THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 25, pp. 21152–21163, June 15, 2012

Characterization of Gut-associated Cathepsin D Hemoglobinase from Tick *Ixodes ricinus* (*Ir*CD1)*[S]

Received for publication, February 5, 2012, and in revised form, April 16, 2012 Published, JBC Papers in Press, April 26, 2012, DOI 10.1074/jbc.M112.347922

Daniel Sojka[‡1], Zdeněk Franta[‡], Helena Frantová[‡], Pavla Bartošová[‡], Martin Horn[§2], Jana Váchová[§], Anthony J. O'Donoghue[¶], Alegra A. Eroy-Reveles[¶], Charles S. Craik[¶], Giselle M. Knudsen[¶‖], Conor R. Caffrey[‖], James H. McKerrow[‖], Michael Mareš[§2], and Petr Kopáček[‡]

From the [‡]Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice, CZ 370 05, Czech Republic, the [§]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, CZ 16610, Czech Republic, the [‖]Sandler Center for Drug Discovery, University of California, San Francisco, California 94158, and the [¶]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94720

Background: Aspartic peptidase activity initiates a multienzyme hemoglobinolysis inside tick guts.
Results: *Ir*CD1 is a structurally unique hemoglobinolytic cathepsin D that is up-regulated in tick gut cells during feeding.
Conclusion: *Ir*CD1 is the major intestinal aspartic peptidase of *I. ricinus*.
Significance: Biochemical and functional characterization of *Ir*CD1 completes our knowledge on initial host hemoglobin degradation inside tick gut cells.

To identify the gut-associated tick aspartic hemoglobinase, this work focuses on the functional diversity of multiple *Ixodes ricinus* cathepsin D forms (*Ir*CDs). Out of three encoding genes representing *Ixodes scapularis* genome paralogs, *Ir*CD1 is the most distinct enzyme with a shortened propeptide region and a unique pattern of predicted post-translational modifications. *Ir*CD1 gene transcription is induced by tick feeding and is restricted to the gut tissue. The hemoglobinolytic role of *Ir*CD1 was further supported by immunolocalization of *Ir*CD1 in the vesicles of tick gut cells. Properties of recombinantly expressed r*Ir*CD1 are consistent with the endo-lysosomal environment because the zymogen is autoactivated and remains optimally active in acidic conditions. Hemoglobin cleavage pattern of r*Ir*CD1 is identical to that produced by the native enzyme. The 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, r*Ir*CD1 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 *Ir*CD1 is a true endopeptidase. The role in hemoglobinolysis was verified with RNAi knockdown of *Ir*CD1 that decreased gut extract cathepsin D activity by >90%. *Ir*CD1 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 h 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 multienzyme 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 (*Ir*CD),[3] supported by the cysteine peptidases cathepsin L (*Ir*CL) and asparaginyl

* This work was supported in part by Grant IAA 600960910 (to P. K.), The Academy of Sciences of the Czech Republic Grant Agency, and by Research Center LC06009 (The Ministry of Education, Youth, and Sports of the Czech Republic). Research at the Institute of Parasitology BC, The Academy of Sciences of the Czech Republic, and the Institute of Organic Chemistry and Biochemistry, The Academy of Sciences of the Czech Republic, supported by Research Projects Z602205728 and 61388963 and research performed at the Sandler Center was supported in part by the Sandler Foundation and the Burroughs Welcome Fund Schistosome Toolbox Collaborative Research Travel Award.

[S] This article contains supplemental Figs. 1–3 and additional references.

[1] Supported by Postdoctoral Project Grants KJB 600960911 from the Grant Agency of The Academy of Sciences of the Czech Republic and P502/11/P682 from the Grant Agency of the Czech Republic. To whom correspondence should be addressed: 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.

[2] Supported by Grant IAA400550705 from Grant Agency ASCR and Grant P207/10/2183 from Grant Agency of the Czech Republic.

[3] The abbreviations used are: *Ir*CD, *I. ricinus* cathepsin D; *Ir*CL, *I. ricinus* cathepsin L; *Ir*CB, *I. ricinus* cathepsin B; *Ir*CC, *I. ricinus* cathepsin C; *Ir*AE, *I. ricinus* asparaginyl endopeptidase/legumain; RACE, rapid amplification of cDNA ends; Ra×r*Ir*CD1, *Ir*CD1-specific polyclonal rabbit antibody; *Bm*AP, *R.* (*B.*) *microplus* aspartic peptidase; Abz, 2-aminobenzoyl; nF, nitrophenylalanine; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; qRT, quantitative RT; r, recombinant; dsRNA, double-stranded RNA.

endopeptidase (legumain; *Ir*AE), is responsible for initiating cleavage of host hemoglobin. Production of shorter secondary hemoglobin fragments is performed primarily by a cathepsin B (*Ir*CB). These peptides are further processed by the exopeptidase activity of *Ir*CB (carboxyl dipeptidase) and *Ir*CC (cathepsin C, an amino dipeptidase). Single amino acids are then released from N or C termini of peptides by leucine or serine monoexopeptidases, respectively.

Here, we demonstrate that out of the three identified *Ixodes* cathepsin D paralogs, the newly characterized and most diverse *Ir*CD1 is responsible for the specific aspartic endopeptidase activity detected from *I. ricinus* female gut extracts. This study 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 below).

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 Applied Science) and oligo(dT) primer and stored at −20 °C. cDNA fragments of *Ir*CD2 and *Ir*CD3 were PCR-amplified and sequenced using primers designed from *Ixodes scapularis* genes ISCW003823 and ISCW023880, respectively (genome dataset IscaW1.1). Full-length *Ir*CD2 and *Ir*CD3 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 and 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 of 2 μl per 25 μl of PCR mixture. The following combinations of gene-specific primers were used for RT-PCR profiling of the three cathepsin D mRNAs: *Ir*CD1 forward 5′-GACAGAAGGCGGACAGTACC-3′ and reverse 5′-CGGAAATTGTGAAGGTGACAT-3′; *Ir*CD2 forward 5′-CCGAGATCCTGCACG-3′ and reverse 5′-GCTCACGATGTACTCTCC-3′; and *Ir*CD3 forward 5′-CCTGACGTTTGTGGCTG-3′ and reverse 5′-TCTTGAGGACGTAGTCGC-3′. Dual-labeled UPL probes and specific primers were designed on line (Roche Applied Science) and used for quantitative RT-PCR assays. *Ir*CD1 forward 5′-GACAGAAGGCGGACAGTACC-3′ and reverse 5′-CGGAAATTGTGAAGGTGACAT-3′ PCR primers were used in combination with probe 78 (Roche Applied Science). *Ir*CD2 forward 5′-GAGCTGCAAGAGCATCGAC-3′ and reverse 5′-TTCGAGCACGAAGTCCTTG-3′ PCR primers were used in combination with probe 44 (Roche Applied Science). The reaction was carried out in triplicate in Rotor-Gene RG3000 PCR cycler (Corbet Research) with the following conditions: 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. 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 *Escherichia coli* bacterial expression system Champion™ pET directional expression kit (Invitrogen) was selected for expression of the *Ir*CD1 zymogen. N-terminal His_6-tagged fusion *Ir*CD1 was prepared by PCR amplification of the *Ir*CD1 cDNA without the signal peptide. PCR primers for directional pET cloning were as follows: forward 5′-CACCGCTTTCAGGATCCCGCT-3′ and reverse 5′-GCAGCGGACGAAGTCGGAA-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 6 M guanidinium hydrochloride (14), and the recombinant *Ir*CD1 (r*Ir*CD1) was purified with Co^{2+} chelating chromatography (Hi-Trap™ IMAC FF, GE Healthcare) in the presence of 8 M urea. A linear gradient of 0.01–0.5 M imidazole was used for elution on an FPLC AKTA purifier (GE Healthcare). 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 against 25 mM Tris/HCl buffer, pH 7.5, 0.15 M NaCl, 1 mM mercaptoethanol containing the following: 1) 4 M urea, 0.4 M L-arginine for 3 h; 2) 2 M urea, 0.4 M L-arginine for 3 h; 3) 0.4 M L-arginine for 3 h; and 4) plain buffer overnight. The refolded r*Ir*CD1 zymogen was purified by FPLC on a Q-Sepharose column (GE Healthcare) equilibrated in 20 mM BisTris, pH 6.5, and eluted using a linear gradient of 0–1 M NaCl. The purified r*Ir*CD1 zymogen was activated in 0.1 M sodium acetate, pH 4.0, for 3 h at 37 °C. Activated r*Ir*CD1 was subsequently purified by FPLC on a Mono S column (GE Healthcare) equilibrated in 50 mM sodium 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 *Ir*CD1-specific polyclonal antibodies (Ra×r*Ir*CD1), a rabbit was repeatedly immunized with purified r*Ir*CD1 according to a previously described protocol (15). To increase specificity, the rabbit antibodies were affinity-purified using a previously described protocol (16). Briefly, purified r*Ir*CD1 zymogen was coupled to the CNBr-activated Sepharose 4B and packed to a column. The isolated Ra×r*Ir*CD1 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, neutral-

ized 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 Co.), 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×r*Ir*CD1 antibodies. Alexa Fluor 488 dye-conjugated goat anti-rabbit antibody (Invitrogen) diluted 1:500 in PBS/Tween and 4′,6′-diamidino-2-phenylindole (DAPI) (Sigma) counterstain were used for fluorescent labeling. Sections were mounted in 2.5% 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 reported previously (18). Briefly, a 281-bp gene-specific DNA fragment of *Ir*CD1 was amplified using primers forward 5′-ATGGGCCCGTTAG-CGCCTCAAAATCGG-3′ and reverse 5′-ATTCTAGACTC-ACGCAAAGCGTTTGACC-3′, containing *ApaI* and *XbaI* restriction sites (underlined) for further cloning into pll10 vector with two T7 promoters in reverse orientation (19). The dsRNA synthesis was performed as described previously. *Ir*CD1 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 an identical volume of GFP dsRNA synthesized under the same conditions from the linearized plasmid pll6 (19). After 24 h 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 and weighed, and the guts were dissected. The level of *Ir*CD1 knockdown was checked on the following levels: 1) transcription (qRT-PCR); 2) protein abundance (immunoblotting using Ra×r*Ir*CD1 antibodies); and 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 version 1.83 software (20). Three-dimensional structure models of *Ir*CD1, *Ir*CD2, and *Ir*CD3 were constructed using the Phyre software version 0.2 (21) with the x-ray structure of porcine pepsinogen (Protein Data Bank entry 3PSG (52)) 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-KPAEFn-FRL; Abz, 2-aminobenzoyl; nF, nitrophenylalanine) in 96-well microplates in a total volume of 100 μl. Recombinant *Ir*CD1 (0.1–0.5 μg) or gut extract (150 μg of proteins) was preincubated 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 2.5–7.0. Assays were performed in triplicate. 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 r*Ir*CD1 was preincubated 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 r*Ir*CD1 (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 BisTris 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 r*Ir*CD1 (0.5 μg) in 50 mM sodium acetate, pH 4.2, in a total volume of 200 μl for 15 min and 4 h 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 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 Fourier transform/MS using an LTQ Orbitrap XL mass spectrometer (Thermo) operating in high resolution mode ($R \sim 10^5$). Cleavage sites were searched by the MS nonspecific 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 preincubated 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 N-terminal ends (23). The fluorescence signal was measured using an Infinity M200 microplate reader at 370 nm excitation and 485 nm emission wavelengths. All measurements were performed in triplicate, and the measured kinetic speeds were normalized per one tick gut (16).

Active Site Labeling of IrCD—Active site labeling of gut extracts and r*Ir*CD1 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 r*Ir*CD1 (0.5 μg of protein) was incubated (20 min at 26 °C) with 0.5 μM FAP-09 in 50 mM sodium acetate, pH 4.0. The competitive labeling was performed after preincubation (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-watt high pressure mercury vapor lamp (at a distance of 20 cm) to allow for photoactivated cross-linking. The reaction mixture was then precipitated with 4 volumes of acetone and

TABLE 1
Basic parameters of the identified *I. ricinus* cathepsin D paralogs (*Ir*CDs)

Name	GenBank™ accession no.	mRNA (bases)	Zymogen[a] (amino acids)	SignalP prediction (amino acids)	Isoelectric point (pH)	Molecular mass[a]
						kDa
*Ir*CD1	EF428204	1304	382	21	4.75	39.45
*Ir*CD2	HQ615697	1776	387	20	6.45	39.35
*Ir*CD3	HQ615698	1456	398	22	7.69	41.32

[a] Data are without the signal peptide.

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 rIrCD1—A highly diversified peptide library consisting of 124 synthetic tetradecapeptides was synthesized using Fmoc (*N*-(9-fluorenyl)methoxycarbonyl) chemistry. Each peptide was purified by HPLC. All peptides had unmodified termini and consisted 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 *Ir*CD1 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 min. 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 was 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 a minimum protein score of 15, minimum peptide score of 10, and maximum expectation value of 0.1 for protein and 0.1 for peptide matches, and resulted in a peptide false discovery rate of 0.17%. Newly formed *Ir*CD1 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 *I. scapularis* and *I. ricinus* ticks. Data mining of the latest *I. scapularis* genome dataset (IscaW1.1, December 2008) identified three cathepsin D paralogs as follows: ISCW013185, ISCW003823, and ISCW023880 tagged as *I. scapularis* cathepsin D1 (*Is*CD1), D2 (*Is*CD2), and D3 (*Is*CD3), respectively. *Is*CD1 is an ortholog of previously identified *I. ricinus* cathepsin D (*Ir*CD1; GenBank™ EF428204) (4). A set of PCR primers was designed to clone cDNA fragments of *Is*CD2 and *Is*CD3 *I. ricinus* homologs. The newly identified *Ir*CD2 and *Ir*CD3 cDNAs were fully sequenced via 5′- and 3′-RACE PCR. Basic parameters of *Ir*CD zymogens, including GenBank™ accession numbers, mRNA length, calculated molecular weight, and theoretical pI are shown in Table 1.

Comparison of Three Identified IrCD Zymogens Reveals Modifications in the Propeptides—The full Clustal X amino acid sequence alignment of the three *Ir*CD zymogens, two other tick hemoglobinolytic cathepsin D precursors longepsin (26) and *Bm*AP (27), extracellular porcine pepsinogen, and lysosomal human cathepsin D can be found in the supplemental File 1. A graphical schematic overview (excluding *Bm*AP and longepsin) demonstrating basic organization of the primary structures is shown in Fig. 1*A*. All aligned *Ir*CDs are synthesized as prepropeptides, with predicted signal peptides for targeting to the endoplasmic reticulum. All three *Ir*CD enzyme core structures are related to other cathepsin D-like enzymes (28) and differ mostly in the propeptide region consisting of the conserved part A and variable part B. Processing of the zymogens most likely involves the removal of propeptide parts that have roles in protein folding, stability, inhibition of the active site, pH dependence of activation, and intracellular sorting (29). All three mature *Ir*CDs have two catalytic residues, Asp^{33} and Asp^{231}, in the conserved motif Asp-Thr-Gly (numbering after mature hCD, see supplemental File 1) and do not incorporate the processing loop of mammalian lysosomal cathepsin Ds. The sequence similarity matrix (supplemental File 1) reveals that *Ir*CD1 is 50% identical to *Ir*CD2 and 49% to *Ir*CD3, whereas *Ir*CD2 and *Ir*CD3 share about 58% identity. The gut-associated tick cathepsins, *Bm*AP from *R.(B.) microplus* (27) and longepsin from *Haemaphysalis longicornis* (26) are 54–58% identical to *Ir*CD1 and *Ir*CD3 and 74% identical to *Ir*CD2. 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 longepsin and *Bm*AP are orthologous to *Ir*CD2, whereas *Ir*CD1 is the most diverse *I. ricinus* cathepsin D. Surprisingly, orthologs of evolutionarily distant ovarian *R. micro-*

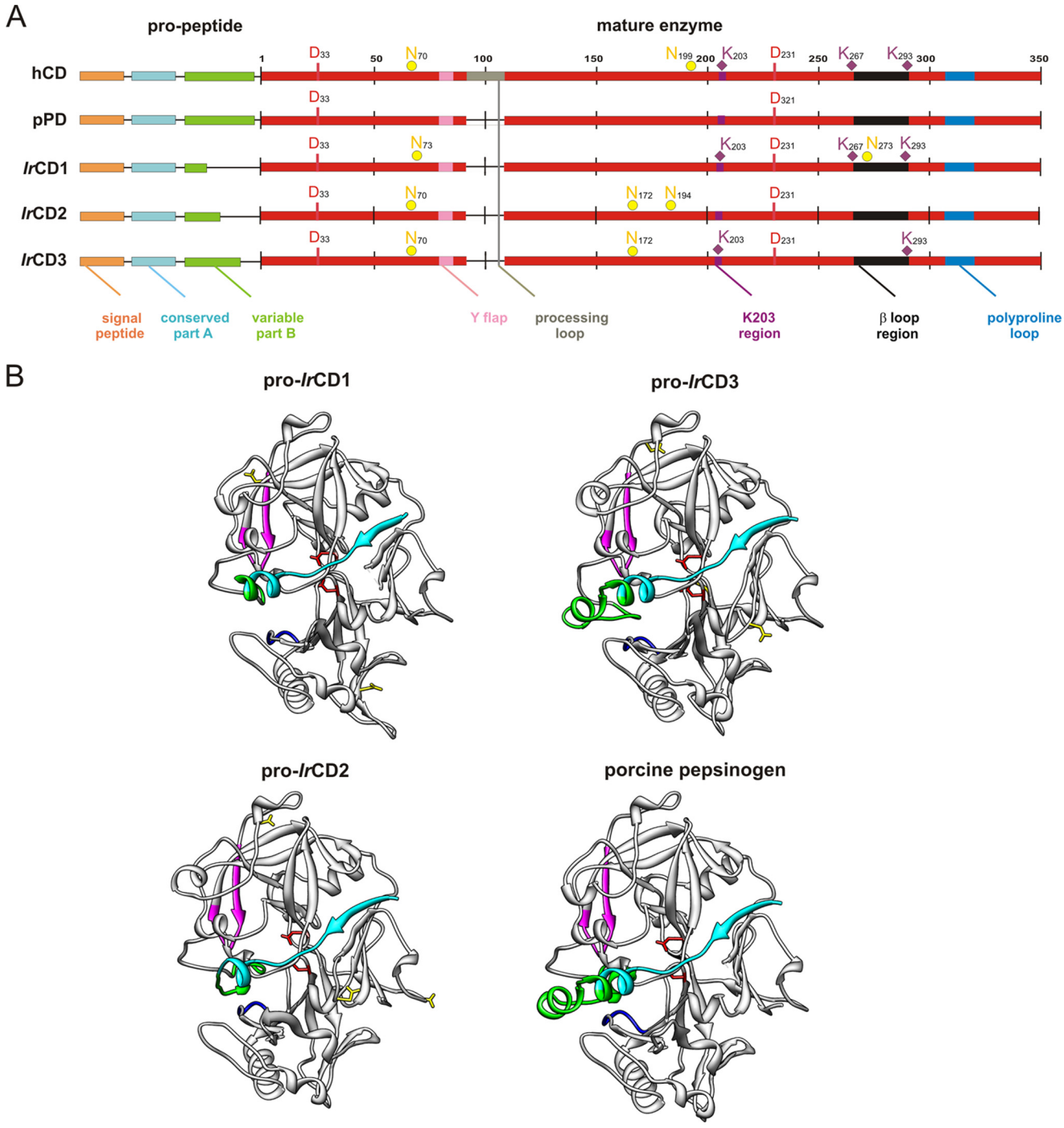

FIGURE 1. **Comparison of identified *I. ricinus* cathepsin D paralogs.** *A,* graphical scheme alignment of *Ir*CD1, *Ir*CD2, *Ir*CD3, lysosomal human cathepsin D (*hCD*), and extracellular porcine pepsinogen (*pPD*). Amino acid numbering is related to mature human cathepsin D. Domains are labeled with colors used in *B*. Depicted residues are shown in single letter amino acid coding. Two catalytic aspartic acid residues (*red* D), predicted *N*-glycosylations, *yellow full circles*, *N*; phosphorylation determinant lysines in the K203 region and in the β-loop region, *violet diamond* labels, *K*. (Note: full Clustal X alignment including two tick aspartic hemoglobinases longepsin and *Bm*AP could be found in supplemental File 1). *B,* tertiary structures of *Ir*CD1, *Ir*CD2, and *Ir*CD3 zymogens (pro-*Ir*CD1–3) modeled using x-ray structure of porcine pepsinogen (Protein Data Bank 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 residues (*red sticks*). Propeptide is divided into conservative part (*blue-green*) and variable (*green*) part, that is reduced in the structure of *Ir*CD1 and *Ir*CD2 compared with *Ir*CD3 and pPD. Y flap region, polyproline 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 *Ir*CD isoforms.

plus yolk cathepsin (30) and the heme-binding aspartic peptidase (tick heme-binding aspartic proteinase) (23) are apparently missing in the *I. scapularis* genome.

Spatial homology models of *Ir*CD1, *Ir*CD2, and *Ir*CD3 zymogens were constructed to compare the structural features of *Ir*CD isoforms (Fig. 1*B*). The x-ray structure of porcine pepsinogen (Protein Data Bank 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 *Ir*CDs contain two loops in the

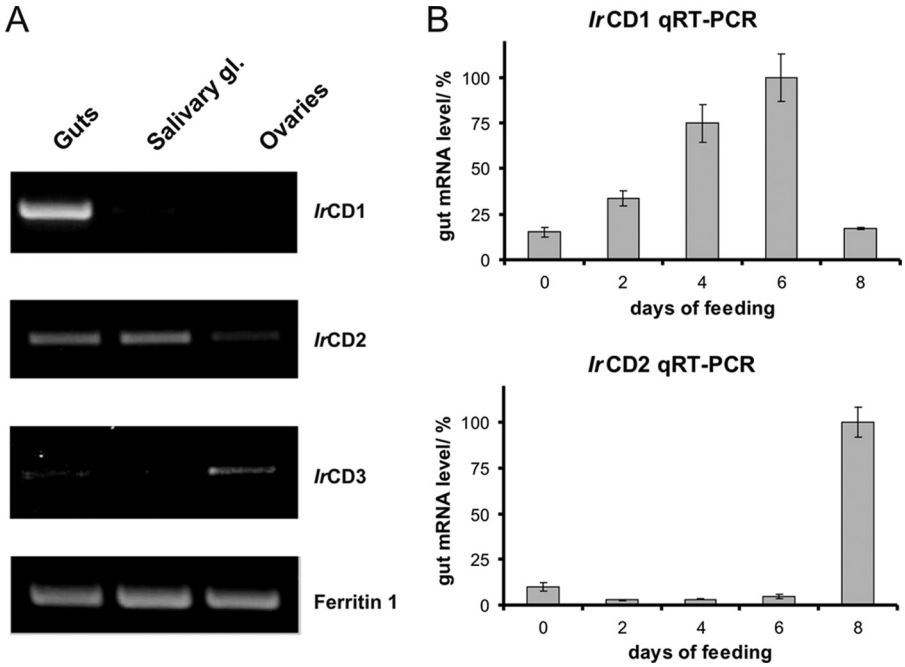

FIGURE 2. **Transcriptional profiling of identified *Ir*CDs.** *A,* transcriptional profiling of *Ir*CD forms in tissues of partially engorged *I. ricinus* females. Two-step RT-PCR was performed with *Ir*CD1, *Ir*CD2, and *Ir*CD3 gene-specific primers and first strand cDNA templates prepared from total RNA isolated from guts, salivary glands (*gl*), and ovaries. The identity of resulting PCR products was confirmed by DNA sequencing. Ferritin 1 primers (15) were used as template loading control. *Ir*CD1 is the only cathepsin D with gut tissue restricted expression. *B,* dynamics of *Ir*CD1 and *Ir*CD2 expression in female guts during feeding. Levels of *Ir*CD1 and *Ir*CD2 mRNAs were determined by qRT-PCR using dual labeled UPL probe 78 and 44 (Roche Applied Science), gene-specific primers, and gut cDNA templates from days 0, 2, 4, 6, and 8 of female feeding. Reactions were done in triplicate, and standard deviations are depicted. The levels of mRNA are normalized to the sample with maximum mRNA level (set to 100%). Relative level of *Ir*CD1 mRNA peaks in the 6th day before rapid engorgement, which is conditioned by fertilization, although the relative level of *Ir*CD2 mRNA peaks in fully fed fertilized females after rapid engorgement.

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 *Ir*CD1 is rearranged; the conserved Gly^{308} and Asp^{310} are mutated to serine and glutamic acid, respectively, and the loop contains three amino acid deletions at Ser^{315}–Pro^{317} (supplemental File 1), which is also present in the *I. scapularis* analog *Is*CD1 (data not shown).

Predicted *N*-glycosylation sites at Asn^{73} and Asn^{273} in the *Ir*CD1 proenzyme are positionally different from the three predicted *N*-glycosylation sites in the *Ir*CD2 (Asn^{70}, Asn^{172}, and Asn^{194}) and the *Ir*CD3 (Asn^{70} and Asn^{172}) proenzymes, respectively (Fig. 1). The positions of lysine residues in the "K203" and "β-loop" regions (Fig. 1*A*) play a critical role in the phosphotransferase recognition patch (32). *Ir*CD1 appears analogous to lysosomal human cathepsin D by possessing conserved Lys^{203}, Lys^{267}, and Lys^{293}. 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, Up-regulated by Feeding, and Localized in Digestive Cell Vesicles—RT-PCR profiling of *Ir*CDs demonstrated that *Ir*CD1 mRNA is restricted to the gut tissue of partially engorged female ticks. *Ir*CD2 is expressed in guts and salivary glands, whereas *Ir*CD3 mRNA is mostly produced in the ovaries (Fig. 2*A*). Thus, *Ir*CD3 was further excluded from studies on hemoglobinolysis. The dynamics of expression of *Ir*CD1 and *Ir*CD2 mRNAs in the gut tissue following feeding were analyzed by qRT-PCR (Fig. 2*B*). *Ir*CD1 mRNA peaks in partially engorged females (day 6 of feeding), whereas *Ir*CD2 mRNA increases in fully fed and detached females (day 8 of feeding). *Ir*CD1 protein abundance in the gut tissue during feeding was monitored by Western blotting using Ra×r*Ir*CD1. The *Ir*CD1 protein signal is detectable in partially and fully fed females (Fig. 3*A*). 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. 3*A*). Indirect immunofluorescence microscopy with Ra×r*Ir*CD1 and Alexa Fluor 488 dye-conjugated secondary antibodies localized *Ir*CD1 in the vesicles of digestive gut cells at the day 6 of feeding (Fig. 3*B*).

rIrCD1 Activates Autocatalytically and Displays Cathepsin D-like Substrate/Inhibitor Specificity—*Ir*CD1 zymogen was expressed in *E. coli* (r*Ir*CD1) and isolated with Co^{2+}-chelating chromatography under denaturing conditions, renatured, and further purified by ion-exchange FPLC. The correctly folded pro-r*Ir*CD1 was autocatalytically 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. 4*A*). Processing of the zymogen was followed by SDS-PAGE to demonstrate that activation was accompanied by autocatalytic proteolytic processing (Fig. 4*C*). Pro-*Ir*CD1 band (47 kDa) was completely converted to a 40-kDa band. This corresponds to the theoretical molecular mass of mature *Ir*CD1 predicted from the amino acid sequence. N-terminal amino acid sequencing of the 40-kDa band identi-

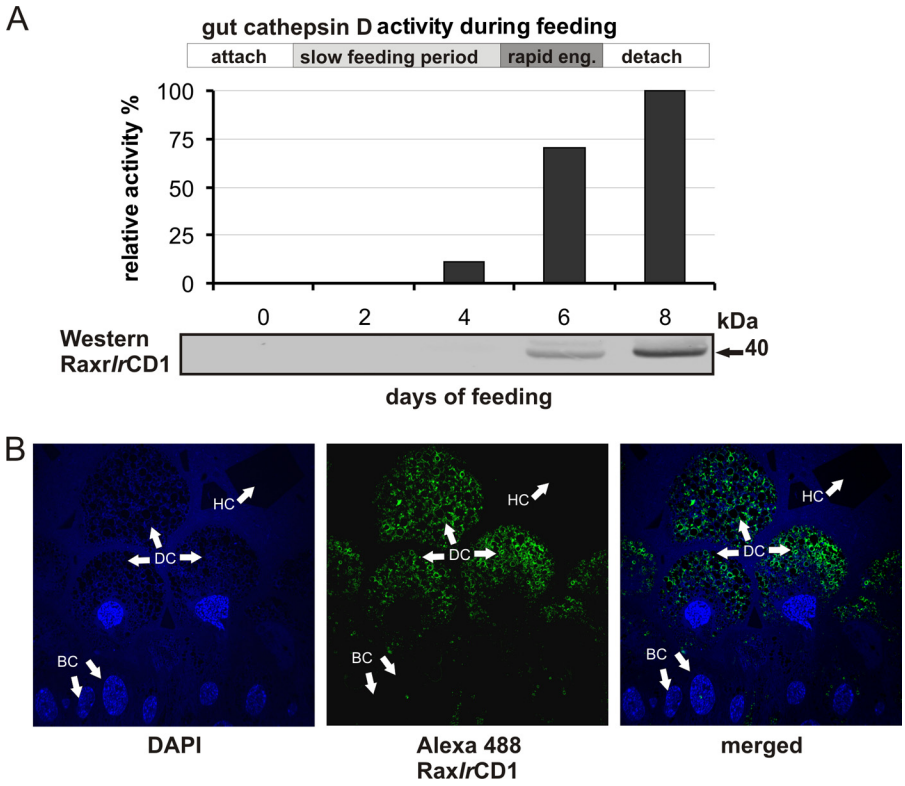

FIGURE 3. ***Ir*CD activity, *Ir*CD1 expressional profiling, and localization with specific antibodies.** *A,* Abz-KPAEFnFRL measured authentic cathepsin D activity in female *I. ricinus* gut extracts during feeding. Time line depicts feeding phases as follows: attachment, slow feeding period, rapid engorgement (*eng*), and detachment. Activity is increasing from the day 4 of feeding, peaks in fully fed females, and than slowly decreases. *Bottom panel*, Western blot, immunodetection with Ra×r*Ir*CD1 of authentic *Ir*CD1 in gut extracts in different days of feeding, signal raises by day 6 and peaks in day 8. *B,* indirect immunofluorescent microscopy, semi-thin section labeled with affinity purified Ra×r*Ir*CD1. The goat anti-rabbit 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.

fied a single sequence, Ile-His-Glu-Gly-Pro-Tyr, that was generated by the cleavage between Lys^{22} and Ile^{23} (pro-*Ir*CD1 numbering). This autoactivation cleavage site is located three 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 r*Ir*CD1 was determined using the synthetic FRET peptide substrate Abz-KPAEFnFRL and bovine hemoglobin (Fig. 4*B*). Both substrates were effectively cleaved at pH 2.5–5.0 with optimal activity at pH 4.0.

Inhibitory specificity of r*Ir*CD1 was determined using a panel of selective peptidase inhibitors (Table 2). r*Ir*CD1 activity was completely inhibited by pepstatin A (34) and potato cathepsin D inhibitor (35), and both specifically inactivate cathepsin D-like peptidases. Partial inhibition was observed with lopinavir (36), which targets the aspartic peptidases of the retropepsin family. r*Ir*CD1 activity was unaffected by Pefabloc (37), leupeptin (38), E-64 (39), and EDTA (40), which inhibit peptidases of serine, cysteine, and metallopeptidase classes, respectively. This inhibition profile confirms that *Ir*CD1 has ligand binding characteristics similar to mammalian cathepsin D.

Recombinant *Ir*CD1 and native *Ir*CD 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. 4*D*) (24). The labeled enzymes migrate on the SDS-PAGE as single bands of 40 and 45 kDa, respectively. The lower molecular weight of the r*Ir*CD1 band can be explained by the absence of *N*-glycosylations at two predicted sites (Asn^{73} and Asn^{273}, Fig. 1*A*; 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.

IrCD1 Has a Preference for Hydrophobic Residues at the P1 and P1′ Substrate Positions—A novel set of short peptidyl substrates and macromolecular hemoglobin was used to determine r*Ir*CD1 cleavage site specificity. r*Ir*CD1 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 time points between 5 and 1200 min 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 min (supplemental File 3). A heat map illustrating the frequency of residues found in the P4 to P4′ sites after 15 min incubation was generated. These data indicated that r*Ir*CD1 has a strong preference for hydrophobic residues in both the P1 and P1′ positions, although 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, r*Ir*CD1 tolerates most amino acids but displays

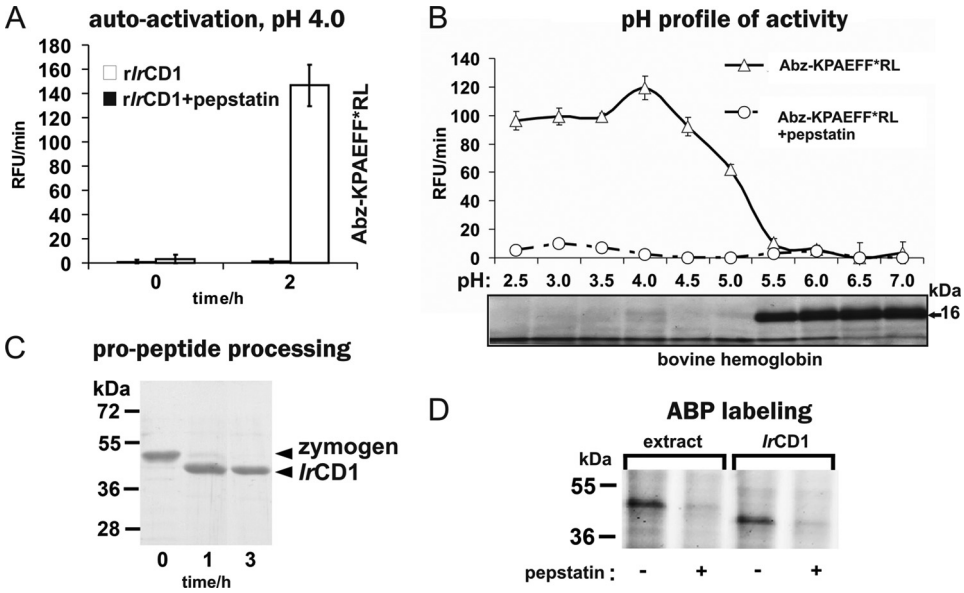

FIGURE 4. **Biochemical characterization and proteomic identification of r*Ir*CD1.** *A,* r*Ir*CD1 zymogen was activated at pH 4.0, and the resulting activity was measured with the Abz-KPAEFnFRL substrate. *B,* pH optimum of mature r*Ir*CD1 was determined with Abz-KPAEFnFRL substrate and compared with the pH profile of r*Ir*CD1 in hemoglobin degradation assay. Fragments were visualized on protein-stained SDS-PAGE (16-kDa band corresponds to hemoglobin monomer). *C,* time dependence of the r*Ir*CD1 zymogen processing at acidic pH analyzed by SDS-PAGE. Pro-r*Ir*CD1 was incubated at pH 4.0 for 1 and 3 h. *Lane 0* contains pro-r*Ir*CD1 before activation. *D,* in-gel visualization of activated r*Ir*CD1 and authentic *Ir*CD1 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 *Ir*CD1 by FAP-09.

TABLE 2

r*Ir*CD1 sensitivity to specific protease inhibitors

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

Inhibitor	Specificity/target protease	Concentration	Inhibition
			%
Pefabloc	Serine proteases (37)	1 mM	3 ± 1
E-64	Cysteine proteases (39)	10 μM	4 ± 2
Leupeptin	Serine and cysteine proteases (38)	10 μM	9 ± 4
EDTA	Metalloproteases (40)	10 μM	2 ± 1
Lopinavir	Aspartic proteases of retropepsin family (36)	10 μM	61 ± 5
Pepstatin	Aspartic proteases of pepsin and retropepsin families (34)	1 μM	99 ± 4
Potato cathepsin D inhibitor (PDI)	Aspartic proteases of cathepsin D type (35)	1 μM	100 ± 2

a preference for tyrosine at P3 and alanine at P2′. Further analysis of the cleavage site location within the peptide substrate indicated that r*Ir*CD1 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. 5*A*). 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. 5*B*). This represents all possible P2 to P2′ residues with Phe in the P1 site. Cleavage sites were characterized by the time at which peptide products were first observed. Although a Phe in the P1 site is the major specificity element for r*Ir*CD1, 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 amino acid 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 or not at all in substrates where the P1′ residue is not optimal. When cleavage occurred at sites with a nonoptimal P1′ residue, such as IF ↓ EI, then residues at P2 and P2′ were often preferred.

Hemoglobin digested by r*Ir*CD1 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. 5*C*. 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 min of incubation with r*Ir*CD1, four cleavage sites were identified, Leu^{29}–Glu^{30}, Phe^{33}–Leu^{34}, and Leu^{109}–Ala^{110} in the α-subunit and Phe^{40}–Phe^{41} in the β-subunit. These initial hemoglobin cleavage sites of r*Ir*CD1 are equal to those reported previously for the gut tissue extract cathepsin D activity (5). The overall identity of cleavage sites found for the r*Ir*CD1 and the authentic enzyme is 62%.

RNAi Confirms the Function of IrCD1 as an Intestinal Aspartic Hemoglobinase—The major contribution of *Ir*CD1 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. The *Ir*CD1 tran-

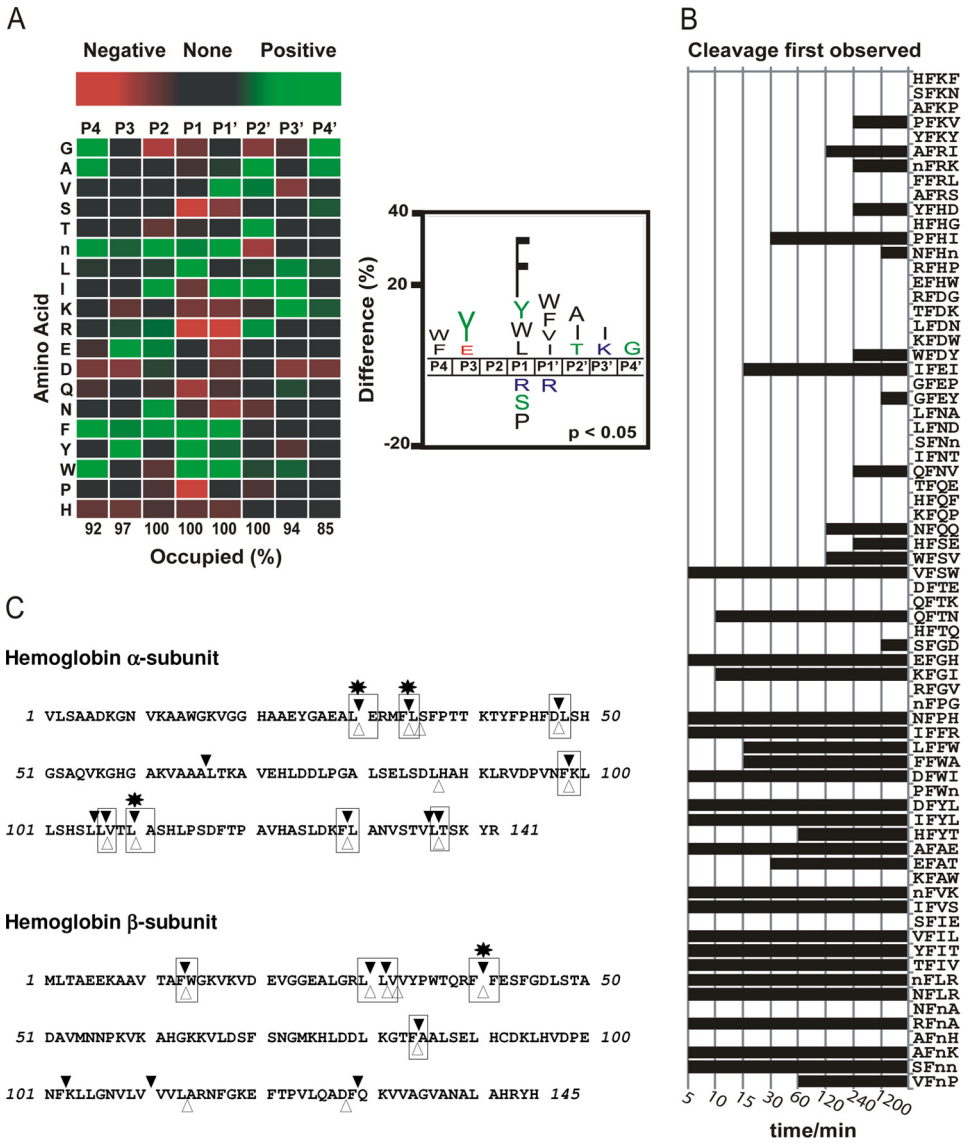

FIGURE 5. **Substrate specificity profiling of r*Ir*CD1-screening of novel highly diversified peptide library consisting of 124 synthetic tetradecapeptides and hemoglobin r*Ir*CD1 cleavage map.** *A*, a heat map and iceLogo reflecting the frequency of residues in the P4 to P4′ positions of 97 unique r*Ir*CD1 cleavage sites. Each amino acid is represented by the single letter code, and *n* corresponds to norleucine. In the heat map *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, whereas 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 r*Ir*CD cleavage products observed between 5 and 1200 min 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 r*Ir*CD1 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 min of reaction) are marked with *asterisks*. Numbering of the α-subunit residues excludes the first mRNA translated methionine of the hemoglobin precursor.

script, as well as protein synthesis and specific cathepsin D activity, was dramatically reduced (Fig. 6) compared with GFP-dsRNA-treated control ticks. Gene specificity of RNAi was verified by qRT-PCR. The level of *Ir*CD1 mRNA in *Ir*CD1-dsRNA-treated tick group was reduced to 16% of the GFP control tick group, although the change in the expressional level of *Ir*CD2 mRNA was still ~90% compared with GFP control (not significant, Fig. 6*A*). In kinetic assays with Abz-KPAEFnFRL and labeled hemoglobin substrates, the overall gut extract cathepsin D activity was reduced in the *Ir*CD1 tick group to 20 and 10%, respectively (Fig. 6*C*). Phenotype markers, mortality, weight, oviposition, and larvae hatching (41) displayed no statistically significant changes in *Ir*CD1-dsRNA-treated ticks (data not shown).

DISCUSSION

Since the 1970s, none of the digestive tick cathepsin D reports (23, 26, 42, 43) reflected the functional diversity of multiple tick cathepsin D paralogs 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

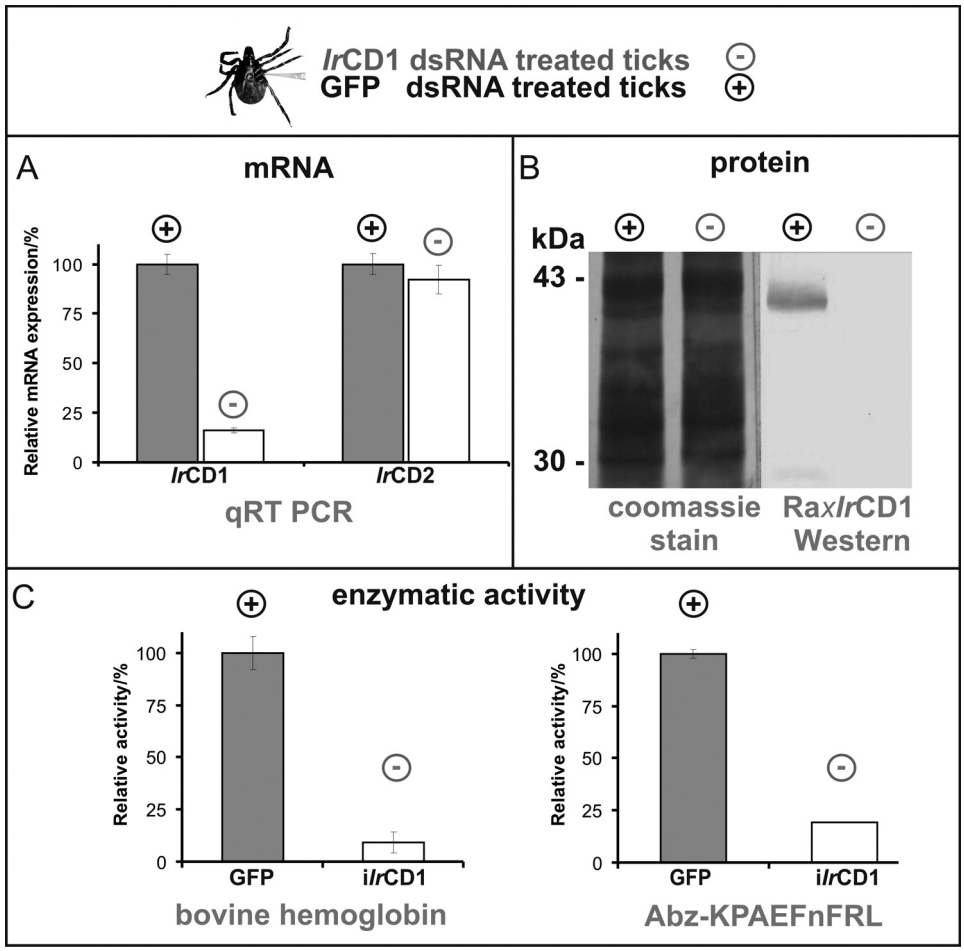

FIGURE 6. **RNAi knockdown of *Ir*CD1.** *I. ricinus* females (25 per each group) were injected with *Ir*CD1 dsRNA (*iIrCD1*, minus, experimental group) and/or GFP dsRNA (*GFP*, plus, control group). Gut tissue of partially engorged ticks was used for RNA/DNA isolation and preparation of tissue homogenates. *A,* effect on mRNA expression levels. Dual-labeled UPL probe 78 (Roche Applied Science) and *Ir*CD1-specific primers were used for qRT-PCR analysis of *Ir*CD1 RNAi effect between the i*Ir*CD1 and GFP tick groups. The presence of *Ir*CD1 mRNA is decreased to 16% in the relative level compared with the GFP control group. For gene specificity of the *Ir*CD1 knockdown, qRT PCR with UPL probe 44 (Roche Applied Science) and *Ir*CD2-specific primers was performed. *Ir*CD2 mRNA level showed no significant decrease in between the two tick groups. All PCRs were performed in triplicate. Relative values are depicted with standard deviations. *B,* effect on protein abundance. SDS-PAGE separated gut extracts of i*Ir*CD1 and GFP ticks were electrotransferred to PVDF membrane (Coomassie Blue-stained lines for loading control). Presence of *Ir*CD1 is determined by Western blot with antibody Ra×r*Ir*CD1. No signal in i*Ir*CD1 tick group line compared with ~40-kDa r*Ir*CD1 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, between GFP and i*Ir*CD1 tick groups. Measuring was performed in triplicates. Relative values are depicted with standard deviations.

*Ir*CD forms most likely evolved by gene duplication and ongoing mutation within the Acari group (supplemental File 2). We hypothesize that the structural modifications of *Ir*CD1 arose with the adaptation of ticks/mites to a blood-feeding lifestyle.

Our results clearly identify *Ir*CD1 as the only *I. ricinus* cathepsin D exclusively expressed in the gut. The hemoglobinolytic role of *Ir*CD1 was previously indicated by missing RT-PCR signals in *I. ricinus* life stages not feeding on hosts (4). Dynamic up-regulation of *Ir*CD1 mRNA was found to be consistent with the rapidly increasing intestinal aspartic peptidase activity during several days of tick feeding (16). The presence of *Ir*CD1 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 *Ir*CD1 shows a distribution through vesicles, most likely lysosomes and endosomes of gut cells. This localization is analogous to previously characterized tick hemoglobinolytic cysteine peptidases *Ir*AE1 (12), *Ir*CB1 (16), *Ir*CL1 (11), and the cathepsin L from *R./B. microplus* (47). We propose that *Ir*CD1 is targeted to endolysosomes via the mannose 6-phosphate pathway due to the presence of an identical recognition patch (Lys^{203}, Lys^{267}, and Lys^{293}; Fig. 1*A*) for the lysosomal *N*-acetylglucosamine-1-phosphotransferase (UDP-GlcNAc) within *Ir*CD1 and the human lysosomal cathepsin D (33). In addition to the difference in the post-translational modification patterns of *Ir*CD1 and the *Bm*AP/longepsin analog *Ir*CD2 (supplemental File 1), the putative role of *Ir*CD2 in the secretory or extracellular processes is supported by *Ir*CD2 expression in both salivary glands and guts, where the expression peaks in fully fed females (Fig. 2*B*).

Our current tick hemoglobinolytic enzyme network model includes primary hemoglobinolytic endopeptidases *Ir*CD1,

*Ir*CL1, and *Ir*AE1 (10). All three peptidases prepared as recombinant proteins are capable of autocatalytic activation in acidic conditions indicating the process to occur upon the acidification of gut cell endosomes (for review see Ref. 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 r*Ir*CD1 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 r*Ir*CD1 and the native *I. ricinus* gut cathepsin D show identical preference for Phe and Leu at P1 (5).

*Ir*CD1 RNAi experiments decreased the overall gut aspartic peptidase activity by >90% demonstrating the major contribution of *Ir*CD1 to the cathepsin D activity of the multienzyme hemoglobinolytic network (5). The decrease excludes hemoglobinolytic activity of papain-like enzymes due to the presence of 10 μM E-64 in the assay. We propose that the remaining 10% of activity in *Ir*CD1 knocked down ticks (Fig. 6*C*) is not sufficient enough to off-set the appearance of obvious RNAi phenotypes with respect to tick weight, mortality, oviposition, and larvae hatching. The missing RNAi phenotype may be better explained by the synchronous operation between *Ir*CD1 and the other initial hemoglobinase *Ir*CL1, which is also remarkable from hemoglobinolytic assays performed with gut extracts (5). The redundancy between *Ir*CD1 and *Ir*CL1 was also indicated by the only changed phenotype marker (tick average weight decreased by 24%) in the *Ir*CL1 RNAi knocked down ticks (11). Recently, Cruz *et al.* (27) confirmed in *B.* (*R.*) *microplus* that *Bm*AP and *Bm*CL are together responsible for the generation of hemoglobin-derived antimicrobial peptides. Because of dsRNA dose and volume limits for each tick female, our current method has not yet allowed us to obtain multiple endo-peptidase RNAi knockdowns to confirm the cathepsin L/D redundancy (data not shown).

To conclude, this report identifies and characterizes *Ir*CD1 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 *Ir*AE (12) and *Ir*CL1 (11), this study completes the biochemical and functional characterization of those primary individual endopeptidases of the network. Although *Ir*CD1 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 *Schistosoma mansoni* (7). However, 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.

Acknowledgment—We thank Miloslav Sanda for the mass spectroscopy.

REFERENCES

1. Mans, B. J. (2011) Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. *J. Innate Immun.* **3,** 41–51
2. Mans, B. J., Louw, A. I., and Neitz, A. W. (2002) Evolution of hematophagy in ticks. Common origins for blood coagulation and platelet aggregation inhibitors from soft ticks of the genus *Ornithodoros*. *Mol. Biol. Evol.* **19,** 1695–1705
3. Tarnowski, B. I., and Coons, L. B. (1989) Ultrastructure of the midgut and blood meal digestion in the adult tick *Dermacentor variabilis*. *Exp. Appl. Acarol.* **6,** 263–289
4. Sojka, D., Franta, Z., Horn, M., Hajdusek, O., Caffrey, C. R., Mares, M., and Kopácek, P. (2008) Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases. *Parasit. Vectors* **1,** 7
5. Horn, M., Nussbaumerová, M., Sanda, M., Kovárová, Z., Srba, J., Franta, Z., Sojka, D., Bogyo, M., Caffrey, C. R., Kopácek, P., and Mares, M. (2009) Hemoglobin digestion in blood-feeding ticks. Mapping a multipeptidase pathway by functional proteomics. *Chem. Biol.* **16,** 1053–1063
6. Caffrey, C. R., McKerrow, J. H., Salter, J. P., and Sajid, M. (2004) Blood 'n' guts. An update on schistosome digestive peptidases. *Trends Parasitol.* **20,** 241–248
7. Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvorák, J., Hsieh, I., Bahgat, M., Dissous, C., and McKerrow, J. H. (2006) A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* **281,** 39316 –39329
8. Williamson, A. L., Brindley, P. J., Knox, D. P., Hotez, P. J., and Loukas, A. (2003) Digestive proteases of blood-feeding nematodes. *Trends Parasitol.* **19,** 417–423
9. Wu, D. D., Wang, G. D., Irwin, D. M., and Zhang, Y. P. (2009) A profound role for the expansion of trypsin-like serine protease family in the evolution of hematophagy in mosquito. *Mol. Biol. Evol.* **26,** 2333–2341
10. Sojka, D., Francischetti, I. M., Calvo, E., and Kotsyfakis, M. (2011) Cysteine proteases from bloodfeeding arthropod ectoparasites. *Adv. Exp. Med. Biol.* **712,** 177–191
11. Franta, Z., Sojka, D., Frantova, H., Dvorak, J., Horn, M., Srba, J., Talacko, P., Mares, M., Schneider, E., Craik, C. S., McKerrow, J. H., Caffrey, C. R., and Kopacek, P. (2011) IrCL1- the hemoglobinolytic cathepsin L of the hard tick *Ixodes ricinus*. *Int. J. Parasitol.* **41,** 1253–1262
12. Sojka, D., Hajdusek, O., Dvorák, J., Sajid, M., Franta, Z., Schneider, E. L., Craik, C. S., Vancová, M., Buresová, V., Bogyo, M., Sexton, K. B., McKerrow, J. H., Caffrey, C. R., and Kopácek, P. (2007) IrAE. An asparaginyl endopeptidase (legumain) in the gut of the hard tick *Ixodes ricinus*. *Int. J. Parasitol.* **37,** 713–724
13. Nijhof, A. M., Balk, J. A., Postigo, M., and Jongejan, F. (2009) Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus* (*Boophilus*) *microplus* and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. *BMC Mol. Biol.* **10,** 112
14. Grunclová, L., Horn, M., Vancová, M., Sojka, D., Franta, Z., Mares, M., and Kopácek, P. (2006) Two secreted cystatins of the soft tick *Ornithodoros moubata*. Differential expression pattern and inhibitory specificity. *Biol. Chem.* **387,** 1635–1644
15. Kopácek, P., Zdychová, J., Yoshiga, T., Weise, C., Rudenko, N., and Law, J. H. (2003) Molecular cloning, expression, and isolation of ferritins from two tick species, *Ornithodoros moubata* and *Ixodes ricinus*. *Insect Biochem. Mol. Biol.* **33,** 103–113
16. Franta, Z., Frantová, H., Konvičková, J., Horn, M., Sojka, D., Mareš, M., and Kopáček, P. (2010) Dynamics of digestive proteolytic system during blood feeding of the hard tick *Ixodes ricinus*. *Parasit. Vectors* **3,** 119
17. Kopácek, P., Weise, C., and Götz, P. (1995) The prophenol oxidase from the wax moth *Galleria mellonella*. Purification and characterization of the proenzyme. *Insect Biochem. Mol. Biol.* **25,** 1081–1091
18. Hajdusek, O., Sojka, D., Kopacek, P., Buresova, V., Franta, Z., Sauman, I.,

Winzerling, J., and Grubhoffer, L. (2009) Knockdown of proteins involved in iron metabolism limits tick reproduction and development. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 1033–1038
19. Levashina, E. A., Moita, L. F., Blandin, S., Vriend, G., Lagueux, M., and Kafatos, F. C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104,** 709–718
20. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface. Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882
21. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web. A case study using the Phyre server. *Nat. Protoc.* **4,** 363–371
22. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera. A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612
23. Sorgine, M. H., Logullo, C., Zingali, R. B., Paiva-Silva, G. O., Juliano, L., and Oliveira, P. L. (2000) A heme-binding aspartic proteinase from the eggs of the hard tick *Boophilus microplus*. *J. Biol. Chem.* **275,** 28659–28665
24. Nussbaumerová, M., Srp, J., Mása, M., Hradilek, M., Sanda, M., Reinis, M., Horn, M., and Mares, M. (2010) Single- and double-headed chemical probes for detection of active cathepsin D in a cancer cell proteome. *ChemBioChem* **11,** 1538–1541
25. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **6,** 786–787
26. Boldbaatar, D., Sikalizyo Sikasunge, C., Battsetseg, B., Xuan, X., and Fujisaki, K. (2006) Molecular cloning and functional characterization of an aspartic protease from the hard tick *Haemaphysalis longicornis*. *Insect Biochem. Mol. Biol.* **36,** 25–36
27. Cruz, C. E., Fogaça, A. C., Nakayasu, E. S., Angeli, C. B., Belmonte, R., Almeida, I. C., Miranda, A., Miranda, M. T., Tanaka, A. S., Braz, G. R., Craik, C. S., Schneider, E., Caffrey, C. R., and Daffre, S. (2010) Characterization of proteinases from the midgut of *Rhipicephalus* (*Boophilus*) *microplus* involved in the generation of antimicrobial peptides. *Parasit. Vectors* **3,** 63
28. Metcalf, P., and Fusek, M. (1993) Two crystal structures for cathepsin D. The lysosomal targeting signal and active site. *EMBO J.* **12,** 1293–1302
29. Horimoto, Y., Dee, D. R., and Yada, R. Y. (2009) Multifunctional aspartic peptidase prosegments. *N. Biotechnol.* **25,** 318–324
30. Logullo, C., Vaz Ida, S., Sorgine, M. H., Paiva-Silva, G. O., Faria, F. S., Zingali, R. B., De Lima, M. F., Abreu, L., Oliveira, E. F., Alves, E. W., Masuda, H., Gonzales, J. C., Masuda, A., and Oliveira, P. L. (1998) Isolation of an aspartic proteinase precursor from the egg of a hard tick, *Boophilus microplus*. *Parasitology* **116,** 525–532
31. Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* **102,** 4431–4458
32. Baranski, T. J., Koelsch, G., Hartsuck, J. A., and Kornfeld, S. (1991) Mapping and molecular modeling of a recognition domain for lysosomal enzyme targeting. *J. Biol. Chem.* **266,** 23365–23372
33. Steet, R., Lee, W. S., and Kornfeld, S. (2005) Identification of the minimal lysosomal enzyme recognition domain in cathepsin D. *J. Biol. Chem.* **280,** 33318–33323
34. Knight, C. G., and Barrett, A. J. (1976) Interaction of human cathepsin D with the inhibitor pepstatin. *Biochem. J.* **155,** 117–125
35. Mares, M., Meloun, B., Pavlik, M., Kostka, V., and Baudys, M. (1989) Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. *FEBS Lett.* **251,** 94–98
36. Hurst, M., and Faulds, D. (2000) Lopinavir. *Drugs* **60,** 1371–1379
37. James, G. T. (1978) Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers. *Anal. Biochem.* **86,** 574–579
38. Aoyagi, T., Takeuchi, T., Matsuzaki, A., Kawamura, K., and Kondo, S. (1969) Leupeptins, new protease inhibitors from *Actinomycetes*. *J. Antibiot.* **22,** 283–286
39. Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M., and Hanada, K. (1982) L-*trans*-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogs as inhibitors of cysteine proteinases including cathepsins B, H, and L. *Biochem. J.* **201,** 189–198
40. Auld, D. S. (1988) Use of chelating agents to inhibit enzymes. *Methods Enzymol.* **158,** 110–114
41. Hajdusek, O., Almazán, C., Loosova, G., Villar, M., Canales, M., Grubhoffer, L., Kopacek, P., and de la Fuente, J. (2010) Characterization of ferritin 2 for the control of tick infestations. *Vaccine* **28,** 2993–2998
42. Akov, S., Samish, M., and Galun, R. (1976) Protease activity in female *Ornithodoros tholozani* ticks. *Acta Trop.* **33,** 37–52
43. Vundla, W. R., Brossard, M., Pearson, D. J., and Labongo, V. L. (1992) Characterization of aspartic proteinases from the gut of the tick, *Rhipicephalus appendiculatus neuman*. *Insect Biochem. Mol. Biol.* **22,** 405–410
44. Mendiola, J., Alonso, M., Marquetti, M. C., and Finlay, C. (1996) Boophilus microplus, Multiple proteolytic activities in the midgut. *Exp. Parasitol.* **82,** 27–33
45. Lara, F. A., Lins, U., Bechara, G. H., and Oliveira, P. L. (2005) Tracing heme in a living cell. Hemoglobin degradation and heme traffic in digest cells of the cattle tick *Boophilus microplus*. *J. Exp. Biol.* **208,** 3093–3101
46. Doherty, G. J., and McMahon, H. T. (2009) Mechanisms of endocytosis. *Annu. Rev. Biochem.* **78,** 857–902
47. Renard, G., Lara, F. A., de Cardoso, F. C., Miguens, F. C., Dansa-Petretski, M., Termignoni, C., and Masuda, A. (2002) Expression and immunolocalization of a *Boophilus microplus* cathepsin L-like enzyme. *Insect Mol. Biol.* **11,** 325–328
48. Luzio, J. P., Parkinson, M. D., Gray, S. R., and Bright, N. A. (2009) The delivery of endocytosed cargo to lysosomes. *Biochem. Soc. Trans.* **37,** 1019–1021
49. Gabay, T., and Ginsburg, H. (1993) Hemoglobin denaturation and iron release in acidified red blood cell lysate. A possible source of iron for intraerythrocytic malaria parasites. *Exp. Parasitol.* **77,** 261–272
50. Liu, P., Marzahn, M. R., Robbins, A. H., Gutiérrez-de-Terán, H., Rodríguez, D., McClung, S. H., Stevens, S. M., Jr., Yowell, C. A., Dame, J. B., McKenna, R., and Dunn, B. M. (2009) Recombinant plasmepsin 1 from the human malaria parasite *Plasmodium falciparum*. Enzymatic characterization, active site inhibitor design, and structural analysis. *Biochemistry* **48,** 4086–4099
51. Pimenta, D. C., Oliveira, A., Juliano, M. A., and Juliano, L. (2001) Substrate specificity of human cathepsin D using internally quenched fluorescent peptides derived from reactive site loop of kallistatin. *Biochim. Biophys. Acta* **1544,** 113–122
52. Hartsuck, J. A., Koelsch, G., and Remington, S. J. (1992) The high resolution crystal structure of porcine pepsinogen. *Proteins* **13,** 1–25