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

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

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 continue 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 worldwide (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). 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 excystation. 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-

The abbreviations used are: ESVs, encystation-specific vesicles; CWP2, cyst wall protein 2; ESCP, encystation-specific cysteine protease; GAPDH, glutaraldehyde phosphate dehydrogenase; GICP2, *G. lamblia* cysteine protease 2; DTT, dithiothreitol; AMC, amino-methylcoumarin; Z, benzoxycarbonyl; GFP, green fluorescent protein.
seven genes and found that *G. lamblia* cysteine protease 2 (GlCP2) 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*—*W.* 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 plasmid DNA was subcloned in the XhoI/XbaI site of 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 × 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 P. pastoris*—The GlCP2 gene was re-synthesized to optimize for yeast codon usage (DNA 2.0). The rGlCP2 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 GlCP2pPicF: CTCGAGAAAGACATCATCATCATCATCATGAGTTGAATCATAT- and the reverse primer GlCP2pPicR: TCTAGATTACTCATGAAAAATTCCAGTATGCC. The 920-bp amplicon was subcloned in the Xhol/XbaI site of the *P. pastoris* expression vector pPic2ZαB (Invitrogen). The plasmid was linearized by digestion with Sal and introduced into *P. pastoris* by electroporation (GenePulser XCell, Bio-Rad) according to the manufacturer’s specifications. Transformants were screened by growth on YPD and 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.

*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: TCTAGATTGGCTTGCCTGCCACCGAGG and CWP2pMalR: GCGCCGGCTTTAATGATGATGA-TGATGCTCTCCCTGATCTTCCTGCGACATAG and inserted into the NotI/XbaI site of the expression vector pCMVhNT (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). [35S]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-arginyln-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.

rGlCP2 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 (radioiodinated 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 GlCP2, 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).

rGlCP2 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.

rGlCP2 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% Me₂SO (from the substrates) or in buffer with NaOAc replacing Tris-HCl (pH 5.5) as referred to 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 Mx3005pTM 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).**

**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 Micro Imaging, 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-GiCP2 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-butanoate) (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).**

**Protease Activity—**Biochemical characterization of total cysteine protease activity found in Giardia lysates was concurrently undertaken to complement the gene expression analysis.
**Giardia CP2 Plays a Role in Encystation**

*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. 2A). 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. 2B). 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 GlCP2, of which peptides were identified in both Peak A and Peak B (Table 1). This was consistent with the observation that GlCP2 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 GlCP2, a resynthesized GlCP2 (rGlCP2) gene (resynthesized to optimize for yeast codon bias) of 34 kDa was expressed heterologously in *P. pastoris* (Fig. 3A). The polyhistidine-tagged rGlCP2 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. 3A). The full-length and mature forms of rGlCP2 had activity on a 10% gelatin zymogram native gel and migrated to an apparent mobility of 60-kDa marker (supplemental Fig. S1).

To compare rGlCP2 activity to that predominantly seen in *Giardia* lysates, the activity profile of rGlCP2 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. 4A). 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. 4B). To determine if full-length rGlCP2 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. 4B). A Western blot of these fractions using an antibody against GlCP2 also demonstrates that two bands in Peak A and one in Peak B are identified as GlCP2 (Fig. 4C). These data are consistent with the biochemical and mass spectrometry evidence that GlCP2 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 rGlCP2 activity against Z-FR-AMC and Z-RR-AMC. Leupeptin and E64 were the most effective inhibitors of rGlCP2 (Table 2). The pH profile of rGlCP2 was elucidated using the peptide substrates Z-FR-AMC and Z-RR-AMC and the macromolecular substrate casein-resorufin. The pH optimum for rGlCP2 was found to be in the neutral range for each of these substrates (Fig. 3B). This is consistent with the localization of GlCP2 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 GlCP2 were found to be 40 μM and 17.5 μM/sec, respectively. The $K_m$ and $k_{cat}/K_m$ of Z-RR-AMC for GlCP2 were found to be 9 μM and 72 μM/sec, respectively.

To characterize the substrate specificity of rGlCP2, 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). rGlCP2 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
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 $\mu$M and 1,473 $\mu$/M/s, respectively.

<table>
<thead>
<tr>
<th>Peak A</th>
<th>Peak B</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVAGLDK</td>
<td>CVAGLDK</td>
</tr>
<tr>
<td>TGGTDECVPYK</td>
<td>TGGTDECVPYK</td>
</tr>
<tr>
<td>VHLATATSYK</td>
<td>DYGDLIPAMMK</td>
</tr>
<tr>
<td>GINDCSHEREQAYAGPD</td>
<td>GINDCSHEREQAYAGPD</td>
</tr>
<tr>
<td>NSWQPDEGDGYFR</td>
<td>NSWQPDEGDGYFR</td>
</tr>
<tr>
<td>SQITTLR</td>
<td>GTCPYK</td>
</tr>
<tr>
<td>GTCPTK</td>
<td>CADGSSK</td>
</tr>
</tbody>
</table>

**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.

**Giardia CP2 Plays a Role in Encystation**

GiCP2 is found in *Giardia* ESVs and can proteolytically process CWP2 to the predicted size found in the cyst wall—The localization of GiCP2 during encystation was determined by episomal expression of a GiCP2-GFP fusion in *Giardia*. Encysting cells were probed with an antibody against cyst wall protein to highlight the ESVs, and GiCP2-GFP was found to localize to ESVs (Fig. 5). Ward et al. (10) previously implicated GiCP2 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 rGiCP2 was incubated with recombinant CWP2 (rCWP2). In the presence of either *Giardia* lysates or rGiCP2, 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-
dase (Fig. 6B). 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. 6C). rCWP2 processing was inhibited by K11777, an endopeptidase inhibitor (17), but not by Y01, a cathepsin C selective inhibitor (18) (Fig. 6C).

**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, GICP2 emerges as the most highly expressed cysteine protease (4, 11, 17, 27). rCWP2 is further degraded to small peptides (data not shown). Interestingly, only rGICP2 Peak B and not Peak A from the anion exchange column exhibited proteolytic activity against rCWP2 in the time frame of this assay (Fig. 6B). rCWP2 processing was inhibited by K11777, an endopeptidase inhibitor (17), but not by Y01, a cathepsin C selective inhibitor (18) (Fig. 6C).

**TABLE 2**

Inhibition of rGICP2 activity against the N-terminally blocked fluorogenic peptides substrate Z-FR-AMC and Z-RR-AMC

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Z-FR-AMC</th>
<th>Z-RR-AMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMSF 1 mM</td>
<td>113</td>
<td>92</td>
</tr>
<tr>
<td>EDTA 10 mM</td>
<td>47</td>
<td>40</td>
</tr>
<tr>
<td>Aprotinin 10 μg/ml</td>
<td>93</td>
<td>31</td>
</tr>
<tr>
<td>Pepstatin A 1 μM</td>
<td>100</td>
<td>112</td>
</tr>
<tr>
<td>Pepstatin A 10 μM</td>
<td>118</td>
<td>126</td>
</tr>
<tr>
<td>Leupeptin 1 μM</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Leupeptin 10 μM</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TLCK 1 μM</td>
<td>38</td>
<td>26</td>
</tr>
<tr>
<td>TLCK 10 μM</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>TPCK 1 μM</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>TPCK 10 μM</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>E64 1 μM</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E64 10 μM</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CA074 1 μM</td>
<td>77</td>
<td>99</td>
</tr>
<tr>
<td>CA074 10 μM</td>
<td>34</td>
<td>85</td>
</tr>
<tr>
<td>Lactacystin 1 μM</td>
<td>100</td>
<td>101</td>
</tr>
<tr>
<td>Lactacystin 10 μM</td>
<td>119</td>
<td>114</td>
</tr>
<tr>
<td>αlantitrypsin 1 μM</td>
<td>90</td>
<td>130</td>
</tr>
<tr>
<td>αlantitrypsin 10 μM</td>
<td>74</td>
<td>146</td>
</tr>
<tr>
<td>ALLN* 1 μM</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>ALLN 10 μM</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>ALLN* 1 μM</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>ALLN 10 μM</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

* ALLN: N′-Acetyl-Leu-Leu-Nle-CHO.
* ALLM: N′-Acetyl-Leu-Leu-Met-CHO.
Giardia CP2 Plays a Role in Encystation

During purification of enzyme activity, in both native lysates and with rGICP2, a unique specificity was seen in fractions containing both the pro enzyme and the mature enzyme not observed in those fractions containing only mature GICP2. As seen in Fig. 4B, Peak A contains predominantly the pro-form 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. norv...
Giardia CP2 Plays a Role in Encystation

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 Medzhiradszky at the UCSF mass spectrometry facility for the mass spectrometry data, and Chris Franklin (UCSF) for assistance in generating some of the figures.

REFERENCES


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

Kelly N. DuBois, Marla Abodeely, Judy Sakanari, Charles S. Craik, Malinda Lee, James H. McKerrow and Mohammed Sajid

doi: 10.1074/jbc.M802133200 originally published online April 29, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M802133200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2008/04/30/M802133200.DC1

This article cites 27 references, 13 of which can be accessed free at http://www.jbc.org/content/283/26/18024.full.html#ref-list-1