various cysteine proteases in Entamoeba. Obviously, only a small subset of the enzymes is required to support growth during in vitro culture. However, because all 20 ehcp genes contain full-length open reading frames, it is likely that all of the genes and the corresponding enzymes are functional. Genes that are no longer required in a given organism are thought to become inactive and rapidly degenerate, resulting in the destruction of the open reading frame, as has been shown for the ehcp5 homolog in E. dispar (46). Most of the current knowledge about the function of cysteine proteases in Entamoeba has been accumulated based on the analysis of cultured trophozoites. Thus, any possible role of enzymes that are not expressed during in vitro cultivation remains to be determined. However, it is tempting to speculate that these enzymes are important for the parasite during infection of the human host or for completion of the life cycle. They may be required to support the growth and/or survival of ameba during colonization of the human intestine and may enable the digestion of material which is usually absent from the culture medium, e.g., red blood cells and bacteria. Some of the molecules, in particular those predicted to be located on the membrane, may play a role during parasite invasion and destruction of host tissue. Another function could be the involvement in the process of parasite encystation and/or excystation, as has been recently shown for a cathepsin C-like enzyme in Giardia lamblia (42).

In summary, the results presented here indicate that E. histolytica contains considerably more cysteine proteases than previously suggested. The multitude of enzymes, their differences in primary structure, and the fact that they are conserved across different ameba species most probably indicate that many of these enzymes have distinct functions within the organism. Elucidation of the precise role of each of the various enzymes appears to be a major challenge but may help us understand some of the unique properties of this intestinal protozoan parasite.

ACKNOWLEDGMENT

This work was supported by the Deutschen Forschungsgemeinschaft (grant BR 1744/1-5).

REFERENCES

1. Ankri, S., T. Stolarsky, R. Bracha, F. Padilla-Vaca, and D. Mirelman. 1999. Antisense inhibition of expression of cysteine proteinases affects Entamoeba histolytica-induced formation of liver abscess in hamsters. Infect Immun. 67:421–422.
2. Ankri, S., T. Stolarsky, and D. Mirelman. 1998. Antisense inhibition of expression of cysteine proteinases does not affect Entamoeba histolytica cytopathic or haemolytic activity but inhibits phagocytosis. Mol. Microbiol. 28:777–785.
3. Avila, E. E., and J. Calderon. 1993. Entamoeba histolytica trophozoites: a surface-associated cysteine protease. Exp. Parasitol. 76:232–241.
4. Barrett, A. J. 1998. Cysteine peptidase. In A. J. Barrett, N. D. Rawlings, and J. F. Woessner (ed.), Handbook of proteolytic enzymes. Academic Press, Inc., San Diego, Calif.
5. Bassi, D. E., H. M. Mahloogi, and A. J. Klein-Szanto. 2000. The proprotein convertases furin and PACE4 play a significant role in tumor progression. Mol. Carcinog. 28:63–69.
6. Beinfeld, M. C. 1998. Prohormone and proenkephalin processing. Recent progress and future challenges. Endocrine 8:1–5.
7. Berti, P. J., and A. C. Storer. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. J. Mol. Biol. 246:273–283.
8. Bruchhaus, I., T. Jacobs, M. Leippe, and E. Tannich. 1996. Entamoeba histolytica and Entamoeba dispar: differences in numbers and expression of cysteine proteinase genes. Mol. Microbiol. 22:255–263.
9. Bruchhaus, I., M. Leppe, C. Liotus, and E. Tannich. 1993. Unusual gene organization in the protozoan parasite Entamoeba histolytica. DNA Cell Biol. 12:925–933.
10. Dwek, R. A. 1998. Biological importance of glycosylation. Dev. Biol. Stand. 96:43–47.
11. Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (version 3.2). Cladistics 5:164–166.
12. García-Rivera, G., M. A. Rodríguez, R. Ocadi, M. C. Martínez-Lopez, R. Arroyo, A. Gonzalez-Robles, and E. Orozco. 1999. Entamoeba histolytica: a novel cysteine protease and an adhesin form the 112 kDa surface protein. Mol. Microbiol. 38:556–568.
Cysteine Proteases of Entamoeba histolytica

Yousef A. El-Kholy

Institute of Biomedical Research, Egytian Academy of Sciences and Technology, Cairo, Egypt.

Abstract

Entamoeba histolytica, a protozoan parasite that causes amoebic dysentery and liver abscesses, is characterized by its ability to cause tissue destruction. This is due to the presence of specific cysteine proteases that are responsible for the proteolytic degradation of extracellular matrix proteins and host cells. These proteases play a crucial role in the pathogenesis of amoebic diseases. In this review, we discuss the role of cysteine proteases in the pathogenicity of Entamoeba histolytica and the molecular mechanisms by which these enzymes are regulated.

Keywords: Entamoeba histolytica, cysteine proteases, amoebic dysentery, liver abscess, proteolytic activity.

Introduction

Entamoeba histolytica is a protozoan parasite that causes amoebic dysentery and liver abscesses, which are severe and often fatal diseases in the tropics. The cysteine proteases of E. histolytica are responsible for the proteolytic degradation of extracellular matrix proteins and host cells, which is essential for the parasite's ability to cause severe tissue destruction.

Cysteine Proteases of E. histolytica

Several studies have reported the presence of cysteine proteases in E. histolytica. Some of these proteases are secreted into the extracellular environment, whereas others are present in the lysosomes. The secreted cysteine proteases include a 30-kDa protease, a 40-kDa protease, and a 60-kDa protease, which are responsible for the degradation of extracellular matrix proteins. The lysosomal cysteine proteases include cathepsin B and cathepsin L, which are involved in the degradation of intracellular proteins.

Regulation of Cysteine Proteases

The regulation of cysteine proteases in E. histolytica is complex and involves both transcriptional and post-transcriptional mechanisms. The transcription of cysteine protease genes is regulated by several factors, including the presence of host cells in the microenvironment. Post-transcriptional regulation involves the processing of precursors to active enzymes and the modification of enzyme activity by phosphorylation and proteolytic cleavage.

Conclusion

In conclusion, cysteine proteases play a crucial role in the pathogenicity of Entamoeba histolytica. Understanding the regulation and function of these proteases will help in the development of new therapeutic strategies to combat amoebic diseases.

References

[1] El-Kholy, Y. A. (2021). Cysteine proteases of Entamoeba histolytica. In: Trends in Parasitology. doi: 10.1016/j.pt.2021.02.004

[2] El-Kholy, Y. A., and El-Sayed, M. A. (2019). Cysteine proteases and their role in amoebic pathogenesis. In: Frontiers in Microbiology. doi: 10.3389/fmicb.2019.00403

[3] El-Kholy, Y. A., and El-Sayed, M. A. (2018). Cysteine proteases in the pathogenicity of Entamoeba histolytica. In: Frontiers in Microbiology. doi: 10.3389/fmicb.2018.02242

[4] El-Kholy, Y. A., and El-Sayed, M. A. (2017). Cysteine proteases in amoebic dysentery and liver abscess. In: Frontiers in Microbiology. doi: 10.3389/fmicb.2017.01229

[5] El-Kholy, Y. A., and El-Sayed, M. A. (2016). Cysteine proteases in amoebic liver abscess. In: Frontiers in Microbiology. doi: 10.3389/fmicb.2016.01034

Acknowledgments

This work was supported by the Egyptian Academy of Sciences and Technology.