THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 26, pp. 18024–18031, June 27, 2008

Identification of the Major Cysteine Protease of *Giardia* and Its Role in Encystation*[S]

Received for publication, March 18, 2008, and in revised form, April 18, 2008 Published, JBC Papers in Press, April 29, 2008, DOI 10.1074/jbc.M802133200

Kelly N. DuBois[‡§1], Marla Abodeely[‡§], Judy Sakanari[‡], Charles S. Craik[¶], Malinda Lee[‡], James H. McKerrow[‡], and Mohammed Sajid[‡2]

From the [‡]Department of Pathology, the Sandler Center for Basic Research in Parasitic Diseases, the [§]Biomedical Sciences Graduate Program, and the [¶]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158

Giardia lamblia **is a protozoan parasite and the earliest branching clade of eukaryota. The *Giardia* life cycle alternates between an asexually replicating vegetative form and an infectious cyst form. Encystation and excystation are crucial processes for the survival and transmission of *Giardia*. Cysteine proteases in *Giardia* have been implicated in proteolytic processing events that enable the continuance of the life cycle throughout encystation and excystation. Using quantitative real-time PCR, the expression of twenty-seven clan CA cysteine protease genes in the *Giardia* genome was measured during both vegetative growth and encystation. *Giardia* cysteine protease 2 was the most highly expressed cysteine protease during both life cycle stages measured, with a dramatic expression increase during encystation. The mRNA transcript for *Giardia* cysteine protease 2 was 7-fold up-regulated during encystation and was greater than 3-fold higher than any other *Giardia* protease gene product. Recombinant *Giardia* cysteine protease 2 was expressed, purified, and biochemically characterized. The activity of the recombinant cysteine protease 2 protein was confirmed to be identical to the dominant cysteine protease activity found in *G. lamblia* lysates. *Giardia* cysteine protease 2 was co-localized with cyst wall protein in encystation-specific vesicles during encystation and processed cyst wall protein 2 to the size found in *Giardia* cyst walls. These data suggest that *Giardia* cysteine protease 2 is not only the major cysteine endoprotease expressed in *Giardia*, but is also central to the encystation process.**

Giardia lamblia is a protozoan parasite that inhabits the upper small intestine of many vertebrate hosts and is the most commonly isolated intestinal parasite world wide (1). In addition to its medical importance, *Giardia* is of interest as a model cell system because it represents the most early branching clade of eukaryota (2, 3). *Giardia* has a simple two-stage life cycle consisting of a vegetative replicating trophozoite and an infectious cyst. Infection is initiated with cyst ingestion by a vertebrate host. After passage through the acidic host stomach, vegetative trophozoites emerge from the cyst by the process of excystation, asexually divide by binary fission, establish the duodenal infection, and give rise to the characteristic symptoms of giardiasis. Trophozoites can form infective cysts that are passed in the host feces and ingested by another host to propagate the life cycle (1).

The process of encystation is a coordinated secretion of cyst wall materials to the periphery of a cell to form the cyst wall (4, 5). In response to environmental cues, trophozoites produce abundant cyst wall proteins that are packaged into encystation-specific vesicles (ESVs).[3] These vesicles grow, mature, and eventually traffic to the plasma membrane of the trophozoite, where cyst wall precursor material is secreted to form the environmentally stable cyst wall (4, 6, 7). The expression of many proteins is up-regulated during the encystation process (4).

Cysteine proteases have been found to be essential to the life cycles of several parasitic organisms, catalyzing diverse processes such as parasite immunevasion, tissue invasion, and encystment/excystment in addition to well established roles in protein processing and catabolism (8, 9). Indeed, in *G. lamblia* (10) indispensable roles for cysteine proteases have been documented in the processes of encystation and excystation. Ward *et al.* (10) validated a role for cysteine endoprotease activity in the excystation process by demonstrating that excystation was inhibited by the addition of small molecule cysteine protease inhibitors to the excystation media. Touz *et al.* (11) implicated a cysteine exoprotease in the process of encystation. Processing of cyst wall protein 2 (CWP2), one of the main cyst wall proteins that form the structure of the cyst wall, was also blocked by cysteine protease inhibitors.

Whereas the role of these chemical knock-out experiments focused attention on clan CA cysteine proteases in *Giardia*, the recent completion of the *Giardia* genome indicated that there are twenty-seven candidate clan CA cysteine protease genes in *Giardia*. To address the question of which gene product(s) was responsible for key events in the life cycle, such as cyst wall processing, we analyzed the transcription levels of all twenty-

* This work was supported, in whole or in part, by National Institutes of Health Grant AI35707 from the Tropical Disease Research Unit. This work was also supported by the Sandler Family Supporting Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S3 and Table S1.

[1] Supported by a NIAID, National Institutes of Health Training Fellowship in Microbial Pathogenesis.

[2] To whom correspondence should be addressed: 1700 4th St., QB3/UCSF 509, San Francisco, CA 94158-2550. Fax: 415-502-8193; E-mail: sajid@ucsf.edu.

[3] The abbreviations used are: ESVs, encystation-specific vesicles; CWP2, cyst wall protein 2; ESCP, encystation-specific cysteine protease; GAPDH, glutaraldehyde phosphate dehydrogenase; *Gl*CP2, *G. lamblia* cysteine protease 2; DTT, dithiothreitol; AMC, amino-methylcoumarin; Z, benzyloxycarbonyl; GFP, green fluorescent protein.

seven genes and found that *G. lamblia* cysteine protease 2 (*Gl*CP2) was in fact the major expressed cysteine protease gene in *Giardia.* We therefore cloned, heterologously expressed, and biochemically characterized this protease, and specifically evaluated its role in encystation.

MATERIALS AND METHODS

Cell Culture, Transfection, and Differentiation—WB isolated *G. lamblia* trophozoites (ATCC number 30957) were maintained in a modified TYI-S-33 medium supplemented with 10% fetal bovine serum (Omega Scientific, Inc.), penicillin-streptomycin (UCSF CCF), vitamins (Invitrogen), and Fungizone (UCSF CCF). The pGFP.pac vector (gift from Theodore Nash, National Institutes of Health; modified by Lei Li from the C. C. Wang laboratory, UCSF) was used to episomally express C-terminal GFP fusion proteins in *Giardia*. The transfection protocol used by Singer *et al.* (12) was followed with modifications: $1–2 \times 10^6$ trophozoites were electroporated with 50 μg of plasmid DNA (GenePulser XCell, Bio-Rad) at 0.45 kV, 950 μF. Transfectants were selected with puromycin dihydrochloride (Sigma) and increased in 5–20 μg/ml increments to a final concentration of 80–120 μg/ml. Trophozoites were induced to encyst as indicated by Abel *et al.* (13).

Transformation and Expression of GlCP2 in Pichia pastoris—The *Gl*CP2 gene was re-synthesized to optimize for yeast codon usage (DNA 2.0). The r*Gl*CP2 gene was amplified by PCR from the pJ31:7972 vector into which the full-length cDNA had previously been cloned and modified to include a polyhistidine tag using the forward primer *Gl*CP2pPicF: CTCGAGAAAAGACATCATCATCATCATCATGAGTTGAATCATATTACTC and the reverse primer *Gl*CP2pPicR: TCTAGATTACTCATCGAAAAATCCAGCATAGGCC. The 920-bp amplicon was subcloned in the XhoI/XbaI site of the *P. pastoris* expression vector pPicZαB (Invitrogen). The plasmid was linearized by digestion with SalI and introduced into *P. pastoris* by electroporation (GenePulser XCell, Bio-Rad) according to the manufacturer's specifications. Transformants were screened by growth on YPD+100 μg/ml zeocin (Invitrogen).

Purification of rGlCP2—P. pastoris was grown under expression induction conditions in a BioFlo 110 Fermentor/Bioreactor (New Brunswick Scientific) for 3 days according to the manufacturer's specifications. Methanol was maintained at 0.5% (calculated by a methanol sensor) by addition of 100% methanol 2×/day. 0.2-μm filtered supernatant was lyophilized. 8 g of lyophilized material was resuspended in 40 ml of ddH2O + 1 mM Pefabloc (Sigma). Solution was filtered at 0.2 μm, dialyzed in 10,000 MWC dialysis tubing (Pierce) against 10 mM Tris-HCl, pH 8.0 at 4 °C, and fractionated by ion exchange chromatography with Fast Flow Q resin (GE Healthcare) followed by dialysis to desalt and a Mono Q anion exchange column (GE Healthcare).

Purification of Cysteine Protease Activity from G. lamblia Lysates—Giardia cells were incubated in a 20 mM Tris-HCl, pH 7.2 and 0.2% Triton X-100 (Sigma-Aldrich) buffer at 4 °C with stirring for 2 h. Debris was pelleted, and supernatant was 0.2-μm filtered and subjected to anion exchange chromatography using a Mono Q column.

Expression and Purification of CWP2—The open-reading frame for CWP2 was amplified from genomic DNA, and a C-terminal polyhistidine tag was added using the primers CWP2pMalF: TCTAGAATGGCTTGCCCTGCCACCGAGG and CWP2pMalR: GCGGCCGCTTTAATGATGATGATGATGATGCCTTCCCTGGATCCTTCTGCGGACAATAG and inserted into the NotI/XbaI site of the expression vector pCMVTnT (Promega). 1 μg of plasmid DNA was used as a template for *in vitro* transcription using the TnT Quick-coupled Transcription/Translation kit according to the manufacturer's specifications (Promega). [^{35}S]rCWP2 was further purified on a nickel-nitrilotriacetic acid column (Qiagen).

Protease Activity Assays—40 μM of the fluorogenic substrates Z-FR-AMC (*N*-carbobenzoxy-phenylalanyl-arginyl-7-amido-4-methylcoumarin) and Z-RR-AMC (*N*-carbobenzoxy-arginyl-arginyl-7-amido-4-methylcoumarin; excitation/emission of AMC: 360 nm/470 nm) (Bachem) were incubated with *Giardia* lysates or recombinant enzyme in Tris-HCl buffer (pH 7.2) or citrate/dibasic sodium phosphate buffers (pH 4.0–8.0) containing 4 mM DTT, 1 mM Pefabloc, and 10 mM EDTA. Subsequent protease activity was measured by monitoring the increase in relative fluorescence units (RFU) over time.

r*Gl*CP2 was incubated for 30 min with 50 μg of casein-resorufin (Molecular Probes) in 200 μl of citrate/dibasic sodium phosphate buffers. 960 μl of 5% (w/v) trichloroacetic acid was added, samples were incubated 10 min, and centrifuged. 400 μl of supernatant were added to 600 μl of 0.5 M Tris, pH 8.8. Hydrolysis was quantified by measuring fluorescence (excitation/emission: 574 nm/584 nm).

Purified rCWP2 was incubated with enzyme in Tris-HCl buffer, pH 7.2, 4 mM DTT at 25 °C. The sample was fractionated by SDS-PAGE, dried, and visualized by phosphorimaging (Typhoon Trio, GE Healthcare).

DCG04 (radio^{125}I-iodinated or BODIPY-labeled), the clan CA cysteine protease inhibitor (14), was incubated with enzyme and 4 mM DTT for 30 min. Proteins were fractionated by SDS-PAGE, dried, and visualized by phosphorimaging (Typhoon Trio, GE Healthcare).

To determine the K_m of *Gl*CP2, the fluorogenic peptide substrates Z-FR-AMC, Z-RR-AMC, and Z-VLK-AMC (Bachem) were incubated with enzyme at a range of concentrations, and the V_{max} units/s was recorded. The non-linear regression and K_m calculations were determined using Prism 4 software (Graphpad).

r*Gl*CP2 was fractionated on a Novex® 10% zymogram (gelatin) gel (Invitrogen) under native conditions as recommended by the manufacturer. The gel was stained with SimplyBlue™ Safestain (Invitrogen) and destained in water to visualize bands of protease activity.

r*Gl*CP2 was fractionated by SDS-PAGE under non-reducing conditions on a 15% Tris-HCl gel. The gel was washed 2× in 20 mM Tris-HCl, 0.2% Triton X-100. The gel was incubated in 20 mM Tris-HCl, 0.2% Triton X-100, 5 mM DTT, and 10 μM Z-FR-MNA (*N*-carbobenzoxy-phenylalanyl-arginyl-4-methoxy-β-naphthylamide) for 2 h at room temperature. Two volumes (compared with substrate) of 2 M coupling reagent (5-nitro-2-salicylaldehyde) was added to the reaction, and the reaction was

incubated for an additional 4 h at room temperature. Fluorescence was visualized on a Typhoon Trio (GE Healthcare).

Positional Scanning Synthetic Combinatorial Library (PS-SCL)—Protease activity was assayed at 25 °C in a buffer containing 20 mM Tris-HCl, pH 7.2, 5 mM DTT, 0.2% Triton X-100 (Sigma-Aldrich), and 1% Me_2SO (from the substrates) or in buffer with NaOAc replacing Tris-HCl (pH 5.5) as referenced in the text. Assays were performed as previously described using 250 μM substrate in each assay (15).

RNA Methods—Total RNA from vegetative or encysting *Giardia* cells was isolated with TRIzol reagent (Invitrogen). 2 μg of RNA was treated with 1 unit of amplification grade DNase I (Sigma). cDNA was synthesized with Superscript III reverse transcriptase according to the manufacturer's specifications (Invitrogen). cDNA samples were stored at −80 °C until use. Control samples were prepared as above using nuclease-free ddH2O in place of RNA.

Real-time PCR—PCR was performed in an Mx3005P™ QPCR system using MxPro™ QPCR software (Stratagene). Amplification was performed in a final volume of 25 μl, containing cDNA from the reverse-transcribed reaction, primer mixture (0.3 μM each of sense and antisense primers), and 12.5 μl of 2× SYBR Green Master Mix (Applied Biosystems). The final mRNA levels of the genes studied were normalized to GAPDH expression using the comparative C_T method (Stratagene).

The sequences of *Giardia* cysteine proteases were obtained from the *Giardia* genome project. For the GenBank™ protein accession numbers and primers see supplemental Table S1.

Microscopy—A confocal microscope (LSM510 META; Carl Zeiss MicroImaging, Inc.) equipped with multiline (458, 477, 488, and 514 nm) Ar, Diode 405 nm, 543 nm HeNe, and 633 nm HeNe visible lasers with a "Plan-Apochromat" 63×/1.40 Oil DIC oil immersion lens (Carl Zeiss MicroImaging, Inc.) was used for fluorescence imaging. Cells were pulsed with oxygen at 37 °C for 1–3 h, fixed in 3% paraformaldehyde (Electron Microscopy Sciences) for 40 min, and mounted with ProLong Gold mounting media (+ or − DAPI) (Molecular Probes). LSM Image Browser software (Carl Zeiss MicroImaging, Inc.) was used for confocal image acquisition and analysis. Adobe Photoshop CS (Adobe Systems, Inc.) was used for subsequent processing.

Antibodies and Reagents—Anti-*Giardia* cyst wall protein polyclonal was used at 1:100 (Waterborne, Inc.) Anti-*Gl*CP2 peptide polyclonal (raised against the peptide SSKVHLATATSYKDYGLDI) was used at 1:500. Inhibitors: phenylmethylsulfonyl fluoride (Sigma), EDTA (Sigma), aprotinin (Sigma), pepstatin A (Calbiochem), leupeptin (Sigma), TLCK (1-chloro-3-tosylamido-7-amino-2-heptanone HCl) (Sigma), TPCK (1-chloro-3-tosylamido-4-phenyl-2-butanone) (Sigma), E64 (Sigma), CA074 (Sigma), lactacystin (Sigma), α-1 antitrypsin (Sigma), calpain inhibitor I (ALLN, Sigma), calpain inhibitor II (ALLM, Sigma). *R. norvegicus* cathepsin C was a gift from John Pederson (Unizyme, Denmark).

Mass Spectrometry—Tryptic digest sample was analyzed by liquid chromatography-mass spectrometry/mass spectrometry. Analyses were performed with an LTQ ion trap (Thermo Scientific) and a QSTAR (Applied Biosystems). The data base search was conducted using Mascot (Matrix Science Inc) on the full NCBI protein data base. Mass accuracy for the QSTAR data: 100 ppm in MS; 0.2 Da in MS/MS. Mass accuracy for the LTQ data: 3 Da in MS; 0.8 Da in MS/MS.

Two-dimensional Gel Electrophoresis—Protein samples were de-salted with Centricon spin columns (Millipore). Two-dimensional gel electrophoresis was performed according to the manufacturer's specifications using the Zoom IPGRunner system (Invitrogen). Gels were silver-stained with the Silver Stain Plus kit (Bio-Rad), and protein spots were excised and trypsin-digested.

RESULTS

There Are Twenty-seven Clan CA Cysteine Proteases in the Genome of Giardia lamblia—Prior to the completion of the *Giardia* genome, only four cysteine proteases from *Giardia* had been identified (10, 11). Three of these genes encode cathepsin B-like cysteine proteases, the fourth a cathepsin C-like protease. Using these papain family enzymes as a query, the *Giardia* genome was mined for additional genes coding for cysteine proteases. In total, twenty-seven clan CA cysteine protease genes were located in the *Giardia* genome, and these could be classified by sequence homology into cathepsin B-like, cathepsin C-like, or cathepsin K/L-like categories (Fig. 1). Three of these (GenBank™ accession number EAA38990) are greater than 95% identical to each other and yet are found in three discrete regions of the genome. There is also a set of four genes greater than 95% identical and assigned to the same GenBank™ accession number (AAK92150).

GlCP2 Is the Most Highly Expressed Cysteine Protease of the Twenty-seven in the Giardia Genome—RT-PCR was used to determine if each of these genes was expressed in the vegetative and encysting stages of the *Giardia* life cycle. It was found that twenty-five of these twenty-seven genes are expressed, while no expression could be seen for the genes with GenBank™ accession numbers EAA37191 and EAA39030 (data not shown). Quantitative RT-PCR was performed to determine the relative levels of gene expression among the expressed *Giardia* cysteine proteases in the vegetative and encysting life stages. Expression was normalized to the expression of the housekeeping gene glutaraldehyde phosphate dehydrogenase (GAPDH), which has been found to have stable expression during the *Giardia* life cycle (5, 16). It is notable that the cathepsin B-like cysteine proteases are more highly expressed than the cathepsin C-like or K/L-like proteases (Fig. 1). Expression of all of the cysteine protease genes was increased marginally during encystation. *Gl*CP2 was the most highly expressed transcript in both vegetative and encysting life stages by 1.6-fold and 3.5-fold, respectively, over the next most highly expressed transcript (*Giardia* cysteine protease 3, with 89% homology to *Gl*CP2). This is consistent with *Gl*CP2 being the only cysteine protease Ward *et al.* (10) biochemically identified from the parasite by affinity purification and N-terminal sequencing. The expression of this gene was also increased by 7-fold during encystation (Fig. 1).

GlCP2 Is Identified in Lysate Fractions Enriched for Cysteine Protease Activity—Biochemical characterization of total cysteine protease activity found in *Giardia* lysates was concurrently undertaken to complement the gene expression analysis.

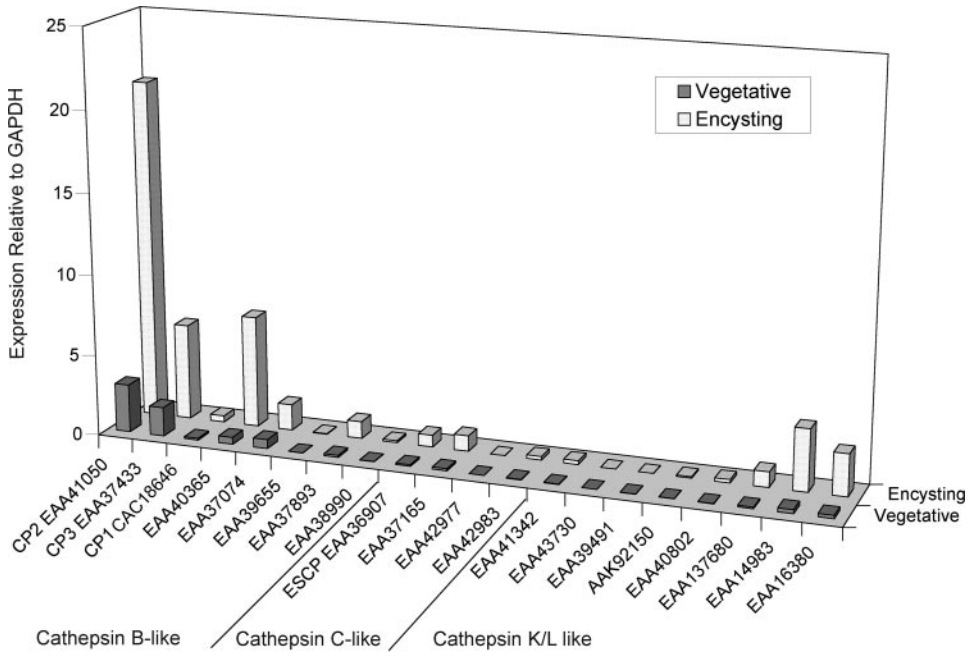

FIGURE 1. **By quantitative RT-PCR, *Gl*CP2 is the dominant transcript of twenty-seven *Giardia* clan CA cysteine protease genes.** Genes were classified by sequence homology into cathepsin B-like (8), cathepsin C-like (4), and cathepsin K/L-like (10) categories. GenBank™ accession number EAA38990 represents three greater than 95% identical genes and GenBank™ accession number AAK92150 represents four genes with greater than 95% homology. Gene expression was measured during vegetative growth and 24 h after induction of encystation. The transcript level is expressed relative to the housekeeping gene GAPDH. Two genes, both cathepsins K/L-like, were not expressed (GenBank™ accession numbers EAA37191 and EAA39030). *Gl*CP2 was the most highly expressed cysteine protease gene in both vegetative and encysting life cycle stages.

Giardia lysates were fractionated by ion exchange chromatography, and each fraction tested against an array of N-terminally blocked fluorescent peptide substrates (data not shown). Two main peaks of cysteine protease activity were resolved against the substrates Z-FR-AMC and Z-RR-AMC (Fig. 2*A*). The first peak eluted (*Peak A*) exhibited activity against both Z-RR-AMC and Z-FR-AMC, while the second peak (*Peak B*) had activity against Z-FR-AMC but far less against Z-RR-AMC. The activity-containing fractions from each peak were subsequently enriched with two additional rounds of ion exchange chromatography, concentrated, probed with a labeled irreversible cysteine protease active site inhibitor and resolved by one-dimensional SDS-PAGE or two-dimensional gel electrophoresis. The active site probe labeled two discrete protein bands in Peak A and only one protein band in Peak B (Fig. 2*B*). Protein bands from one-dimensional SDS-PAGE or spots from two-dimensional gel electrophoresis were subjected to tryptic digest and analyzed using liquid chromatography-mass spectrometry/mass spectrometry. The only cysteine protease identified from these methods was *Gl*CP2, of which peptides were identified in both Peak A and Peak B (Table 1). This was consistent with the observation that *Gl*CP2 was the major cysteine protease transcript expressed by *G. lamblia*.

Expression and Characterization of Recombinant GlCP2—To further analyze the activity and biological role of *Gl*CP2, a resynthesized *Gl*CP2 (r*Gl*CP2) gene (resynthesized to optimize for yeast codon bias) of 34 kDa was expressed heterologously in *P. pastoris* (Fig. 3*A*). The polyhistidine-tagged r*Gl*CP2 was purified by affinity and anion exchange chromatography and was found to autoactivate during the purification process to the mature form of 28 kDa (Fig. 3*A*). The full-length and mature forms of r*Gl*CP2 had activity on a 10% gelatin zymogram native gel and migrated to an apparent mobility of 60-kDa marker (supplemental Fig. S1).

To compare r*Gl*CP2 activity to that predominantly seen in *Giardia* lysates, the activity profile of r*Gl*CP2 by ion exchange chromatography was examined against Z-FR-AMC and Z-RR-AMC. The two activity peaks seen in *Giardia* lysates were reproduced with purified recombinant protein; the peaks of activity represent the pro and mature forms of the protease in Peak A and the mature protease alone in Peak B (Fig. 4*A*). This is consistent with the two protein bands labeled with the cysteine protease active site probe in Peak A, and the single band labeled in Peak B (Fig. 4*B*). To determine if full-length r*Gl*CP2 has activity against a peptide substrate, protein from Peak A and Peak B was fractionated by SDS-PAGE under non-reducing conditions. In-gel activity was tested against the fluorogenic substrate Z-FR-MNA. Two bands of activity could be visualized in Peak A, while only one band of endoprotease activity was resolved in Peak B (Fig. 4*B*). A Western blot of these fractions using an antibody against *Gl*CP2 also demonstrates that two bands in Peak A and one in Peak B are identified as *Gl*CP2 (Fig. 4*C*). These data are consistent with the biochemical and mass spectrometry evidence that *Gl*CP2 is responsible for the activity found in both peaks A and B. Purified peak B was utilized for further biochemical studies.

An array of protease inhibitors was tested for their ability to inhibit r*Gl*CP2 activity against Z-FR-AMC and Z-RR-AMC. Leupeptin and E64 were the most effective inhibitors of r*Gl*CP2 (Table 2). The pH profile of r*Gl*CP2 was elucidated using the peptide substrates Z-FR-AMC and Z-RR-AMC and the macromolecular substrate casein-resorufin. The pH optimum for r*Gl*CP2 was found to be in the neutral range for each of these substrates (Fig. 3*B*). This is consistent with the localization of *Gl*CP2 in non-acidified compartments and its absence in the acidified peripheral vacuoles (PVs) (data not shown). The K_m and k_{cat}/K_m of Z-FR-AMC for *Gl*CP2 were found to be 40 μM and 17.5 μM/s, respectively. The K_m and k_{cat}/K_m of Z-RR-AMC for *Gl*CP2 were found to be 9 μM and 72 μM/sec, respectively.

To characterize the substrate specificity of r*Gl*CP2, a positional scanning synthetic combinatorial library was used to determine the substrate preference of the substrate binding sites for P1-P4 (15) (supplemental Fig. S2). r*Gl*CP2 displays an amino acid preference at subsites P1 and P2 (P1: K>>R, Q, P; P2: L, M, V, F) while subsites P3 and P4 have relaxed specificity. These libraries were tested both at the optimal pH for the

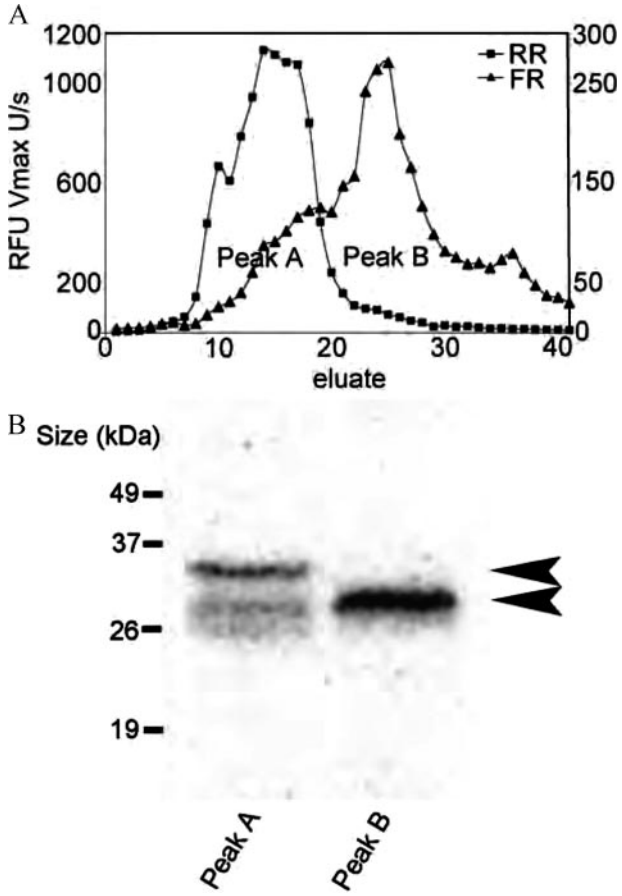

FIGURE 2. **Two distinct cysteine protease activities were resolved by anion exchange chromatography of *Giardia* lysates.** *A*, protease activity from eluates was tested against the N-terminally blocked fluorogenic peptide substrates Z-FR-AMC (FR ▲) and Z-RR-AMC (RR■). The two cysteine protease activity peaks exhibited distinct substrate preferences. Peak A had higher activity against Z-RR-AMC, while Peak B had higher activity against Z-FR-AMC and far less activity against Z-RR-AMC. *B*, eluates from the cysteine protease activity peaks A and B were incubated with the ^{125}I-labeled active site probe DCG04. The probe labeled two protein bands in Peak A while only one band could be seen in Peak B.

TABLE 1

Amino acid sequences of peptide fragments of cysteine protease activity Peak A and Peak B

Sequences were identified by liquid chromatography-mass spectrometry/mass spectrometry of cysteine protease activity Peak A and Peak B eluted from anion exchange chromatography of *Giardia* lysates.

Peak A	Peak B
CVAGLDK	CVAGLDK
TGTTTDECVPYK	TGTTTDECVPYK
	VHLATATSYK
	DYGLDIPAMMK
	GINDCSIEEQAYAGFFDE
	NSWGPDWGEDGYFR
	SGSTTLR
	GTCPTK
	CADGSSK

enzyme (7.2) and at pH 5.5, the conventional pH for this class of enzymes. The substrate specificity did not change over this pH range, though the level of enzyme activity was decreased by approximately 50% at the lower pH (data not shown). Based on the substrate specificity, an ideal substrate (Z-VLK-AMC) was used to measure the K_m and k_{cat}/K_m, which were found to be 19 μM and 1,473 μM/s, respectively.

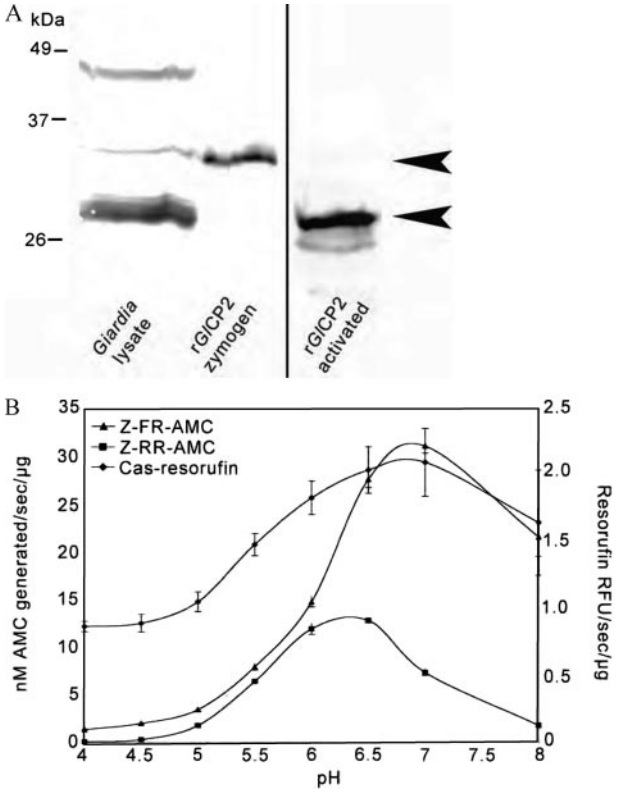

FIGURE 3. **Resynthesized *Gl*CP2 (r*Gl*CP2) was heterologously expressed, purified, and biochemically characterized.** *A*, Western blot probed with anti-*Gl*CP2 peptide antibody showed that r*Gl*CP2 was expressed as a zymogen of 34 kDa. During the process of purification r*Gl*CP2 was autoactivated to the mature form of 28 kDa. *B*, pH profile of r*Gl*CP2 against the fluorogenic substrates Z-FR-AMC (▲) and Z-RR-AMC (■) and against the macromolecular substrate casein-resorufin (♦) demonstrated that the pH optimum was 6–7 for activity against all substrates. The pH optimum r*Gl*CP2 with Z-RR-AMC, Z-FR-AMC, and casein is different (*B*). This may reflect the pH dependence of the side-chain charge of the peptide substrates and the corresponding cleavage sites within casein. Alternatively, the charge at the bottom of the S2 pocket may be pH-dependent as was demonstrated structurally for the homologous cysteine protease cruzain (25).

GlCP2 Is Found in Giardia ESVs and Can Proteolytically Process CWP2 to the Predicted Size Found in the Cyst Wall—The localization of *Gl*CP2 during encystation was determined by episomal expression of a *Gl*CP2-GFP fusion in *Giardia*. Encysting cells were probed with an antibody against cyst wall protein to highlight the ESVs, and *Gl*CP2-GFP was found to localize to ESVs (Fig. 5). Ward *et al.* (10) previously implicated *Gl*CP2 in *Giardia* excystation. However, whether this protease could also play a role in the encystation process was not addressed directly.

Total *Giardia* lysates or purified r*Gl*CP2 was incubated with recombinant CWP2 (rCWP2). In the presence of either *Giardia* lysates or r*Gl*CP2, rCWP2 was processed from its original 39-kDa size to a 26-kDa fragment, the same size of the protein found in the cyst wall and shown to be produced by incubation with a purified fraction of cysteine protease activity containing encystation-specific cysteine protease (ESCP) (Fig. 6, *A* and *B*) (4, 11). rCWP2 was also processed to this 26-kDa fragment in the presence of the endopeptidases trypsin and chymotrypsin, suggesting that the processing of CWP2 is not dependent on protease specificity but instead dependent on the structure of CWP2 and the protein segments accessible to an endopepti-

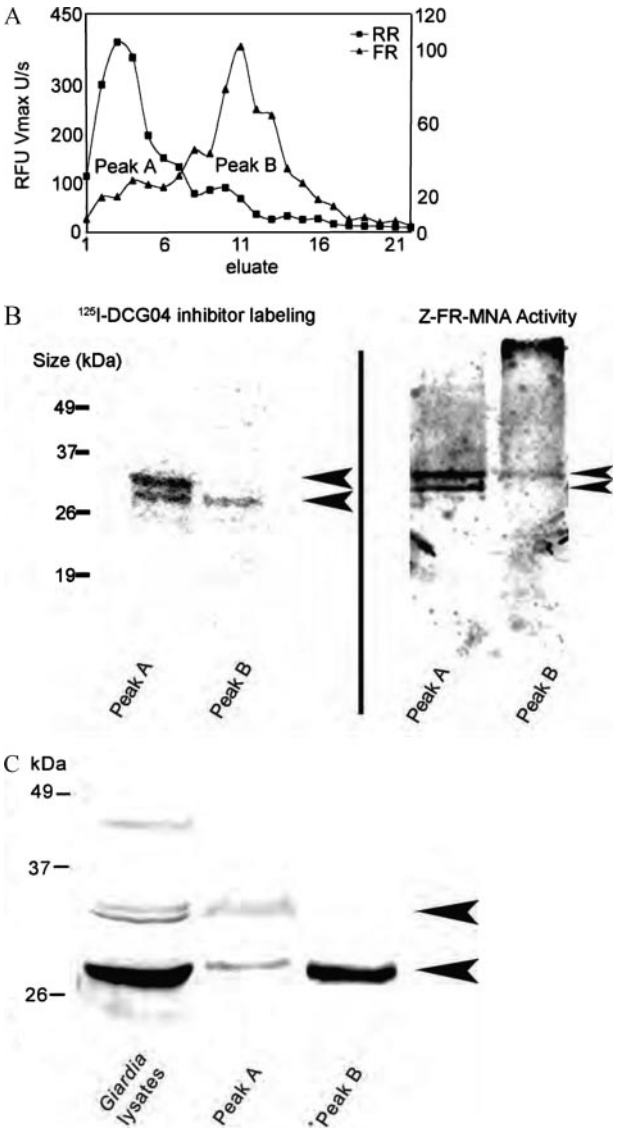

FIGURE 4. **Fractionated r*Gl*CP2 activity was identical to cysteine protease activity from *Giardia* lysates.** *A*, r*Gl*CP2 eluates from anion exchange chromatography were tested against the N-terminally blocked fluorescent peptide substrates Z-FR-AMC (FR▲) and Z-RR-AMC (RR■). As found with protease activity from lysates, two cysteine protease activity peaks were resolved and exhibited distinct substrate preferences. Peak A had higher activity against Z-RR-AMC, while Peak B had higher activity against Z-FR-AMC and far less activity against Z-RR-AMC. *B*, *left panel*, fractions of cysteine protease activity Peaks A and B were incubated with the ^{125}I-labeled active site probe DCG04 and fractionated by SDS-PAGE. The probe labeled two protein bands in Peak A while only one labeled protein band could be seen in Peak B (*arrowheads*). *Right panel*, Peaks A and B proteins were fractionated by SDS-PAGE under native conditions, and the resulting gel was incubated with the fluorogenic peptide substrate Z-FR-MNA. Two bands of fluorescence, indicative of endoprotease activity, were observed in Peak A and one band in Peak B (*arrowheads*). *C*, Western blot probed with anti-*Gl*CP2 peptide antibody demonstrated that two protein bands in Peak A and one protein band in Peak B were identified as *Gl*CP2.

dase (Fig. 6*B*). The exact sequence at the cleavage site has not been experimentally determined, but the 26-kDa fragment is clearly the cyst wall building block (4, 7, 11). *Rattus norvegicus* cathepsin C was also tested and no processing of rCWP2 was seen in the presence of this enzyme (Fig. 6*B*). At high concentrations of r*Gl*CP2, trypsin, or chymotrypsin the 26-kDa fragment of rCWP2 is further degraded to small peptides (data not shown). Interestingly, only r*Gl*CP2 Peak B and not Peak A from the anion exchange column exhibited proteolytic activity against rCWP2 in the time frame of this assay (Fig. 6*B*). rCWP2 processing was inhibited by K11777, an endopeptidase inhibitor (17), but not by Y01, a cathepsin C selective inhibitor (18) (Fig. 6*C*).

TABLE 2

Inhibition of r*Gl*CP2 activity against the N-terminally blocked fluorogenic peptides substrate Z-FR-AMC and Z-RR-AMC

Activity is expressed as percent activity relative to a control reaction.

Inhibitor	% Activity	
	Z-FR-AMC	Z-RR-AMC
PMSF 1 mM	113	93
EDTA 10 mM	47	40
Aprotinin 10 μg/ml	93	31
Pepstatin A 1 μM	100	112
Pepstatin A 10 μM	118	126
Leupeptin 1 μM	1	14
Leupeptin 10 μM	0	1
TLCK 1 μM	38	26
TLCK 10 μM	17	9
TPCK 1 μM	50	16
TPCK 10 μM	25	9
E64 1 μM	0	0
E64 10 μM	0	0
CA074 1 μM	77	99
CA074 10 μM	34	85
Lactacystin 1 μM	100	101
Lactacystin 10 μM	119	114
α1antitrypsin 1 μM	90	130
α1antitrypsin 10 μM	74	146
ALLN[a] 1 μM	5	10
ALLN 10 μM	5	5
ALLM[b] 1 μM	6	13
ALLM 10 μM	5	5

[a] ALLN: *N*-Acetyl-Leu-Leu-Nle-CHO.
[b] ALLM: *N*-Acetyl-Leu-Leu-Met-CHO.

DISCUSSION

Because *G. lamblia* is an early diverging branch of the eukaryotic evolutionary tree, as defined by 16 S ribosomal RNA sequence and protein coding sequences, it is an intriguing model system to investigate the evolution of protein families and their functions (2, 12, 19). The clan CA cysteine protease family has essential functions in numerous organisms including well characterized lysosomal protein degradation and a wide array of other indispensable cellular tasks (20). There are twenty-seven gene sequences for the *Giardia* clan CA cysteine protease family. Twenty-five of these genes are expressed in vegetative and encysting life cycle stages. Of these expressed genes, *Gl*CP2 emerges as the most highly expressed cysteine protease and exhibits developmental regulation, with expression increasing dramatically during encystation (Fig. 1). It is also the only cysteine protease that was identified in protein fractions enriched for cysteine protease activity.

*Gl*CP2 is a cathepsin B-like cysteine protease (10). The protein is produced as a zymogen and is activated by proteolytic removal of an N-terminal propeptide of 51 amino acids. *Giardia* cathepsin B-like cysteine proteases lack the "occluding loop" that is characteristic of cathepsin B-like enzymes of higher eukaryotes (8). This loop endows the protease with exopeptidase activity by stabilizing the free carboxyl at the C terminus of a peptide substrate (21). Therefore, the *Giardia* cathepsin B-like enzyme exhibits only endopeptidase activity (data not shown). Though the mammalian orthologues of this

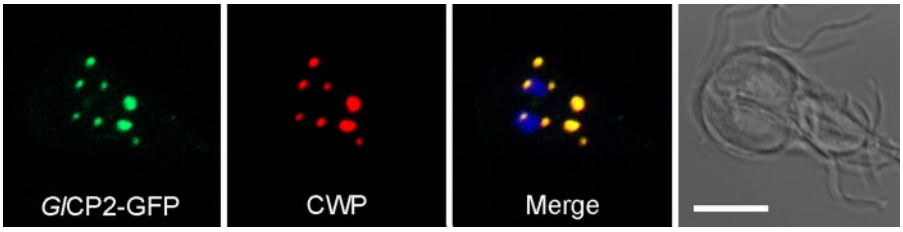

FIGURE 5. **GlCP2 co-localized with cyst wall protein in encystation-specific vesicles (ESVs) during *Giardia* encystation.** A C-terminal GFP fusion of *Gl*CP2 (*green*) was expressed episomally in *Giardia*. Cells were fixed 24 h after induction of encystation and probed with an antibody against cyst wall proteins (*red*). Image represents a three-dimensional projection of confocal images taken along the Z axis. *Gl*CP2 clearly co-localized with CWP in ESVs. *Bar*, 5 μm.

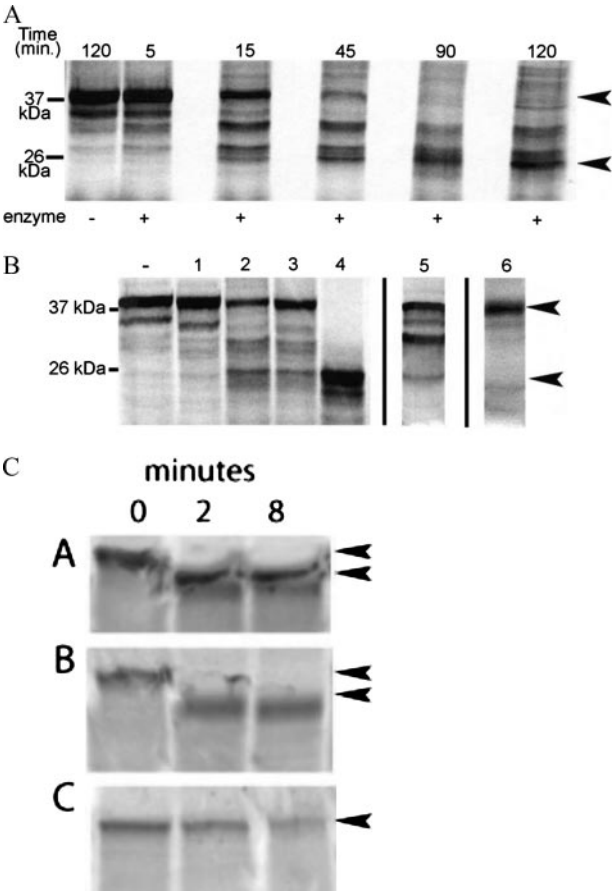

FIGURE 6. **r*Gl*CP2 processes recombinant cyst wall protein 2 (rCWP2) to the predicted 26-kDa size necessary for incorporation into the cyst wall.** *A*, in-solution cleavage of ^{35}S-labeled rCWP2 by purified r*Gl*CP2. rCWP2 was exposed to r*Gl*CP2 in Tris buffer, pH 7.6, and incubated at 25 °C for 0, 5, 15, 45, 90, and 120 min r*Gl*CP2 can cleave the 39-kDa rCWP2 to a 26-kDa fragment (*arrowheads*) in a time-dependent manner. *B*, rCWP2 was incubated for 35 min with various proteases. Peak A from anion exchange purification of r*Gl*CP2 (*lane 1*) and *R. norvegicus* cathepsin C (*lane 6*) did not process rCWP2 from its full-length size of 39 kDa (*arrowhead*). r*Gl*CP2 (*lane 2*) and *Giardia* lysates (*lane 3*) exhibited identical processing patterns of rCWP2. Other endopeptidases, such as trypsin (*lane 4*) and chymotrypsin (*lane 5*), can also process rCWP2 to 26 kDa (*arrowhead*). *Lanes 5* and *6* are set apart as these samples were run on separate gels. *C*, an endopeptidase inhibitor, K11777 (*panel C*) (17) blocks processing of rCWP2, whereas an inhibitor of cathepsin C, Y01, (18) does not (*panel B*). *Panel A* is an uninhibited control.

enzyme are lysosomal enzymes and are optimally active at an acidic pH, *Giardia* cysteine protease activity, and in particular the activity of *Gl*CP2, exhibits optimal substrate degradation at neutral pH (Fig. 3*B*) (20, 22, 23).

During purification of enzyme activity, in both native lysates and with r*Gl*CP2, a unique specificity was seen in fractions containing both the pro enzyme and the mature enzyme not observed in those fractions containing only mature *Gl*CP2. As seen in Fig. 4*B*, Peak A contains predominantly the proform of the protease and exhibits much higher activity against Z-RR-AMC than against Z-FR-AMC. Peak B, which has more mature protease, efficiently cleaves CWP2 (Fig. 6). There are two mechanistic considerations to this substrate specificity difference based on the crystal structure of the proform of homologous cathepsin L proteases (24). The proform of cathepsin L contains a peptide segment that binds to the active site of the enzyme in reverse orientation to substrate. The presence of this inhibitory segment in the proform, therefore, explains the inability of the proform in Peak A to cleave protein substrates like CWP. However, proforms of cathepsin L can "breathe" in solution whereby portions of the prodomain may be in equilibrium between bound and unbound with the active site. Thus small peptide substrates and active sites tags have been reported to be bound by proforms of cathepsins (14). While foci of the prodomain inhibitory peptide can therefore release from the active site, the rest of the segment remains in place, sterically hindering the approach of large protein substrates such as CWP.

As to the difference in substrate specificity between Z-RR-AMC and Z-FR-AMC between the two peaks, there are two possible explanations. First, it is common for the proforms of cathepsin L-like cysteine proteases to auto-activate at lower pH. This is presumably due to disruption of interactions between the prodomain and the active site with increase in hydrogen ion concentration. The Z-RR-AMC substrate may be more likely to produce disruption of the prodomain-enzyme interaction than the less charged Z-FR-AMC. Alternatively, the presence of the proform peptide in the active site may induce different conformation from mature protease whereby a negative charge is now present in the bottom of the S2 pocket favoring the binding of the Z-RR-AMC substrate. This conformational difference in S2 specificity was reported in crystal structures of the related cathepsin L, cruzain, when S2 pocket conformation was observed at different pH values (25).

Previously it was reported that a cathepsin C-like enzyme, encystation-specific cysteine protease (ESCP) was responsible for the essential proteolytic processing of CWP2 from a 39-kDa protein to 26-kDa fragment. This processing step removes a highly basic C-terminal domain, allowing polymerization and formation of the cyst wall (4, 11, 26). However, ESCP has all of the conserved domains of cathepsin C-like proteins including the N-terminal exclusion domain that limits cathepsin C to dipeptidyl exopeptidase activity (20, 22, 23, 27, 28). Therefore, it would not be predicted to possess any endopeptidase activity, such as would be necessary to accomplish CWP2 processing. Recombinant *R. nor-*

vegicus cathepsin C, with the same fold and conserved active site residues as ESCP, did not process CWP2 (Fig. 6*B*). Furthermore, Y01, a specific inhibitor of *Giardi*a cathepsin C does not inhibit cyst wall processing whereas an inhibitor of *G1*CP2 does (Fig. 6*C*). Because Touz *et al.* used ESCP purified from *Giardia* lysates and not recombinant protein, the possibility exists that the enzyme preparation was contaminated with one of the other much more abundant *Giardia* clan CA cysteine endopeptidases. In this study we show that *Gl*CP2 is capable of processing CWP2 to the expected 26-kDa fragment and is found in ESVs with cyst wall protein. *Gl*CP2 is expressed at levels 20-fold higher than ESCP during encystation (Fig. 1) so a small amount of the active *Gl*CP2 could easily contaminate the "purified" preparation of ESCP (7). Indeed, the protease inhibitors demonstrated by Touz *et al.* to interfere with cyst production (E64, ALLN, ALLM) (11) also efficiently inhibit *Gl*CP2 activity (Table 2). The ability of trypsin and chymotrypsin, but not cathepsin C, to accurately process CWP2 suggests that it is the structure of CWP2 that presents specific endoprotease processing sites. The exact cleavage site in *Giardia* has not been experimentally determined. The processing of CWP2 may be redundant, as other cysteine endopeptidases in *Giardia* were localized to the ESVs during encystation, such as EAA37074 (supplemental Fig. S3). However, the high level of *Gl*CP2 expression strongly suggests that *Gl*CP2 is a key proteolytic constituent of the CWP2 processing machinery.

During the proteolytic processing of rCWP by recombinant GlCP2 (Fig. 6*A*), several intermediate products are visualized. This is not unexpected given the multiple potential cleavage sites observed in the segment of CWP, processed to the 26-kDa form. There are at least 7 sites that correspond to the optimal substrate specificity of *Gl*CP2 which is hydrophobic side chains in P2 (F,V,I,L,M) and a positive charge at P1(R,K).[4]

The fact that *Gl*CP2 can, at high concentrations, degrade CWP2 to small peptides suggests that there must be a mechanism in place for regulating the activity of *Gl*CP2 against CWP2 in the ESVs. This may be accomplished by delayed activation of *Gl*CP2, because it was found that the zymogen-containing fractions of r*Gl*CP2 did not exhibit any proteolytic activity against CWP2 in the degradation assay (Fig. 6*B*). It could also be regulated by acidification of the ESVs, as it has been suggested that ESVs fuse with PVs prior to formation of the cyst wall and the activity of *Gl*CP2 toward protein substrates is greatly reduced in an acidic compartment (11) (Fig. 3*B*). The ability of *Gl*CP2 to completely degrade CWP2 under other conditions, as would be expected if *Gl*CP2 were released from ESVs into the extracellular space between the trophozoite and the cyst wall supports a second role for this enzyme in the process of excystation, as was previously postulated by Ward *et al.* (10).

[4] K. N. Dubois (2007) Thesis, University of California.

Acknowledgments—We thank Drs. C. C. Wang, Srini Garlapati, and Lei Li for reporter constructs (UCSF). We thank John Pederson (Unizyme, Denmark) for the gift of the R. norvegicus cathepsin C. We thank Katalin Medzihradsky at the UCSF mass spectrometry facility for the mass spectrometry data, and Chris Franklin (UCSF) for assistance in generating some of the figures.

REFERENCES

1. Adam, R. D. (2001) *Clin. Microbiol. Rev.* **14,** 447–475
2. Hedges, S. B. (2002) *Nat. Rev. Genet.* **3,** 838–849
3. Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., and Peattie, D. A. (1989) *Science* **243,** 75–77
4. Lujan, H. D., Mowatt, M. R., and Nash, T. E. (1997) *Microbiol. Mol. Biol. Rev.* **61,** 294–304
5. Hehl, A. B., Marti, M., and Kohler, P. (2000) *Mol. Biol. Cell* **11,** 1789–1800
6. Reiner, D. S., McCaffery, M., and Gillin, F. D. (1990) *Eur. J. Cell Biol.* **53,** 142–153
7. Lujan, H. D., Mowatt, M. R., Conrad, J. T., Bowers, B., and Nash, T. E. (1995) *J. Biol. Chem.* **270,** 29307–29313
8. Sajid, M., and McKerrow, J. H. (2002) *Mol. Biochem. Parasitol.* **120,** 1–21
9. McKerrow, J. H. (1999) *Int. J. Parasitol.* **29,** 833–837
10. Ward, W., Alvarado, L., Rawlings, N. D., Engel, J. C., Franklin, C., and McKerrow, J. H. (1997) *Cell* **89,** 437–444
11. Touz, M. C., Nores, M. J., Slavin, I., Carmona, C., Conrad, J. T., Mowatt, M. R., Nash, T. E., Coronel, C. E., and Lujan, H. D. (2002) *J. Biol. Chem.* **277,** 8474–8481
12. Singer, S. M., Yee, J., and Nash, T. E. (1998) *Mol. Biochem. Parasitol.* **92,** 59–69
13. Abel, E. S., Davids, B. J., Robles, L. D., Loflin, C. E., Gillin, F. D., and Chakrabarti, R. (2001) *J. Biol. Chem.* **276,** 10320–10329
14. Greenbaum, D., Medzihradszky, K. F., Burlingame, A., and Bogyo, M. (2000) *Chem. Biol.* **7,** 569–581
15. Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Bromme, D., Ellman, J. A., and Craik, C. S. (2006) *J. Biol. Chem.* **281,** 12824–12832
16. Mowatt, M. R., Lujan, H. D., Cotten, D. B., Bowers, B., Yee, J., Nash, T. E., and Stibbs, H. H. (1995) *Mol. Microbiol.* **15,** 955–963
17. Jacobsen, W., Christians, U., and Benet, L. Z. (2000) *Drug. Metab. Dispos.* **28,** 1343–1351
18. Yuan, F., Verhelst, S. H., Blum, G., Coussens, L. M., and Bogyo, M. (2006) *J. Am. Chem. Soc.* **128,** 5616–5617
19. Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K., and Hasegawa, M. (1994) *Mol. Biol. Evol.* **11,** 65–71
20. Turk, B., Turk, D., and Turk, V. (2000) *Biochim. Biophys. Acta* **1477,** 98–111
21. Mort, J. S., and Buttle, D. J. (1997) *Int. J. Biochem. Cell Biol.* **29,** 715–720
22. Turk, V., Turk, B., and Turk, D. (2001) *EMBO J.* **20,** 4629–4633
23. Mottram, J., North, M., Sajid, M., Edit, Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (1998) *Trichomonad and Giardia Cysteine Endopeptidases, Handbook of Proteolytic Enzymes*, pp. 1170–1173, Academic Press, San Diego
24. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J.* **15,** 5492–5503
25. Gillmor, S. A., Craik, C. S., and Fletterick, R. J. (1997) *Protein Sci.* **6,** 1603–1611
26. Gillin, F. D., Reiner, D. S., Gault, M. J., Douglas, H., Das, S., Wunderlich, A., and Sauch, J. F. (1987) *Science* **235,** 1040–1043
27. Que, X., Engel, J. C., Ferguson, D., Wunderlich, A., Tomavo, S., and Reed, S. L. (2007) *J. Biol. Chem.* **282,** 4994–5003
28. Turk, D., Janjic, V., Stern, I., Podobnik, M., Lamba, D., Dahl, S. W., Lauritzen, C., Pedersen, J., Turk, V., and Turk, B. (2001) *EMBO J.* **20,** 6570–6582