Characterization of the gut-associated cathepsin D hemoglobinase from the tick *Ixodes ricinus* (IrCD1)

Daniel Sojka¹, Zdeněk Franta¹, Helena Frantová¹, Pavla Bartošová¹, Martin Horn², Jana Váchová², Anthony J. O’Donoghue⁴, Alegra A. Eroy-Reveles⁴, Charles S. Craik⁴, Giselle M. Knudsen¹,³,⁴, Conor R. Caffrey³, James H. McKerrow³, Michael Mareš² and Petr Kopáček¹

¹ Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice, CZ 370 05, Czech Republic.

² Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, CZ 16610, Czech Republic.

³ Sandler Center for Drug Discovery, University of California San Francisco, San Francisco, CA 94158, USA.

⁴ Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94720, USA.

*Running title: *Ixodes ricinus* gut-associated cathepsin D (IrCD1)

To whom correspondence should be addressed: Daniel Sojka, Institute of Parasitology, Biology Centre ASCR, Branišovská 31, CZ-370 05 České Budějovice, Czech Republic. Tel (420)38-777-5453, Fax (420)38-5310388, E-mail: sojkadan@gmail.com

**Keywords:** aspartic protease, hemoglobin, proteolytic enzymes, protein processing, parasite metabolism, ticks, *Ixodes*

**Background:** Aspartic peptidase activity initiates a multienzyme hemoglobinolysis inside tick guts.

**Results:** IrCD1 is a structurally unique hemoglobinolytic cathepsin D that is upregulated in tick gut cells during feeding.

**Conclusion:** IrCD1 is the major intestinal aspartic peptidase of *I. ricinus*.

**Significance:** Biochemical and functional characterization of IrCD1 completes our knowledge on initial host hemoglobin degradation inside tick gut cells.

**SUMMARY**

In order to identify the gut-associated tick aspartic hemoglobinase, this work focuses on the functional diversity of multiple *Ixodes ricinus* cathepsin D forms (IrCDs). Out of three encoding genes representing *Ixodes scapularis* genome paralogues, IrCD1 is the most distinct enzyme with a shortened propeptide region and a unique pattern of predicted posttranslational modifications. IrCD1 gene transcription is induced by tick feeding and is restricted to the gut tissue. Hemoglobinolytic role of IrCD1 was further supported by immunolocalization of IrCD1 in the vesicles of tick gut cells. Properties of recombinantly expressed rIrCD1 are consistent with the endo-lysosomal environment since the zymogen is auto-activated and remains optimally active in acidic conditions. Hemoglobin cleavage pattern of rIrCD1 is identical to that produced by the native enzyme. Preference for hydrophobic residues at the P1 and P1′ position was confirmed by screening a novel synthetic tetradecapeptidyl substrate library. Outside the S1 - S1′ regions, rIrCD1 tolerates most amino acids but displays a preference for tyrosine at P3 and alanine at P2′. Further analysis of the cleavage site location within the peptide substrate indicated that IrCD1 is a true endopeptidase. The role in hemoglobinolysis was verified with RNAi knock-down of IrCD1 that...
decreased gut extract cathepsin D activity by >90%. IrCD1 was newly characterized as a unique hemoglobinolytic cathepsin D contributing to the complex intestinal proteolytic network of mainly cysteine peptidases in ticks.

Hemathophagy (blood-feeding habit) evolved independently more than 20 times within arthropods (1). Ticks have been adapted to blood-feeding for about 92 million years (2). Hard ticks (Ixodidae) feed only once in each life stage and host blood presents the ultimate source of nutrients for their maturation and reproduction. Females of the major European Lyme disease vector Ixodes ricinus feed for several days and this consists of a slow feeding period (initial 6-9 days) followed by rapid engorgement (12-24 hours prior to detachment). Rapid engorgement occurs ultimately in mated females and detached females deposit a large clutch of eggs and die (3).

Previous mapping of those proteolytic enzymes (peptidases) that digest the blood meal inside the guts of partially engorged I. ricinus females demonstrated the presence of cysteine and aspartic peptidases. Their multi-enzyme complex operating in the acidic compartments of tick gut cells (4,5) is analogous to those found in platyhelminthes (6,7) and nematodes (8). This complex most likely predated the evolution of secreted alkaline trypsin-like hemoglobinases of blood sucking insects (9). Using biochemical assays and PCR cloning systems, a model describing tick hemoglobinolysis was developed (5,10). In this model, an aspartic cathepsin D endopeptidase (IrCD), supported by the cysteine peptidases, cathepsin L (IrCL) and asparaginyl endopeptidase (legumain; IrAE), are responsible for initiating cleavage of host hemoglobin. Production of shorter secondary hemoglobin fragments is performed primarily by a cathepsin B (IrCB). These peptides are further processed by the exopeptidase activity of IrCB (carboxy dipeptidase) and IrCC (cathepsin C – an amino dipeptidase). Single amino acids are then released from N- or C-termini of peptides by leucine or serine mono exopeptidases, respectively.

Here we demonstrate that out of the three identified Ixodes cathepsin D paralogues the newly characterized and most diverse IrCD1 is responsible for the specific aspartic endopeptidase activity detected from I. ricinus female gut extracts. The present report completes our analysis of the initial endopeptidases of the intestinal tick hemoglobinolytic network (11,12).

EXPERIMENTAL PROCEDURES

Tick tissue preparation- I. ricinus ticks were collected and fed on laboratory guinea pigs as described previously (4,12). All animals were treated in accordance with the Animal Protection Law of the Czech Republic No. 246/1992 sb., ethics approval number 137/2008. For tissue preparation, guts, salivary glands and ovaries were dissected from individual partially-engorged females (day 6 of feeding). To prepare gut samples, the luminal contents were carefully removed and remaining tissue was gently washed from the host blood excess in phosphate buffered saline (PBS). Samples were further divided into two halves and pooled for either RNA isolation or tissue extraction. Gut tissue extracts were prepared and stored at -80°C as described previously (5). A smaller number (3 - 4) of dissected tick gut tissues was processed independently for microscopy observations (see bellow).

Isolation of RNA, full cDNA sequencing and RT-PCR- Total RNA was isolated from tissues of I. ricinus via the NucleoSpin® RNA II kit (Macherey-Nagel) and stored at -80°C. First strand cDNA was reverse-transcribed from 0.5 µg of total RNA using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) and oligo(T) primer and stored at -20°C. cDNA fragments of IrCD2 and IrCD3 were PCR amplified and sequenced using primers designed from I. scapularis genes ISCW003823 and ISCW023880, respectively (genome dataset IscaW1.1, www.vectorbase.org). Full length IrCD2 and IrCD3 cDNA sequences were obtained with gene specific primers from the partial PCR amplicons via 3’ RACE PCR modified protocol for SMART™ cDNA Library Construction Kit (Clontech, Takara)
and the 5′ RACE System for Rapid Amplification of cDNA Ends (Invitrogen) as described before (4).

For RT-PCR, the cDNA was diluted 20 fold and used as a template in a ratio 2 µl per 25 µl of PCR reaction mixture. Following combinations of gene specific primers were used for RT-PCR profiling of the three cathepsin D mRNAs: IrCD1 forward 5′GACAGAAAGCGGACAGTACC 3′/reverse 5′CGGAAATTGTGAAGGTGACAT 3′; IrCD2 forward 5′CCGAGATCCTGCACG 3′/reverse 5′GCTCACATGTACTCTCC 3′; IrCD3 forward 5′CCTGACGTTTGTGGCTG 3′/ reverse 5′TCCTGAGGACGTAGTCGC 3′. Dual labeled UPL probes and specific primers were designed online (www.universalprobelibrary.com; Roche) and used for quantitative RT-PCR assays. For IrCD1 forward 5′GACAGAAAGCGGACAGTACC 3′/reverse 5′CGGAAATTGTGAAGGTGACAT 3′ PCR primers were used in combination with probe #78 (Roche). For IrCD2 forward 5′GAGCTGCAAGAGCATCGAC 3′/reverse 5′TTCGAGCACGAAGTCCTTG 3′ PCR primers were used in combination with probe #44 (Roche). The reaction was carried out in triplicates in Rotor-Gene RG3000 PCR cycler (Corbet Research) with following conditions: 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 60°C for 60 sec. Data were analyzed and quantified with the Rotor-Gene 6 analysis software. Relative values were standardized to the PCR amplification of the cDNA for elongation factor 1α (ELF1α) (13) and normalized to the sample with the highest level of expression.

Expression, refolding and purification of recombinant- IrCD1- The E. coli bacterial expression system Champion™ pET directional expression kit (Invitrogen) was selected for expression of the IrCD1 zymogen. N-terminal (6x His) tagged fusion IrCD1 was prepared by PCR amplification of the IrCD1 cDNA without the signal peptide. PCR primers for directional pET cloning were: forward 5′ CACCGCTTTCAGGATCCCGCT 3′; reverse 5′ GCAGCGGACGAGTCGGAA 3′. The product was inserted into the pET100/D-TOPO® expression vector. The sequence verified construct was transformed into BL21 Star™(DE3) E. coli (Invitrogen) and the expression of recombinant protein was performed according to the manual provided with the kit. Inclusion bodies were resolved in buffered 6M Guanidium hydrochloride (14) and the recombinant IrCD1 (rIrCD1) was purified with Co²⁺- chelating chromatography (Hi-Trap™ IMAC FF, GE Healthcare Life-Sciences) in the presence of 8M urea. A linear gradient of 0.01 - 0.5 M imidazole was used for elution on a FPLC AKTA Purifier (GE Healthcare Life-Sciences). The purified protein was refolded using the following protocol: L-arginine was added to the sample to 0.4 M final concentration. This solution was successively dialyzed at 4° C overnight against 25 mM Tris/HCl buffer, pH 7.5, 0.15 M NaCl, 1 mM mercaptoethanol containing: 1) 4 M urea, 0.4 M L-arginine for 3h, 2) 2 M urea, 0.4 M L-arginine for 3h; 3) 0.4 M L-arginine for 3h and 4) plain buffer. The refolded rIrCD1 zymogen was purified by FPLC on a Q Sepharose column (GE Healthcare Life Sciences) equilibrated in 20 mM Bis-Tris pH 6.5 and eluted using a linear gradient of 0-1 M NaCl. The purified rIrCD1 zymogen was activated in 0.1 M Na-acetate pH 4.0 for 3 h at 37°C. Activated rIrCD1 was subsequently purified by FPLC on a Mono S column (GE Healthcare Life Sciences) equilibrated in 50 mM Na-formate pH 3.8 and eluted using a linear gradient of 0-1 M NaCl. The purification and activation processes were monitored by the FRET activity assay and SDS-PAGE (see below).

Preparation of antibodies and indirect immunofluorescence microscopy- To obtain IrCD1 specific polyclonal antibodies (RaxIrCD1) a rabbit was repeatedly immunized with purified rIrCD1 according to a previously described protocol (15). In order to increase specificity the rabbit antibodies were affinity purified using a previously described protocol (16). Briefly, purified rIrCD1 zymogen was coupled to the CNBr-activated...
sepharose 4B and packed to a column. The isolated Ra×rIrCD1 Ig fraction was diluted in PBS and purified over this column, washed with PBS, eluted with 0.2 M L-glycine, 0.15 M NaCl, pH 2.2, neutralized with 1 M Tris-base and stored at -20°C. Reducing SDS-PAGE and Western blot analyses were performed using a previously described protocol (17). To prepare samples for indirect fluorescent microscopy, dissected tissue was fixed and processed using optimized protocol (16). Briefly, the gut tissue was fixed in a solution of formaldehyde and glutaraldehyde, dehydrated in ascending ethanol dilutions, infiltrated in LR White resin (London Resin Company) and polymerized in gelatin capsules (Polysciences). Semi-thin sections (0.5 µm) were transferred onto glass slides, blocked with BSA and low fat dry milk and incubated with Ra×rIrCD1 antibodies. Alexa Fluor 488 dye-conjugated goat anti-rabbit antibody (Invitrogen/Molecular Probes) diluted 1:500 in PBS-Tween and DAPI (4',6'-diamidino-2-phenylindole; Sigma) counterstain were used for fluorescent labeling. Sections were mounted in 2.5% DABCO (1,4-diazabicyclo[2.2.2]octane, Sigma) dissolved in glycerol and examined with the Olympus FW 1000 confocal microscope and consequently processed with the Fluoview (FV10-ASW, Version 1.7 software).

RNAi- The in-detail RNAi protocol was previously reported (18). Briefly, a 281 bp gene-specific DNA fragment of IrCD1 was amplified using primers: forward 5'ATGGGCGCGTTAGCGCTCAAATCG G'3 and reverse 5'ATTCTAGAATCTACGCAAAGCGTTTGAC C'3, containing A Apa I and Xba I restriction sites (underlined) for further cloning into pl10 vector with two T7 promoters in reverse orientation (19). The dsRNA synthesis was performed as described previously. IrCD1 dsRNA (0.5 µl; 3 µg/ µl) was injected into the hemolymph of female ticks using a micromanipulator (Narishige). The control group was injected with identical volume of GFP dsRNA synthesized under the same conditions from the linearized plasmid pl16 (19). After 24 hours of rest in a humid chamber, ticks (25 females and 25 males) were fed on guinea pigs. Partially engorged females were forcibly removed from the host, weighed and the guts were dissected. The level of IrCD1 knockdown was checked on the level of: 1) transcription (qRT-PCR); 2) protein abundance (immunoblotting using Ra×rIrCD1 antibodies); 3) cathepsin D activity (assays of the gut tissue extract using FRET and labeled hemoglobin substrates - see below).

Sequence and in silico structural analysis of IrCDs- The multiple sequence alignment was generated using the Clustal X v. 1.83 software (20). 3D structure models of IrCD1, IrCD2 and IrCD3 were constructed using the Phyre software version 0.2 (21) with the X-ray structure of porcine pepsinogen (PDB entry 3PSGS) as a template. Three-dimensional structures were visualized using the UCSF Chimera program package (22).

IrCD1 activity and inhibition assays-Cathepsin D activity was measured using the FRET peptide substrate Abz-Lys-Pro-Ala-Glu-Phe-Nph-Arg-Leu (single letter code Abz-KPAEFnFRL; Abz, 2-aminobenzoyl; nF, nitrophenylalanine) in 96-well microplates in a total volume of 100 µL. Recombinant IrCD1 (0.1-0.5 µg) or gut extract (150 µg proteins) was pre-incubated for 10 min at 37°C in 150 mM phosphate-citrate buffer, pH 4.0. Hydrolytic activity was continuously measured after addition of substrate (40 µM final concentration) in an Infinite M200 microplate reader (Tecan) at excitation and emission wavelengths set to 330 and 410 nm, respectively. pH profile of activity was determined as stated above in 150 mM phosphate-citrate buffer pH 4.0. Assays were performed in triplicates. The kinetic measurements were performed in the initial linear phase of reaction progress curves and in linear response region to enzyme concentration. For activity assay in the presence of peptidase inhibitors, an aliquot of rIrCD1 was pre-incubated for 15 min at 37°C in the CP buffer pH 4.0 with inhibitors (Table 2). Assay of cathepsin D activity in the tick gut extract was measured in the presence of 10 µM E-64 to
prevent undesired hydrolysis by cysteine cathepsins. For the profiling of cathepsin D activity during feeding on the host, measured activities were normalized per one tick gut as described previously (16).

Hemoglobin degradation: quantification and fragment identification - Bovine hemoglobin (100 µg/ml) was incubated with trrCD1 (0.5 µg) in CP buffer pH 2.5–7.0 in a total volume of 50 µl overnight at room temperature. Hemoglobin digests were separated in 4-12% NuPAGE Bis-Tris gel in NuPAGE MES SDS Running Buffer (Invitrogen) and stained with Coomassie Brilliant Blue. For RP-HPLC analysis, bovine hemoglobin (0.3 mg) was incubated with trrCD1 (0.5 µg) in 50 mM Na-acetate, pH 4.2, in a total volume of 200 µl for 15 min and 4 hr at 35 °C. The reaction mixture was treated with 10 µl of 10% trifluoroacetic acid (TFA) and separated by RP-HPLC on C4 Vydac column (Vydac) equilibrated in 0.1% (v/v) TFA and eluted with a 1%/min gradient of a 99% (v/v) acetonitrile solution in 0.1% (v/v) TFA. The collected peak fractions were analyzed by mass spectrometry. Mass spectra of peptides were measured by FT/MS using an LTQ Orbitrap XL mass spectrometer (Thermo) operating in high-resolution mode (R ~10^5). Cleavage sites were searched by the MS-Non-Specific module of Protein Prospector software (University of California San Francisco) using a mass tolerance of 3 ppm. For quantification of hemoglobin degradation, the tick gut extract was pre-incubated for 10 min with 10 µM E-64 and 1 µM Aza-N-11a (12) to prevent undesired hydrolysis by cysteine cathepsins and asparaginyl endopeptidase. Hemoglobin (10 µg) was incubated with 5 µl of the gut tissue extract in 25 mM sodium acetate, pH 4.2 in a total volume of 35 µl for 1 - 4 h at 37°C. Aliquots of the digest were subjected to derivatization with fluorescamine to quantify the newly formed amino-terminal ends (23). The fluorescence signal was measured using Infinity M200 microplate reader at 370 nm excitation and 485 nm emission wavelengths. All measurements were performed in triplicates and the measured kinetic speeds were normalized per one tick gut (16).

Active-site labeling of trrCD - Active-site labeling of gut extracts and trrCD1 was performed with the cathepsin D specific fluorescently tagged probe in a total volume of 100 µl. The selected probe FAP-09 (24) has a binding core of reversible active-site ligand Val-Val-Sta-Ala-Leu-Gly containing a statin (Sta) residue. An aliquot of the gut tissue extract (100 µg of protein) or purified trrCD1 (0.5 µg of protein) was incubated (20 min at 26°C) with 0.5 µM FAP-09 in 50 mM Na-Acetate pH 4.0. The competitive labeling was performed after pre-incubation (15 min at 26 °C) with 2 µM pepstatin A. The reaction mixture was irradiated in an open tube for 10 min on ice with a 125 W high-pressure mercury-vapor lamp (at a distance of 20 cm) to allow for photoactivated crosslinking. The reaction mixture was then precipitated with four volumes of acetone and boiled in reducing SDS sample buffer, separated by SDS-PAGE (15%) gel and the labeled peptidases were visualized directly in the gel using a Typhoon 9400 Fluorescence Imager (GE Healthcare) with 532 nm excitation (green laser) and the 580 nm emission filter.

Substrate specificity profiling of trrCD1 - A highly diversified peptide library consisting of 124 synthetic tetradecapeptides were synthesized using Fmoc chemistry. Each peptide was purified by HPLC. All peptides had unmodified termini and consist of natural amino acids except methionine and cysteine. Norleucine was included as a substitute for methionine. The peptides were mixed into equimolar pools consisting of 52, 52 and 20 peptides and diluted to 1 µM in 50 mM ammonium acetate, pH 4.0. An equal volume of 100 nM trrCD1 in the same buffer was added to the peptide pools such that the final concentration of each peptide was 0.5 µM. An enzyme-free assay was set up as a control. The assay was incubated at room temperature and aliquots were removed after 5, 10, 15, 30, 60, 120, 240 and 1200 minutes. All reactions were quenched by the addition of pepstatin to a final
concentration of 0.5 µM, evaporated to dryness and reconstituted to the original volume in 0.1% formic acid. 10 µl of each time point were injected onto a 150 × 0.3 mm Magic C18AQ column (Michrom Bioresources) connected to a Thermo Finnigan LTQ ion trap mass spectrometer equipped with a standard electrospray ionization source. Peak lists were generated from the raw files using PAVA software (University of California, San Francisco) and searched against a database consisting of all 124 peptides using Protein Prospector. For estimation of false discovery rate four different decoy databases containing the randomized sequences of the same 124 entries were concatenated to the original 124 entries to create a final database of 620 sequences. Protein Prospector score thresholds were selected to be minimum protein score of 15, minimum peptide score of 10, and maximum expectation values of 0.1 for protein and 0.1 for peptide matches, and resulted in a peptide false discovery rate of 0.17%. Newly formed IrCD1 cleavage products were identified by comparison with a control assay consisting of peptides and buffer. Four residues at either side of the cleaved bond (P4-P4′) were included in the frequency analysis and heatmaps and cleavage signatures were made using iceLogo (25). All possible cleavage sites within the peptide library (n = 1612) served as the negative data set and only amino acids that differ significantly (p <0.05) from the negative dataset are highlighted in the cleavage signature.

RESULTS

Three different cathepsin D enzymes are expressed by Ixodes scapularis and Ixodes ricinus ticks. Data mining of the latest I. scapularis genome dataset (IscaW1.1, December 2008) identified three cathepsin D paralogues: ISCW013185, ISCW003823 and ISCW023880 tagged as I. scapularis cathepsin D1 (IsCD1), D2 (IsCD2) and D3 (IsCD3), respectively. IsCD1 is an ortholog of previously identified I. ricinus cathepsin D (IrCD1; GenBank EF428204) (4). A set of PCR primers was designed to clone cDNA fragments of IsCD2 and IsCD3 I. ricinus homologues. The newly identified IrCD2 and IrCD3 cDNAs were fully sequenced via 5’ and 3’ RACE PCR. Basic parameters of IrCD zymogens including GenBank accession numbers, mRNA length, calculated MW and theoretical pI are shown in Table 1.

Comparison of the three identified IrCD zymogens reveals modifications in the pro-enzymes. The full Clustal X amino acid sequence alignment of the three IrCD zymogens, two other tick hemoglobinolytic cathepsins D precursors longepsin (26) and BmAP (27), extracellular porcine pepsinogen and lysosomal human cathepsin D can be found in the Supplemental file 1. A graphical schematic overview (excluding BmAP and longepsin) demonstrating basic organization of the primary structures is shown in Fig. 1A. All aligned IrCDs are synthesized as pre-pro-enzymes - with predicted signal peptides for targeting to the endoplasmic reticulum. All three IrCD enzyme core structures are related to other cathepsin D-like enzymes (28) and differ mostly in the pro-peptide region consisting of the conserved part A and variable part B. Processing of the zymogens most likely involve the removal of pro-peptide parts that have roles in protein folding, stability, inhibition of the active site, pH dependence of activation and intracellular sorting (29). All three mature IrCDs have two catalytic residues Asp33 and Asp231 in the conserved motif Asp-Thr-Gly (numbering after mature hCD – Supplemental file 1) and do not incorporate the processing loop of mammalian lysosomal cathepsin Ds. The sequence similarity matrix (Supplemental file 1) reveals that IrCD1 is 50% identical to IrCD2 and 49% to IrCD3, while IrCD2 and IrCD3 share about 58% identity. The gut-associated tick cathepsins, BmAP from R.(B.) microplus (27) and longepsin from H. longicornis (26) are 54-58% identical to IrCD1 and IrCD3 and 74% identical to IrCD2. Maximum parsimony analysis was performed using available cathepsin D sequences from ticks aligned with 53 cathepsin D-like molecules from various animal groups. Human pepsinogen was used as outgroup (Supplemental file 2). The analysis resulted in
21 equally parsimonious trees. Cathepsins D from ticks created a monophyletic group within arthropods. The results demonstrate that longeptisin and BmAP are orthologous to IrCD2 while IrCD1 is the most diverse I. ricinus cathepsin D. Surprisingly, orthologs of evolutionarily distant ovarian R. microplus yolk cathepsin (BYC) (30) and the heme-binding aspartic peptidase (THAP) (23), are apparently missing in the I. scapularis genome.

Spatial homology models of IrCD1, IrCD2 and IrCD3 zymogens were constructed to compare the structural features of IrCD isoforms (Fig. 1B). The X-ray structure of porcine pepsinogen (PDB code 3PSGS) was used as a template. All ticks cathepsins D have a conserved bilobal structure with two catalytic aspartic acid residues (red, Fig. 1) on each side of the active site cleft (31). All IrCDs contain two loops in the proximity of the active site cleft for substrate binding. These are designated as 'flap region' and the so called 'polyproline loop' (pink and blue in Fig. 1, respectively) in accordance with nomenclature of mammalian aspartic peptidases (28). However, the polyproline loop of IrCD1 is rearranged: the conserved Gly₃₀₈ and Asp₁₀₀ are mutated to serine and glutamic acid, respectively, and the loop contains three amino acid deletions at Ser₁₉₅-Pro₁₉₇ (Supplemental file 1), which is also present in the I. scapularis analogue IrCD1 (not shown).

Predicted N-glycosylation sites at Asn₇₃ and Asn₁₇₃ in the IrCD1 proenzyme are positionally different to the three predicted N-glycosylation sites in the IrCD2 (Asn₇₀, Asn₁₇₂ and Asn₁₉₄) and the IrCD3 (Asn₇₀ and Asn₁₇₂) proenzymes, respectively (Fig. 1). The positions of lysine residues in the ‘K203’ and β-loop’ regions (Fig. 1A) play a critical role in the phosphotransferase recognition patch (32). IrCD1 appears analogous to lysosomal human cathepsin D by possessing conserved Lys₂₀₃, Lys₂₀₇ and Lys₂₉₃. These residues are important for lysosomal targeting of mammalian cathepsin D via the mannose-6-phosphate pathway (33).

IrCD1 is solely expressed in the gut, upregulated by feeding and localized in digestive cell vesicles. RT-PCR profiling of IrCDs demonstrated that IrCD1 mRNA is restricted to the gut tissue of partially-engorged female ticks. IrCD2 is expressed in guts and salivary glands while IrCD3 mRNA is mostly produced in the ovaries (Fig. 2A). Thus, IrCD3 was further excluded from studies on hemoglobinolysis. The dynamics of expression of IrCD1 and IrCD2 mRNAs in the gut tissue following feeding were analyzed by qRT-PCR (Fig. 2B). IrCD1 mRNA peaks in partially engorged females (day 6 of feeding) while IrCD2 mRNA increases in fully fed and detached females (day 8 of feeding). IrCD1 protein abundance in the gut tissue during feeding was monitored by Western blotting using RaxtIrCD1. The IrCD1 protein signal is detectable in partially and fully fed females (Fig. 3A). Activities monitored with the Abz-KPAEFnFRL substrate in gut extracts prepared from females at different time points (days) of feeding show a rapid increase peaking in fully fed ticks (Fig. 3A). Indirect immunofluorescence microscopy with RaxtIrCD1 and Alexa Fluor 488 dye-conjugated secondary antibodies localized IrCD1 in the vesicles of digestive gut cells at the day 6 of feeding (Fig. 3B).

IrCD1 activates auto-catalytically and displays cathepsin D-like substrate/inhibitor specificity. IrCD1 zymogen was expressed in E. coli (trIrCD1) and isolated with Co²⁺ chelating chromatography under denaturing conditions, renatured and further purified by ion-exchange FPLC. The correctly folded pro-IrCD1 was auto-catalytically activated at pH 4.0. The active enzyme efficiently hydrolyzed the cathepsin D specific FRET substrate Abz-KPAEFnFRL with the quencher 4-nitrophenylalanine group. Maximal activity was detected after 2-3 h of activation at pH 4.0 (Fig. 4A). Processing of the zymogen was followed by SDS PAGE to demonstrate that activation was accompanied by auto-catalytic proteolytic processing (Fig. 4C). Pro-IrCD1 band (47 kDa) was completely converted to a 40 kDa band. This corresponds to the theoretical molecular mass of mature IrCD1 predicted from the amino acid sequence. N-terminal amino acid sequencing of the 40-kDa band identified a...
single sequence, Ile-His-Glu-Gly-Pro-Tyr, that was generated by the cleavage between Lys22 and Ile23 (pro-hCD1 numbering). This auto-activation cleavage site is located 3 amino acids upstream of the homologous mature human cathepsin D sequence (Supplemental file 1). No other processing intermediates or inter-chain processing products were observed.

The pH activity profile of rlrCD1 was determined using the synthetic FRET peptide substrate Abz-KPAEFnFRL and bovine hemoglobin (Fig. 4B). Both substrates were effectively cleaved at pH 2.5-5.0 with optimal activity at pH 4.0.

Inhibitory specificity of rlrCD1 was determined using a panel of selective peptidase inhibitors (Table 2). rlrCD1 activity was completely inhibited by pepstatin A (34) and PDI (potato cathepsin D inhibitor) (35) that both specifically inactivate cathepsin D-like peptidases. Partial inhibition was observed with lopinavir (36) which targets the aspartic peptidases of the retropepsin family. rlrCD1 activity was unaffected by pepstatin (37), leupeptin (38), E-64 (39) and EDTA (40) that inhibit peptidases of serine, cysteine and metallo-peptidase classes, respectively. This inhibition profile confirms that rlrCD1 has ligand binding characteristic similar to mammalian cathepsin D.

Recombinant lrlrCD1 and native lrlrCD in the tick gut tissue extract were visualized using the specific activity-based probe, FAP-09 that binds to the cathepsin D active site (Fig. 4D) (24). The labeled enzymes migrate on the SDSPAGE as single bands of 40 and 45 kDa, respectively. The lower molecular weight of the rlrCD1 band can be explained by the absence of N-glycosylations at two predicted sites (Asn73 and Asn127, Fig. 1A; hCD numbering) in the E. coli expressed zymogen. Labeling with FAP-09 was quenched when the active site had been pre-occupied by pepstatin A as a specificity control.

lrCD1 was incubated with an equimolar mixture of 124 synthetic tetradecapeptides that were designed to have equal representations of all amino acids. Samples of the assay were removed at multiple timepoints between 5 and 1200 minutes and subjected to LC-MS/MS sequencing. The total number of cleavage sites identified was 202, however 97 of these were observed as early as 15 minutes (Supplemental file 3). A heatmap illustrating the frequency of residues found in the P4 to P4′ sites after 15 minutes incubation was generated. These data indicated that rlrCD1 has a strong preference for hydrophobic residues in both the P1 and P1′ positions while the other positions are much less selective. The S1 position appears to be the major determinant of substrate specificity with a preference for Phe>Tyr>Leu/Trp/norleucine but not Ala, Val and Ile. The S1′ pocket is less selective and has an equal preference for all hydrophobic residues except Pro and Leu. Outside the S1 - S1′ regions rlrCD1 tolerates most amino acids but displays a preference for tyrosine at P3 and alanine at P2′. Further analysis of the cleavage site location within the peptide substrate indicated that rlrCD1 is a true endopeptidase with 91% of these cleavages occurring when the S3 to S3′ subsites were occupied and 100% when S2 to S2′ were occupied (Fig. 5A). In order to investigate if a phenylalanine residue in the P1 position was sufficient for cleavage to occur, a list of all tetrapeptides present within the library with phenylalanine in the second position, was generated (Fig. 5B). This represents all possible P2 to P2′ residues with Phe in the P1 site. Cleavage sites were characterized by time at which peptide products were first observed. While a Phe in the P1 site is the major specificity element for rlrCD1, not all Phe-X bonds are cleaved. In general cleavage of Phe-X bonds occurred readily when a hydrophobic residue was present in the P1′ position. In the substrates containing the AFnH, NFnA and SFIE sites (single letter AA codes according to Fig. 5), cleavage does not occur after Phe because each tetrapeptide sequence is situated on the termini of the tetradecapeptide substrate such that no residues are present in either the P3/P4 (NFnA and SFIE) or P3′/P4′ (AFnH) positions. Furthermore, cleavage occurs slowly.

**lrCD1 has a preference for hydrophobic residues at the P1 and P1′ substrate positions.** A novel set of short peptidyl substrates and macromolecular hemoglobin were used to determine rlrCD1 cleavage site specificity.
or not at all in substrates where the P1’ residue is not optimal. When cleavage occurred at sites with a non-optimal P1’ residue, such as IF*EI then residues at P2 and P2’ were often preferred.

Hemoglobin digested by rIrCD1 was resolved by RP-HPLC and peptide fragments were characterized by mass spectrometry. The cleavage sites identified in α- and β-subunits of hemoglobin are indicated in Fig. 5C. In general, these cleavage sites contain hydrophobic residues in the P1 and P1’ positions with a preference for Phe and Leu in P1. After 15 minutes incubation with rIrCD1 four cleavage sites were identified, Leu39-Glu39, Phe39-Leu34 and Leu49-Ala110 in the α-subunit and Phe49-Phe47 in the β-subunit. These are initial cleavage sites detected of digestion corresponding to those reported previously for the authentic cathepsin D activity in gut tissue extract (5). The overall identity of cleavage sites found for the rIrCD1 and the authentic enzyme is 62%.

**RNAi confirms the function of IrCD1 as an intestinal aspartic hemoglobinase.** The major contribution of IrCD1 to the overall gut associated cathepsin D activity and the specific role of this enzyme in the hemoglobinolytic digestive machinery were validated using gene-specific RNAi. IrCD1 transcript, as well as protein synthesis and specific cathepsin D activity, were dramatically reduced (Fig. 6) compared to GFP-dsRNA treated control ticks. Gene specificity of RNAi was verified by qRT-PCR. The level of IrCD1 mRNA in IrCD1-dsRNA treated tick group was reduced to 16% of the GFP control tick group while the change in the expression level of IrCD2 mRNA was still ~90% compared to GFP control (not significant, Fig. 6A). In kinetic assays with Abz-KPAEFnFRL and labeled hemoglobin substrates the overall gut extract cathepsin D activity was reduced in the IrCD1 tick group to 20% and 10%, respectively (Fig. 6C). Phenotype markers — mortality, weight, oviposition and larvae hatching (41) displayed no statistically significant changes in IrCD1-dsRNA treated ticks (data not shown).

**DISCUSSION**

Since 1970s, none of the digestive tick cathepsin D reports (42),(43),(23),(26) reflected the functional diversity of multiple tick cathepsin D paralogues and the coordinated hemoglobinolytic action of cysteine and aspartic peptidases (4,5,44). Therefore identification and full-cloning of all *I. ricinus* cathepsin D genes was the primary concern of this study. Despite the high relative diversity among primary structures phylogenetic analyses confirmed that the three *IrCD* forms most likely evolved by gene duplication and ongoing mutation within the Acari group (Supplemental file 2). We hypothesize that the structural modifications of *IrCD1* arose with the adaptation of ticks/mites to a blood-feeding lifestyle.

Our results clearly identify *IrCD1* as the only *I. ricinus* cathepsin D exclusively expressed in the gut. Hemoglobinolytic role of *IrCD1* was previously indicated by missing RT-PCR signals in *I. ricinus* life stages not feeding on hosts (4). Dynamic upregulation of *IrCD1* mRNA was found to be consistent with the rapidly increasing intestinal aspartic peptidase activity during several days of tick feeding (16). The presence of *IrCD1* in female guts was also previously noted by mass spectrometry (5).

The uptake of blood components by tick gut cells (45) appears to follow similar endocytic mechanisms of fluid phase endocytosis and receptor mediated endocytosis described from mammalian cells (46). Immunolocalization of *IrCD1* shows a distribution through vesicles, most likely lysosomes and endosomes of gut cells. This localization is analogous to previously characterized tick hemoglobinolytic cysteine peptidases *IrAE1* (12), *IrCB1* (16), *IrCL1* (11) and the cathepsin L from R./B. microplus (47). We propose that *IrCD1* is targeted to endolysosomes via the mannose-6-phosphate pathway due to the presence of an identical recognition patch (*Lys*-267, *Lys*-267 and *Lys*-293; Fig. 1A) for the lysosomal N-acetyl-glucosamine-1-phosphotransferase (UDP-GlcNAc) within *IrCD1* and the human lysosomal cathepsin D (33). In addition to the difference in the posttranslational modification patterns of *IrCD1* and the *BmAPlongepsin*
analog \( IrCD2 \) (Supplemental file 1), the putative role of \( IrCD2 \) in the secretory or extracellular processes is supported by \( IrCD2 \) expression in both salivary glands and guts, where the expression peaks in fully-fed females (Fig. 2B).

Our current tick hemoglobinolytic enzyme network model includes primary hemoglobinolytic endopeptidases \( IrCD1 \), \( IrCL1 \) and \( IrAE1 \). All three peptidases prepared as recombinant proteins are capable of auto-catalytic activation in acidic conditions indicating the process to occur upon the acidification of gut cell endosomes [for review see (48)]. Acidification of the large hemoglobin containing vesicles was demonstrated by acridine orange staining of \( R. \) microplus gut cells by Lara et al. (45). Hydrolysis is also facilitated by the spontaneous denaturation of hemoglobin below pH 4.5 (49).

LC-MS characterization of peptide libraries was introduced to resolve secondary preferences at P3, P2, P2′, and P3′ subsites (50) as a more sophisticated method to internally quenched fluorescent peptide libraries (51). Our tetradecapeptidyl library approach revealed that the \( IrCD1 \) is a true endopeptidase with most cleavage occurring when P3 to P3′ residues are occupied. The S1 and S1′ pockets define the primary specificity and a distinct preference for hydrophobic residues at each site is evident. Importantly, hemoglobin cleavage sites by \( IrCD1 \) and the native \( I. \) ricinus gut cathepsin D show identical preference for Phe and Leu at P1 (5).

\( IrCD1 \) RNAi experiments decreased the overall gut aspartic peptidase activity by >90% demonstrating the major contribution of \( IrCD1 \) to the cathepsin D activity of the multi-enzyme hemoglobinolytic network (5). The decrease excludes hemoglobinolytic activity of papain-like enzymes due to the presence of 10 \( \mu \)M E64 in the assay. We propose that the remaining 10% of activity in \( IrCD1 \) knocked-down ticks (Fig. 6C) is not sufficient enough to off-set the appearance of obvious RNAi phenotypes with respect to tick weight, mortality, oviposition and larvae hatching. Missing RNAi phenotype may be better explained by the synchronous operation in between \( IrCD1 \) and other initial hemoglobinase \( IrCL1 \) also remarkable from hemoglobinolytic assays performed with gut extracts (5). The redundancy between \( IrCD1 \) and \( IrCL1 \) was also indicated by the only changed phenotype marker - tick average weight decreased by 24% - in the \( IrCL1 \) RNAi knocked-down ticks (11). Recently, Cruz et al. confirmed in \( B. \) (\( R. \)) microplus that BmAP and BmCL are together responsible for the generation of hemoglobin-derived antimicrobial peptides (27). Due to dsRNA dose and volume limits per tick female our current method has not allowed us to obtain multiple endopeptidase RNAi knock-downs to confirm the cathepsin L/D redundancy, yet (data not shown).

To conclude, this report identifies and characterizes \( IrCD1 \) as the specific gut-associated aspartic peptidase contributing to the peptidase hemoglobinolytic network operating in the digestive epithelium of partially engorged \( I. \) ricinus females. Together with previous reports characterizing \( IrAE \) (12) and \( IrCL1 \) (11), the present study completes the biochemical and functional characterization of those primary individual endopeptidases of the network. Although \( IrCD1 \) plays a greater role in initial processing of hemoglobin (5,10), all three endopeptidases appear to have synergistic roles similar to that described for the intestinal proteolytic network of \( S. \) mansoni (7). However, in order to confirm potential synergies and, possibly, yield phenotypes, combinatorial RNAi and chemical inhibition of multiple protease targets will be required and may help prioritize those individual Ixodid proteases as targets for anti-tick interventions.

REFERENCES:

**FOOTNOTES:**

This work was supported by the postdoctoral projects to D.S. (KJB 600960911 from the Grant Agency ASCR and P502/11/P682 from the Grant Agency of the Czech Republic). Further support was through the grant IAA 600960910 to P. K. (the Grant Agency ASCR) and by Research Center LC06009 (MSMT CR). M.M. and M.H. were supported by grants IAA400550705 (the Grant Agency ASCR) and P207/10/2183 (the Grant Agency of the Czech Republic). Research at the Institute of Organic Chemistry and Biochemistry ASCR was supported by research projects Z602205728 and 61388963. Research performed at the Sandler Center was in part supported by the Sandler Foundation and the Burroughs Welcome Fund Schistosome Toolbox Collaborative Research Travel Award. We would like to thank Miloslav Sanda for the mass spectroscopy.

The abbreviations used are: *IrCD*, *Ixodes ricinus* cathepsin D; *IrCL*, *Ixodes ricinus* cathepsin L; *IrCB*, *Ixodes ricinus* cathepsin B; *IrCC*, *Ixodes ricinus* cathepsin C; *IrAE*, *Ixodes ricinus* asparaginyl endopeptidase/legumain; RACE PCR, rapid amplification of cDNA ends polymerase chain reaction; ELF1α, elongation factor 1α; DAPI, 4',6'-diamidino-2-phenylindole; DABCO, 1,4-diazabicyclo[2.2.2]octane; Ra×rIrCD1, *IrCD1* specific polyclonal rabbit antibodies; *BmAP*, *Rhipicephalus (Boophilus) microplus* aspartic peptidase; RP-HPLC, reverse phase HPLC, Abz, 2-aminobenzoyl; nF, nitrophenylalanine.
FIGURE LEGENDS:

Figure 1- Comparison of identified Ixodes ricinus cathepsin D paralogues.
A. Graphical scheme-alignment of IrCD1, IrCD2, IrCD3, lysosomal human cathepsin D (hCD,) and extracellular porcine pepsinogen (pPD). Amino acid numbering is related to mature hCD. Domains are labeled with colors used in panel B. Depicted residues-single letter AA coding: Two catalytic aspartic acid residues (red Ds), predicted N-glycosylations: yellow full circles, Ns; phosphorylation determinant lysines in the K 203 region and in the β loop region – violet diamond labels, Ks. (Note: Full Clustal X alignment including two tick aspartic hemoglobinases longepsin and BmAP could be found in Supplemental file 1.).
B. Tertiary structures of IrCD1, IrCD2, and IrCD3 zymogens (pro-IRCD1-3) modeled using X-ray structure of porcine pepsinogen (PDB entry 3PSG). Phyre software version 0.2 (21) created models were visualized with UCSF Chimera (22). Pro-cathepsins D (ribbons) are shown with two catalytic aspartic acid residues (red sticks). Pro-peptide is divided into conservative part (blue-green) and variable (green) part, that is reduced in the structure of IrCD1 and IrCD2 compared to IrCD3 and pPD. Y flap region, poly-proline loop in the vicinity of the substrate binding pocket are labeled pink and blue, respectively; predicted sites of N- glycosylations (yellow sticks) differ in the three IrCD isoforms.

Figure 2- Transcriptional profiling of identified IrCDs.
A. Transcriptional profiling of IrCD forms in tissues of partially engorged I. ricinus females. Two-step RT-PCR was performed with IrCD1, IrCD2 and IrCD3 gene specific primers and first strand cDNA templates prepared from total RNA isolated from guts, salivary glands and ovaries. The identity of resulting PCR products was confirmed by DNA sequencing. Ferritin 1 primers (15) were used as template loading control. IrCD1 is the only cathepsin D with gut tissue restricted expression.
B. Dynamics of IrCD1 and IrCD2 expression in female guts during feeding. Levels of IrCD1 and IrCD2 mRNAs were determined by qRT-PCR using dual labeled UPL probe #78 and #44 (Roche), gene specific primers and gut cDNA templates from days 0, 2, 4, 6 and 8 of female feeding. Reactions were done in triplicate, standard deviations are depicted. The levels of mRNA are normalized to the sample with maximum mRNA level (set to 100%). Relative level of IrCD1 mRNA peaks in the sixth day before rapid engorgement, which is conditioned by fertilization, while the relative level of IrCD2 mRNA peaks in fully fed fertilized females after rapid engorgement.

Figure 3- IrCD activity, IrCD1 expressional profiling and localization with specific antibodies.
A. Abz-KPAEFnFRL measured authentic cathepsin D activity in female I. ricinus gut extracts during feeding. Timeline depicts feeding phases: attachment, slow feeding period, rapid engorgement and detachment. Activity is increasing from the day 4 of feeding, peaks in fully fed females and than slowly decreases. Below: Western blot- immunodetection with Ra×rIrCD1 of authentic IrCD1 in gut extracts in different days of feeding- signal raises by day 6 and peaks in day 8. B. Indirect immunoflorescent microscopy - Semi-thin section labeled with affinity purified Ra×rIrCD1. The goat antirabbit IgG conjugated with Alexa-Fluor® 488 was used as the secondary antibody. Note: Signal is localized to intracellular vesicles of digestive cells. Nuclei of cells were counterstained with DAPI. BC - basal epithelial cells, HC - hemoglobin crystal in the gut lumen, DC - digestive gut cells.

Figure 4- Biochemical characterization and proteomic identification of rIrCD1.
A. rIrCD1 zymogen was activated at pH 4.0 and the resulting activity was measured with the Abz-KPAEFnFRL substrate.
B. pH optimum of mature rIrCD1 was determined with Abz-KPAEFnFRL substrate and compared with the pH profile of rIrCD1 in hemoglobin degradation assay. Fragments were visualized on proten stained SDS–PAGE (16 kDa band corresponds to hemoglobin monomer).
C. Time dependence of the rIrCD1 zymogen processing at acidic pH analyzed by SDS–PAGE. Prot-rIrCD1 was incubated at pH 4.0 for 1 and 3 h. Lane 0 contains pro-rIrCD1 before activation.
D. In-gel visualization of activated rIrCD1 and authentic IrCD1 in tick gut extracts by proteomic activity-based probe FAP-09. The enzyme material was incubated with FAP-09, resolved by SDS-PAGE and visualized by a fluorescence scanner.

Cathepsin D inhibitor pepstatin A was used for specificity control in the assays and for the competitive labeling of IrCD1 by FAP-09.

Figure 5- Substrate specificity profiling of rIrCD1 - screening of novel highly diversified peptide library consisting of 124 synthetic tetradecapeptides and hemoglobin rIrCD1 cleavage map.

A. A heat map and iceLogo reflecting the frequency of residues in the P4 to P4’ positions of 97 unique rIrCD1 cleavage sites. Each amino acid is represented by the single letter code and ‘n’ corresponds to norleucine. In the heatmap red boxes identify amino acids that are present at a frequency less than that of the entire library (negative selection) and green boxes identify residues that are enriched (positive selection). Cleavage of substrates occurred only when the P2 to P2’ sites were occupied while 3% and 6% of substrates were cleaved when no amino acid was present in the P3 and P3’ sites, respectively. The same dataset is represented as an iceLogo (25) where only residues that are significantly different from background (p < 0.05) are illustrated.

B. Qualitative assessment of rIrCD cleavage products observed between 5 and 1200 minutes incubation with peptide library. All tetrapeptide sequences within the library that possess Phe in the second position (corresponding to P1) are listed for comparison.

C. Bovine hemoglobin was digested in vitro with rIrCD1 at pH 4.0. The fragments were identified by mass spectrometry and the corresponding cleavage sites are indicated in the hemoglobin sequence with black triangles. The initial cleavage sites (after 15 minutes of reaction) are marked with asterisks. Numbering of the α subunit residues excludes the first mRNA translated methionine of the hemoglobin precursor.

Figure 6- RNAi knock down of IrCD1. I. ricinus females (25 per each group) were injected with IrCD1 dsRNA (iIrCD1 minus, experimental group) and/or GFP dsRNA (GFP – plus, control group) . Gut tissue of partially engorged ticks was used for RNA/DNA isolation and preparation of tissue homogenates.

A. Effect on mRNA expression levels- dual labeled UPL probe #78 (Roche) and IrCD1 specific primers were used for qRT-PCR analysis of IrCD1 RNAi effect in between the iIrCD1 and GFP tick groups. The presence of IrCD1 mRNA is decreased to 16% in the relative level compared to the GFP control group. For gene specificity of the IrCD1 knock-down, qRT PCR with UPL probe #44 (Roche) and IrCD2 specific primers was performed. IrCD2 mRNA level showed no significant decrease in between the two tick groups. All PCR reactions were performed in triplicates. Relative values are depicted with standard deviations.

B. Effect on protein abundance- SDS PAGE separated gut extracts of iIrCD1 and GFP ticks were electrotransferred to PVDF membrane (coomassie blue stained lines for loading control). Presence of IrCD1 is determined by Western blot with antibody RaxrIrCD1 - no signal in iIrCD1 tick group line compared to ~40 kDa rIrCD1 signal in GFP control group line.

C. Effect on gut extract cathepsin D activity- Kinetic assays measured with Abz-KPAEFnFRL and hemoglobin displayed ~80% and > 90% decrease of activity, respectively, in between GFP and iIrCD1 tick groups. Measuring was performed in triplicates. Relative values are depicted with standard deviations.
TABLES

Table 1- Basic parameters of the identified Ixodes ricinus cathepsin D paralogues (IrCDs)

<table>
<thead>
<tr>
<th>name</th>
<th>GenBank a. n.</th>
<th>mRNA [bases]</th>
<th>zymogen* [AA]</th>
<th>SignalP prediction [AA]</th>
<th>isoelectric point [pH]</th>
<th>MW* [kDa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>IrCD1</td>
<td>EF428204</td>
<td>1304</td>
<td>382</td>
<td>21</td>
<td>4.75</td>
<td>39.45</td>
</tr>
<tr>
<td>IrCD2</td>
<td>HQ615697</td>
<td>1776</td>
<td>387</td>
<td>20</td>
<td>6.45</td>
<td>39.35</td>
</tr>
<tr>
<td>IrCD3</td>
<td>HQ615698</td>
<td>1456</td>
<td>398</td>
<td>22</td>
<td>7.69</td>
<td>41.32</td>
</tr>
</tbody>
</table>

* without the signal peptide

Table 2- rIrCD1 sensitivity to specific protease inhibitors

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Specificity/ Target Protease</th>
<th>Concentration</th>
<th>Inhibition [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pefabloc</td>
<td>serine proteases(37)</td>
<td>1 mM</td>
<td>3±1</td>
</tr>
<tr>
<td>E-64</td>
<td>cysteine proteases(39)</td>
<td>10 µM</td>
<td>4±2</td>
</tr>
<tr>
<td>Leupeptin</td>
<td>serine and cysteine proteases(38)</td>
<td>10 µM</td>
<td>9±4</td>
</tr>
<tr>
<td>EDTA</td>
<td>metalloproteases(40)</td>
<td>10 µM</td>
<td>2±1</td>
</tr>
<tr>
<td>Lopinavir</td>
<td>aspartic proteases of retropepsin family(36)</td>
<td>10 µM</td>
<td>61±5</td>
</tr>
<tr>
<td>Pepstatin</td>
<td>aspartic proteases of pepsin and retropepsin families(34)</td>
<td>1 µM</td>
<td>99±4</td>
</tr>
<tr>
<td>PDI</td>
<td>aspartic proteases of cathepsin D type(35)</td>
<td>1 µM</td>
<td>100±2</td>
</tr>
</tbody>
</table>

Recombinant IrCD1 was pre-incubated with the given inhibitor, and remaining activity was measured in a kinetic assay with the FRET substrate Abz-KPAEFnFRL. Mean values ± SD are expressed as the percentage inhibition compared with the uninhibited control.

FIGURES
FIGURE 2

A

<table>
<thead>
<tr>
<th>Gut</th>
<th>Salivary gl.</th>
<th>Ovary</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

B

**IrCD1 qRT-PCR**

![Graph](graph1.png)

**IrCD2 qRT-PCR**

![Graph](graph2.png)
FIGURE 3

A

**gut cathepsin D activity during feeding**

<table>
<thead>
<tr>
<th>attach</th>
<th>slow feeding period</th>
<th>rapid eng.</th>
<th>detach</th>
</tr>
</thead>
</table>

![Western Raxlr/CD1](chart)

**Western Raxlr/CD1**

**days of feeding**

![Relative activity](chart)

B

**DAPI**  
**Alexa 488 Raxlr/CD1**  
**merged**
FIGURE 5

A

Negative  None  Positive

P4  P3  P2  P1  P1'  P2'  P3'  P4'

G  A  V  S  T  R  I  K  E  D  Q  N  F  Y  W

Difference (%)

Occupied (%)

B

Cleavage first observed

HFKF  SPKN  APKD  YFRI  YFPK  YFRL  YFHD  HPFG  PTHI  DFFN  RPFP  RFDP  TDRI  LTMD  KDMD  WDSY  IPEI  GFEI  GFEI  LFN4  SFN4  INFN  QJNV  TFQK  HPQP  KDQP  NFQP  HPSK  WPSS  VFSS  GFTI  QFTI  QFTI  HPTQ  SFGD  EPGL  KFGL  AFQG  nFDG  nFDG  IFPI  FFMP  FWAA  DFWL  STFW  DTYL  ITPL  ITPQ  HIFQ  AFAE  EFAP  EFAP  KFAM  nFVM  lPVS  STPT  VEIL  TVLV  TTVT  NPLK  NPLK  NhNA  AAtA  AAtA  SFNN  VFNP

C

Hemoglobin α-subunit

1 VLGAADKGN VKAWNGKVG HAAEYGAAL ERMGRSTPTT KTVFHPDLSH 50
51 GSAQVKGGM AKVAAALTEA VEHLDLPQA LSELSLJHAK KLRVPVWPLK 100
101 LSHSLTEFL ASLPSDFTP AVHASLDYKFL AVNSTVTLSK YR 141

Hemoglobin β-subunit

1 MLTAEEKAV TAPWQKVD EVQGGEALGL RLVYPWTFQF FESFGDLSTA 50
51 DAVMNPKVK ARGGKVLDST SNGMKHLDDL KGTAALSDEL HCDEKLVDPE 100
101 NFYLLGNYLV VVLRANFGKE FTPVLOADFQ KVAVGVALI AHRH 145

time/min

s 10 30 60 120 240 480 720
FIGURE 6

A. mRNA

B. Protein

C. Enzymatic activity

![Graphs showing relative mRNA expression, protein bands, and enzymatic activity levels for IrCD1 and IrCD2.]