Cryptopain-1, a cysteine protease of *Cryptosporidium parvum*, does not require the pro-domain for folding

B.-K. NA[1]*, J.-M. KANG[1], H.-I. CHEUN[2], S.-H. CHO[2], S.-U. MOON[2], T.-S. KIM[3] *and* W.-M. SOHN[1]

[1] *Department of Parasitology and Institute of Health Sciences, College of Medicine, Gyeongsang National University, Jinju 660-751, Korea*
[2] *Department of Malaria and Parasitic Diseases, National Institute of Health, Centers for Disease Control and Prevention, Seoul 122-701, Korea*
[3] *Department of Parasitology and Inha Research Institute for Medical Sciences, Inha University College of Medicine, Incheon 400-751, Korea*

(*Received 21 August 2008; revised 24 October 2008; accepted 7 November 2008; first published online 18 December 2008*)

SUMMARY

Cryptosporidium parvum is an intracellular protozoan parasite that causes cryptosporidiosis in mammals including humans. In the current study, the gene encoding the cysteine protease of *C. parvum* (cryptopain-1) was identified and the biochemical properties of the recombinant enzyme were characterized. Cryptopain-1 shared common structural properties with cathepsin L-like papain family enzymes, but lacked a typical signal peptide sequence and contained a possible transmembrane domain near the amino terminus and a unique insert in the front of the mature domain. The recombinant cryptopain-1 expressed in *Escherichia coli* and refolded to the active form showed typical biochemical properties of cathepsin L-like enzymes. The folding determinant of cryptopain-1 was characterized through multiple constructs with or without different lengths of the pro-domain of the enzyme expressed in *E. coli* and assessment of their refolding abilities. All constructs, except one that did not contain the full-length mature domain, successfully refolded into the active enzymes, suggesting that cryptopain-1 did not require the pro-domain for folding. Western blot analysis showed that cryptopain-1 was expressed in the sporozoites and the enzyme preferentially degraded proteins, including collagen and fibronectin, but not globular proteins. This suggested a probable role for cryptopain-1 in host cell invasion and/or egression by the parasite.

Key words: *Cryptosporidium parvum*, cysteine protease, cryptopain-1, pro-domain, folding.

INTRODUCTION

Cryptosporidium parvum is an intracellular protozoan parasite responsible for the diarrhoeal illness, cryptosporidiosis, in a number of mammals, including humans (Priest *et al*. 2001; Tzipori and Ward, 2002). *C. parvum* infection is characterized by mucosal injury and villous atrophy with infiltration of the lamina propria by inflammatory cells. Most cases of human infection usually involve self-limiting diarrhoea in immunocompetent or healthy individuals; however, immunocompromised individuals, most notably those with acquired immunodeficiency syndrome (AIDS), can develop prolonged severe diarrhoea that is frequently fatal (Peterson, 1992). Although intensive studies into the biology and biochemistry of the parasite have been conducted in recent years, there are currently no effective vaccines or treatment regimens for *C. parvum* infection (Chen *et al*. 2002). This situation emphasizes the importance of the development of an effective vaccine and/or approach to anti-cryptosporidial chemotherapy. One of the approaches being undertaken for vaccine and chemotherapeutic agent development is the identification and characterization of target molecules with essential roles in the parasite's life cycle.

Cysteine proteases of apicomplexa protozoa, including *Plasmodium* spp. and *Toxoplasma gondii*, play essential roles in the life cycles of these parasites. They appear to be involved in a number of important biological functions, including protein processing, host cell invasion, and nutrient uptake (Shenai *et al*. 2000; Sijwali *et al*. 2001; Que *et al*. 2002; Rosenthal, 2002; Kim *et al*. 2004; Na *et al*. 2004). Cysteine protease inhibitors effectively block the development of the parasites and inhibit host cell invasion (Rosenthal *et al*. 2002; Shaw *et al*. 2002; Teo *et al*. 2007), suggesting that cysteine proteases of the parasites are appropriate new chemotherapeutic targets. Cysteine protease activities have been identified in several developmental stages of *C. parvum*, and are likely to be associated with the excystation and host cell invasion by the parasite (Nesterenko *et al*. 1995; Forney *et al*. 1996); however, the biological and biochemical

* Corresponding author: Department of Parasitology and Institute of Health Sciences, College of Medicine, Gyeongsang National University, Jinju 660-751, Korea. Tel: +82 55 751 8822. Fax: +82 55 759 4022. E-mail: bkna@gnu.ac.kr

Parasitology (2009), **136**, 149–157.
doi:10.1017/S0031182008005350 Printed in the United Kingdom

features of the enzymes are poorly understood. Therefore, improved characterization of cryptosporidial cysteine proteases is necessary for an in-depth understanding of the biology of *C. parvum*, as well as for further investigation of the molecules as chemotherapeutic targets for treatment of cryptosporidiosis.

In the current study, we identified a gene encoding a cysteine protease of *C. parvum*, designated as cryptopain-1, and characterized the biochemical properties of the recombinant protein. Cryptopain-1 was shown to share typical structural and biochemical properties with cathepsin L-like cysteine proteases, but is quite different from other related enzymes in that it does not require a pro-domain for folding. The preferential hydrolysis of collagen and fibronectin suggested a probable biological role for cryptopain-1 in host cell invasion and/or egress by the parasite.

MATERIALS AND METHODS

Identification and cloning of the gene encoding for cryptopain-1

By data-mining the recently updated *C. parvum* genome sequence database (CryptoDB, http://cryptodb.org), we identified a gene encoding a cysteine protease, cryptopain-1 (Gene ID: cgd6_4880). The open reading frame of cryptopain-1 was amplified with forward (5′-ATGGACATA-GGAAACA-ACGTGGAAGAACAT-3′) and reverse (5′-TTA-TATTGATTGAT-TAATCACTGGATACAC-3′) primers with *Taq* DNA polymerase (Takara, Otsu, Japan) and *C. parvum* genomic DNA. The amplified PCR product was gel-purified using a gel extraction kit (Qiagen, Valencia, CA, USA), ligated into the pGEM T-Easy vector (Promega, Madison, WI, USA), and then transformed into *Escherichia coli* DH5α. Sequencing reactions were conducted with a BigDye Terminator Cycle Sequencing Ready Reaction Kit in an ABI 377 automatic DNA sequencer (Applied Biosystems, Foster City, CA, USA). Analysis of the primary structure of the deduced amino acid sequences was conducted with DNASTAR (DNASTAR, Madison, WI, USA), TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0), and Signal P (http://www.cbs.dtu.dk/services/SignalP). TargetP1.1 (http://www.cbs.dtu.dk/services/TargetP) and PSORT (http://www.psort.org) were used to predict protein localization. The nucleotide sequence encoding cryptopain-1 is available in the GenBank database under Accession number DQ156545.

Expression, purification, and refolding of recombinant cryptopain-1

To express the recombinant cryptopain-1, a fragment harbouring a portion of the pro-domain and the entire mature domain of cryptopain-1 was amplified using the primers 5′-GGATCCGAGGAAAATC-AAAGATTTGAAATT-3′, which contained a 5′ *Bam* HI site, and 5′-CTGCAGTTATATTGA-TTGATTAATCACTGG -3′, which harboured a 5′ *Pst* I site. The PCR product was purified, ligated into the pGEM-T Easy vector (Promega), and transformed into *E. coli* DH5α. The resulting plasmid DNA was digested with the appropriate restriction enzymes and ligated into the pQE-30 expression vector (Qiagen), which was then transformed into *E. coli* M15 [pREP4] cells (Qiagen). Selected clones were grown and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). The cells were collected by centrifugation at 10 000 ***g*** for 15 min at 4 °C, and suspended in 8 M urea lysis buffer. The recombinant protein was purified by nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography (Qiagen). The purification and purity of the recombinant protein was monitored by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Refolding of the purified recombinant protein was performed as described previously (Na *et al.* 2008). In brief, 10 ml (1 mg/ml) of the affinity purified recombinant protein was slowly added to 1 litre of 100 mM Tris-HCl (pH 8·0) containing 1 mM ethylenediaminetetraacetic acid (EDTA), 250 mM L-arginine, 5 mM reduced glutathione (GST), and 1 mM oxidized glutathione (GSSG). After gentle stirring at 4 °C for 20 h, the sample was dialysed against 0·01 M Tris-HCl (pH 7·5) and concentrated with Centricon Plus (cut-off 10 kDa; Millipore, Billerica, MA, USA). The refolded cryptopain-1 was processed to form the fully active mature enzyme by adjusting the pH to 5·0 with 3·5 M sodium acetate (pH 2·6) containing 10 mM DTT and incubation at 37 °C for 2 h. The debris was removed by centrifugation and the pH of the sample was readjusted to 7·0 with 1 M Tris-HCl (pH 8·5). For substrate gel analysis, samples were mixed with SDS-PAGE sample buffer lacking 2-mercaptoethanol, and electrophoresed in SDS-polyacrylamide gel co-polymerized with 0·1 % gelatin. The gel was then washed twice with 2·5 % Triton X-100 at room temperature for 30 min, incubated in 100 mM sodium acetate (pH 5·5) supplemented with 10 mM dithiothreitol (DTT) overnight at 37 °C, stained with Coomassie Blue, and destained to identify the proteolytic activity as a clear band on the gel.

Enzyme assay and kinetic analysis

The enzyme activity was assayed fluorometrically as the hydrolysis of benzyloxycarbonyl-L-leucyl-L-arginine 4-methyl-coumaryl-7-amide (Z-LR-MCA; Peptide Institute, Osaka, Japan). Briefly, 10 μl of enzyme solution was added to 190 μl of sodium acetate buffer (pH 5·5) containing 5 μM Z-LR-MCA and 10 mM DTT, and the release of fluorescence (excitation, 355 nm; emission, 460 nm) over 20 min

https://doi.org/10.1017/S0031182008005350 Published online by Cambridge University Press

at room temperature was assessed with a Fluoroskan Ascent FL (Thermo, Vantaa, Finland). For kinetic analysis, the concentration of recombinant cryptopain-1 was determined by active site titration with *trans*-epoxy-succinyl-L-leucylamido(4-guanidino)-butane (E-64). The kinetic parameters of the enzyme for Z-LR-MCA, Z-L-phenylalaninyl-L-arginine-MCA (Z-FR-MCA) and Z-L-arginyl-L-arginine-MCA (Z-RR-MCA) were determined at room temperature using a constant amount of enzyme (25 nM) and varying concentrations of substrate in 100 mM sodium phosphate (pH 6·5) containing 10 mM DTT. Fluorogenic substrates were added and activities were compared in terms of fluorescence as a function of time. The kinetic constants *K*m and *V*max were determined using GraphPad software (GraphPad Software Inc., La Jolla, CA, USA).

Production of antibody for cryptopain-1 and Western blot analysis

Specific antibody against recombinant cryptopain-1 was produced by immunization of BALB/c mice with purified recombinant cryptopain-1 (50 μg) 3 times at 2-week intervals. Two weeks after the final inoculation, the mice were sacrificed and the sera were collected. The immunoglobulin G (IgG) fraction was isolated from the sera with a Protein G-Sepharose column (Amersham Biosciences, Piscataway, NJ, USA). For Western blot analysis, the sporozoites of *C. parvum* were prepared as previously described (Ctrnacta *et al.* 2006). The sporozoites (10^9 cells) were suspended in phosphate-buffered saline (PBS; pH 7·4) which contained a complete protease inhibitor cocktail (Roche, Mannheim, Germany), and were sonicated on ice using 5-sec pulses at 50% output power for up to 3 min in a W-380/385-series Sonicator Ultrasonic Liquid Processor (Misonix Inc., Farmingdale, NY, USA). The sonicated extract of *C. parvum* sporozoites (20 μg) was subjected to SDS-PAGE, and the proteins were transferred to a nitrocellulose membrane (Bio-Rad, Hercules, CA, USA). The membrane was blocked with PBS supplemented with 0·05% Tween 20 (PBST; pH 7·4) and 3% skim milk for 1 h. The membrane was then incubated with antibody raised against the recombinant cryptopain-1 which was diluted 1 : 1000 in PBST at room temperature for 2 h. After several washes with PBST, the membrane was incubated with 1 : 1000 diluted horseradish peroxidase-conjugated goat anti-mouse IgG (Sigma, St Louis, MO, USA). The blot was visualized with 4-chloro-1-naphthol (Sigma) and the reaction was stopped by washing the membrane with distilled water.

Biochemical properties of recombinant cryptopain-1

The optimal pH for the maximum activity of cryptopain-1 was assessed in sodium acetate (pH 4·0–5·5), sodium phosphate (pH 6·0–6·5), and Tris-HCl (pH 7·0–9·0). Recombinant cryptopain-1 (50 nM) was added to each pH buffer supplemented with 5 μM Z-LR-MCA and 10 mM DTT and the enzyme activity was measured as described above. For each pH, the appropriate blank was separately measured as a control. The requirement of reducing conditions for maximum activity of cryptopain-1 was determined by assay of the enzyme activity in 100 mM sodium phosphate (pH 6·5) with different concentrations of GSH, ranging from 0 to 20 mM. The pH stability of recombinant cryptopain-1 was also examined at pH 5·0, 6·5, or 8·0 by incubating the enzyme at 37 °C in the appropriate buffers.

N-terminal amino acid sequencing

Fully processed mature recombinant cryptopain-1 was separated by SDS-PAGE, transferred to a polyvinylidene difluoride (PVDF) membrane, stained with Coomassie Blue, and destained. The protein band was excised and sequenced by Edman sequencing at the Korea Basic Science Institute (Daejeon, Korea).

Requirement of the pro-domain for folding

To investigate the requirement of the pro-domain for the proper folding of cryptopain-1, constructs encoding for the cryptopain-1 mature domain and different lengths of the pro-domain were amplified by PCR using the following primers: −65E CP1 (5′-GGATCCGAGGAAAATCAAAGATTT-3′), −37E CP1 (5′-GGATCCGAAATGAATGAATTTGGT-3′), −10E CP1 (5′-GGATCCGAAAGGGTATTTAAGTCA-3′), CP1 (5′-GGATCCGCAAGCGAATCAGAAGAG-3′), and +14N CP1 (5′-GGATCCAATTGGGTGGAAGCTGGA-3′); in all cases the reverse primer was 5′-CTGCAGTTATATTGATTGATTAATC-3′. The DNA fragments were digested with *Bam* HI and *Pst* I, ligated into pQE-30 expression vector, and transformed into *E. coli* M15 [pREP4] competent cells. Each protein was expressed, purified, and refolded as described above. An enzyme assay and substrate gel analysis of each refolded enzyme was performed as described above.

Degradation of host proteins

Purified collagen (from human placenta), fibronectin (from human plasma), alpha2-macroglobulin (from human plasma), albumin (from human serum), haemoglobin (from human blood), IgA (from human colostrum), IgG (from human serum), and IgM (from human serum) were purchased from Sigma. Each protein (1 mg/ml) was incubated with cryptopain-1 (50 nM) in 50 mM sodium phosphate (pH 6·5) with 1 mM GST for 1 or 2 h at 37 °C. The reactions

https://doi.org/10.1017/S0031182008005350 Published online by Cambridge University Press

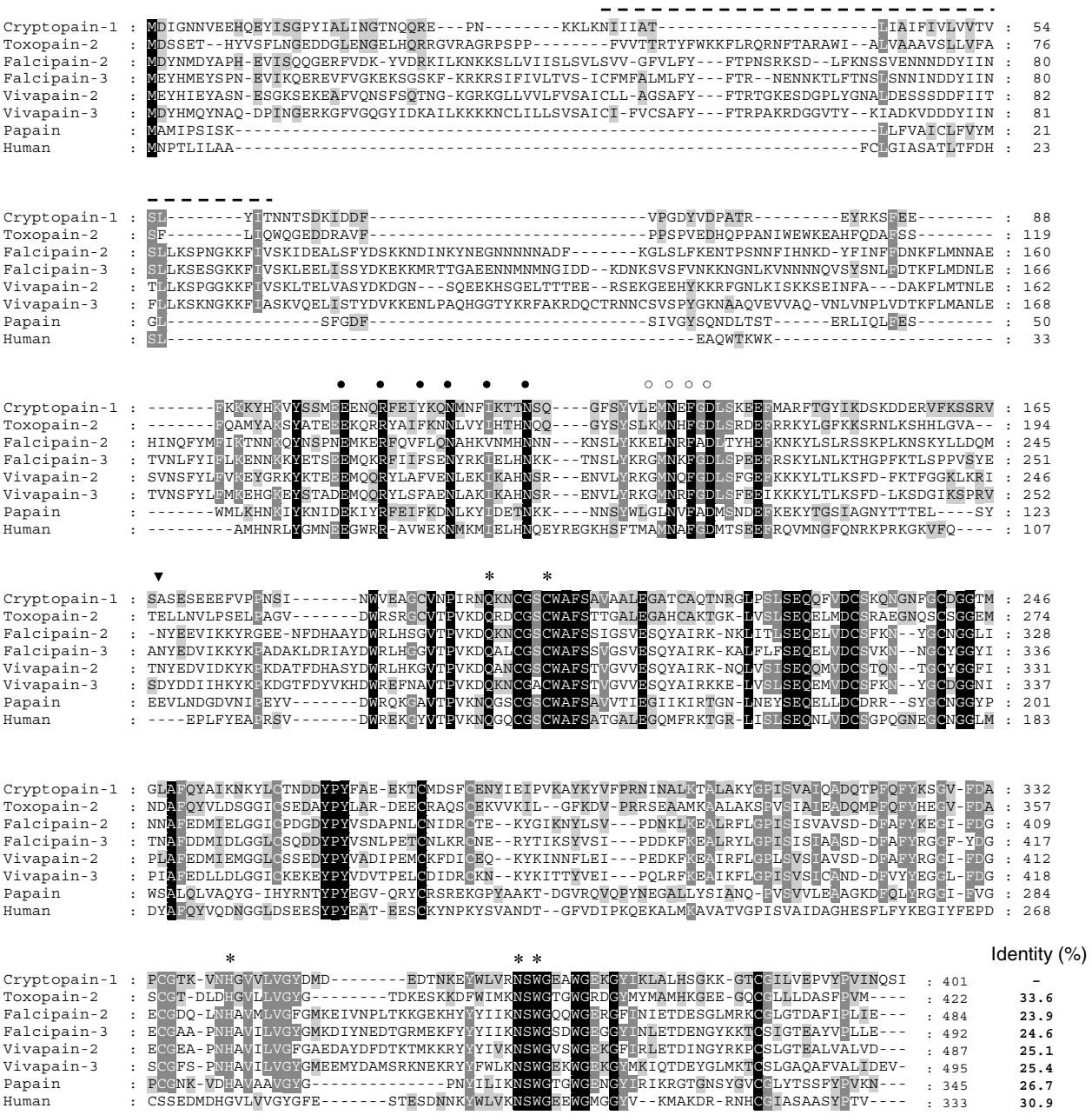


Fig. 1. Multiple sequence alignment. The deduced amino acid sequence of cryptopain-1 (GenBank Accession number DQ156545) was aligned with those of cysteine proteases; toxopain-2 (AY660746), falcipain-2 (AF239801), falcipain-3 (AF282974), vivapain-2 (AY208270), vivapain-3 (AY211736), papain (P00784), and human cathepsin L (M20496). Dashes represent gaps added to maximize alignment. The ERFNIN and GNFD motifs are indicated with closed and open circles, respectively. The dotted line on the cryptopain-1 sequence is the predicted transmembrane domain. Asterisks indicate conserved active-site residues. The arrow-head indicates the position of the mature domain processing site of cryptopain-1.

were terminated by adding reducing sample buffer and analysed by SDS-PAGE.

RESULTS

Cloning and molecular characterization of cryptopain-1

The gene encoding cryptopain-1 consisted of a 1206 bp sequence that encoded for 401 amino acids. The modified ERFNIN and GNFD motifs, characteristic of cathepsin L-like cysteine proteases, were identified in the pro-domain of the enzyme. The Q, C, H, N, and W residues, which are crucial for stabilization of either the thiolate-imidazolium ion pair or the transition states, were well conserved. The 6 cysteine residues that form disulfide bridges critical for maintenance of the tertiary structure were also well conserved (Fig. 1). Primary sequence analysis of cryptopain-1 revealed that a typical signal

https://doi.org/10.1017/S0031182008005350 Published online by Cambridge University Press

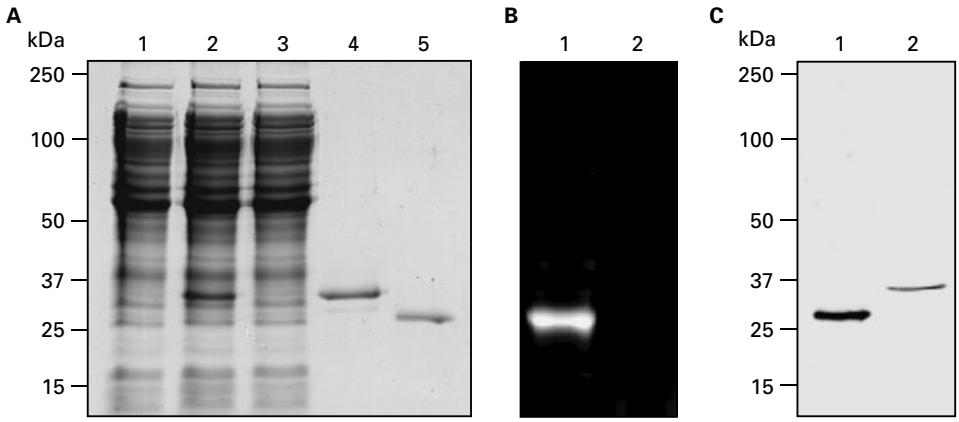


Fig. 2. Expression, purification and refolding of recombinant cryptopain-1. (A) Proteins were analysed by SDS-PAGE and stained with Coomassie Blue. Lane 1, *E. coli* lysate control; lane 2, IPTG-induced *E. coli* lysate; lane 3, unbound fraction of Ni-NTA affinity chromatography; lane 4, bound fraction of Ni-NTA affinity chromatography; and lane 5, fully activated mature recombinant cryptopain-1. (B) Fully activated mature recombinant cryptopain-1 was analysed by substrate gel supplemented with 0·1% gelatin as substrate. Lane 1, without E-64; and lane 2, with E-64 (10 μM). (C) Western blot analysis. Lane 1, recombinant cryptopain-1 (5 μg); lane 2, extract of *Cryptosporidium parvum* sporozoites (20 μg). The blot was probed with the antibody raised against recombinant cryptopain-1.

peptide sequence was absent, but it did contain a 23-amino acid hydrophobic stretch (37–59 amino acids) that was predicted by the PRED-TMR algorithm to represent a transmembrane domain (Pasquier *et al.* 1999). In addition, 4 putative N-glycosylation sites were identified (23NGT25, 60NNT62, 61NTS63, and 398NQS400).

Expression, purification, and refolding of recombinant cryptopain-1

A portion of the pro-domain and the entire mature domain of cryptopain-1 was amplified, cloned into the pQE-30 expression vector, and transformed into *E. coli*. The recombinant cryptopain-1 was expressed in *E. coli* as an insoluble protein with an apparent molecular weight of 34 kDa (Fig. 2A), which was consistent with the estimated molecular mass from the deduced amino acid sequence of cryptopain-1. The recombinant protein purified by Ni-NTA affinity chromatography and refolded under alkaline conditions was processed to the enzymatically-active species of approximately 26 kDa under acidic conditions and its activity was shown to be effectively inhibited by E-64 (Fig. 2A and B). The N-terminal amino acid residues, determined by N-terminal amino acid sequencing of the fully active mature recombinant cryptopain-1, were ASESE (Fig. 1). To determine the expression of cryptopain-1 in *C. parvum* sporozoites, Western blot analysis was performed. The specific antibody for recombinant cryptopain-1 was reacted with a protein with an approximate molecular weight of 35 kDa, which coincided well with the predicted molecular size of pro-form cryptopain-1, in the extract of *C. parvum* sporozoites (Fig. 2C).

Biochemical properties of recombinant cryptopain-1

Cryptopain-1 showed a broad pH optimal range (pH 5·5 to 7·0), with a maximum at pH 6·5 (Fig. 3A) and reducing conditions were required for maximum activity (Fig. 3B). The enzyme was relatively stable at neutral pHs, but relatively unstable in acidic conditions (Fig. 3C). The substrate specificity of cryptopain-1 was characterized with several fluorogenic peptide substrates. Cryptopain-1 readily hydrolysed Z-FR-MCA and Z-LR-MCA, but did not show activity against Z-RR-MCA (Fig. 3D).

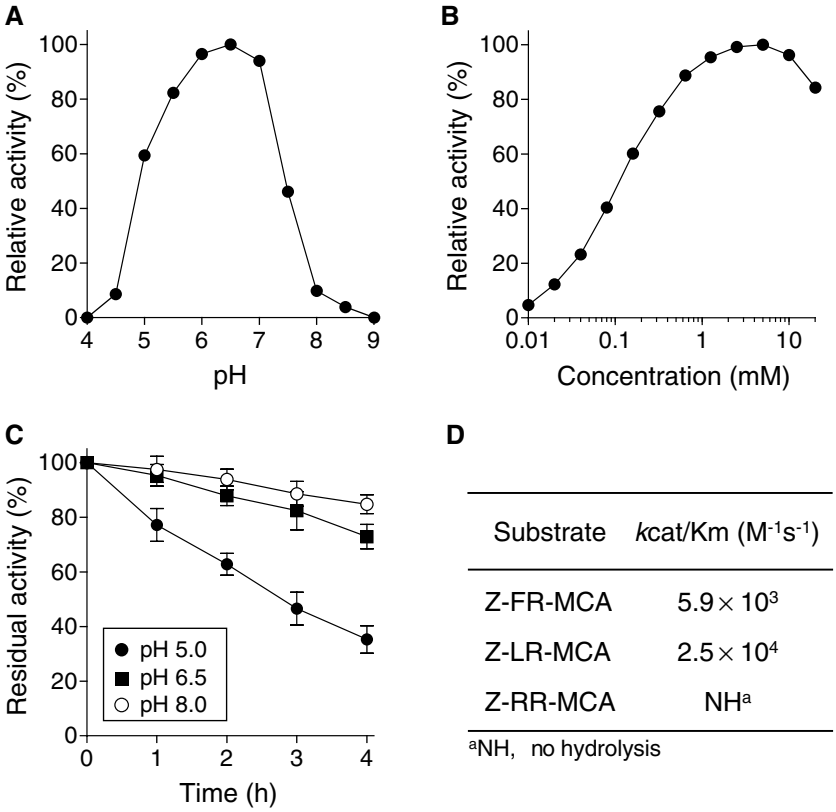
Cryptopain-1 does not require the pro-domain for folding

To better characterize the determinants of folding for cryptopain-1, multiple constructs with or without different lengths of the pro-domain of the enzyme were expressed in *E. coli* (Fig. 4A). All of the constructs were successfully expressed with the sizes expected (Fig. 4B, upper panel). Each recombinant protein was then refolded and activated in acidic pH conditions (Fig. 4B, lower panel). The constructs harbouring full-length mature domain (E1–E4) successfully refolded and showed proteolytic activities, but E5, which did not contain the full-length mature domain, did not show activity (Fig. 4C and 4D).

Degradation of various host proteins

To investigate the probable biological roles of cryptopain-1, its proteolytic activity was determined for several human proteins, including collagen, fibronectin, α2-macroglobulin, albumin, haemoglobin, IgA, IgG, and IgM. High levels of hydrolysis of fibronectin and collagen were observed with no significant

https://doi.org/10.1017/S0031182008005350 Published online by Cambridge University Press



Substrate	*k*cat/Km ($M^{-1}s^{-1}$)
Z-FR-MCA	5.9×10^3
Z-LR-MCA	2.5×10^4
Z-RR-MCA	NH[a]

[a]NH, no hydrolysis

Fig. 3. Characterization of biochemical properties of recombinant cryptopain-1. (A) Optimal pH. Enzyme activity was assayed in various pH buffers ranging from pH 4·0 to 9·0. Maximal activity was shown as 100%. (B) Requirement of reducing conditions. Enzyme activity was assayed in 100 mM sodium phosphate (pH 6·5) with different concentrations of GSH. (C) Enzyme stability. Recombinant cryptopain-1 was incubated in various pH buffers at 37 °C for the indicated time and residual enzyme activity was assayed. (D) Substrate kinetics. Kinetic parameters of cryptopain-1 for different fluorogenic substrates were determined.

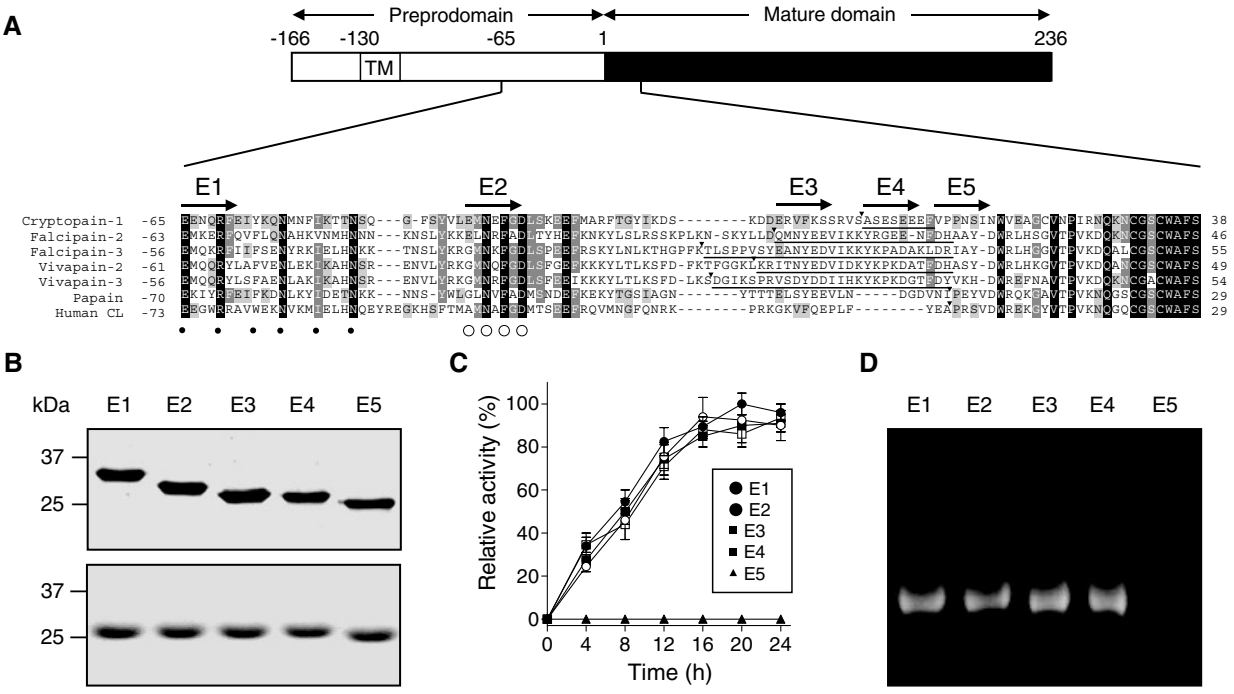


Fig. 4. Cryptopain-1 does not require the pro-domain for folding. (A) Domain organization of pre-pro-cryptopain-1. The schematic shows locations of a predicted transmembrane domain (TM), the mature domain processing site (arrowhead), and an unusual amino-terminal extension of the mature protease (underline). Sequence flanking mature

https://doi.org/10.1017/S0031182008005350 Published online by Cambridge University Press

hydrolytic activity for the other aforementioned protein substrates (Fig. 5).

DISCUSSION

In this study, we identified and characterized cryptopain-1, a cysteine protease of *C. parvum*. Cryptopain-1 is a typical papain-like enzyme which shares similar structural and biochemical characteristics with other cathepsin L-like enzymes, but its pH optimum is slightly neutral compared to other related enzymes. Interestingly, cryptopain-1 has some unusual features, including a large pro-domain, a putative transmembrane domain near the amino terminus, and some unique insertions in the front of the mature domain sequence, which are not found in typical papain-family enzymes. In general, papain-family cysteine proteases show highly similar structural properties, consisting of a signal peptide, a pro-peptide, and a catalytic domain representing the mature active enzyme. All of these proteases are initially synthesized as inactive pro-enzymes harbouring an N-terminal pro-domain which mediates proper folding of the enzymes into their active forms (Smith and Gottesman, 1989). Although the pro-domain sequences of the enzymes are more variable than the mature domain, the two regions of pro-domain (ERFNIN and GNFD motifs) are relatively well-conserved and play an important role in the processing and folding of these enzymes (Smith and Gottesman, 1989; Vernet *et al*. 1995; Yamamoto *et al*. 1999; Shinde and Inouye, 2000). The pro-domains have also been suggested to act as potent inhibitors of their cognate proteases to prevent the disastrous consequences of uncontrolled hydrolytic activity by the enzymes (Baker *et al*. 1992; Carmona *et al*. 1996; Coulombe *et al*. 1996; Guay *et al*. 2000). Mature enzymes were finally produced by cleavage at the mature protease cleavage site. Cleavage usually occurs immediately upstream of the sequence, *X*P where *X* represents a non-polar amino acid commonly about 25 amino acids upstream of the catalytic cysteine residue in typical papain-family

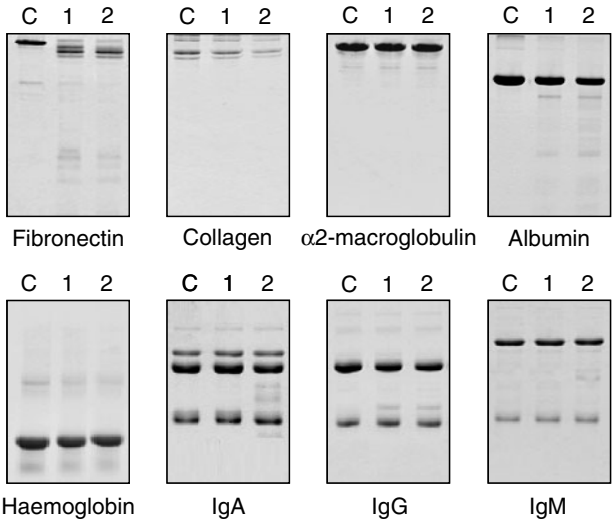


Fig. 5. Degradation of various human proteins by recombinant cryptopain-1. Each macromolecular protein was incubated with fully activated mature recombinant cryptopain-1 (50 nM) in 50 mM sodium phosphate (pH 6·5) in the presence of 1 mM GST for 1 or 2 h at 37 °C and analysed by SDS-PAGE. Lane C, protein without recombinant enzyme; lanes 1 and 2, incubated with recombinant enzyme for 1 and 2 h, respectively.

enzymes (Berti and Storer, 1995). The conserved proline was also found in cryptopain-1 in nearly the same upstream position as in all presently known papain-family proteases. However, the N-terminal amino acid sequencing analysis of fully activated mature cryptopain-1 showed that the enzyme contained an extra amino-terminal extension, with an additional 9 amino acids relative to other papain-family enzymes. This atypical N-terminal extension has also been found in plasmodial cysteine proteases (Shenai *et al*. 2000; Sijwali *et al*. 2001; Na *et al*. 2004) and determined to mediate the correct folding of the enzymes (Sijwali *et al*. 2002; Kumar *et al*. 2004; Pandey *et al*. 2004). Our present result also clearly showed that cryptopain-1 did not require the pro-domain for correct folding to the active form, but rather folding was mediated by a 9 amino acid

domain-processing site of cryptopain-1 was aligned using the Clustal method with corresponding regions of falcipain-2, falcipain-3, vivapain-2, vivapain-3, papain, and human cathepsin L (human CL). Fully conserved amino acids are shaded black, and identities with cryptopain-1 are shaded light or dark grey. The ERFNIN and GNFD motifs are indicated with closed and open circles, respectively. Mature protease domain cleavage sites confirmed in this study and previous studies are indicated by arrow-heads. The amino-terminal extensions are underlined. To better understand the role of the pro-domain of cryptopain-1 for folding, constructs coding for cryptopain-1 mature domain with or without different lengths of the prodomain (E1–E5) were amplified by PCR. Each PCR product was ligated into the pQE-30 vector, transformed into M15 [pREP4] *E. coli* competent cells, and expressed. (B) SDS-PAGE analysis. The recombinant proteins purified with Ni-NTA affinity chromatography were evaluated by SDS-PAGE before (upper panel) and after (lower panel) refolding and activation. (C) Refolding progress. The purified recombinant proteins were refolded in refolding buffer and incubated at 4 °C. Aliquots from each reaction were obtained at the indicated time intervals and assayed for proteolytic activity against Z-LR-MCA. The results are expressed as the percentage of maximum activity. (D) Substrate gel analysis. Each refolded recombinant protein was evaluated by gelatin-substrate SDS-PAGE.

https://doi.org/10.1017/S0031182008005350 Published online by Cambridge University Press

N-terminal extension of the mature enzyme. It is not clear why these apicomplexa protozoan parasites have developed this unique folding mechanism for their cysteine proteases and further comprehensive studies are required to clarify the biological and evolutionary meanings of the atypical folding features of the enzymes. The presence of a transmembrane domain in the pro-domain of cryptopain-1 also remains to be elaborated as to whether this region plays any role in the processing and localization of the enzyme in the parasite.

Because the biological role for cryptopain-1 is not clear, an effort was made to detect or predict the probable role of cryptopain-1, using an immunofluorescence assay of the sporozoites of *C. parvum* to characterize the cellular localization of cryptopain-1 in the parasite. But a clear result, which suggested its expression or localization might be in a specific organelle or cellular compartment of the parasite, was not obtained. *In silico* analyses using several software programs were also not able to predict the probable cellular location of cryptopain-1. Nevertheless, our observation that cryptopain-1 is expressed in the sporozoite stage of the parasite, as confirmed by Western blot analysis, and preferentially hydrolyses proteins such as fibronectin and collagen, which are constituents of the host's extracellular matrix and maintain cellular integrity, suggested a probable role in host cell invasion and/or egression by the parasite. Indeed, it is well-known that parasite-derived proteases play critical roles in the processes of other apicomplexa parasites including *Plasmodium falciparum* and *Toxoplasma gondii* (Blackman, 2000; Hoff and Carruthers, 2002; Que *et al*. 2002; Kim, 2004; Dasaradhi *et al*. 2005). Our observation that only a putative pro-form of the enzyme was detected in *C. parvum* sporozoites also raised the question on *in vivo* processing of the enzyme and should be elucidated further. Although the molecular mechanism of host cell invasion and egression by *C. parvum* is as yet poorly characterized, if cryptopain-1 is involved in host cell invasion and/or egression by the parasite, it is feasible that the enzyme would be a reasonable target for therapeutic drug development for the treatment of cryptosporidiosis. Further comprehensive investigations are also required for in-depth understanding of the biological role of the enzyme in the physiology and life cycle of the parasite.

This work was supported by a grant from the National Institutes of Health (NIH-091-4845-300), National Research and Development Program, Ministry of Health and Welfare, Republic of Korea.

REFERENCES

Baker, D., Silen, J. L. and Agard, D. A. (1992). Protease proregion required for folding is a potent inhibitor of the mature enzyme. *Proteins-Structure Function and Bioinformatics* **12**, 339–344.

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

Blackman, M. (2000). Proteases involved in erythrocyte invasion by the malaria parasite: function and potential as chemotherapeutic targets. *Current Drug Targets* **1**, 59–83.

Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. and Menard, R. (1996). Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry* **35**, 8149–8157.

Chen, X. M., Keithly, J. S., Paya, C. V. and LaRusso, N. F. (2002). Cryptosporidosis. *New England Journal of Medicine* **346**, 1723–1731.

Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S. and Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the pro-segment. *The EMBO Journal* **15**, 5492–5503.

Ctrnacta, V., Ault, J. G., Stejskal, F. and Keithly, J. S. (2006). Localization of pyruvate: NADP+ oxidoreductase in sporozoites of *Cryptosporidium parvum*. *The Journal of Eukaryotic Microbiology* **53**, 225–231.

Dasaradhi, P. V., Mohmmed, A., Kumar, A., Hossain, M. J., Bhatnagar, R. K., Chauhan, V. S. and Malhotra, P. (2005). A role of falcipain-2, principal cysteine proteases of *Plasmodium falciparum* in merozoite egression. *Biochemical and Biophysical Research Communications* **336**, 1062–1068.

Forney, J. R., Yang, S. and Healey, M. C. (1996). Protease activity associated with excystation of *Cryptosporidium parvum* oocysts. *The Journal of Parasitology* **82**, 889–892.

Guay, J., Falgueyret, J. P., Ducret, A., Percival, M. D. and Mancini, J. A. (2000). Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. *European Journal of Biochemistry* **267**, 6311–6318.

Hoff, E. F. and Carruthers, V. B. (2002). Is *Toxoplasma* egress the first step in invasion? *Trends in Parasitology* **18**, 251–255.

Kim, K. (2004). Role of proteases in host cell invasion by *Toxoplasma gondii* and other Apicomplexa. *Acta Tropica* **91**, 69–81.

Kumar, A., Dasaradhi, P. V., Chauhan, V. S. and Malhotra, P. (2004). Exploring the role of putative active site amino acids and pro-region motif of recombinant falcipain-2: a principal hemoglobinase of *Plasmodium falciparum*. *Biochemical and Biophysical Research Communications* **317**, 38–45.

Na, B. K., Shenai, B. R., Sijwali, P. S., Choe, Y., Pandey, K. C., Singh, A., Craik, C. S. and Rosenthal, P. J. (2004). Identification and biochemical characterization of vivapains, cysteine proteases of the malaria parasite *Plasmodium vivax*. *The Biochemical Journal* **378**, 529–538.

Na, B. K., Kang, J. M. and Sohn, W. M. (2008). CsCF-6, a novel cathepsin F-like cysteine protease for nutrient uptake of *Clonorchis sinensis*. *International Journal for Parasitology* **38**, 493–502.

Nesterenko, M. V., Tilley, M. and Upton, S. J. (1995). A metallo-dependent cysteine proteinase of

Cryptosporidium parvum associated with the surface of sporozoites. *Microbios* **83**, 77–88.

Pandey, K. C., Sijwali, P. S., Singh, A., Na, B. K. and Rosenthal, P. J. (2004). Independent intramolecular mediators of folding, activity, and inhibition for the *Plasmodium falciparum* cysteine protease falcipain-2. *Journal of Biological Chemistry* **279**, 3484–3491.

Pasquier, C., Promponas, V. J., Palaios, G. A., Hamodrakas, J. S. and Hamodrakas, S. J. (1999). A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. *Protein Engineering* **12**, 381–385.

Peterson, C. (1992). Cryptosporidiosis in patients with human immunodeficiency virus. *Clinical Infectious Diseases* **15**, 903–909.

Priest, J. W., Xie, L., Arrowood, M. J. and Lammie, P. J. (2001). The immunodominant 17 kDa antigen from *Cryptosporidium parvum* is glycosylphosphatidylinositol-anchored. *Molecular and Biochemical Parasitology* **13**, 117–126.

Que, X., Ngo, H., Lawton, J., Gray, M., Liu, Q., Engel, J., Brinen, L., Ghosh, P., Joiner, K. A. and Reed, S. L. (2002). The cathepsin B of *Toxoplasma gondii*, toxopain-1, is critical for parasite invasion and rhoptry protein processing. *Journal of Biological Chemistry* **277**, 25791–25797.

Rosenthal, P. J. (2002). Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Current Opinions in Hematology* **9**, 140–145.

Shaw, M. K., Roos, D. S. and Tilney, L. G. (2002). Cysteine and serine protease inhibitors block intracellular development and disrupt the secretory pathway of *Toxoplasma gondii*. *Microbes and Infection* **4**, 119–132.

Shenai, B. R., Sijwali, P. S., Singh, A. and Rosenthal, P. J. (2000). Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of *Plasmodium falciparum*. *Journal of Biological Chemistry* **275**, 29000–29010.

Shinde, U. and Inouye, M. (2000). Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Seminars in Cell and Developmental Biology* **11**, 35–44.

Sijwali, P. S., Shenai, B. R., Gut, J., Singh, A. and Rosenthal, P. J. (2001). Expression and characterization of the *Plasmodium falciparum* haemoglobinase falcipain-3. *The Biochemical Journal* **360**, 481–489.

Sijwali, P. S., Shenai, B. R. and Rosenthal, P. J. (2002). Folding of the *Plasmodium falciparum* cysteine protease falcipain-2 is mediated by a chaperone-like peptide and not the prodomain. *Journal of Biological Chemistry* **277**, 14910–14915.

Smith, S. M. and Gottesman, M. M. (1989). Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *Journal of Biological Chemistry* **264**, 20487–20495.

Teo, C. F., Zhou, X. W., Bogyo, M. and Carruthers, V. B. (2007). Cysteine protease inhibitors block *Toxoplasma gondii* microneme secretion and cell invasion. *Antimicrobial Agents and Chemotherapy* **51**, 679–688.

Tzipori, S. and Ward, H. (2002). Cryptosporidosis: biology, pathogenesis and disease. *Microbes and Infection* **4**, 1047–1058.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C. and Thomas, D. Y. (1995). Processing of the papain precursor. *Journal of Biological Chemistry* **270**, 10838–10846.

Yamamoto, Y., Watabe, S., Kageyama, T. and Takahashi, S. Y. (1999). Proregion of *Bombyx mori* cysteine proteinase functions as an intramolecular chaperone to promote proper folding of the mature enzyme. *Archives of Insect Biochemistry and Physiology* **42**, 167–178.