

Cysteine Proteases and Cell Differentiation: Excystment of the Ciliated Protist *Sterkiella histriomuscorum*

Eduardo Villalobo,† Clara Moch, Ghislaine Fryd-Versavel, Anne Fleury-Aubusson, and Loïc Morin*

Laboratoire de Biologie Cellulaire 4, Université de Paris-Sud, Orsay, France

Received 4 April 2003/Accepted 5 August 2003

The process of excystment of *Sterkiella histriomuscorum* (Ciliophora, Oxytrichidae) leads in a few hours, through a massive influx of water and the resorption of the cyst wall, from an undifferentiated resting cyst to a highly differentiated and dividing vegetative cell. While studying the nature of the genes involved in this process, we isolated three different cysteine proteases genes, namely, a cathepsin B gene, a cathepsin L-like gene, and a calpain-like gene. Excystation was selectively inhibited at a precise differentiating stage by cysteine proteases inhibitors, suggesting that these proteins are specifically required during the excystment process. Reverse transcription-PCR experiments showed that both genes display differential expression between the cyst and the vegetative cells. A phylogenetic analysis showed for the first time that the cathepsin B tree is paraphyletic and that the diverging *S. histriomuscorum* cathepsin B is closely related to its *Giardia* homologues, which take part in the cyst wall breakdown process. The deduced cathepsin L-like protein sequence displays the structural signatures and phylogenetic relationships of cathepsin H, a protein that is known only in plants and animals and that is involved in the degradation of extracellular matrix components in cancer diseases. The deduced calpain-like protein sequence does not display the calcium-binding domain of conventional calpains; it belongs to a diverging phylogenetic cluster that includes *Aspergillus* palB, a protein which is involved in a signal transduction pathway that is sensitive to ambient pH.

The exit from dormancy, of which excystment is a particular case, is a widespread process in many eukaryotic and prokaryotic organisms. It is generally accompanied by cellular differentiation, during which numerous intra- and extracellular structures are highly modified, as observed during the germination of fungal conidia and plant seeds (reviewed in references 5, 31, and 39), among others. Cysteine proteases, which are known to be involved in intracellular protein turnover and extracellular matrix remodeling (37), constitute good candidates to take part in this general process. In this paper, we describe the involvement of such proteins in the process of excystment of the ciliate *Sterkiella histriomuscorum*.

S. histriomuscorum (previously called *Oxytricha fallax* and *O. trifallax*) (3) is a free-living ciliated protist from the hypotrich subgroup that is characterized by a highly differentiated pattern of the vegetative cell. This pattern includes a complex oral apparatus and a clustered body ciliature which are duplicated in an orderly process during cell division (22). The *S. histriomuscorum* life cycle is interesting in that the vegetative cell undergoes a complete dedifferentiation process when it is deprived of food, leading in about 3 days to a resting stage which is called a cyst. The mature cyst has lost much water, the whole ciliature and basal bodies have disappeared, numerous autophagic vacuoles (lysosomes) that were particularly active during encystment are present, and a multilayered cyst wall surrounds the spherical cell, which appears to be metabolically

inactive (23). The nature of the cyst wall is unknown, but it may be hypothesized that it includes glycoproteins, as was shown in the case of *Paraurostyla* sp., a related hypotrich species (29, 44).

In presence of either living food or 0.2% milk powder in water, the excystment process takes place in a few hours, leading through a massive influx of water and the resorption of the cell wall to completely differentiated swimming vegetative cells.

Different taxonomic groups, including both unicellular and multicellular organisms, display encystment-excystment behavior, and two previous studies have shown that cysteine proteases are involved in the excystment process. The first described two distinct proteases that modulate excystation in the case of the trematode *Paragonimus westermani* metacercariae (9). The second showed that cathepsin B was the protease responsible for the cyst wall degradation in the diplomonad flagellate *Giardia* (60).

Cysteine proteases constitute a wide superfamily of proteases, including viral, legumain-like, and papain-like proteases. The papain-like group is further divided into families, including calpains and papains, and the papain family includes the bleomycin hydrolase, cathepsin B, and cathepsin L subfamilies (47). Calpains are calcium-dependent cytosolic enzymes, while cathepsins are located mainly in lysosomes. Both were previously known to be involved in intracellular protein turnover in mammals, but it appears that they display a much larger array of functions in protists (26, 37, 40).

During a general study of the genes that are differentially expressed in the cyst and vegetative cells, i.e., during the excystment of *S. histriomuscorum*, we isolated three different cDNAs whose sequences appeared, based on a BLAST analysis, to be highly similar to those of the calpain, cathepsin B,

* Corresponding author. Mailing address: Laboratoire de Biologie Cellulaire 4, Bâtiment 444, Centre d'Orsay, Université de Paris-Sud, 91405-Orsay Cedex, France. Phone 33 69 15 64 84. Fax: 33 69 15 68 03. E-mail: loic.morin@bc4.u-psud.fr.

† Present address: Departamento de Microbiología, Facultad de Biología, Universidad de Sevilla, Sevilla, Spain.

and cathepsin L family genes. Excystation (but not encystation or cell division) was blocked at a particular differentiating stage when cysteine protease inhibitors were added in the nutrient medium, further confirming the involvement of cysteine proteases in the excystment process. The expression pattern was life cycle stage dependent for both cysteine protease genes. The calpain-like gene is not expressed strongly in cysts and early excysting cells, while cathepsins are expressed more in vegetative cells than in resting cysts and early excysting cells.

The whole sequences of the corresponding minichromosomes were determined, and a comparative analysis of the protein sequences shows an important deletion in the occluding loop of cathepsin B, a cathepsin H signature in the cathepsin L-like sequence, and the absence of the calpain III linker domain and calcium-binding motif in the calpain-like sequence. A phylogenetic study confirmed the cathepsin H identification and the diverging status of the cathepsin B and calpain genes.

MATERIALS AND METHODS

Cell culture, encystment, and excystment. *S. histriomuscorum* strain BA was grown at room temperature in commercial mineral water (Volvic [Volvic, France]) with *Tetrahymena pyriformis* or *Chlorogonium* sp. as a food source. Encystment of cells was induced by starvation. Excystment was induced either by adding food (a grown culture of *Aerobacter aerogenes* eubacteria) to the cyst medium or by transferring cysts to a 0.2% milk powder–0.05% yeast extract medium in water. Manipulation of the cysts was performed as described by Adl and Berger (3).

For inhibition experiments, dividing vegetative cells were incubated with 10 μ M inhibitor in feeding medium. To test for encystment, starving vegetative cells were incubated with 10 μ M inhibitor in mineral water. To test for excystment, cysts (about 2,000) were incubated into either food medium or the excystment medium described above with 50 nM to 100 μ M inhibitor. Parallel assays of untreated controls and inhibitor solvent controls were performed. The tested inhibitors were obtained from Sigma and included E64 [*N*-(*trans*-epoxysuccinyl)-L-leucine 4-guanidinobutylamide], CA-074 (*L*-*trans*-epoxysuccinyl-Ile-Pro-OH-propylamide), CA-074Me (*L*-*trans*-epoxysuccinyl-Ile-Pro-Ome-propylamide), MG132 (Z-Leu-Leu-Leu-al), calpain inhibitor II (*N*-acetyl-Leu-Leu-Met-al), and lactacystin.

Assessment of cell division, encystment, and excystment was done with a Leitz dissecting microscope. Excystment rate analysis was carried out after an incubation time of 24 h by counting swimming vegetative cells. For cell structure examination, samples were impregnated as described for the silver proteinate method (54); cysts were then either mounted in glycerol for direct observation or dehydrated and mounted in Canada balsam for conservation.

Differential-display RT-PCR and amplified library. Nucleic acid extraction and purification were conducted as described by Sambrook et al. (48). Differential-display reverse transcription-PCR (RT-PCR) was carried out as described by Villalobo et al. (57). A time-specific amplified cDNA library was prepared as follows. RNAs from 2-h excysting cells were extracted, and double-stranded cDNAs were obtained by following the procedures of the manufacturer (cDNA synthesis kit; Promega). A double-stranded adaptor was ligated to the cDNAs ends, and the excess adaptors were removed on a spin column. cDNAs were then PCR amplified by using the single-strand adaptor (5'-ATGCTTAGGAATTCC GATTTAGCCTCATA-3') as a primer, with 5 U of *Taq* DNA polymerase (Q-BioTaq; Appligene-Quantum) and 35 cycles of denaturation at 94°C for 1 min, annealing at 66°C for 1 min, and polymerization at 72°C for 3 min. The PCR products were run on a 1% (wt/vol) agarose gel in TBE (90 mM Tris, 90 mM boric acid, 2 mM EDTA [pH 8.3]), and fragments smaller than 400 bp were cut away; the gel was then run in the reverse orientation to concentrate the remaining fraction, which was cut out and purified by using the QIAquick gel extraction kit (Qiagen), according to the manufacturer's instructions.

Cloning, screening, and sequencing procedures. cDNAs obtained from either the differential-display RT-PCR or the amplified library, both with an A overhang resulting from PCR, were then cloned in pGEM-T Easy plasmid (Promega) at 4°C overnight and transformed into *Escherichia coli* XL1-Blue MRF' (Stratagene) by electroporation.

Library screening was carried out by hybridization of DNA from plated colo-

nies with a [α -³²P]dATP (Amersham)-labeled probe of cDNAs, constructed with RNAs extracted from 2-h excysting cells. Nucleic acid hybridization was conducted by standard methods. Clones of interest were isolated, and plasmids were extracted and purified by the conventional alkaline lysis method (48). Sequencing reactions with plasmid DNA were done with an automated sequencer (Vistra; Amersham) and the DNA cycle sequencing kit and Texas Red M13 universal primers (Amersham).

Inverse PCR. The ciliate nuclear organization is characterized by the presence of diploid micronuclei, containing the usual diploid chromosomes, and macronuclei which contain expressed genes embedded in minichromosomes. In hypotrich ciliates such as *S. histriomuscorum*, each minichromosome contains an open reading frame flanked by 5' leader and 3' tailer sequences and specific telomeres; this structure is then duplicated several times in order to produce about 1,900 copies of each sequence (43). This particular organization makes it possible to circularize these minichromosomes and to use the inverse-PCR technique to amplify and clone the whole minichromosome.

Total genomic DNA from *S. histriomuscorum* was covalently circularized by ligation, as previously described (57), and then 100 ng of circularized DNA was subjected to PCR in a 50- μ l reaction mixture with the Expand long-template PCR system (Boehringer) according to the manufacturer's procedures and with 35 cycles of denaturation, annealing, and polymerization. The PCR products were run on a 1% (wt/vol) agarose gel, extracted and purified as described above, and then cloned in the pGEM-T Easy plasmid.

The sequences of the outward-facing specific primers used to amplify the macronuclear sequences corresponding to each fragment of interest were as follows: for the cathepsin B gene, 5'-TTGGATTTGGGCAACACCCCTT-3' (upstream) and 5'-TAATCAGGTGTCTACACCCCA-3' (downstream); for the cathepsin L-like gene, 5'-CGTCTTTGATGCTCAACAAGCTGC-3' (upstream) and 5'-TTGAAGAATCCTTGGTCACCCC-3' (downstream); and for the calpain gene, 5'-AAGAGTGACAAGTGGAAATGAC-3' (upstream) and 5'-ATCA GACCAATCTCCCATCC-3' (downstream).

RT-PCR. Total RNAs (2 μ g) from cysts, 1-h excysting cells, and vegetative cells were reverse transcribed by using 3.5 μ M random hexamers (pdN6; Amersham) as primers and 200 U of Moloney murine leukemia virus reverse transcriptase (Eurobio) in a 20- μ l reaction mixture containing 40 U of RNasin (Promega). RNA was incubated at 65°C for 5 min prior to the RT step. After 50 min at 37°C, the enzyme was inactivated at 95°C for 5 min. The PCR step was subsequently carried out with 0.1 and 0.3 μ l of the RT product, a pair of specific primers, and 1.5 U of *Taq* polymerase in a 25- μ l reaction mixture. The sequences of the forward and reverse primers were, respectively, as follows: for the cathepsin B gene, 5'-TTCTCCGACAGATTCGCC-3' and 5'-GTCGCATGAGAAGA CTTG-3'; for the cathepsin L-like gene, 5'-TCAACTGTTGGTTGTGTT-3' and 5'-TGGGTAAGAGTTGCAGTT-3'; and for the calpain gene, 5'-AAACACTC CAGGCTCAAAATG-3' and 5'-AAGAATTCTCAACTCCGGCAT-3'.

Sequence analysis and reconstruction of phylogenetic trees. The gapped BlastX program from the National Center for Biotechnology Information was used for searching protein identities. SMART software (34) was used to search domains in the predicted protein sequences. All sequences except *S. histriomuscorum* sequences were retrieved from GenBank to constitute a data set of 59 sequences for cathepsin B, 349 sequences for cathepsin-L like proteins, and 79 sequences for calpain. Management, formatting of the sequences for the tree-building programs, and construction of distance trees were carried out with the MUST package (41). Sequence alignments were manually performed with the ED program of the package. Cathepsin B and cathepsin L-like sequence data sets were disjoined in order to optimize the node resolution in each subtree for two reasons: first, they constitute two rather diverging branches in the general tree of cathepsin proteins (4, 60), and second, the number of unambiguously aligned characters in each subset (248 and 229, respectively, for cathepsin B and cathepsin L-like sequences) is much higher than it is in the two combined data sets (181). For legibility reasons, the final analysis retained only representatives of the main clusters in each tree, i.e., 37 sequences for cathepsin B, 72 for cathepsin L, and 41 for calpain. The distance matrix was calculated by considering all amino acid differences without weighting the transition probabilities. This matrix served as a basis for phylogenetic reconstructions by the neighbor-joining procedure (46). A 1,000-bootstrap resampling analysis was carried out with the Njboot program. Parsimony and maximum likelihood analyses were conducted for 100 bootstrap resamplings with the PROTPARS and PROML programs from the PHYLIP package (14); for the latter analysis, the Jones-Taylor-Thornton model was used with a constant rate of change among sites.

Nucleotide sequence accession numbers. The nucleic acid and deduced amino acid sequences of the cathepsin B, cathepsin L-like, and calpain-like genes have been deposited in GenBank under accession numbers AY204512, AY204513, and AY204514, respectively.

RESULTS

Isolation of the cysteine protease molecules. From the differential-display RT-PCR gels, we selected several bands that appeared at first to be much more expressed in vegetative and excysting cells than in resting cysts. Subsequent cloning and insert sequencing led to the identification, as determined from BLAST searches, of putative genes of cathepsin L-like and calpain-like proteins for two of the isolated clones. From the 2-h time-specific excysting library, a crude analysis by random sequencing of several clones allowed us to recover a particular clone which was found several times and whose insert appeared to encode a putative cathepsin B protein.

The sequences of these short cDNA inserts (about 400 bp) were used as anchoring sequences to design primers that were utilized with the inverse-PCR technique in order to clone the minichromosomes bearing the cathepsin B, the cathepsin L-like, and the calpain-like genes.

Effects of cysteine protease inhibitors. In order to confirm the involvement of cysteine protease molecules in the excystment process, the effects of several cysteine protease inhibitors were tested with growing, encysting, and excysting cells. We used *L-trans*-epoxysuccinyl peptides such as E64, CA-074, and its membrane-permeable derivative CA-074Me; peptide aldehydes such as MG132 (Cbz-Leu-Leu-leucinal) and calpain inhibitor II (*N*-acetyl-Leu-Leu-Met-aldehyde); and a natural inhibitor with a different structure, lactacystin. No inhibitor, whatever concentration was used (up to 1 mM), was able either to prevent or to delay cell division in a culture of growing cells or to block the encystment process in a nutrient-depleted medium. In contrast, E64, calpain inhibitor II, and MG132 were able to block the excystment at a specific stage after the onset of the cell differentiation process. Efficient inhibitor concentrations were determined by counting swimming vegetative cells, leading to a value of 10 μ M for complete inhibition (Fig. 1). Cell observation during the experiments suggests that excystation started normally, as we witnessed a change of cyst refringency and the onset of contractile vacuole beating. Moreover, blocked cysts that were fixed and protargol impregnated display the ciliary "anlagen" as a whole on the cell cortex; that is, the basal body assembly is almost finished (Fig. 2). Washing the blocked cells with fresh medium lacking the inhibitor reactivated the excystment process and led to normally dividing cells. These results clearly show that the active inhibitors are responsible for a specific blockage of the excystment process and that the inhibition occurs at a late stage of excystation.

E64 is a nonspecific inhibitor of cysteine proteases, calpain inhibitor II and MG132 are relatively nonspecific inhibitors of cysteine proteases and proteasome (33), and CA-074 and CA-074Me are specific inhibitors of cathepsin B (53); lactacystin is a specific inhibitor of the proteasome (33). The excystment process involves cysteine protease activities (apparently with the exception of cathepsin B) and is not dependent of the proteasome activity.

Expression pattern. The RT-PCR patterns observed for the three cysteine proteases are displayed in Fig. 3. Cathepsin B mRNA is strongly expressed in vegetative growing cells which are actively dividing, while a relatively low expression level is observed in cysts and in 1-h excysting cells. Cathepsin L-like mRNA displays a pattern which appears as similar to that of

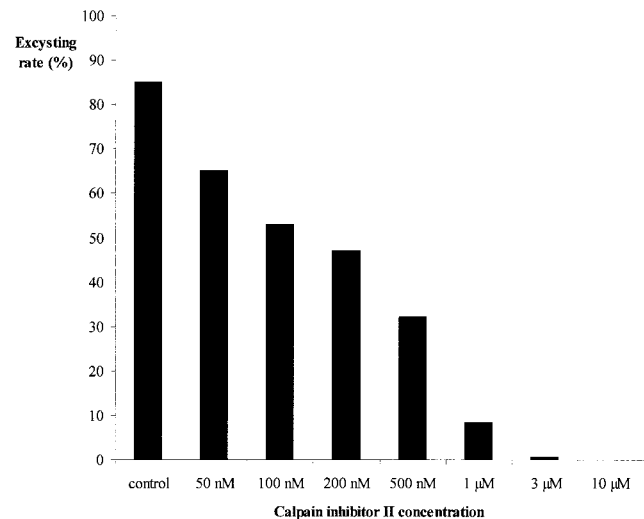


FIG. 1. Dose effect of calpain inhibitor II on the excystment process. About 2,000 cysts were incubated with food and inhibitor (see text for details). The effect was evaluated by counting cysts and vegetative cells in three samples taken from each experiment (i.e., for each inhibitor concentration), and the excysting rate is expressed as the percentage of excysted vegetative cells in each experiment. Total inhibition is observed with a 10 μ M inhibitor concentration; this result was confirmed in all experiments, and the 10 μ M value was retained to test the different inhibitors. Identical results were obtained with the other active inhibitors (data not shown).

cathepsin B, except that the expression level in vegetative cells is hardly greater than in cysts and 1-h excysting cells. Calpain mRNA is apparently absent in cysts and 1-h excysting cells but is strongly expressed in growing cells. These results suggest that the three cysteine proteases are active during the normal vegetative cell growth and that their expression, at least for cathepsin B and calpain, is upregulated during the excystment process; moreover, this expression is increased at least after 1 h of excystment induction, suggesting that upregulation occurs at a late stage after the onset of excystation, in agreement with the above observations concerning the effects of cysteine protease inhibitors.

Structure and phylogeny. The minichromosome bearing the cathepsin L-like gene is around 1.3 kbp long, as confirmed by Southern blot experiments (data not shown). Both ends of the chromosome display extremely AT-rich regions (82% at the 101-bp-long 5' end and 79% at the 109-bp-long 3' end) surrounding a central region of 1,101 bp which is 60% AT rich and probably corresponds to the transcribed gene. The poly(A) tract of the mRNA is inserted after the position 1298, 85 bp downstream from the coding sequence and very close to the 3' telomere sequence. The translated amino acid sequence, using the ciliate nuclear genetic code (TAA and TAG codons code for Gln), discloses a single open reading frame with six TAA codons and 1 TAG codon and a putative methionine start codon located at the end of the 5' AT-rich region of the sequence. The sequence is 366 amino acids long, and a preliminary BLAST analysis shows a strong similarity with cathepsin L-like proteins.

The deduced protein displays the usual structural characteristics of cathepsins, with a strongly conserved active site involv-

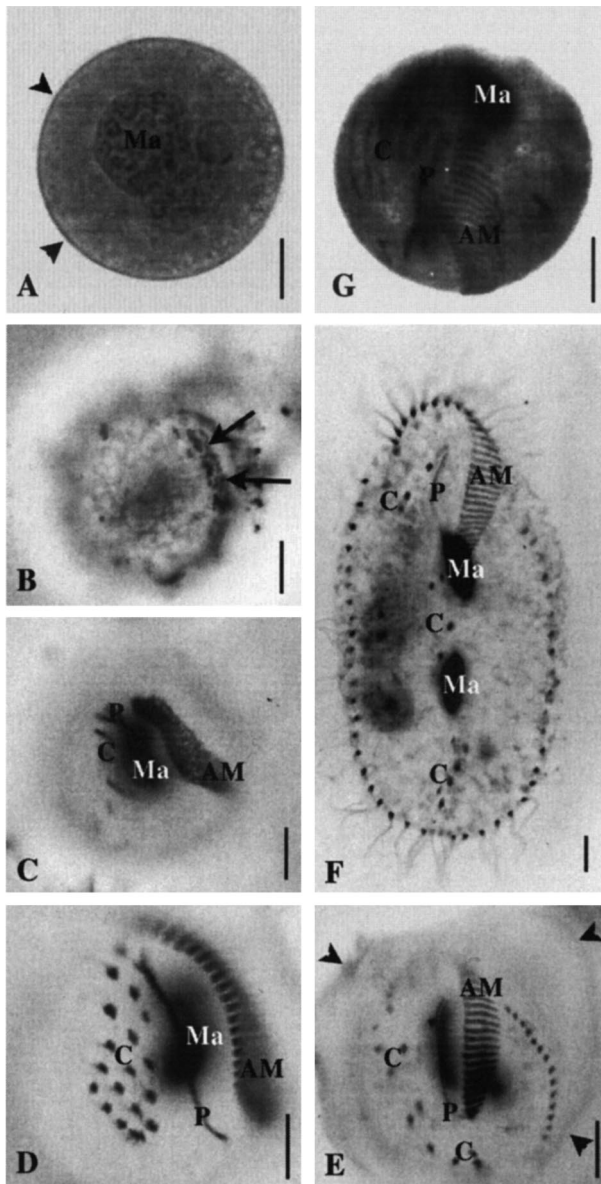


FIG. 2. Excystment process (A to F) and effect of the E64 inhibitor (G) as observed on cells after silver proteinate impregnation and mounting in Citifluor (A and G) or dehydration and mounting in Canada balsam (B to F). Bars, 10 µm. (A to F) Excystment process. The resting cyst (A) is devoid of ciliature and enclosed in a wall (arrowheads). The macronucleus (Ma) is visible in the center of the cell. After induction of excystment, the first cortical event detected by silver proteinate impregnation is the appearance of a field of basal bodies (arrows in panel B). Thousands of basal bodies are quickly assembled and organized into anlagen (C and D), which will differentiate into specific ciliary structures of the young (E) and interphase (F) cell. Several of these anlagen are recognizable very early, and their maturation is easily monitored through the whole process. Adoral membranelles (AM) (or polykineties) and paroral membrane (P) (or haplokineties) represent the two sides of the oral ciliature in the interphase cell (F). The AM differentiate from the field of basal bodies already seen in panel B. This field enlarges (C); basal bodies are progressively organized into transversal rows (D) characteristic of the cell pattern (E and F). The paroral ciliature (P) of the vegetative cell (F) differentiates from a streak of basal bodies situated on the right of the AM anlage (C to E). The ventral cirri (C), which appear as dark points in the interphase cell (F), differentiate from streaks of basal bodies (C) which segment into individual cirri (D) and become pat-

ing cysteine (C), histidine (H), and asparagine (N) residues (Fig. 4A); the propeptide region is 134 amino acids long and contains a signature sequence (Fig. 5) that is shared with mammalian cathepsin H and several plant cathepsins, to the exclusion of the other cathepsin L-like proteins. This signature sequence is known to remain attached to the mature enzyme in mammals (37) and is involved in the aminopeptidase activity of the enzyme. A structural analysis carried out with the SMART software (34) allowed the identification of a signal peptide of 19 amino acids located at the amino-terminal end of the molecule, with a putative cleavage site clearly positioned after a sequence of hydrophobic amino acids (MKTISASVLIILFSSL AMA/DV); the position of the cleavage site is in agreement with the rule (Ala-X-Ala located at positions -3 and -1) for bacterial signal peptidase I (58). This signal peptide is characteristic of molecules entering via the endoplasmic reticulum in the cell secretory pathway (38). Three N-glycosylation sites were detected with the NetNGlyc prediction server (R. Gupta et al., unpublished data), suggesting that a putative mannose-6-phosphate receptor pathway could be used for trafficking to lysosomes.

The neighbor-joining phylogenetic tree of cathepsin L-related proteins displays several paraphyletic clusters whose branching nodes are not well resolved (Fig. 6). Bootstrap values are strikingly dependent on the tree reconstruction method, as exemplified in the well-known plant papain cluster, where parsimony and maximum likelihood values are in conflict with the neighbor-joining values. However, the *S. histriomuscorum* sequence is clearly included within the cathepsin H subgroup, thus confirming the presence of the signature sequence in the propeptide domain. This subgroup involves plant and mammalian cathepsins; to date, no other sequence is available in bank's data sets, and we can make no statement about the relatedness of the ciliate sequence to either plant cathepsins or mammalian cathepsin H. The other available ciliate sequences are clearly different, clustering together (*Paramecium* and *Tetrahymena*) or being located elsewhere in the tree (*Stylonychia*).

The minichromosome bearing the cathepsin B gene is around 1.8 kbp long, as confirmed by Southern blot experiments (data not shown), and includes a 883-bp transcribed gene (47% AT rich) surrounded by a 78% AT-rich 5' region of 187 bp and a 75% AT-rich 3' region of 684 bp. In contrast to the cathepsin H-bearing minichromosome, the 3' region is extremely long. The poly(A) tract of the mRNA is inserted 90 bp after the end of the coding sequence (position 1173), a distance which is equivalent to that observed for the cathepsin H sequence, and far from the 3' telomere sequence. The translated amino acid sequence discloses a single open reading frame, with three TAA codons, and a putative methionine start

terned on the cortex as cell enlarges (E and F). The specific pattern of the interphase cell may be clearly recognized as the cell is still enclosed in its wall (arrowheads in panel E). During the process, the macronucleus elongates (D) and separates into two parts (E and F). (G) Effect of the E64 inhibitor. Inhibition of the excystment occurs as soon as ciliary anlagen are assembled, i.e., between stages C and D: anlagen of the oral apparatus (AM and P) as well as those of the cirri (C) are visible. The macronucleus (Ma) has not separated into two parts.

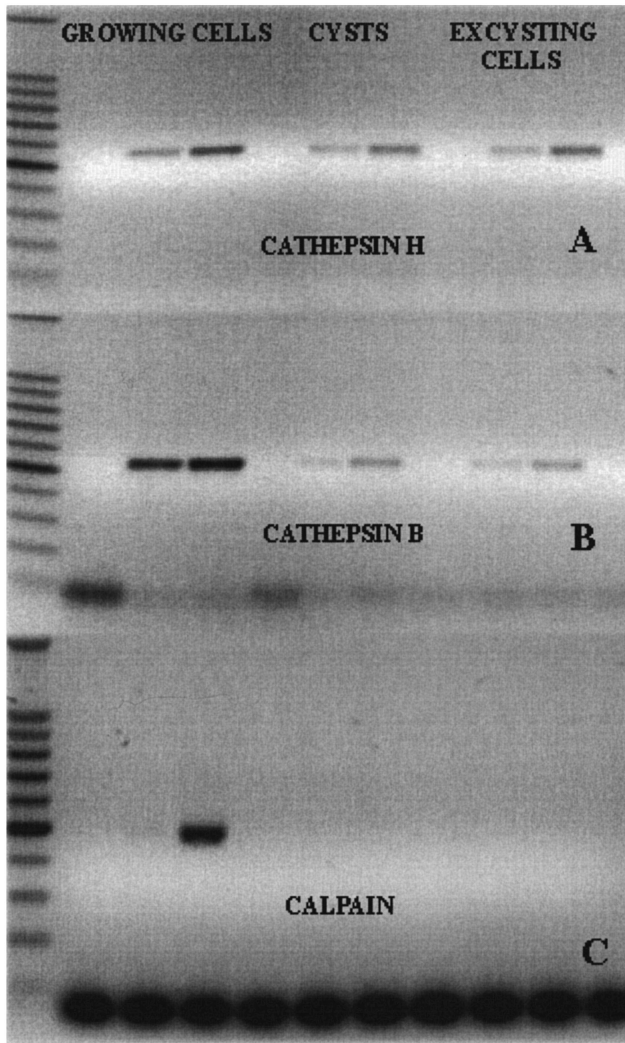


FIG. 3. Gel pictures resulting from cathepsin B, cathepsin H, and calpain RT-PCR experiments. Lanes (from the left): 1, 100-bp marker from Promega; 2, PCR control without cDNA; 5, 8, and 11, RT controls for which mRNA was not reverse transcribed; 3 and 4, 6 and 7, and 9 and 10, PCR carried out with, respectively, 0.1 and 0.3 μ l of cDNA. (A) Cathepsin H results after 35 cycles of PCR with specific primers annealing at 55°C, which show a slightly higher level of expression in vegetative cells. (B) Cathepsin B results after 35 cycles of PCR with specific primers annealing at 55°C, which show a strong differential expression in vegetative cells. (C) Calpain results after 45 cycles of PCR with specific primers annealing at 55°C, which show an apparent absence of expression in cysts and 1-h excysting cells.

codon located at the end of the 5' AT-rich region of the sequence. The sequence is 294 amino acids long, and a preliminary BLAST analysis shows a strong similarity with cathepsin B-like proteins.

The protein displays a conserved active site identical to that observed in cathepsin H (Fig. 4B); the propeptide region is 75 amino acids long, much shorter than the cathepsin H propeptide region. A signal peptide (MFKLVIIGTIVAVAVA/TH) was identified at the amino-terminal end of the molecule, here also following the bacterial rule for signal peptidase I cleavage. The mature peptide displays an important deletion in the so-

A
 MKTISASVLLILLFSSSLAMADVHADYHKIINGELPISTQQVDEIYSKFIQHYKSENSG
 FNPVEHYITKGIDRKATFANKLQQIHKNSDGTNTYKGLNAFSDMTDEFFDYNN
 IKAEQNCSAJINRKSFGNSNANIPTEWDRWTFGVVSPVKNQGGKCGSCWTFSTVGC
 VESHYLLKYGAFRNLSEQQLVDCAGDYDNHGCSGGLPSHAFFEYIKDNGGLALET
 YPYKAANGQCSIQGGQSVGIRGGAVNISLNEDDLKQAIYLHGPVSVAFRVIDGFR
 DYKSGVYAVEGCANGPNDVNHAVLAVGFGTDENKVDYWIKNISWGAAGWDQG
 FFKMKRGMVNMCGIQNCNSYPQDVFDAQQAASFLKLQ

B
 MFKLVIIGTIVAVAVAATHPINEEMVAHIKAKTSLWQPHETTINPFNNMTKIEQLLA
 KCGTYIVPANKEYPGSKIMTVPENFDARQQWGSKIHAIRDQOQCSGWAFGATEA
 FSDRFAINGKDVILSPEDLVSCDTNDYGCNGGYMDVAWEYLADHGAATDSCFPYS
 AGSGFAPACSDKCADGSAMQRFKCAPNSVRQSGKVAIQISEIVSHGPEVGAFTVY
 TDFNYQSGVYTPTTTDDVAGGHAIKILGYGVENGTPYWLCAINSWGPWAGMSGFG
 KIKQGECEGIEDQVFSCDPQL

C
 mpshIKEDGDYQKIIAAKKQYTDSDFFAAASLVNEKKSQLPANQLAEWKNLHWA
 RAKDFIGTDKFTVFDQNTIPNDILQDGLGNYFLSALSIAIEFPKRIRIDIFETEENP
 VGCYAVKVCINGEFRTILVDDQFPCTQDGFSGKIPAFSKTQNKELWVLIKLIW
 AKLNHNVENTITGFASEAFRLTGAPVEFYNHDIYVDNIWDQIWEADNRNICTSA
 GRTNLTSEDYSKIGLVSDHAYAVISVFDIDSKEHGLQLLKLRLNIPWGHKEWMDG
 WSDKSDKWNDDLKQVNFVDADDGVFFISYRDYMNYYRSTTICKVHDHYKHKSI
 KVNTANIDKQIAQGGPKPYQLVKITLQKSKLFTVTVQPTKRFVTQNLPSNDNYDT
 SFVKMILGKHDANKNFYSPFEYIDAKCSNYDNVSIESANEYEAGEYVVLVEVDwrqdi
 ckeyvqaysstgdvelavptgehndfleevlkscaarsktqkafddkgekdayraisivdgsaeygiyyqndsktsatv
 etvfnklenfeilapydqgvngknqvnvtagntiivrrtdraqsvsvysfaihpdtdlyhqtgkqkqiyqgel
 kdijfyisivfgyfvyfenksktllfdakfnfeltnmissegeentpgsqwvnlkpgenkikllimvdpqtawgykf
 syfkleqldsdadvalvrkhgavkkfkykgeeqkmqykvvhffehyvfensdpkllkfnfvlqnlqivdep
 gttfrvelapgtllkrritrdpvmkeskykcsysyefvssgadagvensqdngisqgpkdela

FIG. 4. Amino acid sequences of the isolated cysteine proteases. Capital letters are used for the cysteine protease domain. Catalytic sites are underlined, and characteristic conserved amino acids of these sites (C, H, and N) are in boldface and boxed. Proregions are shaded, and signal peptides are in boldface. Putative N-glycosylation sites are boxed. (A) Cathepsin H sequence. (B) Cathepsin B sequence. (C) Calpain sequence. An internal repeat in the carboxy-terminal domain is in boldface.

called occluding loop, which is involved in the dipeptidyl carboxypeptidase activity of the enzyme; this deletion is also observed in the cathepsin B sequence of *Giardia* (Fig. 7), suggesting that both enzymes display solely the endopeptidase activity. Similar deletions are observed in plant cathepsin Bs and in some cathepsins from *Caenorhabditis elegans*. Only one putative N-glycosylation site was found in the propeptide domain, suggesting a relatively low-affinity binding with mannose-6-phosphate receptors and the possibility that this cathepsin B could be mostly excreted (12).

The neighbor-joining phylogenetic tree of cathepsin B proteins displays two paraphyletic clusters (Fig. 8). The *S. histriomuscorum* sequence is strongly clustered with *Giardia* cathepsin B, thus confirming the shared structural signature of the occluding loop; both molecules are themselves clustered with a monophyletic subgroup involving particular cathepsins, either isolated as excreted tubulointerstitial nephritis antigens (*Homo*, *Mus*, and *Oryctolagus*) or deduced from genome sequencing (*Caenorhabditis*). Altogether, these molecules appear to be related to the plant cathepsins, as opposed to the more conventional animal cathepsin B clustering.

The minichromosome bearing the calpain-like gene is around 3.1 kbp long, as confirmed by Southern blot experiments (data not shown); the variations in the percent GC contents define three regions: a 5'-terminal region (393 bp) and a 3'-terminal region (133 bp) that are extremely AT rich (76 and 80%, respectively) and a central region of 2,511 bp which is 64% AT rich. A single open reading frame is included in the central region, with 7 TAG and 15 TAA codons, and a putative methionine start codon is located at the end of the 5'

Homo sapiens H	EIKHKYL***	**WSEFQNC	*ATKSNYLRG	*****TGPYP	PSVDWR
Sus scrofa H	--R-----*	**-----*	*--G-----*	*****-----	--M---
Mus musculus H	-----F-***	**-----*	*-----*	*****-----	S-M---
Rattus norvegicus H	-----***	**-----*	*-----*	*****-----	S-M---
Sterkiella histriomuscorum	-FFDY-N***	**IKAE----	*--NRKSFGN	****SNANI-	TEW---
Oryza sativa	-FQASR-***	***GAA----	*--LAGNH-M	***RDAPAL-	ETK---
Zea mays	-FRATR-***	***GAA----	*--LTGNH-M	**RAAAVAL-	ETK---
Lolium multiflorum	-FQATK-***	***GAA-T--	*--LAGNHLM	***RDANAL-	ETK---
Hordeum vulgare	-FQATR-***	***GAA-T--	*--LAGNHLM	***RDAAAL-	ETK---
Arabidopsis thalianaAAN31820	-FQRTK-***	***GAA----	*--LKGSHKV	****TEAAL-	ETK---
Prunus armeniaca	-FRQR-***	***GAA----	*--TKGSHEL	****TDAVL-	E-KN--
Arabidopsis thalianaAAM10319	-FQRYK-***	***GAA----	*--LKGSHKI	****TEATV-	DTK---
Pisum sativum	-FRSHR-***	***GAA----	*--LKGSHKI	****TDANL-	DEK---
Nicotiana tabacum	-FRDR-***	***GAA----	*--TKGN-KV	****TNVVL-	ETK---
Lycopersicon esculentum	-FRKHK-***	***GAS----	*--TKGN-KL	****TNVVL-	ETK---
Homo sapiens C	RRSGHS***	RKIIPR-KPAP	*L-AEIQQKI	*****LHL-	T-W---
Homo sapiens F	-FRTI-----*	**NTLLRKEP	*GN-MKQAKS	***VGDLA-	-EW---
Homo sapiens K	-VVQ-MT***	*GLKV-LSH-	**RSNDT-YI	PE**WE-RA-	D---Y-
Homo sapiens L	-F*RQVM***	***NGF--RK	*PR-GKVFQE	P**I, FYEA-	R-----
Homo sapiens O	-F-AI-----*	**R-K-SKF*	*PRY-AEVHM	SI**PNVSL-	LRF---
Homo sapiens S	-VMSLMS***	*SLRV-SQWQ	RNITYKSN**	****PNRIL-	D-----
Homo sapiens W	-FGQL-GYRR	AAGGV-SMGR	EIRSEPEES	*****V-	F-C---
Homo sapiens Z	YRPLRGD***	**GLA-LGR-	**YPRPHEY	***LSPADL-	K-W---

FIG. 5. Sequence alignment of cathepsin L-like sequences in the proregion domain. The alignment shows a signature shared with cathepsin H, the *S. histriomuscorum* sequence, and plant cathepsin H relatives, to the exclusion of the other cathepsin L-like sequences, which are highly divergent in this region. Dashes indicate identical amino acids, and asterisks indicate gaps introduced to fit the alignment.

AT-rich terminal region. The protein sequence is 836 amino acids long, and a preliminary BLAST analysis shows a strong similarity with the large subunit of calpain-like proteins.

The protein displays a cysteine protease domain located at the amino-terminal part of the molecule (Fig. 4C); the conserved active site of this domain is identical to those of cathepsins B and H. There is no signal peptide, suggesting a cytosolic location of the protein. Conventional calpains display a calpain III linker domain followed by a calcium-binding domain at the carboxy-terminal end of the molecule (51), but no such domain was found in the *S. histriomuscorum* sequence. However, an internal repeat was observed in the carboxy-terminal region, and the first repeat displays good similarity with the ciliate *Eufolliculina uhligi* calcium/calmodulin-dependent protein kinase catalytic domains X and XI (36).

The neighbor-joining phylogenetic tree of calpain protein also displays paraphyletic clusters, among which are represented the known mammalian sequences (52) (Fig. 9). The position of the *S. histriomuscorum* sequence is poorly defined: in the neighbor-joining distance analysis, it is related to the *Drosophila melanogaster* sol (small optic lobes) protein, and in the parsimony analysis, it is related to plant calpains (63.7% bootstrap value); in the maximum likelihood analysis, no significant clustering is observed. Nevertheless, whatever the reconstruction method that was used, it clearly belongs to a larger phylogenetic group including plants, trypanosomes, fungal palB, and animal calpain 7, to the exclusion of the more conventional animal calpains. Sequences which so far have been more precisely characterized in this group (sol, palB, cap5.5, capn7, and Dek1) do not display the calcium-binding domain IV of conventional calpains.

DISCUSSION

Inhibition experiments indicate the involvement of cysteine proteases during late stages of the excystment process. These experiments clearly show that inhibitor targets, i.e., cysteine proteases and/or proteasome, are specifically involved in the

excystment process. MG132 and calpain inhibitor II are known to block proteasome, calpain, and cathepsin activities (33), while E64 is a general inhibitor of cysteine proteases (6, 8, 42). Treating cysts with either MG132 or E64 leads to an excystment arrest, a result that clearly suggests that the proteasome is not directly involved in the excystment process. This indication is further reinforced by the fact that lactacystin (a specific inhibitor of proteasome) has no effect on the excystment process. However, this negative effect must be analyzed cautiously: while lactacystin is known to be active in vivo in cultured cells (19, 33), we cannot completely rule out the possibility that it might be actively pumped out of the cell, as is the case with yeast cells (32). Nevertheless, the results suggest that cathepsins and/or calpains are the main proteases involved in the excystment process.

The apparent absence of an effect of CA-074 and CA-074Me is rather surprising, since it was shown previously that cathepsin B is involved in the cyst wall breakdown of *Giardia* cysts (60). We may hypothesize that these inhibitors are unable either to enter the cell or to link their target, due to structural particularities. The first hypothesis can be rejected because the process of cyst wall breakdown implies extracellular secretion of cathepsin B, which then would be directly accessible to the inhibitor; this would result in an incomplete cell wall breakdown. On the other hand, *S. histriomuscorum* and *Giardia* molecules do not display the occluding loop which is specific to conventional cathepsin B proteins. Because histidine residues from this loop are essential for the inhibition specificity of CA-074 (55, 61), the CA-074 and CA-074Me molecules cannot be efficient in *S. histriomuscorum* cysts. Moreover, this should be extended to all cathepsin B molecules that do not possess an occluding loop. Finally, the absence of the occluding loop implies that cathepsin B works as an endopeptidase (the dipeptidyl carboxypeptidase activity is absent) in the process of cyst wall breakdown.

The blockage of excystment occurs at a specific stage: all ciliary anlagen are present in blocked cysts. Many arguments

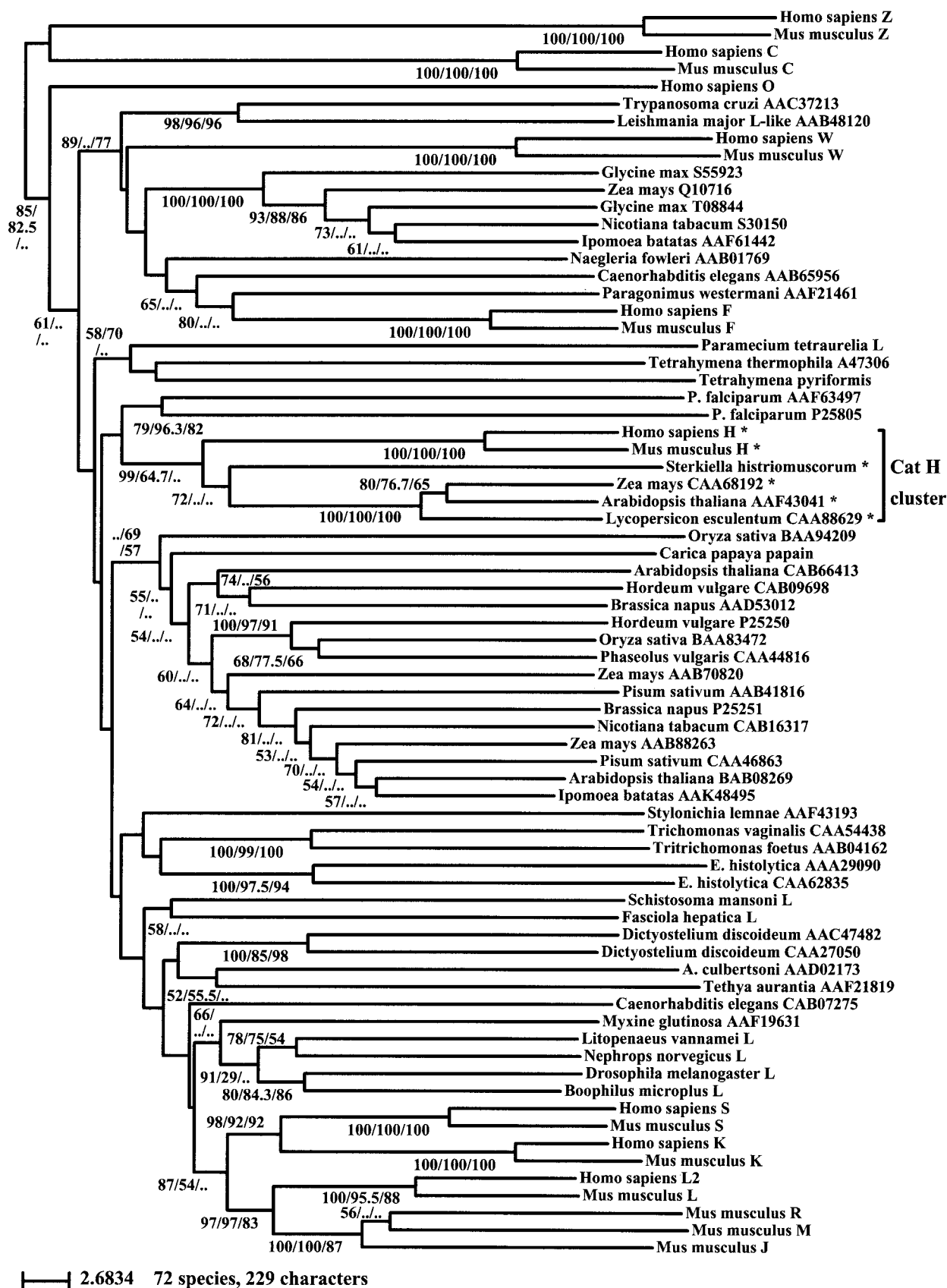


FIG. 6. Neighbor-joining phylogenetic tree of 72 cathepsin L-like sequences, including 229 amino acids. GenBank accession numbers are indicated when no precise identification was published and when several sequences were available for one species. Bootstrap values at the nodes are noted (neighbor-joining/parsimony/maximum likelihood, respectively); values under 50% have been discarded and are indicated by dots. The scale bar indicates 2.6834% of amino acid substitutions.

```

Mus musculus          PY*TIPPCEH HVN*GSRPPC TG*EGDTPRC NKSC
Rattus norvegicus     --*----- --*----- --*-----K- --M-
Homo sapiens          --*S----- --*----- --*-----K- S-I-
Haemonchus contortus AAA29175 --*P-H--G- -G-DTYIGE- R-TAP--P- KRR-
Leishmania major     AAB48119 --*PFD--S- -G-SEKY-- PSTIY--K- -TT-
Caenorhabditis elegans CAB00098 --*PY---- -G-GTHYK-- PSMNYP-DK- ER--
Caenorhabditis elegans CAB03007 ---VSGQSRE ***PGHCLI PK*RDY-N-Q GLR-
Caenorhabditis elegans AAB88058 ---P-A--TS *****GN- PE**SK--S- SM--
Caenorhabditis elegans CAB04322 ---P-R--NS *****DN- VN**LQ--P- RL--
Caenorhabditis elegans AAA98788 ---SFA--TK *****N- PE**ST--S- KTT-
Arabidopsis thaliana AAC44008 --FDNTG-S- *****-G- EP*AYP--K- SRK-
Arabidopsis thaliana AAC24376 --FDNTG-S- *****-G- EP*TYT--K- ERK-
Nicotiana rustica    CAA57522 --FDNEG-S- *****-G- EP*AYP--K- HRK-
Triticum aestivum   CAA46811 --FDQTG-Q- *****-G- EP*AYP--K- QRK-
Triticum aestivum   CAA46810 --FDQVG-K- *****-G- EP*AYP--V- E-K-
Sterkiella histriomuscorum --SAGSGFA* ***** -A- SDK-
Giardia intestinalis --QSGTGAR ***** *****GT- PTK-
Giardia muris        K-FSGMTGDR ***** *****ES- ITH-

```

FIG. 7. Sequence alignment of cathepsin B sequences in the occluding-loop region. The alignment shows the deletion common to *Giardia*, *Sterkiella*, plants, and *Caenorhabditis* sequences. Dashes indicate identical amino acids, and asterisks indicate gaps introduced to fit the alignment. The occluding-loop amino acids are in boldface, and the histidine residues, which are essential for CA-074 binding specificity, are shaded.

suggest that this stage could correspond to the commitment point of morphogenesis. In ciliates, whatever cycle is concerned (division, sexual cycle, regeneration, or excystment), the morphogenesis always proceeds through the same chronology: a new ciliature is assembled in the form of ciliary anlagen at specific locations on the cortex, and then all of these ciliary elements move on the cortex towards their specific location. All of these stages are well known during the fission cycle in many species (reviewed in references 18 and 20). Detailed studies performed with *Paramecium* (1) and *Tetrahymena* (17) have demonstrated that, as in metazoans, ciliates become committed to division at a specific stage; this stage is characterized by the presence of a complete set of ciliary organelles (1) and is thought to result from the classical cyclin/kinase cascade identified in metazoans (see reference 2 for a review). The fact that this stage could also correspond to a commitment point for excystment comes from studies on induced reencystment performed with excysting *Oxytricha* cells (25): excysting cells induced for reencystment before this point remain encysted, thus showing that the process of excystment is blocked. Cells which are induced for reencystment after this point carry on with excystation until the end of cortical morphogenesis but reencyst within the cyst wall, finally yielding two-walled cysts. Cysts blocked at the commitment point have also been described as the standard form of the resting cyst in one *S. histriomuscorum*-related species (16). This point, which marks the irreversibility of the process, is known to be sensitive to calpain inhibitors in metazoans (see reference 49 for a review), suggesting an involvement of calpains and proteasome in cyclin degradation. These data support the idea of the putative involvement of calpains in the process of excystment of *S. histriomuscorum*.

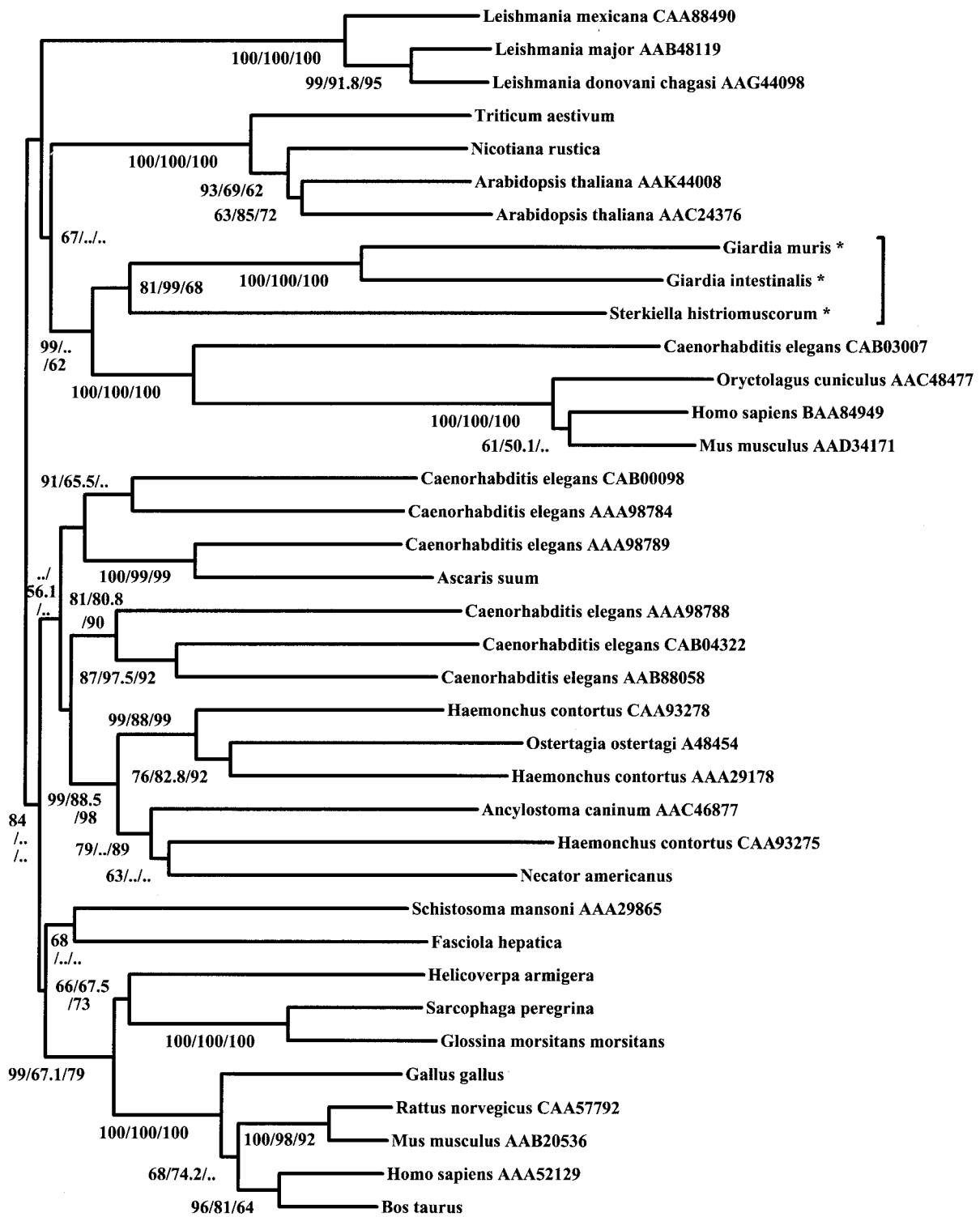
Structural and phylogenetic analyses suggest putative functions for the isolated cathepsins and calpain. The above discussion strongly suggests that cathepsin B is involved in the cyst wall breakdown process; while the mRNA differential-display technique may produce false-positive results, the isolated cathepsin H and calpain-like genes are not the conventional and abundant cathepsin L and capn1 (μ) or capn2 (m) calpain genes that would have been isolated in a cDNA random

screening. We thus hypothesize that the three isolated molecules are the main inhibitor targets involved in the excystment process. If we consider their gene expression patterns, only the cathepsin B and calpain genes display a strong differential expression between the cyst and growing cell stages. Vegetative cells divide and excystment proceeds, apparently without any problem, whatever inhibitor has been added. This observation is similar to that for *Giardia*, in which cathepsin B is essential for the excystation to be achieved and the corresponding gene is strongly expressed in vegetative cells (60). This is not the case of *Entamoeba*, for which lactacystin blocks in vivo the excystment process (19). As was previously reported for *Saccharomyces*, we assume that during the vegetative growth, either the inhibitor is actively pumped out of the cell or inhibition is not complete (15, 32).

Cathepsin B, which we know enters the secretory pathway and could be mostly excreted, would be involved in the breakdown of the cyst wall, in a way similar to that observed for *Giardia*. It is noteworthy that the cluster most related to the *S. histriomuscorum* cathepsin B sequence is made up of sequences of tubulointerstitial nephritis antigen, a protein which is accumulated in the basement membrane and is able to promote the adhesion of epithelial cells (28, 30), i.e., a protein which is excreted and involved in cell wall remodeling. Concerning the related plant cluster, a recent study describes a molecule involved in cell wall remodeling by cleavage of cell wall structural proteins in *Phleum pratense* (24).

Cathepsin H also enters the secretory pathway and should be directed mostly to lysosomes through the mannose-6-phosphate receptor pathway. However, some of it also could be excreted in order to take part in the cell wall breakdown process. Cathepsin L-like proteins are involved in intracellular protein degradation throughout their lysosomal activity; outside lysosomes, their activity contributes to extracellular matrix remodeling and apoptosis (56). Cathepsin H expression is increased in several cancer diseases and could be involved in the degradation of extracellular matrix components, leading to cell proliferation, migration, and metastasis (59). On the other hand, plant cathepsin H-like proteins are strongly expressed during leaf senescence and seedling germination, but also following water stress (7, 21), and are regulated by the plant hormones gibberellic acid and abscisic acid (45). These data suggest that secreted *S. histriomuscorum* cathepsin H could participate in the cyst wall breakdown, even if there is no strong differential expression of this molecule during the excystment process.

The calpain-like molecule does not display EF hands (which are characteristic of the calcium-binding domains) in its carboxy-terminal part, but an internal repeat similar to a calcium/calmodulin-dependent protein kinase domain has been found in this region. In the case of the conventional m-calpain, only one EF hand of the calcium-binding domain is strictly required for calcium-induced activation (13, 27). Moreover, the calpain-like tra-3 molecule displays a calcium-dependent protease activity, but lacks the EF hand domains (50). Although there is no formal evidence, these observations suggest that calcium ions could be involved in some way in the activation of the isolated *S. histriomuscorum* calpain. From other experiments, we also know that treatment of ionophore-permeabilized cysts with calcium ions is able to set off the excystment process,



3.0818 37 species, 248 characters

FIG. 8. Neighbor-joining phylogenetic tree of 37 cathepsin B sequences, including 248 amino acids. GenBank accession numbers are indicated when no precise identification was published and when several sequences were available for one species. Bootstrap values at the nodes are noted (neighbor-joining/parsimony/maximum likelihood, respectively); values under 50% have been discarded and are indicated by dots. The scale bar indicates 3.0818% of amino acid substitutions.

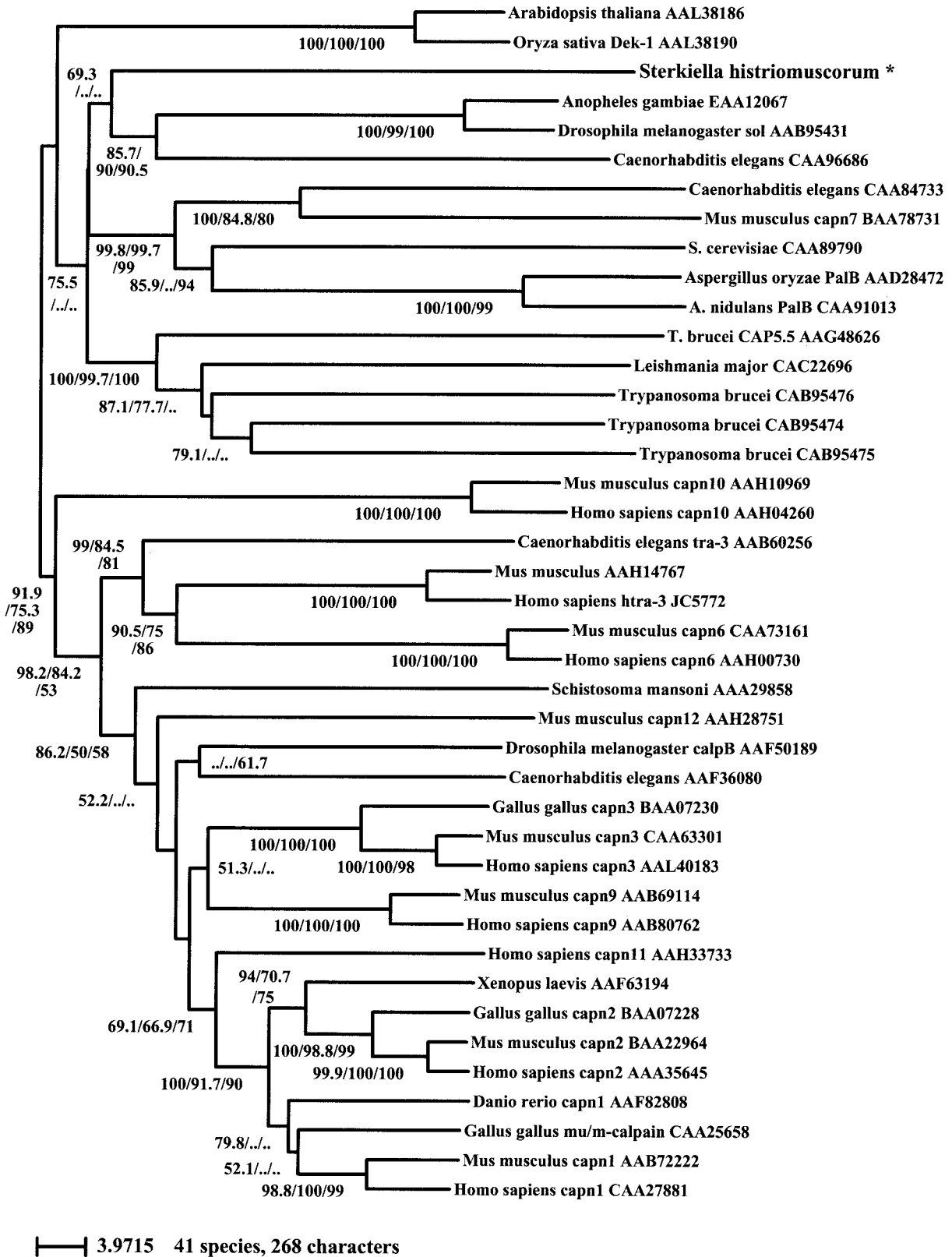


FIG. 9. Neighbor-joining phylogenetic tree of 41 calpain sequences, including 268 amino acids. GenBank accession numbers are indicated when no precise identification was published and/or when several sequences were available for one species. Bootstrap values at the nodes are noted (neighbor-joining/parsimony/maximum likelihood, respectively); values under 50% have been discarded and are indicated by dots. The scale bar indicates 3.9715% of amino acid substitutions.

without any addition of some other nutriment (R. Lescasse, G. Fryd-Versavel, and A. Fleury, unpublished data). Thus, a release of calcium ions in the cyst cytoplasm could be a signal for calpain-like activation. In other respects, RT-PCR experiments have shown that the *S. histriomuscorum* calpain-like molecule is not expressed during the first stages of the excystment process, suggesting that a signal cascade might be involved in the gene expression. Phylogenetic considerations suggest a relationship with the sol protein from *D. melanogaster* and with the palB protein from *Aspergillus*. While the exact function of the sol protein is unknown, it has been shown that palB belongs to a signal transduction cascade which is sensitive to ambient pH (10, 51). At alkaline pH, the transcription factor pacC is proteolytically processed to a functional form which activates alkaline-expressed genes and represses acid-expressed genes. PalB is involved in the two-step proteolytic processing of pacC (11). On another hand, it is noteworthy that the *S. histriomuscorum* calpain-like molecule is phylogenetically more related to the plant Dek-1 protein than to conventional animal calpains. Dek-1 is developmentally regulated and is involved in plant signal transduction through maintenance of aleurone cell fate close to the outer endosperm cell layer (35). Together, these observations suggest that the *S. histriomuscorum* calpain-like protein could be involved in a signal transduction pathway yet to be defined and could be activated during the excystment process.

In conclusion, cysteine protease genes isolated from samples of mRNAs expressed during the excystment stage of the life cycle of *S. histriomuscorum* are not the conventional cathepsin B and L, capn1 (μ), or capn2 (m) calpain genes. Inhibitors affect only the excystment process, to the exclusion of the dividing vegetative cell cycle. We speculate that cathepsins are involved in the cell wall breakdown process and that the calpain-like protein might be involved in a signal transduction pathway in relation to a cytoplasmic release of calcium ions; future work will focus on the characterization of the elements of this putative pathway.

ACKNOWLEDGMENTS

Special thanks go to Anne Baroin-Tourancheau for technical help and valuable advice.

This work was supported by the Centre National de la Recherche Scientifique and the Université de Paris-Sud. E.V. was the recipient of a Marie Curie postdoctoral fellowship from the EC (contract number ERBFMBICT 983187).

REFERENCES

- Adl, S. M., and J. D. Berger. 1991. Timing of oral morphogenesis and its relation to commitment to division in *Paramecium tetraurelia*. *Exp. Cell Res.* **192**:497–504.
- Adl, S. M., and J. D. Berger. 1996. Commitment to division in ciliate cell cycles. *J. Eukaryot. Microbiol.* **43**:77–86.
- Adl, S. M., and J. D. Berger. 1997. Timing of life cycle morphogenesis in synchronous samples of *Sterkiella histriomuscorum*. I. The vegetative cell cycle. *Eur. J. Protistol.* **33**:99–109.
- Berti, P. J., and A. C. Storer. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.* **246**:273–283.
- Bewley, J. D. 1997. Seed germination and dormancy. *Plant Cell* **9**:1055–1066.
- Boudjelal, M., Z. Wang, J. J. Voorhees, and G. J. Fisher. 2000. Ubiquitin/proteasome pathway regulates levels of retinoic acid receptor gamma and retinoid X receptor alpha in human keratinocytes. *Cancer Res.* **60**:2247–2252.
- Campalans, A., M. Pages, and R. Messegueur. 2001. Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond (*Prunus amygdalus*). *Tree Physiol.* **21**:633–643.
- Castilleja, A., N. E. Ward, R. B. Epstein, A. P. Kudelka, D. M. Gershenson, C. L. Efferson, C. A. O'Brian, and C. G. Ioannides. 2002. Treatment with HER-2 phosphorylation agonists enhance tumor ability to stimulate epitope specific CTL in vitro. *Oncol. Rep.* **9**:929–935.
- Chung, Y. B., Y. Kong, I. J. Joo, S. Y. Cho, and S. Y. Kang. 1995. Excystment of *Paragonimus westermani* metacercariae by endogenous cysteine protease. *J. Parasitol.* **81**:137–142.
- Denison, S. H., M. Orejas, and H. N. Arst, Jr. 1995. Signaling of ambient pH in *Aspergillus* involves a cysteine protease. *J. Biol. Chem.* **270**:28519–28522.
- Diez, E., J. Alvaro, E. A. Espeso, L. Rainbow, T. Suarez, J. Tilburn, H. N. Arst, Jr., and M. A. Penalva. 2002. Activation of the *Aspergillus* PacC zinc finger transcription factor requires two proteolytic steps. *EMBO J.* **21**:1350–1359.
- Dong, J. M., and G. G. Sahagian. 1990. Basis for low affinity binding of a lysosomal cysteine protease to the cation-independent mannose 6-phosphate receptor. *J. Biol. Chem.* **265**:4210–4217.
- Dutt, P., J. S. Arthur, P. Grochulski, M. Cygler, and J. S. Elce. 2000. Roles of individual EF-hands in the activation of m-calpain by calcium. *Biochem. J.* **348**:37–43.
- Felsenstein, J. 2001. PHYLIP (phylogeny inference package), version 3.6 (alpha 2). University of Washington, Seattle.
- Fleming, J. A., E. S. Lightcap, S. Sadis, V. Thoroddsen, C. E. Bulawa, and R. K. Blackman. 2002. Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. *Proc. Natl. Acad. Sci. USA* **99**:1461–1466.
- Fleury, A., and G. Fryd-Versavel. 1990. Précisions sur la dynamique morphogénétique de *Keranopsis muscicola* (cilié hypotriche). *J. Protozool.* **37**:33A.
- Frankel, J. 1962. The effects of heat, cold, and p-fluoro-phenyl-alanine on morphogenesis in synchronized *Tetrahymena pyriformis* GL. *C. R. Trav. Lab. Carlsb.* **33**:1–52.
- Frankel, J. 1989. Pattern formation: ciliate studies and models. Oxford University Press, New York, N.Y.
- Gonzalez, J., G. Bai, U. Frevert, E. J. Corey, and D. Eichinger. 1999. Proteasome-dependent cyst formation and stage-specific ubiquitin mRNA accumulation in *Entamoeba invadens*. *Eur. J. Biochem.* **264**:897–904.
- Grassé, P. P. 1994. *Traité de zoologie. Anatomie, systématique, biologie*, tome II. Infusoirs ciliés. Masson, Paris, France.
- Griffiths, C. M., S. E. Hosken, D. Oliver, J. Chojecki, and H. Thomas. 1997. Sequencing, expression pattern and RFLP mapping of a senescence-enhanced cDNA from *Zea mays* with high homology to oryzain gamma and aleurain. *Plant Mol. Biol.* **34**:815–821.
- Grimes, G. W. 1972. Cortical structure in nondividing and cortical morphogenesis in dividing *Oxytricha fallax*. *J. Protozool.* **19**:428–445.
- Grimes, G. W. 1973. Differentiation during excystment and excystment in *Oxytricha fallax*. *J. Protozool.* **20**:92–104.
- Grobe, K., M. Poppelmann, W. M. Becker, and A. Petersen. 2002. Properties of group I allergens from grass pollen and their relation to cathepsin B, a member of the C1 family of cysteine proteinases. *Eur. J. Biochem.* **269**:2083–2092.
- Hashimoto, K. 1963. Formation of ciliature in excystment and induced reexcystment of *Oxytricha fallax* Stein. *J. Protozool.* **10**:156–166.
- Hertz-Fowler, C., K. Ersfeld, and K. Gull. 2001. CAP 5.5, a life-cycle-regulated, cytoskeleton-associated protein is a member of a novel family of calpain-related proteins in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **116**:25–34.
- Hosfield, C. M., T. Moldoveanu, P. L. Davies, J. S. Elce, and Z. Jia. 2001. Calpain mutants with increased Ca²⁺ sensitivity and implications for the role of the C(2)-like domain. *J. Biol. Chem.* **276**:7404–7407.
- Ikeda, M., T. Takemura, S. Hino, and K. Yoshioka. 2000. Molecular cloning, expression and chromosomal localization of a human tubulointerstitial nephritis antigen. *Biochem. Biophys. Res. Commun.* **268**:225–230.
- Izquierdo, A., A. Martin-Gonzalez, and J. C. Gutierrez. 2000. Resting cyst wall glycoproteins isolated from two colpodid ciliates are glycine-rich proteins. *Cell. Biol. Int.* **24**:115–119.
- Kalfa, T. A., J. D. Thull, R. J. Butkowski, and A. S. Charonis. 1994. Tubulointerstitial nephritis antigen interacts with laminin and type IV collagen and promotes cell adhesion. *J. Biol. Chem.* **269**:1654–1659.
- Koornneef, M., L. Bentsink, and H. Hilhorst. 2002. Seed dormancy and germination. *Curr. Opin. Plant Biol.* **5**:33–36.
- Lee, D. H., and A. L. Goldberg. 1996. Selective inhibitors of the proteasome-dependent and vacuolar pathways of protein degradation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **271**:27280–27284.
- Lee, D. H., and A. L. Goldberg. 1998. Proteasome inhibitors: valuable new tools for cell biologists. *Trends Cell. Biol.* **8**:397–403.
- Letunic, I., L. Goodstadt, N. J. Dickens, T. Doerks, J. Schultz, R. Mott, F. Ciccarelli, R. R. Copley, C. P. Ponting, and P. Bork. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30**:242–244.
- Lid, S. E., D. Gruis, R. Jung, J. A. Lorentzen, E. Ananiev, M. Chamberlin, X. Niu, R. Meeley, S. Nichols, and O. A. Olsen. 2002. The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of

- maize grains encodes a membrane protein of the calpain gene superfamily. *Proc. Natl. Acad. Sci. USA* **99**:5460–5465.
36. **Markmann-Mulisch, U., B. Reiss, and M. Mulisch.** 1999. Cell type-specific gene expression in the cell cycle of the dimorphic ciliate *Eufolliculina uhligi*. *Mol. Gen. Genet.* **262**:390–399.
37. **McGrath, M. E.** 1999. The lysosomal cysteine proteases. *Annu. Rev. Biophys. Biomol. Struct.* **28**:181–204.
38. **Nothwehr, S. F., and J. I. Gordon.** 1990. Targeting of proteins into the eukaryotic secretory pathway: signal peptide structure/function relationships. *Bioessays* **12**:479–484.
39. **Osherov, N., and S. M. Gregory.** 2001. The molecular mechanisms of conidial germination. *FEMS Microbiol. Lett.* **199**:153–160.
40. **Perrin, B. J., and A. Huttenlocher.** 2002. Calpain. *Int. J. Biochem. Cell Biol.* **34**:722–725.
41. **Philippe, H.** 1993. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acid Res.* **21**:5264–5272.
42. **Pierre, P., I. Shachar, D. Matza, E. Gatti, R. A. Flavell, and I. Mellman.** 2000. Invariant chain controls H2-M proteolysis in mouse splenocytes and dendritic cells. *J. Exp. Med.* **191**:1057–1062.
43. **Prescott, D. M.** 1994. The DNA of ciliated protozoa. *Microbiol. Rev.* **58**:233–267.
44. **Rios, R. M., J. Martin, A. Torres, and C. Fedriani.** 1992. Production and characterization of three polyclonal antibodies raised against cyst wall proteins of a hypotrichous ciliate. *J. Protozool.* **39**:584–588.
45. **Rogers, J. C., D. Dean, and G. R. Heck.** 1985. Aleurain: a barley thiol protease closely related to mammalian cathepsin H. *Proc. Natl. Acad. Sci. USA* **82**:6512–6516.
46. **Saitou, N., and M. Nei.** 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
47. **Sajid, M., and J. H. McKerrow.** 2002. Cysteine proteases of parasitic organisms. *Mol. Biochem. Parasitol.* **120**:1–21.
48. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
49. **Santella, L., K. Kyojuka, L. De Riso, and E. Carafoli.** 1998. Calcium, protease action, and the regulation of the cell cycle. *Cell Calcium* **23**:123–130.
50. **Sokol, S. B., and P. E. Kuwabara.** 2000. Proteolysis in *Caenorhabditis elegans* sex determination: cleavage of TRA-2A by TRA-3. *Genes Dev.* **14**:901–906.
51. **Sorimachi, H., S. Ishiura, and K. Suzuki.** 1997. Structure and physiological function of calpains. *Biochem. J.* **328**:721–732.
52. **Sorimachi, H., and K. Suzuki.** 2001. The structure of calpain. *J. Biochem.* **129**:653–664.
53. **Towatari, T., T. Nikawa, M. Murata, C. Yokoo, M. Tamai, K. Hanada, and N. Katunuma.** 1991. Novel epoxysuccinyl peptides. A selective inhibitor of cathepsin B, in vivo. *FEBS Lett.* **280**:311–315.
54. **Tuffrau, M.** 1967. Perfectionnement et pratique de la technique d'imprégnation au protargol des Infusoires ciliés. *Protistologica* **3**:91–98.
55. **Turk, B., D. Turk, and V. Turk.** 2000. Lysosomal cysteine proteases: more than scavengers. *Biochim. Biophys. Acta* **1477**:98–111.
56. **Turk, D., M. Podobnik, T. Popovic, N. Katunuma, W. Bode, R. Huber, and V. Turk.** 1995. Crystal structure of cathepsin B inhibited with CA030 at 2.0-Å resolution: a basis for the design of specific epoxysuccinyl inhibitors. *Biochem. J.* **34**:4791–4797.
57. **Villalobo, E., C. Moch, R. Perasso, and A. Baroin-Tourancheau.** 2001. Searching for excystment-regulated genes in *Sterkiella histriomuscorum* (Ciliophora, Oxytrichidae): a mRNA differential display analysis of gene expression in excysting cells. *J. Eukaryot. Microbiol.* **48**:382–390.
58. **Von Heijne, G.** 1990. The signal peptide. *J. Membr. Biol.* **115**:195–201.
59. **Waghray, A., D. Keppler, B. F. Sloane, L. Schuger, and Y. Q. Chen.** 2002. Analysis of a truncated form of cathepsin H in human prostate tumor cells. *J. Biol. Chem.* **277**:11533–11538.
60. **Ward, W., L. Alvarado, N. D. Rawlings, J. C. Engel, C. Franklin, and J. H. McKerrow.** 1997. A primitive enzyme for a primitive cell: the protease required for excystation of *Giardia*. *Cell* **89**:437–444.
61. **Yamamoto, A., K. Tomoo, T. Hara, M. Murata, K. Kitamura, and T. Ishida.** 2000. Substrate specificity of bovine cathepsin B and its inhibition by CA074, based on crystal structure refinement of the complex. *J. Biochem.* **127**:635–643.