Research

Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases

Daniel Sojka*[1], Zdeněk Franta[1,3], Martin Horn[2], Ondřej Hajdušek[3], Conor R Caffrey[4], Michael Mareš[2] and Petr Kopáček[1]

Address: [1]Institute of Parasitology, Biology Centre, Academy of Sciences of the Czech Republic, České Budějovice, CZ-370 05, The Czech Republic, [2]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Praha, CZ-166 10, The Czech Republic, [3]Faculty of Science, University of South Bohemia, České Budějovice, CZ-370 05, The Czech Republic and [4]Sandler Center for Basic Research in Parasitic Diseases, University of California San Francisco, San Francisco, CA 94158, USA

Email: Daniel Sojka* - dsojka@seznam.cz; Zdeněk Franta - zdeny@paru.cas.cz; Martin Horn - horn@marilyn.uochb.cas.cz; Ondřej Hajdušek - hajdus@paru.cas.cz; Conor R Caffrey - caffrey@cgl.ucsf.edu; Michael Mareš - mares@marilyn.uochb.cas.cz; Petr Kopáček - kopajz@paru.cas.cz

* Corresponding author

Abstract

Background: Ticks are vectors for a variety of viral, bacterial and parasitic diseases in human and domestic animals. To survive and reproduce ticks feed on host blood, yet our understanding of the intestinal proteolytic machinery used to derive absorbable nutrients from the blood meal is poor. Intestinal digestive processes are limiting factors for pathogen transmission since the tick gut presents the primary site of infection. Moreover, digestive enzymes may find practical application as anti-tick vaccine targets.

Results: Using the hard tick, *Ixodes ricinus*, we performed a functional activity scan of the peptidase complement in gut tissue extracts that demonstrated the presence of five types of peptidases of the cysteine and aspartic classes. We followed up with genetic screens of gut-derived cDNA to identify and clone genes encoding the cysteine peptidases cathepsins B, L and C, an asparaginyl endopeptidase (legumain), and the aspartic peptidase, cathepsin D. By RT-PCR, expression of asparaginyl endopeptidase and cathepsins B and D was restricted to gut tissue and to those developmental stages feeding on blood.

Conclusion: Overall, our results demonstrate the presence of a network of cysteine and aspartic peptidases that conceivably operates to digest host blood proteins in a concerted manner. Significantly, the peptidase components of this digestive network are orthologous to those described in other parasites, including nematodes and flatworms. Accordingly, the present data and those available for other tick species support the notion of an evolutionary conservation of a cysteine/aspartic peptidase system for digestion that includes ticks, but differs from that of insects relying on serine peptidases.

Background

Ticks are important vectors of infectious agents causing diseases in human and domestic animals [1]. The castor bean tick *Ixodes ricinus* transmits Lyme disease caused by *Borrelia burgdorferi* spirochetes and tick borne encephalitis caused by the tick-borne encephalitis virus [2].

Blood-feeding and -digestion are essential activities for ticks. Blood provides a rich source of proteins and nutrients for anabolic processes such as vitellogenesis and egg production [3]. Unlike other blood-feeding arthropods, ticks are believed to digest blood intracellularly – in the endo/lysosomal vesicles of gut cells [4] at pH values well below the pH 6.3 – 6.5 of the gut contents [5,6]. Digestive gut cells use both receptor-mediated and fluid-phase endocytoses to uptake the liquid blood meal from the gut lumen [7]. Lara et al. [8] showed that the digestive cells of *Boophilus microplus* have separated endocytic pathways for two major proteins of host blood – serum albumin and hemoglobin. The requirement for receptor-mediated endocytosis might be directly linked to the detoxification of released heme groups during intracellular digestion of hemoglobin. Most of the toxic heme forms a unique type of heme aggregate ultimately accumulated inside specialized organelles called hemosomes [9]. The virtual absence of extracellular digestive enzymes in ticks enables the gut lumen to serve as a major storage organ [4].

In spite of the above studies, our understanding of the molecular proteolytic machinery involved in digesting host proteins in the tick gut is still rather fragmented. Previous studies have tended to focus on individual enzymes in particular species; all however, have identified either cysteine or aspartic peptidases; e.g., a cysteine class cathepsin L in *B. microplus* [10], two forms of a cathepsin L in *Haemaphysalis longicornis* [11] and the aspartic peptidase, cathepsin D (termed longepsin) in *H. longicornis* [12]. Also, cysteine-class asparaginyl endopeptidases (AE, legumains) have been characterized in *I. ricinus* [13] and *H. longicornis* [14].

The data thus far from different tick species raise the hypothesis that tick intestinal digestion relies on an evolutionarily conserved network of cysteine and aspartic peptidases characterized in other parasites, including platyhelminths [15,16] and nematodes [17]. It comprises mainly cysteine peptidases cathepsin B, L, C, asparaginyl endopeptidase/legumain and an aspartic peptidase cathepsin D. To address this hypothesis we focused on a defined feeding phase of a single tick species, namely partially engorged females of *I. ricinus*. Two-pronged profiling strategy involving biochemical assays and PCR-based cloning displayed a simultaneous expression and activity of the above listed peptidase types in the tick digestive tissue. An improved global insight increases the possibilities

for practical interventions involving vaccines and offers a better understanding of vector-pathogen interactions at the primary interface, namely the tick gut.

Results

Functional profiling of multiple peptidase activities in the gut of *I. ricinus*

Gut tissue extract prepared from partially engorged *I. ricinus* females (5th day of feeding) was tested on degradation of the biologically relevant substrate hemoglobin. Proteolysis was analyzed using a fluorescence assay incorporating AMC-hemoglobin. Optimal proteolysis was measured at acidic pH between 3.0 and 4.5 (Fig. 1). No substantial degradation occurred above pH 6.0. At optimum pH (~4.0), hemoglobinolytic activity was inhibited by the small molecule inhibitors E64 and pepstatin that selectively target cysteine and aspartic peptidases, respectively. A combined application of both compounds resulted in nearly complete blockage of the activity (~97% inhibition), the individual treatment showed about 80% inhibition (~83% and ~78% for E64 and pepstatin, respectively). In contrast, inhibitors of serine peptidases and metallopeptidases Pefabloc and EDTA, respectively, were ineffective (data not shown).

Next, we focused on dissecting the major component peptidases in gut extracts responsible for the acidic degradation of blood meal using peptidase selective substrates and inhibitors (Fig. 2). At pH 4.0, hydrolytic activity cleav-
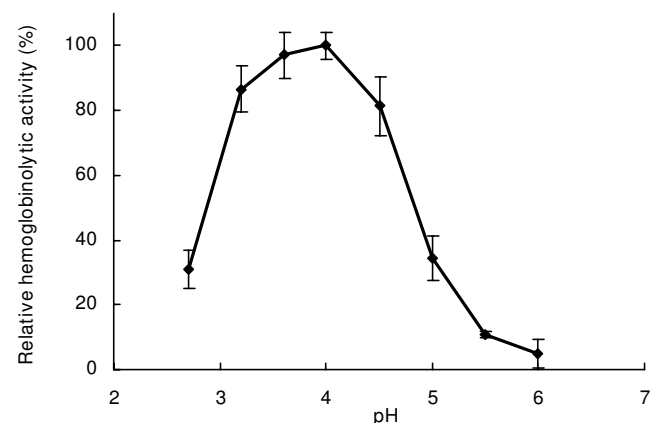

Figure 1
The pH profile of hemoglobinolytic activity in the *I. ricinus* gut tissue extracts. Gut tissue was dissected from partially engorged tick females (the 5-th day of feeding), washed from the gut contents and extracted. Fluorescently labeled AMC-hemoglobin was digested *in vitro* with gut extract at various pH values. The relative rate of degradation of the substrate was determined using the measurement of fluorescence in a continuous hemoglobinolytic assay. The error bars indicate standard deviations of the mean of triplicates.
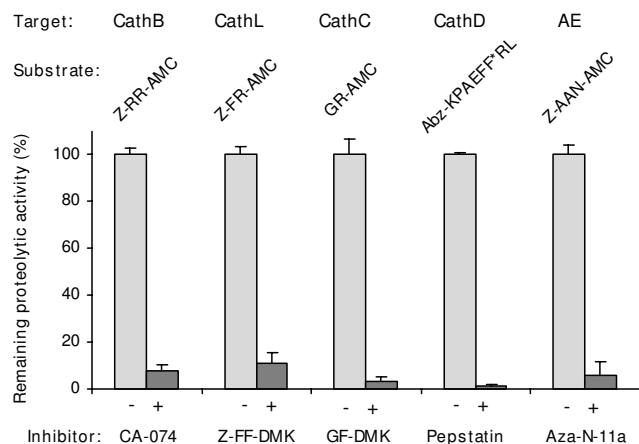
Figure 2
Activity profiling of *I. ricinus* gut peptidases. Peptidolytic activities in the gut tissue extract of partially engorged tick females (the 5-th day of feeding) were demonstrated *in vitro* with selective peptide substrates (structure of the fluorogenic substrates is indicated). The activity for individual substrates was suppressed in the presence of selective peptidase inhibitors to obtain diagnostic responses indicative of a protease type. Values are expressed as percent inhibition of the control activities. The identified major activities (Target) correspond to papain-type peptidases cathepsin B, L and C (CathB, CathL and CathC, respectively), cathepsin D-like aspartic peptidase (CathD) and asparaginyl endopeptidase (AE). The assay was performed at pH 4.0, an optimum pH for hemoglobin degradation by the gut extract. The activity of AE and CathL was measured in the presence of CA-074 inhibitor to prevent an interference with the activity of CathB. The error bars indicate standard deviations of the mean of triplicates.

ing the substrate Z-Arg-Arg-AMC (i.e., suggestive of cathepsin B activity) was inhibited 90% by the cathepsin B inhibitor, CA-074. Likewise, activity against Z-Phe-Arg-AMC was inhibited 90% by Z-Phe-Phe-DMK. Dipeptidyl peptidase activity of cathepsin C was measured with Gly-Arg-AMC and inhibited > 95% by Gly-Phe-DMK. Asparaginyl endopeptidase activity as measured with Z-Ala-Ala-Asn-AMC was inhibited > 95% by the azapeptide, Aza-N-

11a. Cathepsin D-like activity measured with Abz-Lys-Pro-Ala-Glu-Phe-Nph-Ala-Leu was effectively inhibited (~98%) by pepstatin.

Thus, five significant endo- and exopeptidase activities of the cysteine and aspartic classes of peptidases were profiled in the gut tissue of *I. ricinus*. The identified activities include (i) the CA clan (papain-type) cysteine peptidases: cathepsins B, L, and C and a Clan CD asparaginyl endopeptidase (legumain), and (ii) a Clan AA aspartic peptidase activity: cathepsin D.

Genetic screening of gut tissue identifies cDNAs encoding one aspartic and four cysteine peptidases

Single stranded cDNA derived from the gut of *I. ricinus* was used as a template for identification and cloning of genes encoding cysteine and aspartic peptidase precursors. Based on results of the functional screening of peptidase activities, multiple protein and cDNA alignments were performed to identify conserved domains in cathepsins B, L, D and C as shown for the schistosomal orthologues in Table 1. Degenerate primers derived from these motifs are listed together with expected PCR product lengths, optimal annealing temperatures, number of sequenced clones and number of identified isoforms (Table 1). Two isoforms of cathepsin B and one form each of cathepsins L, C and D were identified.

Hybridization screening of the gut cDNA library with radio-labeled PCR amplicons had been previously used to obtain the full-length coding sequence for IrAE [13]. Here the same approach succeeded in identifying full length cDNA sequences for cathepsins B, L and D. The full cDNA sequence of cathepsin C was generated by overlapping PCR, 5' and 3' RACE PCR fragments.

For the sake of consistency, we adopted a nomenclature for these enzymes previously used for schistosomal peptidases [15,18] – a nomenclature already applied to the *I. ricinus* asparaginyl endopeptidase (IrAE) [13]. Thus, we designated the novel peptidases as IrCB for cathepsin B (form 1), IrCL for cathepsin L, IrCC for cathepsin C and IrCD for cathepsin D.

Table 1: Degenerate PCR primers used for identification of genes encoding gut-associated peptidases of *I. ricinus*

peptidase	cathepsin B	cathepsin L	cathepsin C	cathepsin D
forward primer	5'CKTGYGGDTCTTGTTGGG 3'	5'CAGTGCGGVTCCTGCTGGGC3'	5'GTCGACACKCCYGCBAACTGCAC3'	5'TCACCCCAGCCKNCGTTCCA3'
reverse primer	5'GGCTCCTCTAAYYTBTGGGT 3'	5'GGCGATTCCGCACTGGTTGT3'	5'CCRTAGWAACCWCCMACATA3'	5'GAACACGTCNCCBARDATCCA3'
Sm* GenBank	CAC85211	CAA83538	CAA83543	AAB63442
Sm* forward domain	[117]SSCGSCWA[124]	[123]LCGSCWA[129]	[20]ADTPANC[26]	[85]GSSNLWV[91]
Sm* reverse domain	[313]WNSDWGD[319]	[304]NMCGIA[309]	[337]YIGGYY[342]	[355]WILGDIF[361]
annealing temperature	55°C	53°C	62°C	53°C
product length	608 bp	554 bp	998 bp	855 bp
sequenced clones	12	8	8	8
isoforms	2	1	1	1

- orthologues in *Schistosoma mansoni

The three clan CA cysteine peptidases, namely IrCB, IrCL and IrCC could be clearly classified using multiple sequence alignments followed by a phylogenetic analysis (Fig. 3). The GenBank blast program blastp [19] search of the IrCD sequence revealed the closest relation (55% identity) as longepsin, the aspartic peptidase from *H. longicornis* [12].

I. ricinus *cathepsin B (IrCB)*

The nucleotide and the deduced amino acid sequences of IrCB (form 1) enzyme precursor are shown in Fig. 4. The cDNA sequence [GenBank:EF428206] is 1073 bp long and contains one open reading frame encoding a polypeptide of 337 amino acid (AA) residues. Use of the SignalP 3.0 server [20] predicts a signal peptide cleavage between G^{17} and R^{18}. The pro-enzyme has a theoretical mass of 35.725 Da and an isoelectric point 5.76. The catalytic residues C^{113}, H^{282} and two other active site residues

Q^{107} and N^{302} were found in positions typical for the C1 peptidase (papain) family. The protein has three potential *N*-glycosylation sites predicted by the NetNGlyc 1.0 Server [21]. The occluding loop responsible for the putative exopeptidase (peptidyl dipeptidase) activity is predictable between C^{191} and C^{211}. Also, the domain ^{297}YWLVANSWxxDWGD^{310} accords to a domain previously described as being associated with the hemoglobinase activity of cathepsin B in blood feeding helminths [22].

I. ricinus *cathepsin L (IrCL)*

Cathepsin L is another member of the papain family of peptidases that we have identified from *I. ricinus* guts (Fig. 5). The cDNA sequence [GenBank:EF428205] is 1151 bp long and contains one open reading frame encoding an enzyme precursor of 316 AA residues. The C1 family active site residues of IrCL were identified as Q^{137}, C^{143},
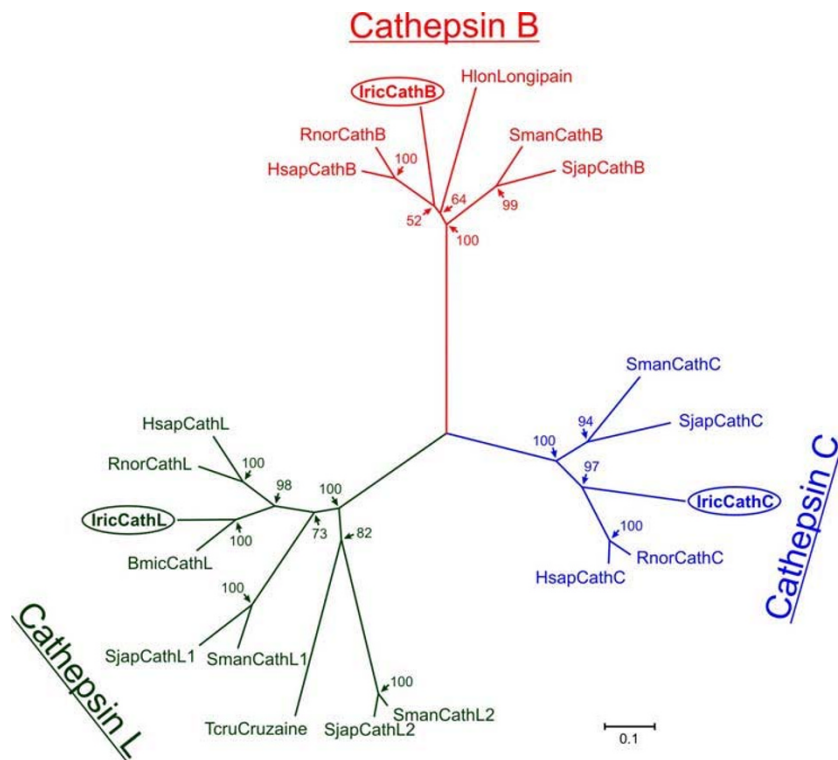

Figure 3
Phylogenetic relation of *Ixodes ricinus* gut-associated cysteine peptidases (clan CA, family C1) to selected representatives of the papain family. The tree was reconstructed by the Neighbor-joining method using amino acid sequences spanning across the homologous domains of mature enzymes. The horizontal bar represents a distance of 0.1 substitutions per site. Numbers at the branches display bootstrap support. **Cathepsins B**: *Homo sapiens* (AAH95408), *Rattus norvegicus* (P00787), *Ixodes ricinus* (ABO26563), *Haemaphysalis longicornis* (BAF43801), *Schistosoma mansoni* (P25792), *Schistosoma japonicum* (P43157); **Cathepsins L**: *H. sapiens* (M20496), *R. norvegicus* (AAH63175), *I. ricinus* (ABO26562), *Boophilus microplus* (AF227957), *S. mansoni* (CathL1 (Z32529), CathL2 (U07345)), *S. japonicum* (CathL1 (U38476), CathL2 (U38475)), *Trypanosoma cruzii* (P25779); **Cathepsins C**: *H. sapiens* (X87212), *R. norvegicus* (D90404), *I. ricinus* (ABV29335), *S. mansoni* (Z32531), *S. japonicum* (U77932).

Figure 4
Nucleotide and deduced amino acid sequence of *Ixodes ricinus* cathepsin B, form 1 (IrCB). The PCR primers used are upperlined and named as in Table 1 and 2. The predicted 17 AA signal peptide is underlined, in italics. Three potential *N*-glycosylation sites are double underlined and the respective asparagine residues are in bold; Active site residues C[113], H[282], Q[107] and N[302] are bold and shaded. Occluding loop responsible for cathepsin B exo-peptidase activity is bold, and italics.

H[262] and N[282]. Predicted signal peptide is 16 AA long and the molecular weight of the pro-enzyme without the signal peptide is 35.403 Da with a theoretical isoelectric point 5.79. The proenzyme has two potential *N*-glycosylation sites both within the mature peptidase.

l. ricinus *cathepsin C (IrCC)*
Cathepsin C belongs to the clan CA papain family, but unlike B and L it acts solely as an exopeptidase. It has been shown to sequentially remove dipeptides from the substrate N-terminus (therefore alternatively named as dipep-

tidyl peptidase I). The cDNA sequence of IrCC [GenBank: EU128750] is 1597 bp long and encodes an enzyme precursor of 465 AA residues (Fig. 6). Calculated molecular weight of the proenzyme without signal peptide is 49.812 Da and the theoretical isoelectric point is 6.88. Based on the structure of human cathepsin C, the active site residues are Q[250], C[256], H[409] and N[431]. The predicted signal peptide counts for 20 AA residues and the proenzyme has four potential *N*-glycosylation sites, one close to the predicted mature N-terminus.

```
   1 - agtgcgctcgggagtccggtcggagtcttttgttggagtgtttaccaatcgtgcaggtca -  60
   1 -                                                           M  -   1

  61 - tgctgcggtgcctcgtgttgtgtctcctcgttgctgccgcagctgcggtcagctaccagg - 120
   2 -  L   R   C   L   V   L   C   L   L   V   A   A   A   A   A   V   S   Y   Q   E  -  21

 121 - aagttcttggcgctgaatggtcggctttcaaggcaaagcatgggaagtcttatgtgtctg - 180
  22 -  V   L   G   A   E   W   S   A   F   K   A   K   H   G   K   S   Y   V   S   E  -  41

 181 - aaactgaggaggtgttcaggctgaagatctacatggagaacaggcacaagattgccaagc - 240
  42 -  T   E   E   V   F   R   L   K   I   Y   M   E   N   R   H   K   I   A   K   H  -  61

 241 - acaacgagaagtatgccaggggggaggtgccctacagcatggccatgaacgagtttggag - 300
  62 -  N   E   K   Y   A   R   G   E   V   P   Y   S   M   A   M   N   E   F   G   D  -  81

 301 - acatgcttcaccacgaattcgtgagcacccgcaatggcttcaagcgaaactacaaggatc - 360
  82 -  M   L   H   H   E   F   V   S   T   R   N   G   F   K   R   N   Y   K   D   Q  - 101

 361 - agcccagggagggatctacatacctggagccagaaaacattgaggacttcagcctgccca - 420
 102 -  P   R   E   G   S   T   Y   L   E   P   E   N   I   E   D   F   S   L   P   K  - 121
                                              CathL-RT-S
 421 - agaccgtggactggaggaccaagggcgcggtcacgcccgtcaagaaccagggacagtgtg - 480
 122 -  T   V   D   W   R   T   K   G   A   V   T   P   V   K   N   Q   G   Q   C   G  - 141
           CathL-S
 481 - gatcctgctgggccttcagtgctaccggatccctggaaggacagcacttccgcaagagtg - 540
 142 -  S   C   W   A   F   S   A   T   G   S   L   E   G   Q   H   F   R   K   S   G  - 161

 541 - gctccatggtctccctgagcgaacagaacctggtcggctgctccactgactttggaaaca - 600
 162 -  S   M   V   S   L   S   E   Q   N   L   V   G   C   S   T   D   F   G   N   N  - 181

 601 - acggttgtgagggtggtctcatggacgatgccttcaaatacatcagggccaacaagggca - 660
 182 -  G   C   E   G   G   L   M   D   D   A   F   K   Y   I   R   A   N   K   G   I  - 201

 661 - ttgacaccgaaaagagctacccctacaatggcacggacggcacttgccacttcaagaaga - 720
 202 -  D   T   E   K   S   Y   P   Y   N   G   T   D   G   T   C   H   F   K   K   S  - 201

 721 - gcaccgttgggggccaccgacagcggctttgtggacatcaaggaaggctcggagacgcagc - 780
 202 -  T   V   G   A   T   D   S   G   F   V   D   I   K   E   G   S   E   T   Q   L  - 221
                                             CathL-RT-AS
 781 - tgaagaaggccgtggccaccgttggacccatctctgttgccattgatgccagccacgagt - 840
 222 -  K   K   A   V   A   T   V   G   P   I   S   V   A   I   D   A   S   H   E   S  - 241

 841 - ctttccagttctactctgacggtgtgtacgacgagccagagtgtgacagcgagtccctgg - 900
 242 -  F   Q   F   Y   S   D   G   V   Y   D   E   P   E   C   D   S   E   S   L   D  - 261

 901 - accacggtgttcttgtggttggttacggcaccctcaacggcactgactactggttcgtca - 960
 262 -  H   G   V   L   V   V   G   Y   G   T   L   N   G   T   D   Y   W   F   V   K  - 281

 961 - agaacagttggggaaccacctggggtgatgaaggctacatccgcatgagcaggaacaaga - 1020
 282 -  N   S   W   G   T   T   W   G   D   E   G   Y   I   R   M   S   R   N   K   K  - 302
                  CathL-AS
1021 - agaaccagtgtgggaatcgcttccagtgctagcatacctcttgtgtaaccttccgtatctg - 1080
 303 -  N   Q   C   G   I   A   S   S   A   S   I   P   L   V   *            - 316

1081 - cactttcaccccattgtcatcactggtcctgttgtgcattcaataaacatctttgtacag - 1140
1141 - tgcccacgcacaaaa                                             - 1155
```

Figure 5
Nucleotide and deduced amino acid sequence of *Ixodes ricinus* cathepsin L (IrCL). The depictions of PCR primers, putative 16 AA signal peptide, two potential *N*-glycosylation sites and active site residues C^{143}, H^{262}, Q^{137} and N^{282} are as in Fig 4.

I. ricinus cathepsin D (IrCD)

A cathepsin-D-like aspartic peptidase precursor of 322 amino acid residues (Fig. 7) is encoded by a 1304 bp long cDNA sequence [GenBank:EF428204]. The predicted signal peptide counts for 21 AA residues. The active site contains two catalytic aspartic acid residues D^{79} and D^{270} within the conserved D-T-G motifs and Y^{111}. Molecular weight of the proenzyme without the signal peptide is 39.451 Da and the theoretical isoelectric point is 4.75. There are only two possible *N*-glycosylation sites in the IrCD proenzyme.

Differential expression of peptidases during development and in tissues

Gene-specific PCR primer sets for the newly identified cDNAs and IrAE cDNA (Table 2) were used to amplify the relevant peptidase genes from different tick developmental stages and tissues. Semi-quantitative RT-PCR of whole-body homogenates revealed that IrCL, IrCC and IrAE are abundantly present in all developmental stages including eggs (Fig. 8A). In contrast, messages for IrCB and IrCD were absent from tick eggs. No apparent differences were observed for any enzyme expression between un-fed and

```
   1 - cattcgagcagaagacgtcgcccgcgcagcatgtccacggttcccgctcttctgctgctg - 60
   1 -                                     M  S  T  V  P  A  L  L  L  L   - 10
                                                    CathC-S
  61 - ctgccgctctgcctggggctggtggctgccgacaccccgtccaactgcacatacgaggac - 120
  11 - L  P  L  C  L  G  L  V  A  A  D  T  P  S  N  C  T  Y  E  D   - 30
 121 - gtccggggctggtggaccttcgtggagggacccgtaccggcactaaagacgtccgatgc - 180
  31 - V  R  G  W  W  T  F  V  E  G  P  R  T  G  T  K  D  V  R  C   - 50
 181 - gacgactggaagtcagaagtcgctggtggcgagaccagaaaagtgcggctgctccttgat - 240
  51 - D  D  W  K  S  E  V  A  G  G  E  T  R  K  V  R  L  L  L  D   - 70
 241 - ttccccaacttggctacggattccgatggtaacgttggaacctggaccctcatctacaat - 300
  71 - F  P  N  L  A  T  D  S  D  G  N  V  G  T  W  T  L  I  Y  N   - 90
 301 - cagggcttcgaggtgactctgcaataccgaaggtacttcgccttctccaactacacggac - 360
  91 - Q  G  F  E  V  T  L  Q  Y  R  R  Y  F  A  F  S  N  Y  T  D   - 110
 361 - gacggacacgggagggtggtcagctactgtggcagcaccaaggtgggctggtcacacgac - 420
 111 - D  G  H  G  R  V  V  S  Y  C  G  S  T  K  V  G  W  S  H  D   - 130
 421 - gtcctcggtcacaactgggcctgcttcgagggtcgcaaggatgatggtctgggagagaag - 480
 131 - V  L  G  H  N  W  A  C  F  E  G  R  K  D  D  G  L  G  E  K   - 150
                                                    CathC-RT-S
 481 - ccgcttcaagttggcagacatgaagggctcggccgccacccgggaccctacaacaccgaa - 540
 151 - P  L  Q  V  G  R  H  E  G  L  G  A  T  R  D  P  Y  N  T  E   - 170
 541 - gtgttcgcgcagctcatcaacgcagcgcagaactcctggaaggccaaggttcacccacag - 600
 171 - V  F  A  Q  L  I  N  A  A  Q  N  S  W  K  A  K  V  H  P  Q   - 190
 601 - ttcaaggagaagacactggacgaaatgatgatgctccgcggaggacatcgcaacaagcat - 660
 191 - F  K  E  K  T  L  D  E  M  M  M  L  R  G  G  H  R  N  K  H   - 210
 661 - ctgcggagtcctggtgctgccaaggtgaccgaagaacagcggcgagaggcccaggccctg - 720
 211 - L  R  S  P  G  A  A  K  V  T  E  E  Q  R  R  E  A  Q  A  L   - 230
 721 - cccgaagagttcgactggaggaacgtgagcggcgtcaactacgtgtcccccgtccggaac - 780
 231 - P  E  E  F  D  W  R  N  V  S  G  V  N  Y  V  S  P  V  R  N   - 250
 781 - caagaggcctgcgggagctgctacagcttcgcctccatggccatgctcgaggcccgactc - 840
 251 - Q  E  A  C  G  S  C  Y  S  F  A  S  M  A  M  L  E  A  R  L   - 270
 841 - cgcatcgccaccaagaacagggtgaagaagatattttcgccgcaagacatcgtctcctgc - 900
 271 - R  I  A  T  K  N  R  V  K  K  I  F  S  P  Q  D  I  V  S  C   - 290
                           CathC-RT-AS
 901 - tcgccctactctcaggcttgcgagggcggcttcccgtacctgatcggaggcaaatacggc - 960
 291 - S  P  Y  S  Q  G  C  E  G  G  F  P  Y  L  I  G  G  K  Y  G   - 310
 961 - caggactttggcgtcgtcgaggaaagctgcaaccccctaccggggcgcagacgggccctgc - 1020
 311 - Q  D  F  G  V  V  E  E  S  C  N  P  Y  R  G  A  D  G  P  C   - 330
                                             CathC-AS
1021 - aagaccaacacgacgtgtcccaggcattacgtcggcaactaccgctacgttggcggatac - 1080
 331 - K  T  N  T  T  C  P  R  H  Y  V  G  N  Y  R  Y  V  G  G  Y   - 350
1081 - tacggaggctgcaacgaagagctgatgcgcctggccccttgtccgcggcggtcccgtggcc - 1140
 351 - Y  G  G  C  N  E  E  L  M  R  L  A  L  V  R  G  G  P  V  A   - 370
1141 - gtgggcttcgaggtgtaccctgacttccagatgtaccaggggcggcgtgtaccggcacacg - 1200
 371 - V  G  F  E  V  Y  P  D  F  Q  M  Y  Q  G  G  V  Y  R  H  T   - 390
1201 - ggcgttcaccgcagcctgaacctgggctcgcccttttgaccccttcgagctgaccaaccac - 1260
 391 - G  V  H  R  S  L  N  L  G  S  P  F  D  P  F  E  L  T  N  H   - 410
1261 - gccgtgctcgtgaccggctacggggtggacaaggaaacgggtctcaagtactggagcgtc - 1320
 411 - A  V  L  V  T  G  Y  G  V  D  K  E  T  G  L  K  Y  W  S  V   - 430
1321 - aagaactcgtggggcccccggctggggcgagagcggctacttccgcatccttcgcggcgcc - 1380
 431 - K  N  S  W  G  P  G  W  G  E  S  G  Y  F  R  I  L  R  G  A   - 450
1381 - gacgagtgcgggatcgagagtctcgccgtcgaggtcaccccgatcccgtgaacggagctc - 1440
 451 - D  E  C  G  I  E  S  L  A  V  E  V  T  P  I  P  *             - 466
1441 - gggcccgcccagtcgccgtcacccggagagaagtcctcgtcgtgggtgatgcgtgtaatg - 1500
1501 - taatgccctcaccagaggaaccttaaggatcaaatataaaaatgaattcctaaa        - 1555
```

Figure 6
Nucleotide and deduced amino acid sequence of *Ixodes ricinus* cathepsin C (IrCC). The depictions of PCR primers, putative 20 AA signal peptide, four potential *N*-glycosylation sites and active site residues C^{256}, H^{409}, Q^{250} and N^{431} are as in Fig 4.

Figure 7
Nucleotide and deduced amino acid sequence of *Ixodes ricinus* cathepsin D (IrCD). The depictions of PCR primers, putative 21 AA signal peptide, two potential *N*-glycosylation sites and active site residues D^{79}, Y^{111}, D^{270} are as in Fig 4.

freshly attached females suggesting that the enzyme messages are not changed in the initial feeding phase. Once partially engorged, it is possible to reliably dissect individual organs of females for RT-PCR tissue profiling (Fig. 8B) and the data demonstrate that all the peptidases of interest are co-expressed in the gut towards the end of the slow feeding period [7] what indicates their simultaneous action in a putative cascade or network. Moreover, IrCB, IrAE and IrCD seem to be strictly gut-specific, whereas messages for IrCL and IrCC were also found in other tick tissues. Negative controls in which the template cDNA

Table 2: Gene specific primers used for RT-PCR expression profiling of *I. ricinus* peptidases

peptidase	forward primer	reverse primer	annealing temperature	product length
IrAE	5'TCGGTGACGCTGAGAAGACTGAA3'	5'TAGATTATGCCCGATGACTGTTGG3'	55°C	239 bp
IrCB	5'CGACGGCGGATACCCGGCGGCAGC3'	5'CCGAGTACGTCACCGCTGTGGTGC3'	60°C	360 bp
IrCL	5'GTCACGCCCGTCAAGAACC3'	5'TGGCATCAATGGCAACAGAGA3'	62°C	382 bp
IrCC	5'CCACCCGGGACCCCTACAACACC3'	5'GGAAGCCGCCCTCGCAACC3'	68°C	447 bp
IrCD	5'GGACCCCGCCCCAAGACT3'	5'CGATGCCGCCGAACAGGA3'	61°C	459 bp

was replaced by sterile distilled water gave no PCR products (data not shown).

Discussion

The present functional and genetic profiling in the gut of the hard tick *I. ricinus* has identified a number of peptidase activities and genes. The overall goal was to have a better global understanding of the component peptidases for one important tick species in a particular feeding phase, namely partially engorged *I. ricinus* females, in contrast to the present fragmented picture regarding individual enzymes in a variety of different tick species. Also, this broad approach offers both the possibility to compare entire digestive systems with other hematophagous para-

Figure 8
Stage and tissue expression profiles of *Ixodes ricinus* cysteine and aspartic peptidases. Messenger RNA levels of individual enzymes were determined by semi-quantitative two-step RT PCR. **Panel A**: Expression of peptidase mRNAs in whole body homogenates of *I. ricinus* eggs, unfed larvae, unfed nymphs, males, unfed females and females attached for 1 day on the guinea pigs. **Panel B**: Expression of peptidase mRNAs in tissues dissected from partially engorged females (the 5-th day of feeding). The abbreviations used are as in the text. IrFer shows the mRNA amplification of tick ferritin used as template loading control. (For details, see Methods).

sites as well as to investigate the potential of one or more component peptidases as molecular vaccines.

Initial activity profiling of *I. ricinus* gut extracts using a biologically relevant protein substrate (AMC hemoglobin) indicated that hemoglobinolysis is optimal at acid pH, a finding in accordance with data presented for other tick species [5,23,24]. This suggested that proteolysis is mediated by peptidases belonging to the aspartic and/or cysteine peptidase classes which are known to operate optimally at acid pH [25]. This conclusion was further supported by the sensitivity of hemoglobinolysis to class-selective peptidase inhibitors. Accordingly, we functionally scanned gut extracts for individual peptidase activities with a battery of diagnostic, small molecule substrates and inhibitors. These studies revealed the presence of four cysteine peptidase activities, cathepsins B, C, L and AE, and an aspartic peptidase activity, cathepsin D.
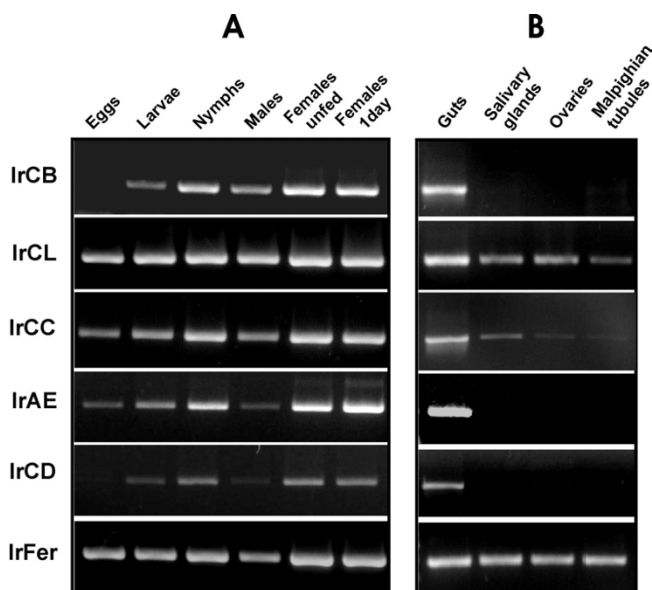
To identify the peptidase genes putatively responsible for the activities measured in gut extracts, we next screened gut-derived cDNA with degenerate primers designed to amplify individual peptidases. Five cysteine and one aspartic peptidases were classified: IrAE [13], IrCB, IrCL, IrCC and IrCD. Interestingly, from approximately 10 sequenced amplicons of each peptidase cDNA, only IrCB presented as two different isoforms. The finding is comparable with the data for hematophagous flukes, *Schistosoma mansoni* [26] and *Trichobilharzia regenti* [27], both of which have more than one cathepsin B isoforms.

Sequence comparison with human cathepsin B reveals that IrCB has the signature 'occluding loop' necessary for its exopeptidase (specifically, peptidyl dipeptidase) activity [28]. Regarding murine cathepsin C, this enzyme was previously shown to process and activate granulocyte serine peptidases by the removal of N-terminal dipeptides [29]. Thus, it would be of interest to test the potential competence of cathepsin C to processes cubulin-like serine peptidases inducing lysis of host blood cells [30]. The primary structure of the IrCD precursor is homologous to longepsin from *H. longicornis* [12] with two conserved Asp-Thr-Gly (DTG) catalytic site motifs either side of the substrate binding groove, a structure not shared by the more evolutionary distinct yolk-processing tick cathepsin D [31] and tick heme-binding aspartic peptidase [32] in

the eggs of *B. microplus*. Finally, a search through the available EST database (NCBI Blast with a limitation to tick ESTs) indicated the existence of several isoforms for cathepsins B, L, AE and cathepsin D, but only one form of cathepsin C in the whole body derived cDNA of the closely related tick species, *Ixodes scapularis* [33]. Certainly, these preliminary data need to await final contig assembly and gene annotations and show the need for tick gut transcriptome projects.

With the exception of the eggs, host blood is taken up and processed to provide energy and nutrients for the transition from larva to nymph and finally, to adult male or female [4]. By RT-PCR, enzymes under study are expressed in all the feeding developmental stages indicating their simultaneous action in digestive cells. Notably, however, IrCB and IrCD are not expressed in eggs, suggesting a function specifically associated with blood digestion. In support of this notion, the tissue-specific RT-PCR demonstrated that both peptidases are expressed solely in the gut. Likewise, IrAE is also restricted to the gut but, being also found in eggs, must have an additional function(s) not associated with the blood meal.

Conclusion

The combined biochemical and genetic analyses presented in this study demonstrate that *I. ricinus* expresses a suite of gut-associated cysteine and aspartic peptidases in order to catabolize ingested host proteins as a nutrient source. The data accord with previous results for enzyme activities in different tick species [5,24]. The particular combination of cysteine and aspartic peptidases comprising AE, and cathepsins B, C, D and L, operating at acidic pH and localized to the gut, is remarkably similar to those found in phylogenetically distant nematodes [17,34,35] and platyhelminths [15,16,34]. Indeed, cysteine and aspartic peptidases also contribute to amino acid acquisition in protozoa such as *Plasmodium* [36,37]. Therefore, and as noted by Delcroix et al. [16] for the platyhelminth *S. mansoni*, digestive systems based on cysteine and aspartic peptidases are widespread in invertebrates and stand in contrast to those systems utilizing serine peptidases (e.g., in insects and vertebrates). The present report extends this observation to include arthropods, specifically, ixodid ticks.

On an applied note, gut-associated peptidases may prove useful as vaccine targets. Other ixodid gut proteins, such as Bm86/Bm95 or Bm91, are suitable antigens for vaccination strategies (reviewed in de la Fuente and Kocan [38]). With this goal in mind, the detailed molecular and cellular characterization of the *I. ricinus* peptidases will be the subject of future reports.

Methods

Animals

I. ricinus ticks were collected by flagging in woodland localities around České Budějovice in the Czech Republic. Adult males and females were kept separately in glass vials in wet chambers with humidity of about 95% and temperature 26°C. If not stated otherwise, the females were allowed to feed naturally for 5 days on laboratory guinea pigs, carefully removed by forceps and referred to as partially engorged ticks in experiments described below. Laboratory animals were treated in accordance with the Animal Protection Law of the Czech Republic no. 246/1992 Sb.

Materials

The 7-amino-4-methylcoumarin (AMC)-conjugated bovine hemoglobin was prepared according to Partanen et al. [39]. All peptidyl AMC-substrates were from Bachem, Abz-Lys-Pro-Ala-Glu-Phe-Nph-Ala-Leu (Abz, aminobenzoic acid; Nph, 4-nitrophenylalanine) substrate was prepared as described in Máša et al. [40]. Peptidase inhibitors were from Bachem, Gly-Phe-diazomethyl ketone (DMK) was prepared as described in Green and Shaw [41]. The aza-peptide Michael acceptor (CBz-Ala-Ala-(aza-Asn)-CH = CH-COOEt) further referred to as Aza-N-11a [42] was kindly donated by Dr. J.C. Powers of the School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia.

Preparation of the tick gut tissue extract

For an experiment, tissues were dissected from 10 partially engorged *I. ricinus* females. The gut contents were carefully removed with a special care not to disrupt the epithelium. Cleaned guts were washed in phosphate-buffered saline solution (PBS) and pooled. The gut tissue extract (150 μg protein/ml) was prepared by homogenization of the gut tissue (without contents) in 1 ml of 0.1 M Na-acetate, pH 4.5, 1% CHAPS, 2.5 mM DTT using teflon-glass homogenizer on ice. The homogenate was centrifuged for 10 min at 10000 × g, the supernatant was filtered with Micropure-0.22 Separator (Millipore) and stored at -80°C.

Quantification of hemoglobin degradation

Hemoglobinolytic activity was assayed using AMC-hemoglobin as a fluorogenic substrate [43]. Digestion of fluorogenic AMC-hemoglobin (0.5 μg supplemented with 2 μg of bovine hemoglobin) was performed at 35°C with the gut tissue extract (20-fold diluted stock solution) in 0.1 M Na-citrate-phosphate, pH 2.5–8.0 including 2.5 mM DTT, 25 mM NaCl and 0.05% Tween 20 in a reaction mixture of 100 μl. The proteolytic fragmentation of AMC-hemoglobin results in an increase of the fluorescence intensity that was continuously monitored to determine the relative reaction rate. The fluorescence signal was measured using a GENios Plus reader (TECAN) at 360 nm excitation

and 465 nm emission wavelengths. For the hemoglobino-lytic assay in the presence of a peptidase inhibitor, an aliquot of the extract was preincubated (15 min at 35°C) in the same buffer pH 4.0 with 10 μM E64, 10 μM pepstatin, 1 mM Pefabloc or 1 mM EDTA.

Profiling component gut peptidases with substrates and inhibitors

Peptidase activities were identified and characterized by hydrolysis of the following fluorogenic substrates: 25 μM Z-Arg-Arg-AMC for cathepsin B [44], 25 μM Z-Phe-Arg-AMC for cathepsin L [44], 30 μM Gly-Arg-AMC for cathepsin C [45], 40 μM Abz-Lys-Pro-Ala-Glu-Phe-Nph-Ala-Leu for cathepsin D [40], and 30 μM Z-Ala-Ala-Asn-AMC for asparaginyl endopeptidase [46] as described previously [40,47]. The activity measurement was performed at 35°C using an aliquot of the gut tissue extract (20 to 200-fold diluted stock solution) in 0.1 M Na-acetate, pH 4.0 including 2.5 mM DTT (for cysteine peptidases) and 25 mM NaCl (for cathepsin C). For the activity assay in the presence of peptidase inhibitors, an aliquot of the extract was pre-incubated (15 min at 35°C) in the same buffer with the inhibitor: 10 μM CA-074 for cathepsin B [48], 10 μM Z-Phe-Phe-DMK for cathepsin L [49], 1 μM Gly-Phe-DMK for cathepsin C [41], 10 μM pepstatin for cathepsin D [50] or 1 μM Aza-N-11a for asparaginyl endopeptidase [42]. Hydrolytic activity was continuously measured after addition of substrate in a fluorescence reader GENios Plus at 320 nm excitation and 420 nm emission wavelengths (for Abz-containing substrate) or at 360 nm excitation and 465 nm emission wavelengths (for AMC-containing substrates). Assays of asparaginyl endopeptidase and cathepsin L were measured in the presence of 10 μM CA-074 to prevent confounding hydrolysis by cathepsin B [16,47].

cDNA synthesis

Tissues (gut, salivary glands, ovaries and Malpighian tubules) were dissected from partially engorged females in a wax filled Petri dish with phosphate-buffered saline (PBS) under a binocular dissection microscope. The whole body homogenates from different developmental stages were prepared by crushing the appropriate number of eggs, larvae, nymphs, males, unfed females and females removed from guinea pigs one day after attachment using mortar and pestle and repeated freezing under liquid nitrogen. For total RNA isolation, the samples were further homogenized in a micro-tube with a plastic pestle in the TRI Reagent® solution (Sigma) at 1 ml per 50–100 mg of wet tissue and processed according to the instructions provided with the TRI-reagent kit (Sigma). Isolated total RNA was stored at -80°C and further used for preparing single stranded cDNA templates using Superscript II (Invitrogen) and oligo(dT) primers, following the instructions

provided by the manufacturer or for RT-PCR experiments described below.

Designing PCR oligonucleotide primers

Degenerate primers were designed from conserved domains of *Schistosoma mansoni*, *S. japonicum*, mosquito, rat and human peptidases. Protein and nucleotide sequences downloaded from the NCBI GenBank web site were used for multiple ClustalW alignments in DNAstar MegAlign software (Lasergene). The oligonucleotides are listed in Table 1.

PCR and rapid amplification of cDNA ends (5'- and 3'-RACE)

Mastercycler gradient (Eppendorf) was used to optimize PCR amplifications. Amplicons were purified, ligated into vector plasmid pCR 4-TOPO using the TOPO TA® Cloning Kit (Invitrogen) and transformed into *E. coli* TOP 10 cells (Invitrogen). Clones containing ligated PCR products were sequenced using an automated sequencer model ABI Prism 3130 XL and the BigDye® Terminator sequencing kit (Applied Biosystems) with appropriate sequencing primers. Sequence data were compared by blastn [51] against the NCBI GenBank database records. To obtain complete cDNA sequences, 3' RACE PCR was performed using a modified protocol for SMART™ cDNA Library Construction Kit (Clontech, BD Biosciences) described previously [52]. The N-terminal sequences including signal peptides and the 5' untranslated regions were determined using the Invitrogen 5'RACE system and instructions provided by the manufacturer.

cDNA library construction and screening

The protocol for construction of the *I. ricinus* gut-derived cDNA library with SMART™ cDNA Library Construction Kit (Clontech, BD Biosciences) and the Gigapack® III Gold packaging extract (Stratagene) as well as the method of cDNA library screening by [P32]dATP radio-labeled gene-specific probes have been described previously [13].

Phylogenetic analysis

The primary sequences used for phylogenetic analysis comprised the conserved domains spanning across the mature enzyme sequences without pro-domains. Sequences were obtained from the MEROPS database [53] and aligned in the program ClustalX 1.81 [54]. The alignment was manually checked using the BioEdit program [55]. Tree reconstruction employed the Neighbor Joining (NJ) method [56] in the program MEGA 2.1 [57]. Nodal supports were calculated with 1000 replications.

Semi-quantitative RT-PCR

Gene specific PCR primer pairs (listed in Table 2) were designed for each peptidase type with DNAstar PrimerSelect software (Lasergene). Two-step RT-PCR was per-

formed using total RNA templates (prepared as described above; 50 ng/μl final concentration) and the Enhanced Avian HS RT-PCR Kit (Sigma) according to the protocol provided by the manufacturer. Amplification of the ferritin mRNA, previously shown to be presented in all tick tissues and expressed independently of feeding [58], was used as a loading control.

Abbreviations

AA, amino acid; Abz, aminobenzoic acid; AMC, 7-amino-4-methyl-coumarin; bp, base pairs; DMK, Gly-Phe-diazomethyl ketone; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; EST, expressed sequence tag; NCBI, National Center for Biotechnology Information; Nph, 4-nitrophenylalanine; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription polymerase chain reaction.

Authors' contributions

DS designed the degenerate primers, identified and isolated the genes described in this study from the cDNA library, and participated in all aspect of this manuscript including its conception and drafting. ZF was involved in cloning and sequencing of full length cDNAs by 5- and 3'-RACE PCR and conducted the RT-PCR expression profiling. MH and MM designed and performed the substrate/inhibitor based profiling and characterization of endogenous activities in the tick gut extract. OH contributed to the present manuscript by alignments and phylogenetic analyses. CRC interpreted the data in relation to the digestive system of flatworms and nematodes, included the evolutional aspect of this work and contributed to the manuscript drafting. PK and MM were responsible for the overall conception and coordination of the study and participated in drafting the manuscript. All authors have read and approved the final version of this manuscript.

References

1. Hoogstraal H: **Argasid and nuttalliellid ticks as parasites and vectors.** *Adv Parasitol* 1985, **24:**135-138.
2. Nutall PA: **Pathogen-tick-host interactions:** *Borrelia bugdorferi* **and TBE virus.** *Zentralbl Bakteriol* 1999, **289:**492-505.
3. Grandjean O: **Blood digestion in** *Ornithodoros moubata* **Murray sensu stricto Walton (Ixodoidea: Argasidae) females. I. Biochemical changes in the midgut lumen and ultrastructure of the midgut cells, related to intracellular digestion.** *Acarologia* 1984, **25:**147-165.
4. Sonenshine DE: **Biology of Ticks.** *Volume 1*. New York: Oxford University Press; 1991.
5. Reich CI, Zorzopulos J: *Boophilus microplus*: **characterization of larval proteases.** *Exp Parasitol* 1978, **44:**1-6.
6. Coons LB, Rosell-Davis R, Tarnowski BI: **Bloodmeal digestion in ticks.** In *Morphology, Physiology, and Behavioral Biology of Ticks* Edited by: Sauer JR, Hair JA. Chichester, England: Ellis Horwood Ltd; 1986:248-279.
7. Coons LB, Alberti G: **The Acari-Ticks.** In *Microscopic Anatomy of Invertebrates, Chelicerate Arthropoda Volume 8B*. Edited by: Harrison FW, Foelix R. NewYork: Wiley-Liss; 1999:267-514.
8. Lara FA, Lins U, Bechara GH, Oliveira PL: **Tracing heme in a living cell: hemoglobin degradation and heme traffic in digest cells of the cattle tick** *Boophilus microplus*. *J Exp Biol* 2005, **208:**3093-3101.
9. Lara FA, Lins U, Paiva-Silva G, Almeida IC, Braga CM, Miguens FC, Oliveira PL, Dansa-Petretski M: **A new intracellular pathway of haem detoxification in the midgut of the cattle tick** *Boophilus microplus*: **aggregation inside a specialized organelle, the hemosome.** *J Exp Biol* 2003, **206:**1707-1715.
10. Renard G, Garcia JF, Cardoso FC, Richter MF, Sakanari JA, Ozaki LS, Termignoni C, Masuda A: **Cloning and functional expression of a** *Boophilus microplus* **cathepsin L-like enzyme.** *Insect Biochem Mol Biol* 2000, **30:**1017-1026.
11. Mulenga A, Sugimoto C, Ingram G, Ohashi K, Onuma M: **Molecular cloning of two** *Haemaphysalis longicornis* **cathepsin L-like cysteine proteinase genes.** *J Vet Med Sci* 1999, **61:**497-503.
12. Boldbaatar D, Sikalizyo Sikasunge C, Battsetseg B, Xuan X, Fujisaki K: **Molecular cloning and functional characterization of an aspartic protease from the hard tick** *Haemaphysalis longicornis*. *Insect Biochem Mol Biol* 2006, **36:**25-36.
13. Sojka D, Hajdušek O, Dvořák J, Sajid M, Franta Z, Schneider EL, Craik CS, Vancová M, Burešová V, Bogyo M, Sexton KB, McKerrow JH, Caffrey CR, Kopáèek P: **IrAE – An asparaginyl endopeptidase (legumain) in the gut of the hard tick** *Ixodes ricinus*. *Int J Parasitol* 2007, **37:**713-724.
14. Alim MA, Tsuji N, Miyoshi T, Islam MK, Huang X, Motobu M, Fujisaki K: **Characterization of asparaginyl endopeptidase, legumain induced by blood feeding in the ixodid tick** *Haemaphysalis longicornis*. *Insect Biochem Mol Biol* 2007, **37:**911-922.
15. Caffrey CR, McKerrow JH, Salter JP, Sajid M: **Blood 'n' guts: an update on schistosome digestive peptidases.** *Trends Parasitol* 2004, **20:**241-248.
16. Delcroix M, Sajid M, Caffrey CR, Lim KC, Dvořák J, Hsieh I, Bahgat M, Dissous C, McKerrow JH: **A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite.** *J Biol Chem* 2006, **281:**39316-39329.
17. Williamson AL, Brindley PJ, Knox DP, Hotez PJ, Loukas A: **Digestive proteases of blood-feeding nematodes.** *Trends Parasitol* 2003, **19:**417-423.
18. Dalton JP, Clough KA, Jones MK, Brindley PJ: **Characterization of the cathepsin-like cysteine proteinases of** *Schistosoma mansoni*. *Infect Immun* 1996, **64:**1328-1334.
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
20. **SignalP 3.0 server** [http://www.cbs.dtu.dk/services/SignalP/]
21. **NetNGlyc 1.0 Server** [http://www.cbs.dtu.dk/services/NetNGlyc/]
22. Baig S, Damian RT, Peterson DS: **A novel cathepsin B active site motif is shared by helminth bloodfeeders.** *Exp Parasitol* 2002, **101:**83-89.
23. Akov S: **Blood digestion in ticks.** In *Physiology of Ticks* Edited by: Obenchain F, Galun R. Oxford: Pergamon Press; 1982:197-211.
24. Mendiola J, Alonso M, Marquetti MC, Finlay C: *Boophilus microplus*: **multiple proteolytic activities in the midgut.** *Exp Parasitol* 1996, **82:**27-33.
25. Barrett AJ, Rawlings D, Woessner JF: **Proteolytic enzymes.** In *Handbook of Proteolytic Enzymes* Edited by: Oxford: Academic Press. Barrett AJ, Rawlings D, Woessner JF; 1998:801-805.
26. Sajid M, McKerrow JH, Hansell E, Mathieu MA, Lucas KD, Hsieh I, Greenbaum D, Bogyo M, Salter JP, Lim KC, Franklin C, Kim JH, Caffrey CR: **Functional expression and characterization of** *Schistosoma mansoni* **cathepsin B and its trans-activation by an**

endogenous asparaginyl endopeptidase. *Mol Biochem Parasitol* 2003, **131**:65-75.

27. Dvořák J, Delcroix M, Rossi A, Vopálenský V, Pospíšek M, Šedinová M, Mikeš L, Sajid M, Sali A, McKerrow JH, Horák P, Caffrey CR: **Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition.** *Int J Parasitol* 2005, **35**:895-910.
28. Krupa JC, Hasnain S, Nagler DK, Menard R, Mort JS: **S2' substrate specificity and the role of His110 and His111 in the exopeptidase activity of human cathepsin B.** *Biochem J* 2002, **361**:613-619.
29. Pham CTN, Ley TJ: **Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B in vivo.** *Proc Nat Acad Sci* 1999, **96**:8627-8632.
30. Miyoshi T, Tsuji N, Islam MK, Huang X, Motobu M, Alim MA, Fujisaki K: **Molecular and reverse genetic characterization of serine proteinase-induced hemolysis in the midgut of the ixodid tick *Haemaphysalis longicornis*.** *J Insect Physiol* 2007, **53**:195-203.
31. Logullo C, Vaz IS, Sorgine MH, Paiva-Silva GO, Faria FS, Zingali RB, De Lima MF, Abreu L, Oliveira EF, Alves EW, Masuda H, Gonzales JC, Masuda A, Oliveira PL: **Isolation of an aspartic proteinase precursor from the egg of a hard tick, *Boophilus microplus*.** *Parasitology* 1998, **116**:525-532.
32. Sorgine MH, Logullo C, Zingali RB, Paiva-Silva GO, Juliano L, Oliveira PL: **A heme-binding aspartic proteinase from the eggs of the hard tick *Boophilus microplus*.** *Biol Chem* 2000, **275**:28659-28665.
33. Hill CA, Wikel SK: **The *Ixodes scapularis* Genome Project: an opportunity for advancing tick research.** *Trends Parasitol* 2005, **21**:151-153.
34. Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP: **Proteinases and associated genes of parasitic helminths.** *Adv Parasitol* 1999, **43**:161-266.
35. Oliver EM, Skuce PJ, McNair CM, Knox DP: **Identification and characterization of an asparaginyl proteinase (legumain) from the parasitic nematode, *Haemonchus contortus*.** *Parasitology* 2006, **133**:237-244.
36. Rosenthal PJ: **Cysteine proteases of malaria parasites.** *Int J Parasitol* 2004, **34**:1489-1499.
37. Liu J, Istvan ES, Gluzman IY, Gross J, Goldberg DE: ***Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems.** *Proc Nat Acad Sci* 2006, **103**:8840-8805.
38. de la Fuente J, Kocan KM: **Strategies for development of vaccines for control of ixodid tick species.** *Parasite Immunol* 2006, **28**:275-283.
39. Partanen S, Storch S, Löffler HG, Hasilik A, Tyynelä J, Braulke T: **A replacement of the active-site aspartic acid residue 293 in mouse cathepsin D affects its intracellular stability, processing and transport in HEK-293 cells.** *Biochem J* 2003, **369**:55-62.
40. Máša M, Marešová L, Vondrášek JR, Horn M, Ježek J, Mareš M: **Cathepsin D propeptide: mechanism and regulation of its interaction with the catalytic core.** *Biochemistry* 2006, **45**:15474-15482.
41. Green GD, Shaw E: **Peptidyl diazomethyl ketones are specific inactivators of thiol proteinases.** *J Biol Chem* 1981, **256**:1923-1928.
42. Ekici ÖD, Götz MG, James KE, Li ZZ, Rukamp BJ, Asgian JL, Caffrey CR, Hansell E, Dvořák J, McKerrow JH, Potempa J, Travis J, Mikolajczyk J, Salvesen GS, Powers JC: **Aza-peptide Michael acceptors: a new class of inhibitors specific for caspases and other clan CD cysteine proteases.** *J Med Chem* 2004, **47**:1889-1892.
43. Demoz M, Castino R, Follo C, Hasilik A, Sloane BF, Isidoro C: **High yield synthesis and characterization of phosphorylated recombinant human procathepsin D expressed in mammalian cells.** *Protein Expr Purif* 2006, **45**:157-167.
44. Barrett AJ, Kirschke H: **Cathepsin B, Cathepsin H, and Cathepsin L.** *Methods Enzymol* 1981, **80**:535-561.
45. McGuire MJ, Lipsky PE, Thiele DL: **Purification and characterization of dipeptidyl peptidase I from human spleen.** *Arch Biochem Biophys* 1992, **295**:280-288.
46. Kembhavi AA, Buttle DJ, Knight CG, Barrett AJ: **The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays.** *Arch Biochem Biophys* 1993, **303**:208-213.
47. Grunclová L, Horn M, Vancová M, Sojka D, Franta Z, Mareš M, Kopáček P: **Two secreted cystatins of the soft tick *Orntho-

doros moubata: differential expression pattern and inhibitory specificity.** *Biol Chem* 2006, **387**:1635-1644.
48. Murata M, Miyashita S, Yokoo C, Tamai M, Hanada K, Hatayama K, Towatari T, Nikawa T, Katunum N: **Novel epoxysuccinyl peptides. Selective inhibitors of cathepsin B, in vitro.** *FEBS Lett* 1991, **25**:307-310.
49. Baricos WH, Zhou Y, Mason RW, Barrett AJ: **Human kidney cathepsins B and L. Characterization and potential role in degradation of glomerular basement membrane.** *Biochem J* 1988, **252**:301-344.
50. Knight CG, Barrett AJ: **Interaction of human cathepsin D with the inhibitor pepstatin.** *Biochem J* 1976, **155**:117-125.
51. **blastn** [http://www.ncbi.nlm.nih.gov/BLAST/]
52. Rego RO, Kovář V, Kopáček P, Weise C, Man P, Šauman I, Grubhoffer L: **The tick plasma lectin, Dorin M, is a fibrinogen-related molecule.** *Insect Biochem Mol Biol* 2006, **36**:291-299.
53. **MEROPS database** [http://merops.sanger.ac.uk]
54. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
55. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41**:95-98.
56. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-425.
57. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17**:1244-1255.
58. Kopáček P, Ždychová J, Yoshiga T, Weise C, Rudenko N, Law JH: **Molecular cloning, expression and isolation of ferritins from two tick species–*Ornithodoros moubata* and *Ixodes ricinus*.** *Insect Biochem Mol Biol* 2003, **233**:103-113.