Charles University

Faculty of Science

Study programme: Parasitology

Mgr. Hana Dvořáková

Acid peptidases of schistosomes and haematophagous monogeneans

Kyselé peptidázy schistosom a krevsajících monogeneí

Doctoral thesis/Dizertační práce

Supervisor/školitel: RNDr. Libor Mikeš, Ph.D.

Prague, 2022

Author's declaration:

I declare that the submitted thesis is my own work and that I properly cited all scientific literature used and that neither this work nor its substantial part have been utilised to achieve the same or any other academic degree.

Prohlášení autora:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu a tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Praha,

Mgr. Hana Dvořáková

Supervisor´s declaration:

I declare that data presented in the Ph.D. thesis result from the research activities of Mgr. Hana Dvořáková and her colleagues, co-authors of the enclosed publications. Her involvement and contributions were substantial and are clearly stated in the text.

Prohlášení školitele:

Prohlašuji, že data prezentovaná v této dizertační práci jsou výsledkem vědecké činnosti Mgr. Hana Dvořáková a jejích kolegů a spoluautorů z přiložených publikací. Její zapojení a podíl byly významné a jsou jasně uvedeny v textu.

Praha,

RNDr. Libor Mikeš, Ph.D.

Acknowledgement:

First, I would like to thank to my supervisor Libor Mikeš for the opportunity to work on a very interesting topic, for guidance during my long studies and especially for strengthening my nervous system. I want also to thank Conor Caffrey for help and valuable advice during my internship. My great thanks go to Petr Horák for an excellent leadership of Helminthology lab and to all my ex-colleagues for creation a friendly atmosphere and for all the fun we had. I am deeply indebted to my family that they have been standing by me all the time, not just at the study. I'm grateful to have you all. Finally, I would like to remember my grandmother and grandfather, who are unfortunately no longer with us. They kept asking me if I would ever finish my studies. Grandma and grandpa, here we go.

Table of contents:

Abstract 5

Abstrakt 7

Introduction 8

Literature review 8

1.1 Experimental models 8

1.1.1 Eudiplozoon nipponicum 8

1.1.2 Trichobilharzia regenti 9

1.2 Peptidases 10

1.2.1 Classification of peptidases 11

1.2.2 The substrate specificity of a peptidase 11

1.2.3 Regulation of peptidase activity 12

1.2.4 Cathepsins 15

1.2.5 Non-peptidase homologs 22

1.3 Digestion of blood by haematophagous helminths 23

Aims of the thesis 27

Results 28

Concluding remarks 85

References 105

Abstract

Blood is a complex nutrient-rich mixture. No wonder that haematophagy has been adopted as a feeding strategy by many invertebrates, including many parasitic helminths. In general, processing of haemoglobin (and other blood proteins) in blood-feeding helminths relies on an evolutionary conserved network of cysteine and aspartic peptidases (e.g., cathepsins L, B and D). However, some helminth taxa have been neglected from this point of view - very little information has been available about the occurrence of these enzymes in haematophagous monogeneans. Therefore, the presented thesis focuses on the molecular and biochemical characteristics of peptidases that maybe potentially involved in blood processing by the monogenean *Eudiplozoon nipponicum* (Heteronchoinea, Diplozoidae), an ectoparasite inhabiting gills of common carp. We show that the most abundant haemoglobinolytic endopeptidase activities in soluble protein extracts and excretory/secretory products of *E. nipponicum* belong to the cysteine and aspartic classes, with cathepsin L-like activity predominating over cathepsin B-like activity and supplemented with cathepsin D-like activity (paper 1). Additionally, we found that *E. nipponicum* adults express a variety of cathepsins L with different structural characteristics and probably different functions. The properties and localisation of two highly expressed EnCLs suggest their role in the digestion of host blood proteins. Surprisingly, a single cathepsin B is expressed by the parasite. Due to its localisation in the vitelline cells of the parasite, it is most likely not a digestive peptidase (paper 2). An aspartic peptidase cathepsin D of *E. nipponicum* was also localised in the haematin cells of the worm's digestive tract, which indicated its participation in the digestion of host blood proteins (manuscript 3, in preparation).

My further research was focused on detailed characterization of isoforms of a digestive peptidase cathepsin B1 (TrCB1) from the gut of the bird schistosome *Trichobilharzia regenti*. Comparison of substrate preferences of the two active forms showed that TrCB1.4 had narrow substrate specificity and after processing it was less effective toward protein substrates when compared to TrCB1.1. Moreover, the study showed that self-processing of both active isoforms could be facilitated by sulphated polysaccharides. Recombinant TrCB1.6wt (paralog with a Cys^{29}-to-Gly substitution in the active site) was processed by the active TrCB1 forms and by an asparaginyl endopeptidase. Although TrCB1.6wt lacked hydrolytic activity, its endopeptidase, but not carboxydipeptidase, activity could be restored by Gly^{29} to Cys site-directed mutagenesis. The active forms and the inactive TrCB1.6wt paralog formed complexes with chicken cystatin. Thus, the inactive paralog retains the ability to bind cystatins, and in addition, it is

highly expressed in schistosomula. This suggests its role in sequestration of cystatin inhibitors of host origin in order to protect the activity of parasite's cysteine peptidases (paper 4).

Abstrakt

Krev je komplexní směs bohatá na živiny. Není proto divu, že si hematofágii osvojilo mnoho bezobratlých, včetně mnoho parazitických helmintů. Obecně platí, že zpracování hemoglobinu (a i jiných krevních proteinů) je u krevsajících helmintů založené na evolučně konzervované síti cysteinových a aspartických peptidáz (např. katepsiny L, B and D). Nicméně, některé skupiny helmintů byly z tohoto hlediska opomíjeny – velmi málo informací bylo známo o výskytu těchto enzymů u hematofágních monogeneí. Předkládaná práce se proto zaměřila na molekulární a biochemické vlastnosti peptidáz, které se mohou potenciálně podílet na zpracování krve u monogenea *Eudiplozoon nipponicum* (Heteronchoinea, Diplozoidae), ektoparazita z žaber kapra obecného. V homogenátu a exkrečně/sekrečních produktech *E. nipponcum* převažovala aktivita cysteinových aspartických endopeptidáz, s aktivitou typu katepsinu L převažující nad aktivitou typu katepsinu B a doplněné o aktivitu typu cathepsinu D (článek 1). Dále jsme zjistili, že dospělec *E. nipponicum* exprimuje širokou škálu katepsinů L, které mají odlišné strukturální vlastnosti a tím pravděpodobně i odlišné funkce. Vlastnosti a lokalizace dvou vysoce exprimovaných EnCL naznačují jejich roli při trávení proteinů hostitelské krve. Je překvapivé, že pouze jediný katepsin B je exprimován tímto parazitem. Vzhledem k jeho lokalizaci ve vitelínových buňkách parazita, se pravděpodobně nejedná o trávicí peptidázu (článek 2). Aspartická peptidáza katepsin D byla také lokalizována v hematinových buňkách trávicího traktu červa *E. nipponicum*, což napovídá o jeho účasti na trávení proteinů hostitelské krve (rukopis 3, v přípravě).

Dále jsem se zabývala detailní charakteristikou isoforem trávicí peptidázy katepsin B1 (TrCB1) ze střeva ptačí schistosomy *Trichobilharzia regenti.* Srovnání preferencí k substrátu u dvou aktivních forem ukázalo, že TrCB1.4 má úzkou substrátovou specificitu a po aktivaci byl méně účinný vůči proteinovým substrátům než TrCB1.1. Práce navíc ukázala, že autoaktivace obou aktivních isoforem může být usnadněna pomocí sulfatovaných polysacharidů. Rekombinantní TrCB1.6wt (paralog se substitucí Cys^{29} na Gly v aktivním místě) byl aktivován pomocí aktivních forem TrCB1 a asparaginyl endopeptidázy. Ačkoliv TrCB1.6wt postrádá hydrolytickou aktivitu, jeho endopeptidázová, ale nikoliv karboxydipeptidázová, aktivita může být obnovena řízenou mutací Gly^{29} na Cys. Aktivní formy i inaktivní TrCB1.6wt paralog tvoří komplex s kuřecím cystatinem. To znamená, že neaktivní paralog si zachovává schopnost vázat cystatin a navíc je neočekávaně vysoce exprimován u schistosomuly. To naznačuje jeho roli v sekvestraci cysteinových inhibitorů hostitele za účelem ochrany aktivních peptidáz parazita (článek 4).

Introduction

The present thesis consists of two main parts: the literature review and the results of our research in the form of three original papers published in peer-reviewed journals and one manuscript in preparation. The literature review provides a brief insight into the biology of experimental models and an overview of the past and recent discoveries related to helminth peptidases. A particular regard is given to cysteine and aspartic peptidases, which are indispensable tools employed by many parasites (including blood-feeding helminths) during processing of blood proteins.

Literature review

1.1 Experimental models

In our research we used two blood-feeding helminths, a fish ectoparasite *Eudiplozoon nipponicum* (Monogenea, Diplozoidae**)** and an avian schistosome *Trichobilharzia regenti* (Trematoda, Schistosomatidae), to investigate their peptidases.

1.1.1 Eudiplozoon nipponicum

The East-Asian monogenean *E. nipponicum* Goto (1891) (Monogenea, Heteronchoinea, Diplozoidae) was first described under the name *Diplozoon nipponicum* based on specimens found on *Carassius vulgaris* obtained from a pond in the Shimura district, Japan (Goto, 1891). The species has become invasive and was first recorded in France in 1983 (Denis et al., 1983). Since then it has expanded throughout the Europe and currently constitutes a regular part of the parasite fauna of the common carp (*Cyprinus carpio*) even in the Czech Republic (Matějusová et al., 2001). Just recently, a genetic analysis and morphological comparison of diplozoids found on *Cyprinus carpio* and species of *Carassius* captured in Japan indicated that they may belong to different species (Nishihira and Urabe, 2020). According to the authors, the name *E. nipponicum* should be applied to the species associated with the type of host (*Carassius*) and the similar species infecting *C. carpio* (Eurasian type) should be established as a new species named *Eudiplozoon kamegaii*. However, the published data are not fully convincing to us because the conclusion was based mainly on the comparison of ITS sequences and was not supported by CoxI data sets. Additional samples and data from different localities in the world should be collected to justify this claim. Therefore, I will keep the name *E. nipponicum* throughout the thesis.

E. nipponicum is a haematophagous ectoparasite which inhabits the gills of the host fish. Its pathogenic effects are associated mainly with a mechanical damage to gill filaments accompanied by a risk of secondary bacterial and mycotic infections or hypochromic microcytic anaemia (Buchmann and Bresciani, 2006; Kawatsu, 1978).

The monoxenous lifecycle of *E. nipponicum* includes an egg, a free-swimming larva called oncomiracidium, post-oncomiracidial stage called diporpa (sexually immature stage on the host gill) and adult worm (**Fig. 1**). Diplozoids are unique in their pairing strategy, in which the worm does not become sexually mature until two diporpae meet and fuse permanently while creating an X-like shaped pair ("permanent copula", see Fig. 1). After somatic fusion, the two individuals grow together and their nervous, muscular, alimentary and also genital systems are fused reciprocally (Bovet, 1967; Khotenovsky, 1977; Pečínková et al., 2007).

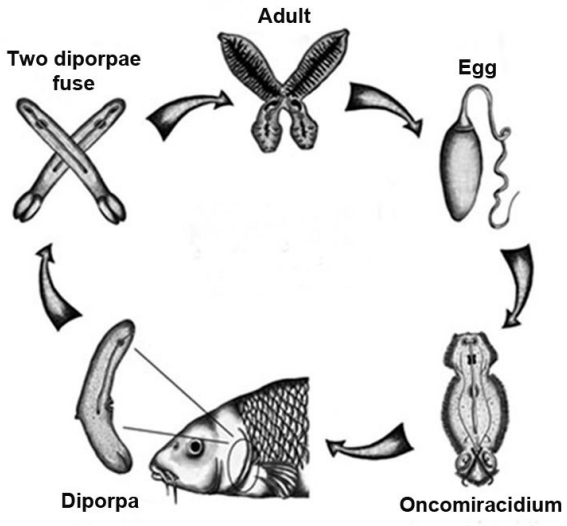


Fig.1. Lifecycle of members of the family Diplozoidae (Drawing by L. Jedličková)

1.1.2 Trichobilharzia regenti

Trichobilharzia regenti belongs to the family Schistosomatidae, a group of digenean trematodes parasitizing in the blood systems of birds or mammals (including humans) (Horák et al., 1998). It has an indirect lifecycle (**Fig. 2**) employing snails of the genus *Radix* as intermediate hosts and anatid birds as definitive hosts. The intermediate host becomes infected by ciliated miracidia that hatch from the eggs and swim in the water while seeking a snail. After a period of a sexual multiplication in the snail, the other aquatic larval stages, cercariae, emerge and infect the definitive host by direct penetration of the skin. Then the cercariae transform to schistosomula which migrate to their final location, where maturation, mating and egg production take place (for review see Horák et

al., 2019, 2015, 2002). This avian schistosome uses the peripheral and the central nervous system as an unusual migratory route to the final localisation in the blood vessels around the nasal cavity of the birds (Horák et al., 1999). Besides the skin of birds, the infectious larvae (cercariae) can also penetrate the skin of non-specific hosts (mammals including humans). Although this fluke is unable to reach adulthood in mammals, repeated infection of the skin can lead to an allergic reaction called cercarial dermatitis or "swimmer´s itch" (Horák et al., 2015; Kouřilová et al., 2004b, 2004a; Lichtenbergová et al., 2008; Macháček et al., 2018). Most parasites are trapped and eliminated already in the skin by this inflammatory reaction. In primary infections, the larvae may escape and migrate to the CNS and cause neuromotor disorders or leg paralysis in experimental mice, much like in the susceptible avian hosts. Those pathologies sometimes result in the death of the hosts (Horák et al., 1999; Hrádková and Horák, 2002; Kolářová et al., 2001; Lichtenbergová et al., 2008).

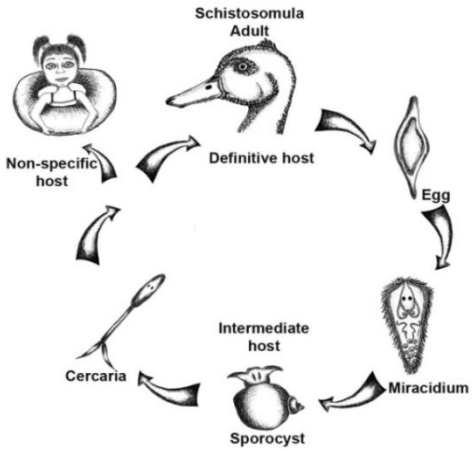


Fig. 2. Lifecycle of avian schistosomes of the genus *Trichobilharzia* (Drawing by L. Jedličková)

1.2 Peptidases

Peptidases (also called proteolytic enzymes, proteases or proteinases) are a class of enzymes, which hydrolyse the peptide bonds that link amino acids in a polypeptide chain. In general they play a major role in different key biological processes, for example cell death, DNA replication, immune responses, tissue remodelling, cell proliferation, protein processing and digestion (Rawlings and Salvesen, 2013). In parasites, they are often involved in invasion of the host, migration through tissues, pathogenesis, immune interactions, processing of nutrients including ingested blood, etc (Caffrey et al., 2013; Dalton et al., 2006; McKerrow, 2018a; McKerrow et al., 2006).

1.2.1 Classification of peptidases

Peptidases are usually classified according to three main criteria: "the place of action", catalytic type, and sequence/structural similarities. Depending on the location of the cleavage site within a substrate, peptidases are divided into endopeptidases, which target internal peptide bonds, and exopeptidases which cleave one or a few amino acids from the N- (aminopeptidases) or C-terminus (carboxypeptidases) of the polypeptide chain (Rawlings and Salvesen, 2013).

According to specific chemical groups responsible for catalysis of a peptide bond, peptidases are classified into six specific types: serine (S), cysteine (C), threonine (T), aspartic (A), glutamic (G) and metallopeptidases (M). The nucleophilic attack of serine, cysteine and threonine peptidases is mediated via the reactive group of an amino acid side chain, either a hydroxyl group of serine and threonine or a sulfhydryl group of cysteine peptidases. The aspartic, glutamic and metallopeptidases utilize an activated water molecule as a nucleophile. The water molecule is bound by the side chains of aspartic residues of aspartic peptidases, by glutamine/glutamic acid catalytic dyad of glutamic peptidases or by one or two metal ions (zinc, cobalt or manganese) of metallopeptidases. There are also peptidases for which the catalytic mechanism remains to be determined; these are classified as peptidases of unknown type and also peptidases with protein nucleophiles of mixed catalytic type (Rawlings and Salvesen, 2013; Rawlings et al., 2018).

The peptidases are also grouped according to their molecular structure and homology into families and clans. This hierarchical classification was established by Rawlings and Barrett in 1993 and represents the structure of the MEROPS peptidase database. In this, each peptidase is assigned to a family based on significant similarities in amino acid sequences. The family is expressed by the abbreviation of the catalytic type followed by a number (for example family C1). Groups of peptidase families can be arranged into "clans" based on similarities of tertiary structures. The clan name consists of two letters: the first defines the catalytic type and the second consists of letters from A-Z (for example clan CA) (Rawlings et al., 2018; Rawlings and Barrett, 1993).

1.2.2 The substrate specificity of a peptidase

The substrate specificity may provide a clue to the biological function of the peptidase and can be exploited in designing of selective peptidase inhibitors or efficient substrates. The general nomenclature of the cleavage site positions of the substrate was formulated by Schechter and Berger (1967). The authors defined that the active site of a peptidase is flanked on both sides by several peptidase subsites (binding pockets); each of

which can capture aside chain of one amino acid of the substrate. The amino acids towards the C-terminal of the substrate are on the prime side, and are numbered P1′, P2′, P3′, P4′ … Pn'. Residues towards the N-terminus are on the non-prime side and are numbered P1, P2, P3, P4 … Pn. P_n and P_n' residues of the cleaved substrate interact with complementary peptidase subsites called S_n and S_n'. The catalytic cleavage subsite is S1, and thus the cleavage of a peptide bond occurs between P1 and P1′amino acids of the substrate (**Fig.3**).

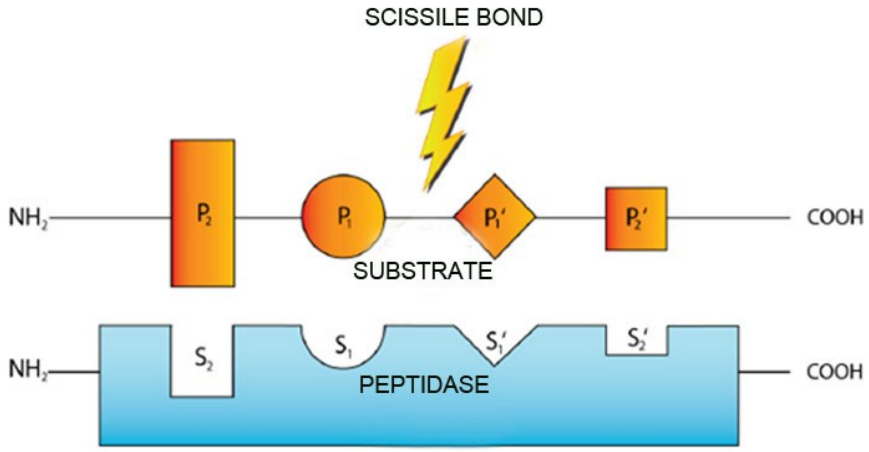


Fig.3. Schematic representation of a peptide substrate interaction with the subsites of a peptidase, according to Schechter and Berger (1967). The peptide amino acid residues (P) and subsites (S) are numbered and designated prime or non-prime based on their position relative to the targeted peptide bond.

1.2.3 Regulation of peptidase activity

Peptidase activity must be strictly regulated at different levels since uncontrolled proteolysis could lead to pathophysiological states and eventual death of the cell/organism. The first regulation occurs already at the level of the gene expression, for example by epigenetic changes in the promoter regions of a peptidase gene (Yan and Boyd, 2007). Alternative splicing of mRNA may also happen, which may affect the rate of the protein expression, lead to undesirable extracellular secretion or change of target compartment (Baici et al., 2006; Müntener et al., 2004).

The action of peptidases could be also controlled by production of an inactive precursor (zymogen or pro-enzyme) with an N-terminal pro-peptide extension that blocks the active site. Orientation of the pro-peptide at the active site is opposite to the binding substrate, thus preventing uncontrolled hydrolysis of the peptide bond between the pro-peptide and the core of the parent enzyme (Coulombe et al., 1996). Apart from acting as an intramolecular inhibitor, the pro-peptide can also participate in correct intracellular targeting (Huete-Pérez et al., 1999; Mcintyre et al., 1994; Wiederanders et al., 2003) or assist in folding of the mature enzyme (Cappetta et al., 2002; Tao et al., 1994). Activation

of the precursor occurs by cleavage and removal of the pro-peptide from the active site. However, the pro-peptide may remain a potent inhibitor of the peptidase active site after release (Fox et al., 1992; Guay et al., 2000; Hánová et al., 2018; Horn et al., 2011). Partial selectivity of the pro-peptide for the parent enzyme makes it a tool for design of new inhibitors (Carmona et al., 1996; Chowdhury et al., 2002; Horn et al., 2011). Proteolytic cleavage of the pro-peptide can be autocatalytic (Menard et al., 1998; Rowan et al., 1992) or involve other trans-activating peptidases (Nishimura et al., 1989). For example, it was suggested that asparaginyl endopeptidase (legumain) of *S. mansoni* (SmAE) trans-activates other schistosomal peptidases (cathepsin D, L, F, B1 and C), since their zymogens contain a hypothetical cleavage site for asparaginyl endopeptidase between the pro-sequence and the catalytic domain (Dalton and Brindley, 1996). This claim was supported by an in vitro experiment when recombinant SmAE was able to trans-activate *S. mansoni* cathepsin B1 (SmCB1) (Sajid et al., 2003). Autocatalytic activation often requires additional factors such as drop in pH or presence of strongly negatively charged sulphated polysaccharides, such as glycosaminoglycans (GAGs) (e.g., heparin) (**Fig.4**). For example, GAGs are essential for the autocatalytic activation of SmCB1 (Jílková et al., 2014) as well as for cathepsin B1 of *T. regenti* (**paper 4** Dvořáková et al., 2020) and they were also reported to accelerate or facilitate activation of other cathepsins (Beckham et al., 2006; Caglič et al., 2007; Fairhead et al., 2007; Vasiljeva et al., 2005). In addition to their role in activation, GAGs were shown to stabilize cathepsin B (partial protection against alkaline-pH-induced inactivation) at neutral pH (Almeida et al., 2001), to protect the serine peptidase cathepsin G against inhibition by its inhibitor (Ermolieffs et al., 1994) or to modulate the collagenolytic activity of cathepsins K, L and S (Li et al., 2002; Li et al., 2004).

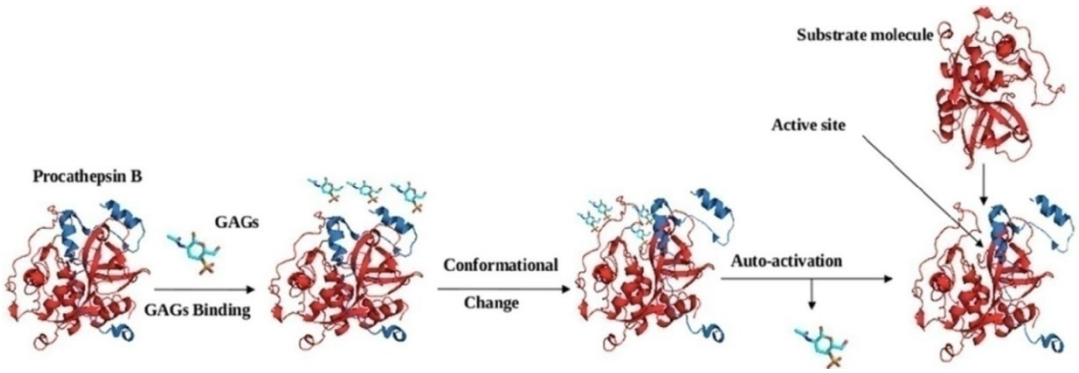


Fig.4. Mechanism of human pro-cathepsin B auto-activation in the presence of glycosaminoglycans (GAGs). GAG molecules bind to a pro-cathepsin B molecule (pro-peptide in blue and the mature domen in red) in the occluding loop and the propeptide via electrostatic interactions. GAGs binding induce conformational change in pro-cathepsin B, which unmasks the active site and allows the access of an another complex pro-cathepsin B-GAG (Caglič et al., 2007; Verma et al., 2016).

An essential regulating factor of the proteolytic activity of peptidases is also the environment. In particular, mammalian cysteine cathepsins require reducing and slightly acidic environment, such as found in the lysosomes, in order to be optimally active. When cathepsins are outside the lysosomes or extracellularly, they can be usually, with the exception of cathepsin S, rapidly irreversibly inactivated at neutral pH (Kirschke et al., 1989). This represents a control mechanism that protects cells against uncontrolled proteolysis when the enzyme is accidentally released from lysosomes (Turk et al., 1995). On the other hand, many parasitic cysteine cathepsins function in a broader chemical environment and have a pH optimum closer to neutral pH, which corresponds with their role in extracellular processes (Sajid and McKerrow, 2002). Cysteine peptidases are also readily inactivated by oxidation of the active site cysteine and require reducing environment for full activity (Dickinson, 2002). While glutathione is provided as a reducing agent *in vivo,* it is often necessary to add mercaptoethanol or dithiothreitol (DTT) during *in vitro* experiments (Szabelski et al., 2005).

One of the major control mechanisms governing the peptidase activity is the presence of peptidase inhibitors. Peptidase inhibitors are classified into two categories, reversible or irreversible inhibitor. An irreversible inhibitor forms a stable complex with the enzyme. Both, inhibitor and peptidase molecules get deformed and as a result, the peptidase is permanently inactivated. A reversible inhibitor does not cause any permanent changes in the enzyme and can be easily removed by dilution or dialysis. Majority of peptidase inhibitors are competitive. The competitive inhibitor mimics the substrate and binds to the active site of the peptidase. In non-competitive inhibition, inhibitor binds to the enzyme at a location other than the active site, inducing usually change in the shape of the enzyme molecule, thereby deforming its active site which is no longer functional (Farady and Craik, 2010; Rawlings and Salvesen, 2013). Peptidase inhibitors are also divided according to their molecular weight. The low molecular weight inhibitors (less than 1 kDa) have been identified in many natural resources from which they were subsequently isolated. The first discovered was the non-selective inhibitor of the cysteine cathepsins - E-64 (trans-epoxysuccinyl-l-leucylamido-(4-guanidino)butane) isolated from the extract of a solid culture of the fungus *Aspergillus japonicus* (Hanada et al., 1978). The structure of the natural inhibitor E-64 was modified to develop new selective inhibitors of cathepsin B, e.g., CA-030 (N-[L-3-trans-ethoxycarbonyloxirane-2-carbonyl]-L-isoleucyl-L-proline) and CA-074 (N-[L-3-trans-(propylcarbamoyl)oxirane-2- carbonyl]-L-isoleucyl-L-proline) (Murata et al., 1991; Towatari et al., 1991). Those inhibitors were designed to take the advantage of an extra structural element termed occluding loop, which occupies

the S′ subsites of cathepsin B (see also section 1.2.4.1 Cysteine cathepsins). The crystal structures of cathepsin B with CA-074 or CA-030 have shown that the carboxyl group of proline interacts with the His^{110} and His^{111} (human cathepsin B numbering) in the occluding loop similarly to substrate binding (Turk et al., 1995; Yamamoto et al., 1997).

High molecular weight inhibitors are also known as protein inhibitors. They are divided into families and clans based on similarities in primary, secondary and tertiary structures of inhibitory units. The MEROPS database currently contains 40 clans covering 107 families of protein inhibitors that express a wide variety of inhibition mechanisms (Rawlings et al., 2018).

1.2.4 Cathepsins

The term "cathepsin" originated from the Greek word “kathepsein”, which means “to digest” or "boil down" and dates back more than 90 years (Willstätter and Bamann, 1929). Nowadays, the cathepsins represent a very heterogeneous group of enzymes of three peptidase classes - serine (cathepsin A and G), aspartic (cathepsin D and E) and the more numerous cysteine peptidases. Considering the content of our papers presented, I will deal only with cysteine and aspartic cathepsins (particularly cathepsin D) in the following chapters.

1.2.4.1 Cysteine cathepsins

Based on their primary and tertiary structures, cysteine cathepsins are members of clan CA and are widely distributed in living organisms. More specifically, they belong to the C1 family of papain-like enzymes, the largest and the best characterized family of cysteine peptidases (Rawlings and Salvesen, 2013). Members of the C1 family share structure and evolutionary history with the papain (from the plant *Carica papaya*),of which proteolytic activity has been known since the 1880s (Martin, 1884). Besides papain, the family contains mainly mammalian cathepsins, but also homologous enzymes present in other vertebrates, invertebrates, plants and viruses (Rawlings and Salvesen, 2013). In human, eleven cysteine cathepsins (B, C, F, H, K, L, O, S, V, X and W) were confirmed by bioinformatic analysis of the draft sequence of the human genome (Rossi et al., 2004).

Apart from the housekeeping function of lysosomal protein degradation, other specific functions of cysteine cathepsins, have been discovered. Moreover, their presence in other compartments (such as the nucleus, cytoplasm and plasma membrane) and extracellular localisation were also found. For example, it was shown that the catalytically active cathepsin L localized to the nucleus plays a role in regulation of cell-cycle

progression (Goulet et al., 2004) and proteolytic processing of the N-terminus of the histone H3 tail (Duncan et al., 2008).

When secreted, cathepsins can be also very harmful to their environment. Thus, cysteine cathepsins are involved in a number of diseases such as cancer, Alzheimer's disease, multiple sclerosis, osteoporosis etc. (see in Kramer et al., 2017; Reiser et al., 2010; Siklos et al., 2015).Genetic defects in cathepsins may also cause severe physiological disorders. For example, mutation in human cathepsin C gene was shown to cause Papillon-Lefèvre syndrome, characterized by palmo-plantar hyperkeratosis and severe early-onset periodontitis (Hart et al., 1999).

Moreover, the papain-like family includes most of the parasite cysteine peptidases that play important roles in the development and nutrition, as well as in the passage through the tissues and in the defence against the host immune system (Dalton et al., 2006). Thus, cysteine peptidases seem to be good candidates for anti-parasitic treatment. As with many other pathogens, parasites resistant against available anti-parasitic drugs have emerged and, therefore, there is an urgent need to identify additional strategies and suitable molecular targets in these parasites for the development of new drugs. One promising inhibitor that has emerged over the last two decades is the peptidomimetic vinyl sulphone K11777 (4-methyl-N-((S)-1-oxo-3-phenyl-1-(((S)-1-phenyl-5-(phenylsulfonyl)pentan-3-yl)amino)propan-2-yl)piperazine-1-carboxamide), an irreversible cysteine peptidase inhibitor (Palmer et al., 1995). K11777 (also known as K777) has been efficacious in preclinical studies on mice and dogs infected by the protozoan parasite *Trypanosoma cruzi* (the causative agent of Chagas disease) (Barr et al., 2005; Doyle et al., 2007; Engel et al., 1998). Subsequent studies have described its therapeutic benefit in other protozoan infections (Caffrey et al., 2000; Chaparro et al., 2018; Meléndez-López et al., 2007; Ndao et al., 2013), also against the blood fluke *Schistosoma mansoni* (Abdulla et al., 2007) and the hookworm *Ancylostoma ceylanicum* (Vermeire et al., 2012). Importantly, K11777 inhibitor was found to be safe in rodents, dogs, and non-human primates. Safety testing included 7-, 14- and 28-day chronic dosing (McKerrow, 2018b).

Cysteine cathepsins are also generally immunogenic and have been exploited as vaccine targets to prevent diseases, or at least decrease morbidity or reduce egg outputs (see review Hewitson and Maizels, 2014; Loukas et al., 2011). They have also been investigated as serodiagnostic markers (Cornelissen et al., 1999; Li et al., 1996).

The mature cysteine cathepsins are generally small (Mw 20-30 kDa) and mostly monomeric, with the exception of ~200 kDa tetrameric cathepsin C (Turk et al., 2001). The overall fold of cathepsins is highly conserved and consists of two domains, the N-terminal

domain referred to as the left (L-) and the C-terminal domain as right (R-). The two-domain interface opens at the top, forming the active site cleft. The active site cysteine residue is located in a structurally conserved α-helix of the L-domain whereas the histidine is located within the β-barrel residues of the R-domain (see review McGrath, 1999). These two catalytic residues form the thiolate-imidazolium ion pair, which is essential for proteolytic activity of the enzymes. Moreover, the imidazolium ring of the catalytic His is frequently orientated by an Asn residue, thereby stabilizing the ion pair. First, sulphur from the active site Cys residue attacks the carbon of the reactive peptide bond, producing a tetrahedral intermediate stabilized by the so-called oxyanion hole containing a highly conserved Gln residue. The tetrahedral intermediate then accepts a proton from the imidazolium group, resulting in formation of an acyl enzyme intermediate and release of the C-terminal portion of the substrate. In a second reaction, the acyl enzyme intermediate is deacylated by a water molecule and the remaining portion of the substrate is released (**Fig. 5**) (Buttle and Mort, 2004).

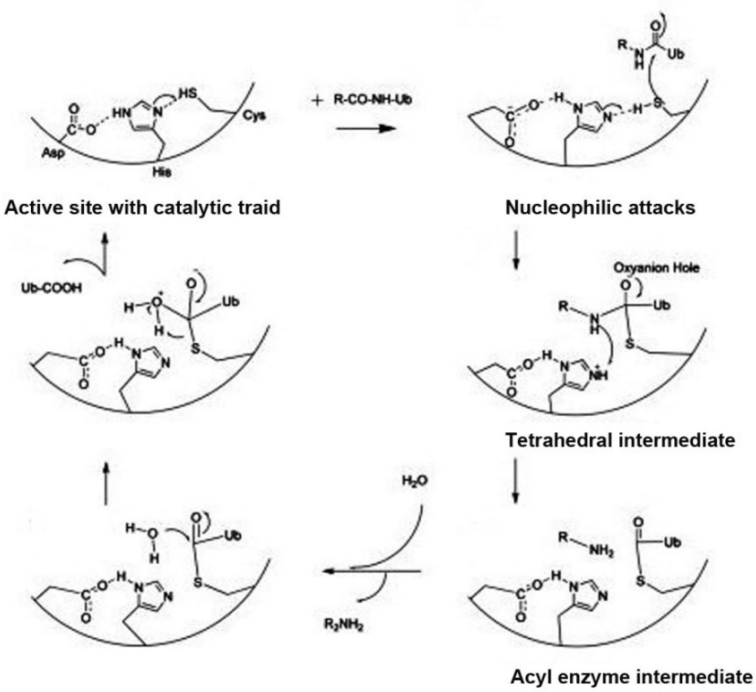


Fig.5. Substrate hydrolysis by cysteine cathepsins (modified from Hanpude et al., 2015)

Most of the cysteine cathepsins are endopeptidases; some, however, evolved into exopeptidases (cathepsins B, C, H and X). While cathepsin B and H possess both exo- and endo- activities, cathepsin C and X are strict exopeptidases. The crystal structures revealed that exopeptidases contain additional structural elements that restrict binding of substrate into the active site (Guncar et al., 2000, 1998; Musil et al., 1991; Nägler et al., 1999; Turk et al., 2001). For example, the exopeptidase activity of cathepsin B is exerted by a special

structure called occluding loop, an insertion of ca. 20-30 amino acids residues located on the prime side of the catalytic site. It blocks binding of substrate beyond the S2 ′ position and thereby enables cathepsin B to act as a peptidyl-dipeptidase (Illy et al., 1997; Musil et al., 1991). The binding of the occluding loop is pH dependent and is regulated by the protonation of the side chains of His^{110} and His^{111} present in the loop (human cathepsin B numbering). At low pH (<5) these groups are protonated and the occluding loop in bound into the active site. At high pH (>5), deprotonation of these groups causes a displacement of the occluding loop and the enzyme exhibits endopeptidase activity (Musil et al., 1991).Removal of the occluding loop abolishes exopeptidase activity and the binding of CA-074, but increases endopeptidase activity as well as the susceptibility to macromolecular inhibitors such as cystatin (Illy et al., 1997).

The crystal structures of the complexes with small-molecule, substrate-mimicking inhibitors provided support for only three well-defined substrate binding sites, S2, S1 and S1′, where the amino acid residues of the substrate interact with the main- and side-chain atoms at the surface of the enzyme. Among these only the S2 site deserves the term pocket. Interaction of the substrate with the enzyme outside this region involves only a side-chain interaction. Therefore, these sites were recommended to be termed as areas in which the substrate residues individually find their most favourable binding position (Turk et al., 2003, 1998). In general, cysteine cathepsins prefer large hydrophobic amino acids at P2 position. Although the structurally similar human cathepsins L, K, and S have similar binding sites and specificities, cathepsin L favours aromatic residues, particularly Phe. A similar specificity is characteristic of cathepsin V, a close cathepsin L homolog. On the other hand, cathepsins K and S will take any hydrophobic residue (Choe et al., 2006). Cathepsin K also exhibits a unique preference for a proline residue at P2 among mammalian cathepsins (Lecaille et al., 2002). Mammalian cathepsin B, as an endopeptidase, also favours large hydrophobic amino acids in the S2 subsite, but also accepts Arg at this position which is an exception among other cysteine cathepsins (Choe et al., 2006). This is due to the presence of a Glu residue lying at the bottom of the S2 subsite of cathepsin B, acting as an acceptor for the positive charge on Arg side-chain (Hasnains et al., 1993).This ability is true for mammalian cathepsins B and for cathepsins B of many other organisms, but there are also cathepsins B which do not have this ability - for example cathepsin B of *Leishmania major* (Chan et al., 1999) or some cathepsins B of *T. regenti* (TrCB1.4 and TrCB1.6) (Dvořák et al., 2005; **paper 4** Dvořáková et al., 2020). On the contrary, cathepsins L with “cathepsin B-like” substrate specificity also exist (Gillmor, et al., 1997; **paper 2** Jedličková et al., 2018).

An important group of inhibitors of the cysteine cathepsins are cystatins. They are tight-binding inhibitors which block the active site of the target peptidase on an equimolar basis and the interaction is reversible (Rawlings et al., 2004). Cystatins differ in their inhibitory specificity against various cathepsins. For example, the presence of the occluding loop was proposed to be responsible for the generally lower affinities of cystatins to cathepsin B, compared with other endopeptidases of the papain-like family with open active sites (Naudin et al., 2010; Nycander et al., 1998; Pavlova et al., 2000). Removal of the flexible part of the occluding loop from cathepsin B increases its affinity for cystatin C about 40-fold (Illy et al., 1997) and reduction of electrostatic interaction between the loop and the body of the mature human cathepsin B (mutating Asp^{244} to Asn in order to break the salt bridge with Arg^{116}) results in an increase of its affinity to cystatin by an order of magnitude (Podobnik et al., 1997).Besides their occurrence in mammals and plants, cystatins have also been detected in several species of parasites. Studies to date suggest a variety of functions from cysteine peptidase regulation within the parasite itself to modulation of host immune response (Hartmann et al., 2002; Ilgová et al., 2017; Kang et al., 2014; Morales et al., 2004; Schönemeyer et al., 2001). For example, a stefin (type I cystatin) produced by the sanguinivorous fish parasite *Eudiplozoon nipponicum* (**Fig. 6**), with atypical asparaginyl endopeptidase-binding domain has putative roles in the regulation of blood digestion, modulation of antigen presentation, and in the regeneration of host tissues (Ilgová et al., 2017).

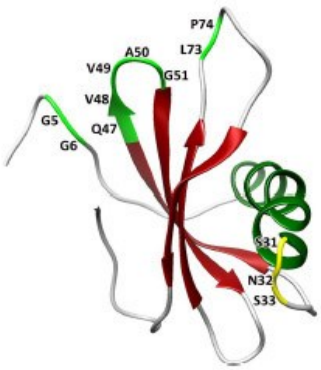


Fig. 6. Predicted 3D structure for the *E. nipponicum* stefin. The structure comprises a four-strand, anti-parallel beta-sheet, and a five-turn alpha-helix. Motifs critical for binding to papain- like peptidases (G5G6, Q47VVAG51 and L73P74) are highlighted in light green, while the asparaginyl endopeptidase-binding domain (S31N32S33) is highlighted in yellow (Ilgová et al., 2017).

Although the repertoire of helminth cysteine cathepsins is less extensive than that in their mammalian hosts, these parasites typically own several paralogs of these peptidases

(comprising either allelic variants and/or genes) (Cancela et al., 2008; Dvořák et al., 2005; Guiliano et al., 2004; Ranjit et al., 2008) that may improve the efficiency and redundancy in degrading a variety of physiologic substrates, as well as possibly avoiding targeted immune responses (Caffrey et al., 2013). Above that, these gene paralogs and/or isoforms are often expressed in various life stages, thus reflecting the parasite's specific needs during its life cycle. For example, characterisation of several distinct cathepsins L of the liver fluke *Fasciola hepatica* had shown that the newly excysted juvenile flukes secrete cathepsins L3and B to penetrate the host´s intestinal wall, while the immature stages use both L1 and L2 to migrate through the host´s liver tissue, and adults employ cathepsin L1 jointly with cathepsins L2 and L5 to digest host proteins (McVeigh et al., 2012).

In multicellular parasites such platyhelminths and nematodes, the digestive tract is a major source of cysteine cathepsins (e.g. Brindley et al., 1997; Caffrey et al., 2004; **papers 1 and 2** Jedličková et al., 2018, 2016, Williamson et al., 2004, 2003). Together with cathepsin D aspartic endopeptidases (see also sections1.2.4.1 Aspartic peptidases), cysteine cathepsins form part of a network that generates absorbable dipeptides or amino acids from ingested host proteins (see also section 1.3 Digestion in haematophagous helminths).

Besides general catabolic functions and protein processing, cysteine peptidases may be key to helminth immune evasion (e.g. Macháček et al., 2016; O'Neill et al., 2001; Shin et al., 2001), skin and tissue/cellular invasion (e.g. Dolečková et al., 2009; McGonigle et al., 2008; Na et al., 2006), excystment (e.g. Chung et al., 2005; S. Li et al., 2004), moulting (e.g. Guiliano et al., 2004; Lustigman et al., 2004), reproduction and development (e.g. **paper 2** Jedličková et al., 2018; Kang et al., 1995; Michel et al., 1995; Yang et al., 2011).

1.2.4.1 Aspartic peptidases (cathepsin D)

According to MEROPS classification (Rawlings et al., 2018), cathepsin D belongs to the clan AA family A1 (also known as pepsin-like enzymes) of aspartic peptidases. The peptidases of the family A1 are endopeptidases and are found predominantly in eukaryotes but examples are also known from bacteria (Rawlings and Bateman, 2009; Simões et al., 2011). Most members of the family have a low pH optimum, but a few show activity under neutral conditions, e.g., kidney renin (Inagami and Murakami, 1977). Examples of the enzymes from mammals include the digestive enzymes pepsin and chymosin and intracellular cathepsins D and E.

In the early studies, it was accepted that cathepsin D mainly participates in rather non-specific degradation of intracellular proteins within the acidic endosomal-lysosomal compartments of a diverse range of mammalian cells. The data obtained over last three

decades, however, demonstrated that enzymatic action of cathepsin D is not restricted only to lysosomes, but also has important consequences in regulation of programmed cell death - apoptosis. Cathepsin D has also been studied from the perspective of its role in cancer development and in other pathological processes. The complex roles that mammalian cathepsin D plays both in physiology and pathology are reviewed for example in (Benes et al., 2008; Masson et al., 2010). In parasites, cathepsin D contributes mainly to the degradation of host proteins (Sojka et al., 2016) (see also1.3 Digestion in haematophagous helminths).

The A1 family comprises proteins with a three-dimensional structure close to that of pepsin. Crystal structures of enzymes from this family show that the molecules are bi-lobed, with the active site located between the lobes. In spite of that, one lobe has evolved from the other by gene duplication, but very little amino acid sequence similarity remained. However, the three-dimensional structures are very similar. Although the two lobes are structurally almost similar, the lobe on the right-hand side has an extra β-hairpin loop (termed the "flap") capping the active site and carries residues that are important for specificity, notably Tyr^{137} and Thr^{139} (in pepsin numbering) which can interact with the substrate. The active site is formed by a dyad of Asp residues, one from each of the lobes of the enzyme. Each Asp residue occurs in a very similar motif within the sequence Xaa-Xaa-Asp-Xbb-Gly-Xbb (the "DTG" motif), in which Xaa is a hydrophobic residue and Xbb is either Thr or Ser (Rawlings and Barrett, 2013).

Unlike cysteine peptidases, the aspartic peptidases do not form a covalent intermediate during cleavage of peptide bonds. Aspartic peptidases use a water molecule as a nucleophile which is activated by one of the Asp residue by attracting the hydrogen atom of water. The other Asp residue polarizes the carbonyl group on the peptide making it easier to attack. The water molecule then attacks the carbonyl peptide bonds and the bond is thus cleaved (**Fig. 7**) (Paul et al., 2016).

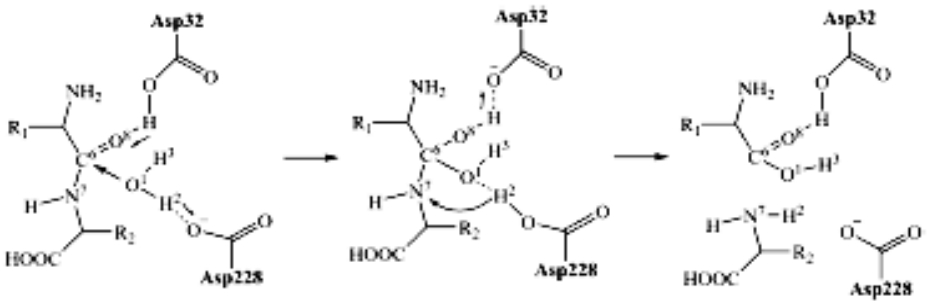


Fig.7. Catalytic mechanisms of aspartic peptidases (Paul et al., 2016).

Cathepsin D accommodates up to eight amino acid residues in the binding cleft of the active site (approximately from P5 to P3′). There is a preference for hydrophobic residues around the cleaved bond (P1 and P1′position) (Masa et al., 2006). Important differences between cathepsin D and other aspartic peptidases were found in the P2 position (with optimal Glu) and P2′ (with preference for basic residues) of the substrate (Dunn et al., 1995).

A non-specific inhibitor of all aspartic peptidases of the A1 family is pepstatin A that was isolated from various species of actinomycetes. It is a hexapeptide containing the amino acid statine in the sequence Isovaleryl-Val-Val-Sta-Ala-Sta (Umezawa et al., 1970). Despite the importance of pepsin-like enzymes, a little information is available about specific endogenous molecules which regulate their activity. The only known endogenous mammalian cathepsin D-specific inhibitors are sphingolipids which modulate the peptidolytic activity of human cathepsin D (Žebrakovská et al., 2011). Recently, it was described that cathepsin D could be inhibited by a fragment derived from its own cleaved pro-peptide. After activation, a free pro-peptide fragment interacts with an exosite located outside the active-site cleft, and thereby inhibits the parental enzyme through a novel allosteric mechanism (Hánová et al., 2018).

In various blood-feeding platyhelminths and nematodes, the cathepsin D-like molecule has been proved as one of the most crucial enzymes in the digestive cascade (Brindley et al., 2001; Brinkworth et al., 2001; Delcroix et al., 2006; Kang et al., 2019; Williamson et al., 2002; Wong et al., 1997), usually being involved in the initial processing of host blood proteins (see also section Biochemistry of helminth digestion). In addition, it was found that *S. japonicum* cathepsin D (SjCD) is able to effectively cleave human IgG by removing Fc fragments or degrade C3 factor of the complement which might be useful for immune evasion (Verity et al., 2001). Moreover, cathepsin D-like molecules were detected in other tissues (for example reproductive) or in other developing stages of helminths suggesting their indispensable role in the host–parasite relationship (Jolodar et al., 2004; Kang et al., 2019; Suttiprapa et al., 2009; Verity et al., 1999).

1.2.5 Non-peptidase homologs

Many of the peptidase families contain proteins that have peptidase-like sequences but appear to lack the residues essential for catalytic activity. Examples of homologs with amino acid mutations in the catalytic site are also known from many parasite peptidases (Caffrey et al., 2000; Dvořák et al., 2005; He et al., 2005; Holt et al., 2004; **paper 2** Jedličková et al., 2018; Mccoubrie et al., 2007; Mendoza-Palomares et al., 2008;

Merckelbach et al., 1994; Pillay et al., 2010, **paper 4** Dvořáková et al. 2020, manuscript 3 in preparation Dvořáková and Mikeš). In fact, of 259 proteins from parasitic protozoa in the family C1 (papain-like cysteine peptidase), 16% do not appear to have the classical Cys-His catalytic dyad and 8% of helminth family C1 peptidases have substitutions at these positions (Atkinson et al., 2009). Some of these proteins still appear to be enzyme-like, and the active site might be catalytically active, as in the case of Cys-to-Ser mutation (McCoubrie et al., 2007; Pillay et al., 2010) It has also been suggested that this mutation in cysteine peptidases of parasites might be a way to moderate oxidation of the active site in case the parasites have to deal with an oxygen-rich environment, such as blood or external environment (Atkinson et al., 2009). However, in some cases the substitutions affect the proteolytic ability of an enzyme, which is the case of cathepsin B isoforms of *T. regenti* TrCB1.6 and probably TrCB1.5, too, (**paper 3** Dvořáková et al., 2020) or of asparaginyl endopeptidases of *S. mansoni*, Sm32 (SmAE) (Caffrey et al., 2000). The biological significance of these inactive homologs is not entirely known but they might regulate the activity of active forms by competing for substrates or inhibitors (Dvořák et al., 2005; **paper 4** Dvořáková et al., 2020; Holt et al., 2004, 2003; Merckelbach et al., 1994). Moreover, they may alter host immune response (to the benefit of the parasite) through the inhibition of host complement pathways by binding the complement components C1q, mannose-binding lectin or properdin (Bergström et al., 2009; Holt et al., 2004; Reynolds et al., 2014).

1.3 Digestion of blood by haematophagous helminths

Since blood provides a rich source of energy and easily accessible building materials for processes such as growth, development and reproduction, it is not surprising that during the evolution, haematophagy evolved independently in many different groups of organisms. In blood-feeding parasites, haemoglobin and other blood proteins are usually cleaved by the action of aspartic peptidases homologs of cathepsin D (clan AA), with a contribution of cysteine peptidases of the clan CA. The cleavage to absorbable dipeptides and liberation of free amino acids is accomplished by exopeptidases, cysteine dipeptidases and metalloaminopeptidases and/or serine and cysteine carboxypeptidases. This multi-enzyme digestive pathway has been described for such different organisms as the protozoan malaria parasite (Liu et al., 2006; Rosenthal, 2004), for lower invertebrates such as mites (Mendiola et al., 1996; Sojka et al., 2008) and blood-feeding helminths (Delcroix et al., 2006; **paper 1** Jedličková et al., 2016; **paper 2** Jedličková et al. 2018, Williamson et al., 2004) (**Fig. 8**). A very different picture occurs in insects (and some other invertebrates)

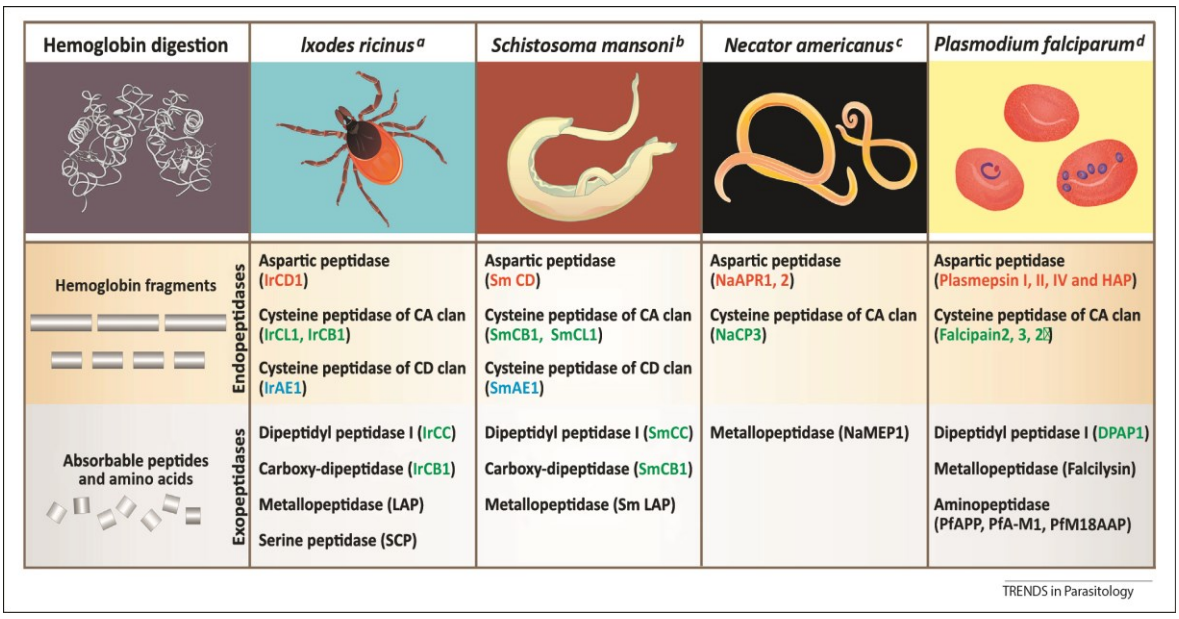


Fig.8.Comparison of haemoglobinolytic systems in haematophagous parasites (Sojka et al., 2013).

which involved serine peptidases (primarily members of the trypsin family) in digestion (Lehane et al., 1996; Terra and Ferreira, 1994). The reason for this switch remains unknown, but one might envision that the change reflected adaptation to the terrestrial life with higher oxygen tension, since cysteine peptidases are more easily inactivated than serine peptidases in an oxygen-rich environment (McKerrow, 2018a).

In the blood fluke *Schistosoma mansoni*, the extracellular digestion occurs probably in acidic microenvironments generated between the lamellae of gastrodermis where aspartic and cysteine peptidases have been localized, and where the pH value of approximately 4.5 would allow both classes of the enzymes to cleave host proteins efficiently (Brindley et al., 2001; Delcroix et al., 2006; Sajid et al., 2003).

Several gut-associated peptidases have been identified in adult *S. mansoni*. These include cysteine peptidases of the CA clan-cathepsin B1 with its dual endopeptidase and peptidyl dipeptidase activity (SmCB1), cathepsin L1 (SmCL1, later determined as cathepsin F - SmCF), L2 and L3 (SmCL2 and 3) and dipeptidyl peptidase I (cathepsin C, SmCC), cysteine peptidase of the CD clan asparaginyl endopeptidase (legumain, SmAE), aspartatic of the clan AA - cathepsin D (SmCD) and clan MF leucine metalloaminopeptidase (SmLAP) (Bogitsh et al., 2001; Brindley et al., 2001; Butler et al., 1995; Caffrey et al., 2000; Dvořák et al., 2009; Hola-Jamriska et al., 1999; McCarthy et al., 2004; Sajid et al., 2003). A combination of the class-specific peptidase inhibitors with siRNA showed that the major endopeptidases in the schistosome gut function in a cooperative network and a large number of peptidases provide some redundancy.

Haemoglobin is initially cleaved into large fragments by the action of SmCD. Although SmCD is most effective, even if it is absent some cleavage still occurs. Therefore, some redundancy exists with the cysteine peptidases for this cleavage. The products resulting from initial cleavage are further degraded by SmCB1 and SmCLs. On the other hand, for albumin degradation, cathepsin B appears to play the primary role in initiating the cleavage but again with some redundancy. The resulting albumin fragments are further digested with endopeptidases SmCD, SmCB1 and SmCLs (Delcroix et al., 2006). Host protein degradation is completed by the exopeptidase activities of SmCB1, SmCC (Butler et al., 1995) and SmLAP (McCarthy et al., 2004). A putative role of the SmAE might be in trans-activation of other endopeptidase zymogens (cathepsin B, L and D) (Dalton and Brindley, 1996; Sajid et al., 2003) and/or in cooperation with cathepsin B and L in direct degradation of substrates (Delcroix et al., 2006). Although most of the research has focused on digestion in *S. mansoni*, it is likely that a similar system of orthologous peptidases facilitating nutrition and migration exists in other schistosome species, including the bird schistosome *Trichobilharzia regenti* (Dvořák et al., 2005; **paper 4** Dvořáková et al., 2020). Another trematode parasite, the liver fluke *Fasciola hepatica*, combines extracellular digestion of haemoglobin in the gut performed by a group of cathepsin L peptidases with intracellular degradation of haemoglobin peptides within gastrodermal epithelial cells by cathepsin C and other aminopeptidases (Robinson et al., 2008).

A number of peptidases of all types, which are homologs to those involved in haemoglobin digestion in *S. mansoni,* have also been found in the gut of many haematophagous nematodes (Williamson et al., 2003). It was also proven that haemoglobin is degraded by a multi-enzyme cascade in the alimentary tract of these parasites (Williamson et al., 2004). For example, a cathepsin D-like peptidase from the canine hookworm *Ancylostoma caninum* (Ac-APR-1) and an orthologous peptidase from the human hookworm *Necator americanus* (Na-APR-1) cleaved haemoglobin at numerous distinct positions including the hinge region, a step that would facilitate unravelling of haemoglobin tetramer and further proteolysis by other digestive enzymes (Williamson et al., 2002). Moreover, it has been shown that cathepsin B from the intestine of *A. caninum* achieved its optimal activity when it acted after cathepsin D from the same parasite (Williamson et al., 2004), so it is a good reason to place them both in the enzymatic cascade that digests haemoglobin in this parasite. Interestingly, a second family of aspartic peptidases has been identified in the gut of blood-feeding nematodes. As these aspartic peptidases more closely resemble mammalian pepsin than cathepsin D, it has been suggested that this family of nematode-specific enzymes be termed the “nemepsins,” for

"nematode pepsin-like enzymes".(Williamson et al., 2003). These include Pep1 from *Haemonchus contortus* (Longbottom et al., 1997), Na-APR-2 from adult *N. americanus* (Williamson et al., 2003) and a similar peptidase from infective larvae of *Strongyloides stercoralis* (Gallego et al., 1998). Unlike fasciolid flukes, blood-feeding strongyle nematodes use a group of cathepsins B rather than cathepsins L for digestion of blood proteins (Jasmer et al., 2004; Ni et al., 2012; Ranjit et al., 2008).

In contrast to other blood-feeding parasites, current knowledge of bioactive molecules involved in digestion and other processes by haematophagous monogeneans is poor. The digestion in monogeneans has been believed to occur probably inside a specialized type of intestinal cells (hematin cells) within the lysosomal cycle (Konstanzová et al., 2015; Smyth and Halton, 1983; Valigurová et al., 2011). However, the enzymes of intracellular origin may be released from the digestive cells into the lumen of the gut during exocytosis of residual material and thus may also contribute to the extracellular phase of digestion in monogeneans. This supposition was confirmed by the fact that cathepsin L1 and L3 of a diplozoid monogenean *E. nipponicum* were localised not only in the haematin cells but also in gut lumen of the worm's digestive tract (and some of the enzymes possess a signal sequence) (**paper 2** Jedličková et al., 2018) and in excretory/secretory products of the worms (**paper 1** Jedličková et al., 2016; Vorel et al., 2021). So, the combination of extracellular and intracellular digestion appears to be similar to the liver fluke *F. hepatica.*

It has been also assumed that the processing of host blood proteins in diplozoid monogeneans depends on a network of cysteine peptidases **(papers 1 and 2** Jedličková et al., 2018, 2016) in cooperation with aspartic peptidases (cathepsins D) (**paper 1** Jedličková et al., 2016; Dvořáková and Mikeš - **manuscript 3 in preparation**), where cysteine endopeptidases of cathepsin L type probably play a major role (**paper 2** Jedličková et al., 2018).

Aims of the thesis

This work focuses on determination and characterisation of acid endopeptidases employed by a haematophagous monogenean *Eudiplozoon nipponicum* during processing of host blood proteins. The main aim was to verify the hypothesis that multiple-enzyme cascade is employed during the digestion, similarly to other blood-feeding helminths (e.g., schistosomes). In this context, the thesis also deals with characterisation of the paralogs of the digestive peptidase cathepsin B1 from the bird schistosome *Trichobilharzia regenti*.

Particular aims:

- Biochemical characterisation of peptidolytic properties in soluble protein extracts and excretory/secretory products of *E. nipponicum*.
- Preparation of recombinant forms of selected peptidases (cathepsins B and D) from *E. nipponicum*.
- Biochemical and functional characterisation of recombinant cathepsins B and D and their immunolocalisation in adult *E. nipponicum.*
- Detailed characterisation of the isoforms of cathepsin B1 of *T. regenti*. In particular, comparison of substrate specificities of active forms (TrCB1.1 and TrCB1.4), conditions of their activation, and identification of the possible biological roles of inactive paralogs employing heterologously expressed wild type TrCB1.6 and its site-direct mutant (retrograde Gly^{29}-to-Cys^{29} mutation).

Results

The results of this thesis are summarized in three published papers and one manuscript in preparation. These publications are presented bellow; each is preceded by a short summary that recapitulates the most important findings of the experimental work.

Publication No. 1

Major acid endopeptidases of the blood-feeding monogenean Eudiplozoon nipponicum (Heteronchoinea:Diplozoidae).
Jedličková, L., Dvořáková, H., Kašný, M., Ilgová, J., Potěšil, D., Zdráhal, Z., Mikeš, L., 2016. Parasitology143,494–506. https://doi.org/10.1017/S0031182015001808.

In blood-feeding helminths, acid endopeptidases are key molecules in crucial processes such as digestion, skin and tissue invasion, interactions with host immune system, excystment, moulting, reproduction and development. However, current knowledge of monogenean peptidases involved in digestion and other processes has been poor. This publication represents the first direct evidence of the presence and activities of acid endopeptidases in *Eudiplozoon nipponicum* (Monogenea, Diplozoidae), an invasive haematophagous parasite of common carp. We found that cysteine peptidase activities prevailed in soluble protein extracts and excretory/secretory products of adult *E. nipponicum*, with cathepsin L-like activity predominating over cathepsin B-like activity. In vitro degradation of haemoglobin in the presence of worm protein extracts and excretory/secretory products was completely inhibited by a combination of cysteine and aspartic peptidase inhibitors, and only partially by particular cathepsin L, B and D inhibitors. Using PCR with degenerate primers and RACE-PCR, the complete sequences of two cathepsins L were amplified from cDNA of *E. nipponicum* and bioinformatically annotated. Mass spectrometry revealed several tryptic peptides in excretory/secretory products matching to the two translated sequences of cathepsin L genes. Based on our results, it seems that the processing of blood in *E. nipponicum* probably relies on an evolutionary network of cysteine and aspartic peptidases, similar to the proteolytic cascades in other blood-feeding helminths.

My contribution to the work was the detection of endopeptidase activity in samples of soluble protein extracts and excretory/secretory products of *E. nipponicum***,** in vitro cleavage/inhibition tests with haemoglobin, gaining cathepsin L sequences (design of degenerate primers, PCR) and preparation of the manuscript**.**

Publication No. 1

Major acid endopeptidases of the blood-feeding monogenean Eudiplozoon nipponicum (Heteronchoinea:Diplozoidae).

Jedličková, L., Dvořáková, H., Kašný, M., Ilgová, J., Potěšil, D., Zdráhal, Z., Mikeš, L.

(2016)

Parasitology143, 494–506. https://doi.org/10.1017/S0031182015001808

Major acid endopeptidases of the blood-feeding monogenean *Eudiplozoon nipponicum* (Heteronchoinea: Diplozoidae)

LUCIE JEDLIČKOVÁ[1]†, HANA DVOŘÁKOVÁ[1]†, MARTIN KAŠNÝ[1,2], JANA ILGOVÁ[2], DAVID POTĚŠIL[3], ZBYNĚK ZDRÁHAL[3] *and* LIBOR MIKEŠ[1]*

[1] *Department of Parasitology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague 2, Czech Republic*
[2] *Department of Botany and Zoology, Faculty of Science, Masaryk University, Kamenice 753/5 (A31), 625 00 Brno, Czech Republic*
[3] *RG Proteomics, Central European Institute of Technology, Masaryk University, Brno, Kamenice 753/5 (A26), 625 00 Brno Bohunice, Czech Republic*

(*Received 22 September 2015; revised 30 November 2015; accepted 30 November 2015*)

SUMMARY

In parasitic flatworms, acid endopeptidases are involved in crucial processes, including digestion, invasion, interactions with the host immune system, etc. In haematophagous monogeneans, however, no solid information has been available about the occurrence of these enzymes. Here we aimed to identify major cysteine and aspartic endopeptidase activities in *Eudiplozoon nipponicum*, an invasive haematophagous parasite of common carp. Employing biochemical, proteomic and molecular tools, we found that cysteine peptidase activities prevailed in soluble protein extracts and excretory/secretory products (ESP) of *E. nipponicum*; the major part was cathepsin L-like in nature supplemented with cathepsin B-like activity. Significant activity of the aspartic cathepsin D also occurred in soluble protein extracts. The degradation of haemoglobin in the presence of ESP and worm protein extracts was completely inhibited by a combination of cysteine and aspartic peptidase inhibitors, and diminished by particular cathepsin L, B and D inhibitors. Mass spectrometry revealed several tryptic peptides in ESP matching to two translated sequences of cathepsin L genes, which were amplified from cDNA of *E. nipponicum* and bioinformatically annotated. The dominance of cysteine peptidases of cathepsin L type in *E. nipponicum* resembles the situation in, e.g. fasciolid trematodes.

Key words: cysteine peptidase, aspartic peptidase, protease, haematophagous monogenea, cathepsin L, cathepsin B, cathepsin D, fish parasite, common carp.

INTRODUCTION

Monogeneans of the family Diplozoidae are blood-feeding freshwater ectoparasites inhabiting the gills of cyprinid fishes. They can be significantly virulent to their hosts, causing mechanical damage to gill filaments accompanied by a risk of secondary bacterial and mycotic infections or hypochromic microcytic anaemia (Kawatsu, 1978; Buchmann and Brescini, 2006). The alimentary tract of these worms is morphologically adapted to blood uptake, although the mechanisms of blood processing in monogeneans are largely unknown. The foregut is composed of an oral opening with prominent buccal suckers, with an eversible pharynx that leads via an oesophagus into several dead-end side caeca (Smyth and Halton, 1983; Valigurová *et al*. 2011; Konstanzová *et al*. 2015). Both sections of the gut are morphologically and functionally well differentiated. Numerous gland cells with proposed extracellular digestive function open into the foregut lumen. The chemical nature of their secretion is unknown, but it has been expected that they may produce haemolysins or peptidases, thus could be involved in preliminary degradation of complex foodstuff (e.g., blood cells) to smaller components suitable for endocytosis (Smyth and Halton, 1983; Hodová *et al*. 2010; Valigurová *et al*. 2011). The basic principles of the digestive process have been described for a few monogenean species and they were based mainly on ultrastructural and histochemical studies (Jennings, 1959; Tinsley, 1973; Halton and Stranock, 1976; Hodová *et al*. 2010; Valigurová *et al*. 2011; Konstanzová *et al*. 2015). According to these records, the terminal phase of digestion occurs inside specialized types of gut cells within the lysosomal cycle (Smyth and Halton, 1983). From this view, the digestion in diplozoids is more similar to haematophagous mites such as ticks (Sonenshine, 1991), than to other blood-feeding platyhelmints with their extracellular digestion (Dalton *et al*. 2004). The protein part of haemoglobin is broken down into peptides and amino acids (aa); the toxic haem is oxidized to haematin, exocytosed into the gut lumen and regurgitated by the worm through its oral opening (Llewellyn, 1954; Jennings, 1959; Smyth and Halton, 1983; Konstanzová *et al*. 2015).

* Corresponding author. Department of Parasitology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague 2, Czech Republic. E-mail: mikes@natur.cuni.cz
† These authors contributed equally to this work.

Parasitology, Page 1 of 13. © Cambridge University Press 2016
doi:10.1017/S0031182015001808

Current knowledge of monogenean peptidases involved in digestion and other processes is poor. So far, only two reports have been published concerning peptidases of marine monogeneans *Neobenedenia* spp. feeding on epidermal tissues and mucus of their hosts. A gene encoding cysteine peptidase cathepsin L in *Neobenedenia melleni* was characterized and cloned (Rao and Yang, 2007) and some activity of serine peptidases was found in homogenates of adult *Neobenedenia girellae* by zymography (Hirazawa *et al.* 2006). Their functions are unknown.

In general, processing of blood in sanguinivorous helminths and mites relies on an evolutionary conserved network of cysteine and aspartic peptidases (e.g., cathepsins B, L, F, C, legumain and cathepsin D) identified in trematodes – *Schistosoma mansoni, Schistosoma japonicum, Fasciola hepatica* (Caffrey and Ruppel, 1997; Dalton *et al.* 2003; Sajid *et al.* 2003; Caffrey *et al.* 2004; Delcroix *et al.* 2006; Robinson *et al.* 2008), nematodes – *Haemonchus contortus, Ancylostoma caninum, Necator americanus* (Williamson *et al.* 2003*a*, 2004), and in ticks – *Ixodes ricinus, Boophilus microplus* (Renard *et al.* 2000; Sojka *et al.* 2008). We suppose that similar molecular tools for protein hydrolysis are employed inside the digestive tract of monogeneans.

Our study represents the first direct evidence of the presence of acid endopeptidases in a haematophagous monogenean – *Eudiplozoon nipponicum*; these were identified and partially characterized by biochemical, proteomic and molecular methods.

METHODS

Parasites

Living adult worms of *E. nipponicum* were carefully removed from gills of common carps (*Cyprinus carpio*) immediately after slaughter in a commercial facility of Rybářství Třeboň, Plc. The fish originated from localities (ponds) in South Bohemia, Czech Republic.

Sample preparation

Freshly collected worms were repeatedly gently washed in sterilized tap water, placed in Eppendorf tubes, frozen immediately on dry ice and stored at −80 °C. Soluble protein extracts (*sol*PE) were prepared by homogenization of worms in 10 mM acetate buffer pH 5 using motor-driven teflon pellet pestle followed by 3 cycles of sonication on ice (10 W, 30 s) and centrifugation at 16 000 ***g*** for 20 min. Supernatants were collected. Excretory/secretory products (ESP) were obtained by incubation of washed adult live worms in 10 mM phosphate buffer pH 7·4 or 10 mM acetate buffer pH 6 for 3–4 h at room temperature (RT) in Eppendorf tubes. The ability of the ectoparasitic worms to survive for several hours in selected low osmolarity buffers was verified by previous overnight incubation. Full mobility and no obvious damage of the worms were observed. ESP were concentrated on Amicon Ultra filters (MWCO 10 kDa; Millipore). Protein concentration in all samples was measured by Quant-iT Protein Assay Kit (Invitrogen). The samples were stored at −80 °C.

Living adult worms were washed several times in sterile tap water and then placed in TRIzol® Reagent (Invitrogen) for stabilization of RNA and processed or stored at −80 °C. After homogenization, total RNA was extracted following the TRIzol protocol and concentration measured in NanoDrop 1000 (Thermo Scientific). First-strand cDNA synthesis was carried out with 0·3 μg of total RNA using oligo-dT18 primer from SuperScript™ III First-Strand Synthesis kit (Invitrogen).

Peptidolytic activities, pH optima and inhibition assays

Enzyme activities in *sol*PE and ESP were measured with a set of synthetic fluorogenic peptide substrates (Bachem): Z-Phe-Arg-AMC (FR) was used for detection of papain-like cysteine peptidases (acidic pH) and trypsin-like serine peptidases (neutral to basic pH), Z-Arg-Arg-AMC (RR) selectively for cathepsin B and Z-Ala-Ala-Asn-AMC (AAN) for legumain (asparaginyl endopeptidase). Abz–Lys–Pro–Ala–Glu–Phe–Nph–Arg–Leu (KPAEFnFRL), the specific substrate for cathepsin D, was kindly provided by Dr. Martin Horn, Institute of Organic Chemistry and Biochemistry, AS CR Prague (Abz = 2-aminobenzoyl; Nph and nF = nitrophenylalanine). All assays were performed in black 96-well flat bottom plates (Nunc, Thermo Scientific). Samples (1 and 5 μg of protein per well of *sol*PE and ESP, respectively) were pre-incubated (5 min, RT) in various buffers of pH in the range 2–10 for substrates FR, RR, AAN and pH 2–6 for KPAEFnFRL (0·1 M phosphate buffer pH 2; 0·1 M citrate pH 3; 0·1 M citrate/0·2 M phosphate pH 4–7; 0·2 M Tris–HCl pH 8; 0·2 M glycine/NaOH pH 9–10). Final volume was 100 μL. All buffers contained 2 mM dithiothreitol (DTT, Sigma-Aldrich), except for KPAEFnFRL substrate. The reactions were started by addition of particular peptidyl substrate (final concentration 50 μM) in 100 μL of the same buffer. Kinetics of the release of free fluorophors was detected by Infinite M200 flurometer (TECAN) at 28 °C for 60 min in 1 min intervals. Excitation/emission wavelengths for AMC substrates and nF substrate were 355/460 nm and 330/410 nm, respectively. Controls contained equal volume of appropriate buffer instead of the sample. Measurements were performed in triplicates. The quantification of cysteine and aspartic peptidase activities was performed in particular pH optima with FR and KPAEFnFRL substrates, respectively.

Table 1. Sequences of primers used for amplification of *E. nipponicum* cathepsin L genes.

Primer name	Primer sequence
cDNA fragment PCR with degenerate primers	
Cys forward primer[a]	5′-CAA/GGGNCARTGYGGITCNTGC/TTGG-3′
Asn reverse primer[b]	5′-CCANSA/TRTTYTTIACRATCCAA/GTA-3′
RACE–PCR with specific primers	
M-CL1-RACE5′	5′-CCG ACT TGG ATC ACG CCG TAT TGT TGG T-3′
M-CL1-RACE3′	5′-GAC CTT CCA ATG ACC CTG TCG TGG AGA A-3′
M-CL3-RACE5′	5′-GTG CTG CCC CTC TAG TGA TCC TGT TGT-3′
M-CL3-RACE3′	5′-CCA TGG TGT GCT GGT TGT AGG CTA TGG A-3′
GeneRacer™Forw	5′-CGA CTG GAG CAC GAG GAC ACT GA-3′
GeneRacer™Rev	5′-GCT GTC AAC GAT ACG CTA CGT AAC G-3′

[a] **TGC** nucleotide triplet encoding cysteine.
[b] **RTT** nucleotide triplet encoding asparagine.

Results were expressed as nM of substrate cleaved per minute in the presence of 1 μg of *sol*PE total protein.

Peptidase inhibitors (50 μM final concentration, Sigma-Aldrich) were used for inhibition of particular peptidase activities: E-64 (L-*trans*-epoxysuccinyl–leucylamido [4-guanidino] butane), an irreversible inhibitor of papain-like cysteine peptidases; CA-074 (N-(L-3-*trans*-propylcarbamoyloxirane-2-carbonyl)-L-isoleucyl-L-proline]), an irreversible inhibitor of cathepsin B; iCL (Arg–Lys–Leu–Leu–Trp–NH_2), a reversible inhibitor of cathepsin L; pepstatin A (isovaleryl–Val–Val–Sta–Ala–(3S,4S)-4-amino-3-hydroxy-6-methylheptanoic acid), inhibitor of aspartic peptidases. Inhibitors were mixed with samples prior to the addition of substrates and incubated for 15 min. Inhibition tests were performed in pH optimum of peptidase activity for each particular substrate. All measurements were performed in triplicates, repeated at least three times. Values in graphs are expressed as means with standard deviations.

Degradation of haemoglobin

Bovine haemoglobin (10 μg, Sigma Aldrich) diluted in 20 μL of particular buffers in the range of pH 3-6 (see above) containing 2 mM DTT was incubated with 3 μg of ESP or 1·5 μg of *sol*PE proteins for 16 h at 30 °C. Control reactions contained only haemoglobin. The contribution of individual aspartic or cysteine peptidase activities to degradation of haemoglobin was evaluated by inhibition assays at pH 3 and pH 5. Pepstatin, E-64, CA-074 and iCL were used at concentrations of 10 μM. Resulting hydrolysates were mixed with reducing electrophoretic sample buffer and separated by SDS–PAGE in 4–15% gradient precast gels (Bio-Rad) which were stained with Coomassie Brilliant Blue R-250 (CBB) and scanned on GS-800 Calibrated Densitometer (Bio-Rad).

Active site-labelling of cysteine peptidases

Fluorescent Green BODIPY-DCG-04 (Greenbaum *et al*. 2002), an analogue of E-64 inhibitor which binds irreversibly to the active site of papain-like peptidases, was incubated with ESP (15 μg of total protein) for 1 h (20 μM DCG-04, 5 mM DTT, 5 mM $MgCl_2$) at RT in the dark. Controls were incubated for 30 min with cysteine peptidase inhibitors (10 μM) E-64, iCL or CA-074 prior to addition of the probe. SDS–PAGE of labelled samples was performed as described above. Fluorescent signal was recorded on a fluorescence scanner (Molecular Imager FX, Bio-Rad) using excitation/emission wavelengths 488/530 nm. Finally, the gels were stained by Silver Stain Plus™ Kit (Bio-Rad).

Amplification of partial DNA sequences of cysteine peptidases

cDNA obtained by reverse transcription from adult *E. nipponicum* was amplified by PCR using PPP Master Mix (Top-Bio) and degenerate primers (Table 1) designed according to the consensus sequences corresponding to the coding sequences of cysteine peptidases of several parasitic species (especially according to the active site with a cysteine residue – Cys forward primer and with an asparagine residue – Asn reverse primer) (Eakin *et al*. 1990; Heussler and Dobbelaere, 1996; Li *et al*. 2006). The PCR protocol was as follows: one cycle 5 min initial denaturation at 94 °C, then 35 cycles of denaturation at 94 °C for 1 min, primer annealing 45 °C for 1 min, extension at 72 °C for 1 min and finally one cycle 10 min final extension at 72 °C. Control reactions were performed using the same primers and cDNA obtained from gills of a non-infected carp. The amplified gene fragments of the expected length (ca. 500 bp) were electrophoresed and isolated from agarose gel (1·5%) with Gel Extraction Kit (Qiagen), sub-cloned into pCR2·1-TOPO cloning vector (Invitrogen) and transformed into TOP10 *E. coli* (Invitrogen). pCR2·1-TOPO constructs were isolated using Qiaprep Purification Kit (Qiagen) and sequenced with M13 forward and reverse primers (DNA Sequencing Laboratory, Faculty of Science, Charles University in Prague).

Obtained partial DNA sequences were compared with NCBI database using BLASTX (http://blast.ncbi.nlm.nih.gov/).

Rapid amplification of cDNA ends (RACE–PCR)

In order to amplify the 5′/3′ ends and to obtain the entire gene sequences, two pairs of gene-specific primers (M-CL1-RACE5′ + M-CL1-RACE3′, M-CL3-RACE5′ + M-CL3-RACE3′; see Table 1) were designed according to the obtained ca 500 bp sequence fragments (see above). First-strand cDNA for RACE prepared according to manufacturer's instructions of GeneRacer™Kit (Invitrogen) was used as a template. The PCR reaction (25 μL) contained 12·5 μL of the EmeraldAmp PCR master mix (Clontech), 1 μL of the RACE cDNA (100 ng μL^{-1}), 1 μL of the M-CL1/3-RACE5′/3′ (10 μM), 3 μL of the GeneRacer primers Fwd/Rev (10 μM), 7·5 μL of ddH_20. The profile of amplification of 5′/3′ ends was: 1× 94 °C = 120 s, 5 × (94 °C = 30 s + 72 °C = 60 s), 5 × (94 °C = 30 s + 70 °C = 60 s), 35 × (94 °C = 30 s + 60 °C = 90 s + 70 °C = 90 s) and finally 1 × 72 °C = 10 min. PCR products from 5′ and 3′ RACE were cloned into pCR2·1-TOPO cloning vector as described above and submitted for sequencing with M13 forward and reverse primers (DNA Sequencing Laboratory, Faculty of Science, Charles University in Prague).

Sequence analysis

The obtained full-length cDNA sequences were blasted against sequences available in the GenBank database (NCBI) by BLASTX (http://blast.ncbi.nlm.nih.gov/). The presence of a signal sequence was predicted by the SignalP 4·1 Server (http://www.cbs.dtu.dk/services/SignalP/). The theoretical position of a pro-region was determined by multiple sequence alignment of cathepsins L from other organisms: human cathepsin L [GenBank: NP_001903·1], cathepsin L of *Clonorchis sinensis* [GenBank: ABK91809·1], cathepsin L of *C. carpio* [GenBank: BAD08618·1] and cathepsin L of *Neobenedenia melleni* [GenBank: ABK62794·1]. Multiple sequence alignments of aa sequences were carried out using the program Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Molecular mass and theoretical pI of the deduced proteins were determined by the Compute pI/Mw software available at the Expasy site (http://www.expasy.ch/tools/pitool.html). Detection of potential N-glycosylation sites was performed using an online tool at NetNGlyc 1·0 Server (http://www.cbs.dtu.dk/services/NetNGlyc).

Protein identification from 1D gel bands by LC–MS/MS

ESP were separated by SDS–PAGE as described above and stained by CBB. Bands (range 25–40 kDa) were manually excised, destained and incubated each with trypsin (Promega). LC–MS/MS analyses were done using RSLCnano system (Thermo Fisher Scientific) on-line connected to Impact II Q-TOF mass spectrometer with CaptiveSpray nanoBooster ion source (Bruker). Prior to LC separation, tryptic digests were online concentrated and desalted using trapping column (100 μm × 30 mm) filled with 3·5 μm X-Bridge BEH 130 C18 sorbent (Waters). After washing of trapping column (0·1% FA), the peptides were separated (300 nl/min) using Acclaim Pepmap100 C18 column (2 μm particles, 75 μm × 500 mm; Thermo Fisher Scientific) by the 0·1% FA in water : 0·1% FA in 80% ACN gradient program.

MS data were acquired in a data-dependent strategy with 3 s long cycle time. Mass range was set to 150–2200 m/z and precursors were selected from 300 to 2000 m/z. Acquisition speed of MS and MS/MS scans was 2 and 4–16 Hz, respectively. Speed of MS/MS spectra acquisition was based on precursor intensity (low and high absolute thresholds were 10 000 and 100 000 cts, respectively). Mascot (version 2.4·1) MS/MS ion searches were done against in-house database with sequences of expected recombinant proteins. Search results obtained against in-house database were checked for false positive identifications using database search against the whole SwissProt database (version 2014_07). Mass tolerances for peptides and MS/MS fragments were 15 ppm and 0·05 Da, respectively. Oxidation of methionine, deamidation (N, Q) and propionamide (C) as variable modifications and three enzyme miss cleavages were set for all searches.

RESULTS

Activity profiling

The results of fluorometric analyses of peptidolytic activities in *sol*PE and in ESP, pH optima and inhibition tests are shown in Fig. 1. The highest activity in both sample types was attributed to papain-like cysteine peptidases in acid buffers of pH 3–5 (optimum pH 5) with FR substrate (Fig. 1A and B). It dropped down to only ca. 20% in pH 6 and vanished above pH 7. Inhibitors E-64 and iCL suppressed the activity to zero and CA-074 by ca. 25 and 80% in the case of *sol*PE and ESP, respectively (Fig. 1a and b).

With cathepsin B-selective substrate RR, the activity in *sol*PE peaked in pH 5 and it was negligible in pH 6 and above (Fig. 1C). However, in ESP the highest activity appeared in pH 6, dropped to ca. 50% in pH 7 (Fig. 1D). In the case of both sample types, E-64 and iCL inhibited the peptidolytic activities nearly to zero. CA-074 reduced the activities in ESP by ca. 75% and by ca. 95% in *sol*PE (Fig. 1c and d).

The activity of aspartic cathepsin D with was found in the range of pH 2–4 (optimum in pH 3) in *sol*PE of worms in the presence of KPAEFnFRL

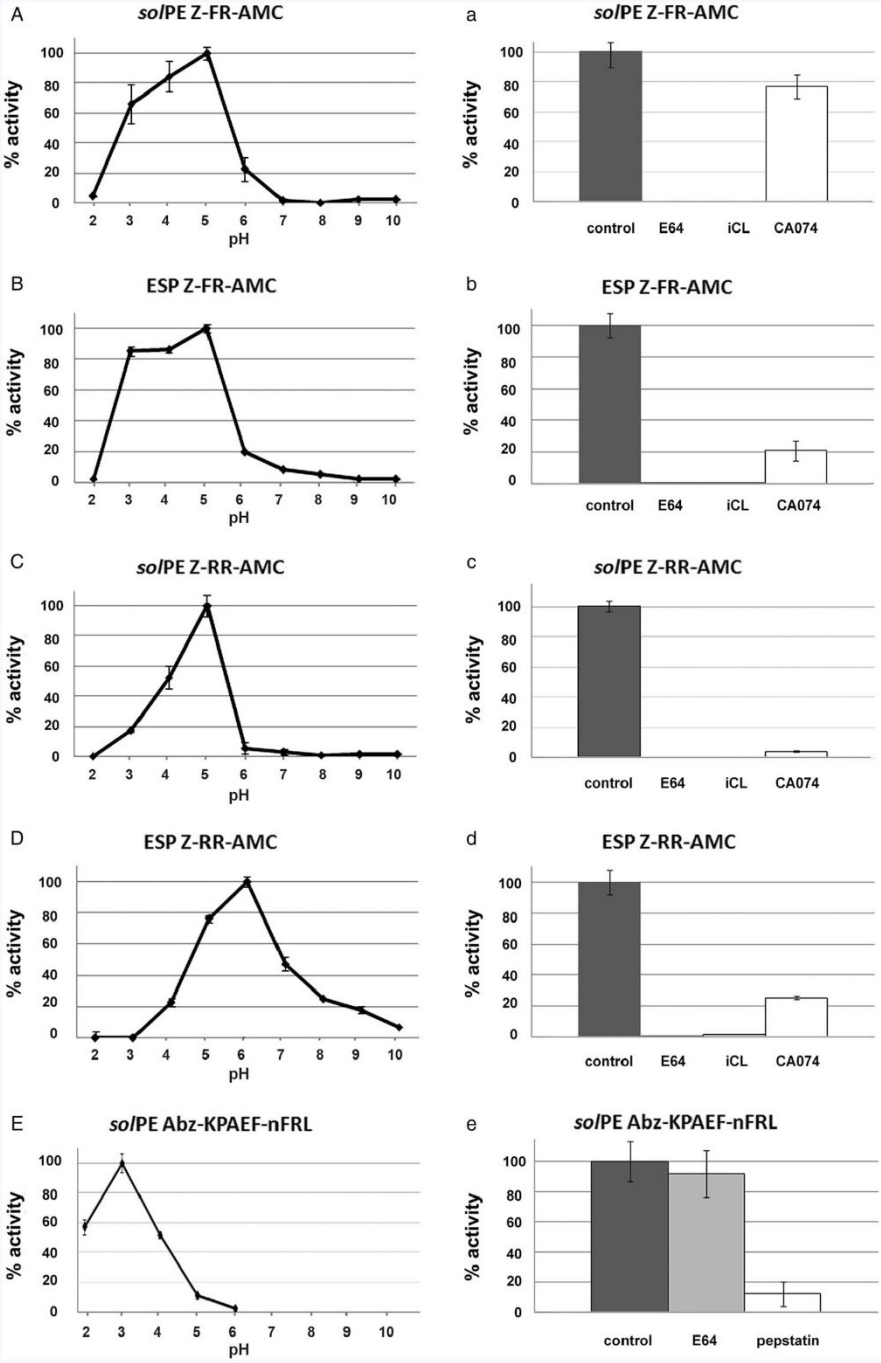


Fig. 1. The pH profiles and inhibition assays of acid endopeptidase activities. The pH optima of peptidolytic activities in soluble protein extracts (*sol*PE) and E/S products (ESP) of *E. nipponicum* and their inhibition at corresponding pH optimum were measured with a spectrum of fluorogenic substrates: (A, a, B, b) Z-FR-AMC; (C, c, D, d) Z-RR-AMC; (E, e) KPAEF-nFRL and with selective peptidase inhibitors (E-64, CA-074, iCL, pepstatin A). The values are expressed as percentage of maximum activity in the sample (100% at pH optimum).

(Fig. 1E); this was inhibited by ca. 90% by pepstatin A (Fig. 1e). In ESP, the activity of cathepsin D was recorded occasionally in some samples (not shown). The ratio of cathepsin L + B to cathepsin D activity in *sol*PE was approx. 4:3 (0·45 *vs* 0·33 nM min^{-1} μg^{-1} of protein). No activity of cysteine asparaginyl endopeptidase (legumain) was detected with AAN substrate in any sample (not shown). There was no difference in activity profiles between the samples of ESP collected at pH 7·4 and pH 6 (not shown). Therefore, a pooled mixture of samples was used for the experiments.

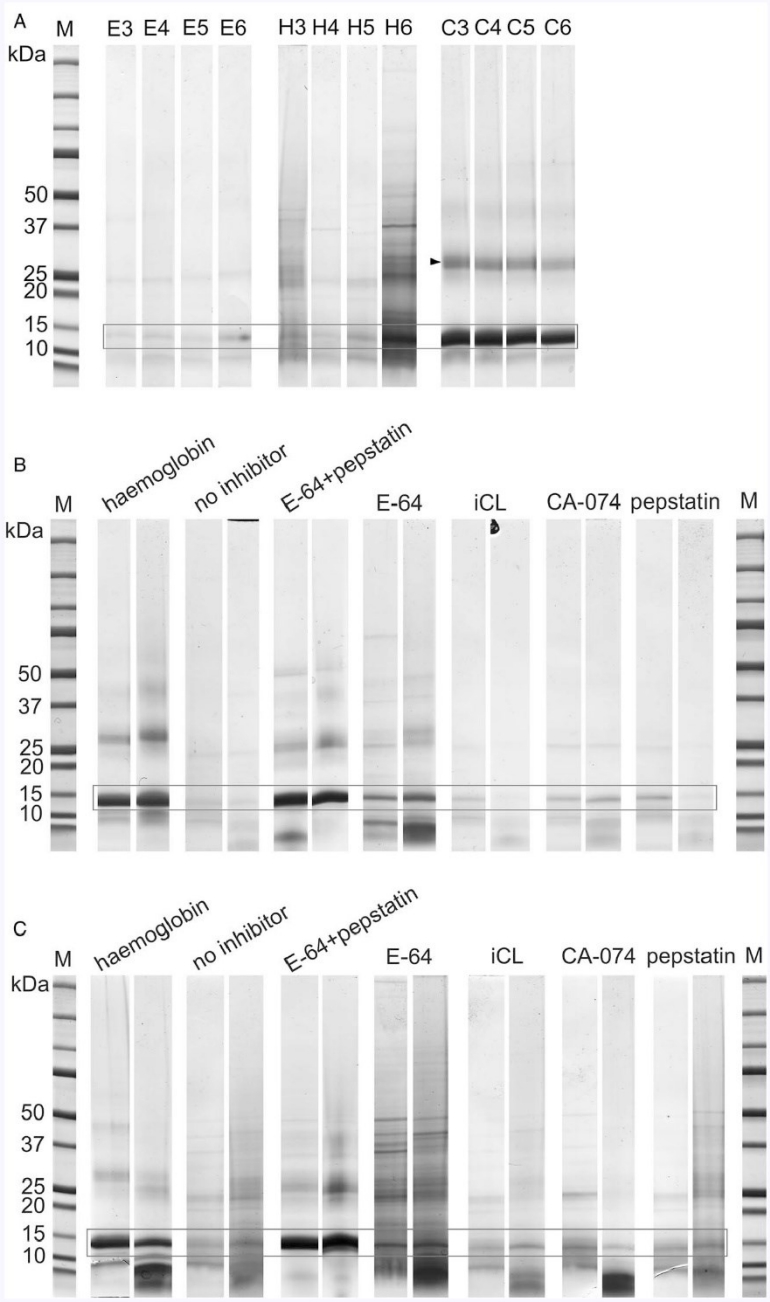


Fig. 2. The pH profile and inhibition of haemoglobin degradation by ESP and *sol*PE. Electrophoretograms of bovine haemoglobin (Hb) after overnight incubation with E/S products and soluble protein extracts of *E. nipponicum*. (A) Degradation in the presence of ESP (lanes E3-E6) and *sol*PE (H3-H6) at pH values 3-6. Controls of Hb in particular pH (C3-C6) contained neither ESP nor *sol*PE. (B) Inhibition of haemoglobinolysis in the presence of ESP in pH 5 and pH 3. (C) Inhibition of heamoglobinolysis in the presence of *sol*PE in pH 5 and pH 3. Headings above each pair of lanes (left lane = pH 5, right lane = pH 3) indicate the type of sample: haemoglobin = controls of Hb without ESP or solPE; no inhibitor = controls of Hb with ESP or *sol*PE; other samples contained inhibitors as labelled. M = markers of molecular size. Bands of Hb monomers are boxed in grey. Arrowhead points to alpha/beta heterodimers of Hb.

Degradation of haemoglobin

ESP as well as *sol*PE efficiently degraded bovine haemoglobin during overnight (16 h) incubation within the pH range of 3–5 and less at pH 6 (Fig. 2A). No substantial degradation occurred above pH 7 (not shown). At pH 5 and 3, haemoglobinolytic activity was markedly (but not completely) inhibited by the general inhibitor of papain-like cysteine peptidases E-64. Lower level of inhibition

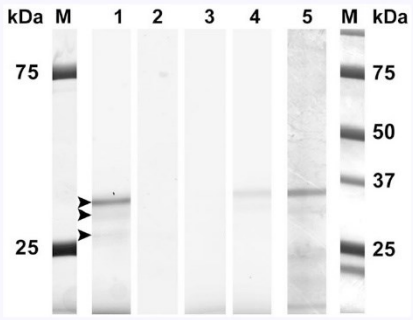


Fig. 3. Active site-labelling of papain-like peptidases in ESP by DCG-04 affinity probe. Electrophoretogram of ESP incubated with fluorescent DCG-04 and inhibitors of particular cysteine peptidases; signal (lanes 1–4) was recorded on a fluorescence scanner, excitation/emission wavelengths 488/530 nm. Lane 1 - control without any inhibitor; lane 2 - E-64; lane 3 - iCL; lane 4 - CA-074; lane 5 - silver staining of the gel. Arrowheads point to the three protein bands labelled by the probe. M = markers of molecular size.

was observed with CA-074, iCL and pepstatin alone; the latter two were less effective at pH 3 in the case of ESP. Complete inhibition at both pH values was reached only when using a mixture of E-64 and pepstatin (Fig. 2B and C). No marked differences in inhibition profiles were observed between the samples containing ESP and *sol*PE.

Active site-labelling of cysteine peptidases

Using the irreversible affinity probe DCG-04, the presence of papain-like cysteine peptidases was demonstrated in the samples of ESP. After SDS–PAGE, three protein bands of molecular sizes between 25 and 35 kDa were demonstrated in the gels. The interaction of DCG-04 with the active sites of cysteine peptidases was efficiently inhibited by the general inhibitor of cysteine peptidases (E-64) and by the inhibitor of cathepsin L (iCL), and only partially by the inhibitor of cathepsin B (CA-074) (Fig. 3).

Amplification of E. nipponicum *cathepsin* L genes

Two amplified DNA fragments (ca. 500 bp) obtained by PCR using *E. nipponicum* cDNA as a template and primers shown in Table 1 were identified as partial sequences of cathepsin L-like cysteine peptidases. In control PCR reactions with DNA template from a non-infected carp and with the same set of primers, a sequence encoding cathepsin L was also amplified. This showed 94·8% sequence identity with cathepsin L gene from *C. carpio* [GenBank: BAD08618·1] and only 55·19 and 49·35% with the fragments amplified from cDNA of *E. nipponicum*.

Applying RACE–PCR on the basis of fragments mentioned above, whole sequences of two *E. nipponicum* cathepsin L genes were obtained, termed EnCL1 and EnCL3. Their nucleotide sequences have been deposited in the NCBI GenBank database under accession numbers [GenBank: KP793605] and [KP793606]. The EnCL1 gene has an open reading frame (ORF) of 954 bp encoding a proenzyme consisting of 317 aa (Fig. 4). Deduced propeptide region contains 93 aa and the catalytic domain (mature enzyme) 224 aa residues. No signal leader sequence has been identified. The calculated theoretical molecular weights of the zymogen/mature protein are 35/24·4 kDa and pI 5·86/6·08. EnCL1 exhibits the highest similarity (53%) to cathepsin L from the Chinese liver fluke *Clonorchis sinensis* [GenBank: ABK91809·1]. The second obtained cathepsin L sequence from *E. nipponicum* was termed EnCL3, as it has the greatest similarity to cathepsin L3 precursor of *Schistosoma mansoni* [GenBank: ABV71063·1]. The ORF of EnCL3 is made up of 1107 bp and encodes a pre-proenzyme of 368 aa (Fig. 4). The sequence is composed of a signal leader sequence (24 aa), an unusually long propeptide (120 aa) and a mature (catalytic) domain (224 aa). Expected Mr/pI values of the zymogen and mature enzyme are 38·0/4·8 and 24·1/4·15, respectively.

The conserved 'ERFNIN' motif typically present in prosequences of various papain-like cysteine peptidases, e.g. cathepsins L, K, S, but not in cathepsin B (Karrer *et al*. 1993), was found in EnCL3; this was slightly modified in the case of EnCL1 (ERFNVN). Another sequence motif, 'GNFD', that may be involved in intra-cellular trafficking and intramolecular processing of some papain-like cysteine peptidases (Vernet *et al*. 1995; Dvořák *et al*. 2009), is modified in the prosequences of EnCL1 and EnCL3 to 'ANLD' and 'TNFD', respectively. The catalytic domains of both enzymes include the typical catalytic triad (EnCL1/3 numbering: Cys_{118}, His_{264}, N_{284}/Cys_{169}, His_{314}, N_{335}) of papain-like cysteine peptidases and the relatively conserved flanking sequences around the catalytic residues (QGQCGS**C**WAFS, LD **H**A/GVL, YWIVK**N**S/TW) (Turk *et al*. 2000) (Fig. 4). No potential N-linked glycosylation sites have been found throughout the sequences of both proenzymes.

The sequence identity between EnCL1 and EnCL3 is 55·21% on the aa level. Identities among *E. nipponicum* cathepsins L and selected aa sequences of cathepsins L from other organisms are summarized in Table 2. EnCL1 shares the highest number of identical aa positions (52·58%) with cathepsin L of *C. sinensis* [GenBank: ABK91809·1] and EnCL3 with cathepsin L3 of *S. mansoni* [GenBank: ABV71063·1] (52·54%). Identities of EnCL1 and EnCL3 with the only published whole sequence of cathepsin L from a monogenean, *N. melleni*, [GenBank: ABK62794·1] are only 44·01 and 41·27%, respectively.

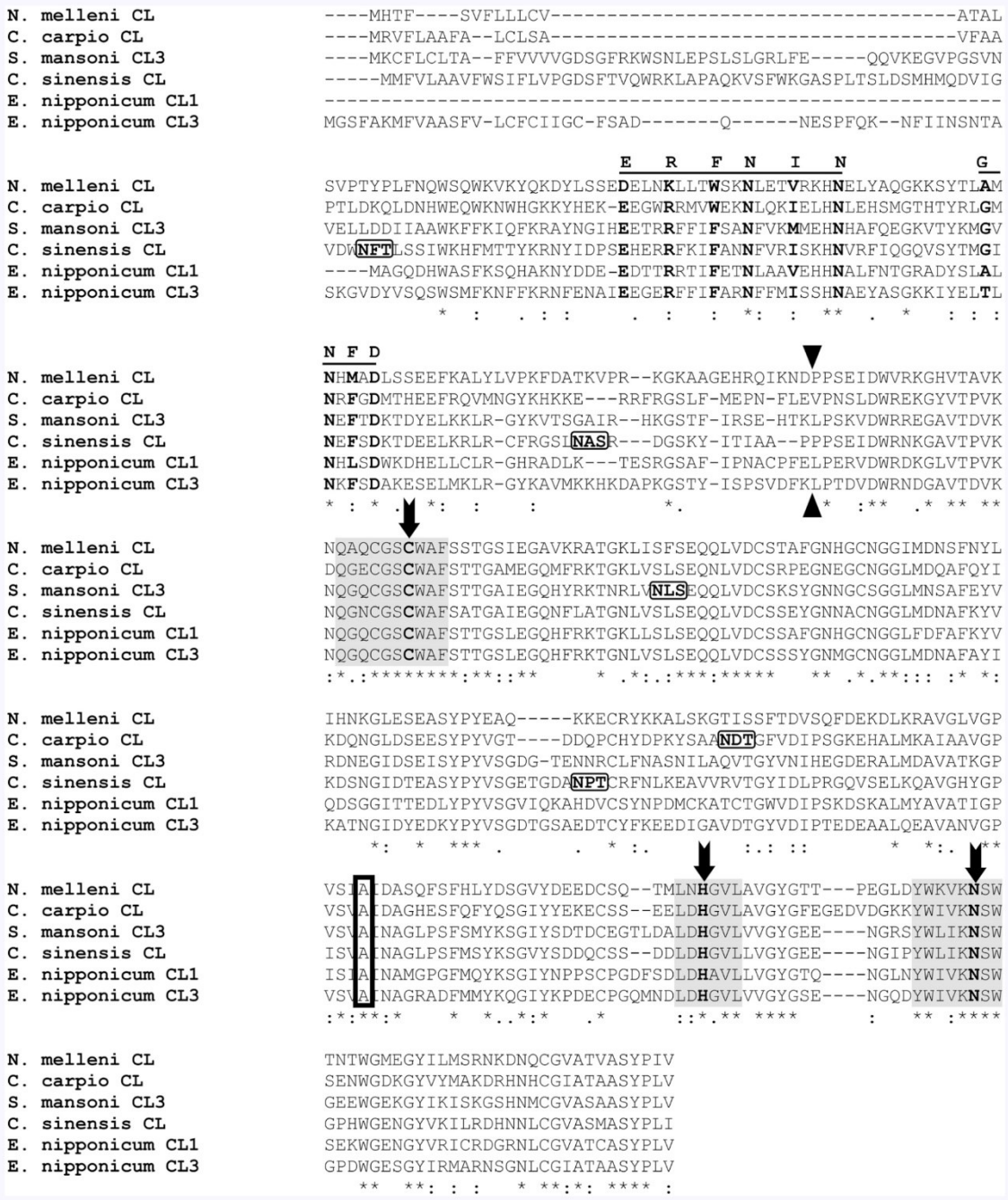


Fig. 4. Multiple sequence alignment of *Eudiplozoon nipponicum* cathepsins L with orthologs from other organisms. *Neobenedenia melleni* [GenBank:ABK62794·1]; *Cyprinus carpio* [GenBank:BAD08618·1]; *Schistosoma mansoni* CL3 [GenBank:ABV71063·1]; *Clonorchis sinensis* [GenBank:ABK91809·1]; *Eudiplozoon nipponicum* CL1 [GenBank: KP793605]; *E. nipponicum* CL3 [GenBank:KP793606]. The position of pro-region cleavage site is marked by arrowheads. ERFNIN- and GNFD-like motifs are in bold and indicated by underlined headings. Catalytic triad of the active site (C, H, N) is marked by arrows. Conserved motifs around active site residues are shaded in light grey. Tripeptides of potential N-glycosylation sites are boxed. Amino residue (A) at the base of the S2 pocket is boxed in black.

Mass spectrometry analysis of ESP of E. nipponicum

Both cathepsins L1 and L3, the genes of which have been amplified, were confirmed in ESP of adult *E. nipponicum* by mass spectrometry. Fifteen peptide sequences from trypsin-digested protein bands matched to the sequence of EnCL1 [GenBank: KP793605] and three peptides matched to EnCL3 [GenBank: KP793606] (Fig. 5). EnCL1

Table 2. Amino acid identities between cathepsins L of *E. nipponicum* and orthologues from other organisms.

	EnCL1	EnCL3
EnCL1	100	55·21
EnCL3	55·21	100
CsCL	52·58	51·12
SmCL3	50·80	52·54
CcCL	50·32	50·30
NmCL	44·01	41·27

Identities expressed in [%]. EnCL1, *Eudiplozoon nipponicum* [GenBank:KP793605]; EnCL3, *E. nipponicum* [GenBank:KP793606]; CsCL, *Clonorchis sinensis* [GenBank:ABK91809·1]; SmCL3, *Schistosoma mansoni* [GenBank:ABV71063·1]; CcCL, *Cyprinus carpio* [GenBank:BAD08618·1]; NmCL, *Neobenedenia melleni* [GenBank:ABK62794·1].

Fig. 5. Identification of cathepsins L1 and L3 in ESP of *E. nipponicum* by mass spectrometry. Each grey line corresponds to a single identified peptide; the sequences of some peptides partially overlap with others. (A) Fifteen tryptic peptides identified in ESP matched to the sequence of cathepsin L1 [GenBank: KP793605]. (B) Three peptides matched to the sequence of cathepsin L3 [GenBank: KP793606].

peptides were identified in nine of ten bands between ca. 24–35 kDa; EnCL3-derived peptides were recorded only above 35 kDa.

DISCUSSION

In contrast to the comprehensive knowledge concerning peptidases from various groups of blood-feeding parasites, only very limited information is available about proteolytic enzymes of haematophagous monogeneans. These neglected helminths in terms of biochemical studies have been thought to digest haemoglobin and other proteins intracellularly in the acidic lysosomes/phagolysosomes within the specialized types of gut cells, unlike other blood-feeding helmints such as schistosomes (Dalton *et al.* 2004), but similarly to ticks (Smyth and Halton, 1983; Sonenshine, 1991). However, it is still not clear whether the first phase of digestion in monogeneans may take place in the extracellular space of the lumen of the gut, with the participation of secreted proteases.

We aimed to identify and partially characterize dominant cysteine and aspartic endopeptidases which can be potentially involved in blood digestion or other processes in the adults of *E. nipponicum*. *sol*PE of worms and ESP were compared in terms of peptidolytic activities. We expected that ESP may be a potential source of peptidases released into the gut lumen (either from gut epithelium or accessory glands that are opened into the front part of the digestive system), and from there excreted outside the worm by continuous regurgitation of the rests of digestion (Smyth and Halton, 1983; Hodová *et al.* 2010; Valigurová *et al.* 2011). However, the presence of peptidases originated in other tissues of the worm cannot be excluded, especially in the case of *sol*PE from whole worms.

Activity profiling revealed that the major peptidase activity can be attributed to the cysteine class, clan CA, family C1 (MEROPS classification). This activity tested with Z-FR-AMC as a substrate at pH 5 was fully inhibited by E-64 (general inhibitor of papain-like cysteine peptidases) and the iCL, but only partially by an inhibitor specific for cathepsin B (CA-074). It is known, that in a reducing environment (presence of DTT), cathepsin L can also be inhibited by CA-074 (Steverding, 2011; Steverding *et al.* 2012). Besides, although iCL (the pentapeptide RKLLW) has been stated as a highly potent iCL (Brinker *et al.* 2000), it was able to diminish significantly the activity of recombinant cathepsin B2 of *Trichobilharzia regenti* in an unrelated previous experiment (Jedličková, unpublished). With the substrate Z-RR-AMC, which is often considered as specific for cathepsin B, the relatively high peptidase activity in ESP was also fully inhibited by E-64 and iCL, but only partially by CA-074 (at pH 6). This brought some confusion on the interpretation of the relative proportion of cathepsin L/B activities.

Cathepsin B- and cathepsin L-like peptidases have been usually distinguished by their ability to degrade the substrate with an arginine in P2 position. Human cathepsin B is able to hydrolyse the substrate, whereas cathepsin L is not – a situation which is not always applicable to orthologous enzymes in other organisms (Sajid and McKerrow, 2002; Choe *et al.* 2006). The substrate specificity of cysteine cathepsins is thought to be determined by interactions in the S2 pocket, particularly the

glutamic acid residue in cathepsin B (Glu_{205} in human CB) and the equivalent alanine residue in cathepsin L (Ala_{205} in human CL) localized at the bottom of the S_2 pocket. The Glu residue can accommodate and stabilize the polar guanidino group of Arg in the substrate, but Ala in this position cannot bind to Arg (Sajid and McKerrow, 2002). This was confirmed for mammalian cathepsins L/B and for peptidases of numerous organisms (including parasites). However, many other parasite peptidases do not possess 'mammalian' residues in this position. For example, the ability to cleave Z-RR-AMC was described for cathepsin L-like peptidase 'cruzain' from *Trypanosoma cruzi* (Gillmor *et al*. 1997) and cathepsin L of *Entamoeba histolytica* (Brinen *et al*. 2000). On the other hand, cathepsin B from *Leishmania major* (Chan *et al*. 1999) and cathepsin B1·4 from *T. regenti* (Dvořák *et al*. 2005) cannot hydrolyse this 'typical' cathepsin B substrate. According to the sequence data of EnCL1 and EnCL3, cathepsins L identified in the ESP of *E. nipponicum* have Ala (Ala_{234} in EnCL1, Ala_{285} in EnCL3) at the bottom of the S2 pocket, and therefore should likely not be able to cleave substrates with Arg in P2 position; however, in the light of the facts mentioned above, this must be verified experimentally with recombinant enzymes.

Cysteine peptidases in both *sol*PE and ESP of *E. nipponicum* exhibited the greatest activity at pH 5 with the substrate Z-FR-AMC; in the presence of Z-RR-AMC the optima were pH 5 and pH 6 for *sol*PE and ESP, respectively. The latter discrepancy could be caused by the presence of different other peptidases in *sol*PE which is a more complex sample in terms of protein composition in comparison to ESP. The results corresponded, e.g. to pH optima measured for cysteine peptidases in secretions of adult *S. mansoni* (Dalton *et al*. 1996); it has been suggested that the activity against Z-FR-AMC at lower pH is predominantly due to cathepsin L, while cathepsin B activity is less significant under these conditions. The presence of at least two different cathepsins L in our samples might explain the high activity in a broader range of acidic pH 3–5. Similarly to our results, the complex sample of protein extract from the gut of *I. ricinus* ticks showed the activity in the presence of the Z-FR-AMC at acidic pH 4-6 (Horn *et al*. 2009). The fact that this substrate was not hydrolysed at pH > 7 suggests that trypsin-like serine peptidases were not present in the samples.

Summarizing the results, we dare to assume that the majority of cysteine peptidase activity in the samples was of cathepsin L-like nature. In addition, a smaller proportion of the activity was cathepsin B-like and sensitive to inhibition by iCL.

To verify the presence of other peptidases in *sol*PE and ESP, substrates suitable for detection of cysteine asparaginyl endopeptidase (legumain) and aspartic cathepsin D were used. Surprisingly, no activity of legumain was detected. Legumain has been found in other blood-feeding helminths as well as ticks (Dalton *et al*. 1995; Caffrey *et al*. 2000; Oliver *et al*. 2006; Sojka *et al*. 2007; Abdul *et al*. 2007). Its function is connected not only with hydrolysis of ingested host proteins, but it also plays an important role in activation of other peptidases by cleaving their prosequences (Sajid *et al*. 2003; Caffrey *et al*. 2004; Sojka *et al*. 2008). There is a certain possibility that this enzyme occurred in our samples as an inactive zymogen or an inactive variant. Significant activity of cathepsin D (ca. ¾ of that of cathepsins L + B) was observed in *sol*PE, whereas only a minute activity was usually recorded in some samples of ESP (not shown). Also cathepsin D acts in the digestive process in other blood-feeding parasites (Brindley *et al*. 2001; Verity *et al*. 2001; Banerjee *et al*. 2002; Williamson *et al*. 2003*b*; Boldbaatar *et al*. 2006; Sojka *et al*. 2012) and we expect its involvement in blood digestion by monogenean parasites, too. The non-presence or low abundance of activities of these two crucial enzymes in some samples could be also possibly explained by alleged instability of their molecules and low production into the gut lumen.

In vitro degradation of haemoglobin by *E. nipponicum* peptidases is optimal at acidic pH, which corresponds with the pH optima of cysteine and aspartic peptidases. The reaction was completely blocked by a mixture of E-64 and pepstatin, less by E-64 and only partially by individual inhibitors. These results clearly confirmed that cysteine cathepsins L + B and aspartic cathepsin D are involved in haemoglobinolysis in *E. nipponicum*.

Affinity labelling with the active-site specific probe DCG-04 confirmed active cysteine peptidases in ESP, proved in three bands (between ca. 25–37 kDa), indicating the presence of various cysteine peptidases or intermediates of their activation by limited proteolysis (the sizes in gel approximately agree with theoretical MWs of proenzymes and mature parts of EnCL1 and EnCL3). The binding of the probe to all three bands was fully inhibited by both E-64 and iCL, whereas CA-074 inhibited only the band of lowest MW. This can be considered as another proof of the presence of both cathepsins L and B in ESP of *E. nipponicum*.

PCR with degenerate primers and RACE-PCR led to the amplification of *E. nipponicum* gene specific DNAs which were sequenced and identified as cathepsin L genes (named EnCL1 and EnCL3) [GenBank: KP793605, KP793606]. The pro-peptide of *En*CL1 and *En*CL3 contains some highly conserved regions typical for cathepsins L. One such a region is a variation of the $EX_3RX_2(V/I)FX_2NX_3IX_3N$ ('ERFNIN') motif (present as 'ERFNVN' in EnCL1) (Karrer *et al*. 1993). The residues in this motif are probably important in stabilizing the globular structure of the propeptide (Coulombe *et al*. 1996; Groves *et al*. 1998).

Another conserved motif is the GXNXFXD ('GNFD') present as 'ANLD' in EnCL1 and as 'TNFD' in EnCL3, with residues supposed to be involved in the control of intramolecular processing of cysteine peptidase precursors and/or participate in folding (Vernet *et al*. 1995; Dvořák *et al*. 2009). Processing is believed to be facilitated by a low pH. When pH is lowered, protonation of Asp_{36} (papain numbering) may cause a conformational change of the cysteine peptidase precursor in which the propeptide is bound less tightly into the active site, making it more susceptible for clip-off (Coulombe *et al*. 1996; Jerala *et al*. 1998). Amino acid differences in these conserved motifs in EnCL1 and EnCL3 should not affect the function.

An asparagine residue, found between the propeptide and mature domain of gut cysteine peptidases – cathepsins from other blood-feeding parasites, e.g. schistosomes and ticks, is responsible for recognition and cleavage by legumain for trans-activation (Sajid *et al*. 2003; Sojka *et al*. 2007). However, this Asn is absent in both EnCL1 and EnCL3, suggesting that pro-EnCL1 and pro-EnCL3 are rather activated by autocatalysis.

The analysis of EnCL1 sequence using SignalP 4·1 software did not indicate the presence of an N-terminal signal sequence necessary for vesicle or extracellular targeting. The absence of such signal sequence has been described for a few cathepsins L from other organisms, e.g. the assassin bug *Rhodnius prolixus* cathepsin L-like protein (Lopez-Ordoñez *et al*. 2001). Additionally, it has been reported that human cathepsin L can be directed to a secretion pathway due to an aa sequence (-SXPXV) located at the carboxy (C) terminus of the enzyme (Chauhan *et al*. 1998). Since the secretion of EnCL1 was supported by our result from mass spectrometry analysis, it seems that it may be realized in an alternative way; a non-signal peptide triggered protein secretion was also predicted by SecretomeP 2·0 software tool (http://www.cbs.dtu.dk/services/SecretomeP/), but just at threshold values of the analysis. So far, it is not clear whether some cryptic (unknown) intramolecular signals are involved in secretion of this enzyme and only speculations can be made on another possibility – extracellular transport in exosomes. On the other hand, EnCL3 possesses a typical N-terminal signal sequence (24 aa). EnCL3 zymogen has an atypical long pro-peptide (120 aa) similarly as the human blood fluke *S. mansoni* SmCL3 (130 aa) [GenBank: ABV71063·1]. It has been suggested that these extensions may have other functions to the proenzyme, for example in protein trafficking or as binding sites for other proteins (Dvořák *et al*. 2009).

Both EnCL1 and EnCL3 have no potential N-linked glycosylation sites in their sequences, similarly to cathepsin L of the monogenean parasite *N. melleni* [GenBank: ABK62794·1]. In general, glycosylation with mannose 6-phosphate has been shown to be an important sorting signal for routing mammalian proteins into lysosomes. However, there are several experimental evidences for mannose 6-phosphate independent trafficking of proteins into lysosomes (Ni *et al*. 2006; Braulke and Bonifacino, 2009). According to this, it would be possible that EnCL occurrence in ESP is not connected with the discharge of phagolysosome residual content into the gut lumen or that the enzyme may be targeted into lysosomes in an alternative manner.

Both cathepsins L were confirmed in *E. nipponicum* ESP in areas of various molecular sizes by means of mass spectrometry. Individual bands may represent partial processing or autolysis of the enzymes. According to the results, it seems that EnCL1 represents an abundant peptidase in ESP of *E. nipponicum*. In addition, the dominant protein spots found in *E. nipponicum sol*PE separated by 2D electrophoresis were also identified by mass spectrometry as EnCL1 (data not shown).

Digestion of proteins in blood-feeding parasites represents a major proteolytic process. We showed that the most abundant haemoglobinolytic endopeptidase activities in *E. nipponicum* belong to the cysteine class, with cathepsin L-like activity predominating over cathepsin B-like activity. Proteomic investigations confirmed this fact. Besides, significant involvement of aspartic cathepsin D was well documented. The possibility of aimed secretion of proteolytic enzymes to the external environment by this ectoparasite of carp gills appears to be purposeless and therefore improbable. From this point of view and based on our results, we believe that cathepsins L, B and D are involved in processing of ingested host's blood in *E. nipponicum*.

ACKNOWLEDGEMENTS

We are grateful to Dr C. Caffrey and Dr D. Greenbaum, Center for Discovery and Innovation in Parasitic Diseases, UCSF, USA, who kindly provided the probe Green BODIPY-DCG-04. The MS part of the work was carried out with the support of Proteomics Core Facility ('CEITEC – Central European Institute of Technology'). We also thank Mr Stanislav Sojka, Rybářství Třeboň, Plc., for arrangements concerning collection of parasites.

FINANCIAL SUPPORT

This research was enabled due to the grants of the Czech Science Foundation (L.M., Grant no. P506/12/1258), (Z.Z., Grant no. P206-12-G151); Grant Agency of the Charles University in Prague (L.J., Grant no. 502313); Masaryk University institutional support (M.K., J.I., Grant no. MUNI/A/1484/2014); Charles University institutional/departmental support (M.K., Grant no. UNCE 204017; L.J., H.D., M.K., L.M., Grant no. PRVOUK P41; L.J., H.D., Grant no. SVV 260202/2015); and the European Regional Development Fund (Z.Z., Grant no. CZ.1·05/1·1·00/02·0068).

REFERENCES

Abdul, A. M., Tsuji, N., Miyoshi, T., Khyrul Islam, M., Huang, X., Motobu, M. and Fujisaki, K. (2007). Characterization of asparaginyl endopeptidase, legumain induced by blood feeding in the ixodid tick *Haemaphysalis longicornis*. *Insect Biochemistry and Molecular Biology* **37**, 911–922.

Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M. and Goldberg, D. E. (2002). Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 990–995.

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 Biochemistry and Molecular Biology* **36**, 25–36.

Braulke, T. and Bonifacino, J. S. (2009). Sorting of lysosomal proteins. *Biochimica et Biophysica Acta* **1793**, 605–614.

Brindley, P. J., Kalinna, B. H., Wong, J. Y., Bogitsh, B. J., King, L. T., Smyth, D. J., Verity, C. K., Abbenante, G., Brinkworth, R. I., Fairlie, D. P., Smythe, M. L., Milburn, P. J., Bielefeldt-Ohmann, H., Zheng, Y. and McManus, D. P. (2001). Proteolysis of human hemoglobin by schistosome cathepsin D. *Molecular and Biochemical Parasitology* **112**, 103–112.

Brinen, L. S., Que, X., McKerrow, J. H. and Reed, S. L. (2000). Homology modeling of *Entamoeba histolytica* cysteine proteinases reveals the basis for cathepsin L-like structure with cathepsin B-like specificity. *Archives of Medical Research* **31**, S63–S64.

Brinker, A., Weber, E., Stoll, D., Voigt, J., Müller, A., Sewald, N., Jung, G., Wiesmüller, K. H. and Bohley, P. (2000). Highly potent inhibitors of human cathepsin L identified by screening combinatorial pentapeptide amide collections. *European Journal of Biochemistry/FEBS* **267**, 5085–5092.

Buchmann, K. and Brescini, J. (2006). *Monogenea*. *In Fish Diseases and Disorders*, Volume 1, 2nd Edn. pp. 297–344. CABI, Wallingford.

Caffrey, C. R. and Ruppel, A. (1997). Cathepsin B-like activity predominates over cathepsin L-like activity in adult *Schistosoma mansoni* and *S. japonicum*. *Parasitology Research* **83**, 632–635.

Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Salter, J. P., Sajid, M., Lucas, K. D., Franklin, C., Bogyo, M. and McKerrow, J. H. (2000). Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Letters* **466**, 244–248.

Caffrey, C. R., McKerrow, J. H., Salter, J. P. and Sajid, M. (2004). Blood "n" guts: an update on schistosome digestive peptidases. *Trends in Parasitology* **20**, 241–248.

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

Dalton, J. P., Hola-Jamriska, L. and Brindley, P. J. (1995). Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* **111**, 575–580.

Dalton, J. P., Clough, K. A., Jones, M. K. and Brindley, P. J. (1996). Characterization of the cathepsin-like cysteine proteinases of *Schistosoma mansoni*. *Infection and Immunity* **64**, 1328–1334.

Dalton, J. P., Neill, S. O., Stack, C., Collins, P., Walshe, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D., Khaznadji, E., Moire, N., Brennan, G., Mousley, A., Kreshchenko, N., Maule, A. G. and Donnelly, S. M. (2003). Fasciola hepatica cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *International Journal for Parasitology* **33**, 1173–1181.

Dalton, J. P., Skelly, P. and Halton, D. W. (2004). Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. *Canadian Journal of Zoology*. **82**, 211–232.

Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvořák, J., Hsieh, I., Bahgat, M., Dissous, C. and McKerrow, J. H. (2006). A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *Journal of Biological Chemistry* **281**, 39316–39329.

Dvořák, J., Delcroix, M., Rossi, A., Vopálenský, V., Pospíšek, M., Šedinová, M., Mikeš, L., Sajid, M., Sali, A., McKerrow, J. H., Horák, P. and Caffrey, C. R. (2005). Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. *International Journal for Parasitology* **35**, 895–910.

Dvořák, J., Mashiyama, S. T., Sajid, M., Braschi, S., Delcroix, M., Schneider, E. L., McKerrow, W. H., Bahgat, M., Hansell, E., Babbitt, P. C., Craik, C. S., McKerrow, J. H. and Caffrey, C. R. (2009). SmCL3, a gastrodermal cysteine protease of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* **3**, e449.

Eakin, A. E., Bouvier, J., Sakanari, J. A., Craik, C. S. and McKerrow, J. H. (1990). Amplification and sequencing of genomic DNA fragments encoding cysteine proteases from protozoan parasites. *Molecular and Biochemical Parasitology* **39**, 1–8.

Gillmor, S. A., Craik, C. S. and Fletterick, R. J. (1997). Structural determinants of specificity in the cysteine protease cruzain. *Protein Science: a Publication of the Protein Society* **6**, 1603–1611.

Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K. F. and Bogyo, M. (2002). Chemical approaches for functionally probing the proteome. *Molecular & Cellular Proteomics: MCP* **1**, 60–68.

Groves, M. R., Coulombe, R., Jenkins, J. and Cygler, M. (1998). Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. *Proteins* **32**, 504–514.

Halton, D. W. and Stranock, S. D. (1976). The fine structure and histochemistry of the caecal epithelium of *Calicotyle kröyeri* (Monogenea: Monopisthocotylea). *International Journal for Parasitology* **6**, 253–263.

Heussler, V. and Dobbelaere, D. (1996). Cloning of a protease gene family in Fasciola hepatica using the polymerase chain reaction (PCR). *Schweizer Archiv für Tierheilkunde* **138**, 125–132.

Hirazawa, N., Umeda, N., Hatanaka, A. and Kuroda, A. (2006). Characterization of serine proteases in the monogenean *Neobenedenia girellae*. *Aquaculture* **255**, 188–195.

Hodová, I., Matějusová, I. and Gelnar, M. (2010). The surface topography of *Eudiplozoon nipponicum* (Monogenea) developmental stages parasitizing carp (*Cyprinus carpio* L.). *Central European Journal of Biology* **5**, 702–709.

Horn, M., Nussbaumerová, M., Šanda, M., Kovářová, Z., Srba, J., Franta, Z., Sojka, D., Bogyo, M., Caffrey, C. R., Kopáček, P. and Mareš, M. (2009). Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. *Chemistry & Biology* **16**, 1053–1063.

Chan, V. J., Selzer, P. M., McKerrow, J. H. and Sakanari, J. A. (1999). Expression and alteration of the S2 subsite of the *Leishmania major* cathepsin B-like cysteine protease. *Biochemical Journal* **340**, 113–117.

Chauhan, S. S., Ray, D., Kane, S. E., Willingham, M. C. and Gottesman, M. M. (1998). Involvement of carboxy–terminal amino acids in secretion of human lysosomal protease cathepsin L. *Biochemistry* **37**, 8584–8594.

Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Brömme, D., Ellman, J. A. and Craik, C. S. (2006). Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *Journal of Biological Chemistry* **281**, 12824–12832.

Jennings, J. B. (1959). Studies on digestion in the monogenetic trematode *Polystoma integerrimum*. *Journal of Helminthology* **33**, 197–204.

Jerala, R., Zerovnik, E., Kidric, J. and Turk, V. (1998). pH-induced conformational transitions of the propeptide of human cathepsin L. A role for a molten globule state in zymogen activation. *Journal of Biological Chemistry* **273**, 11498–11504.

Karrer, K. M., Peiffer, S. L. and DiTomas, M. E. (1993). Two distinct gene subfamilies within the family of cysteine protease genes. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3063–3067.

Kawatsu, H. (1978). Studies on the anemia of fish - IX. Hypochromic microcytic anemia of crusian carp caused by infestation with a trematode, *Diplozoon nipponicum*. *Bulletin of the Japanese Society of Scientific Fisheries* **44**, 1315–1319.

Konstanzová, V., Koubková, B., Kašný, M., Ilgová, J., Dzika, E. and Gelnar, M. (2015). Ultrastructure of the digestive tract of *Paradiplozoon homoion* (Monogenea). *Parasitology Research* **114**, 1485–1494.

Li, A. H., Moon, S.-U., Park, Y.-K., Na, B.-K., Hwang, M.-G., Oh, C.-M., Cho, S.-H., Kong, Y., Kim, T.-S. and Chung, P.-R. (2006). Identification and characterization of a cathepsin L-like cysteine protease from *Taenia solium* metacestode. *Veterinary Parasitology* **141**, 251–259.

Llewellyn, J. (1954). Observations on the food and the gut pigment of the Polyopisthocotylea (Trematoda: Monogenea). *Parasitology* **44**, 428–37.

Lopez-Ordoñez, T., Rodriguez, M. H. and Hernández-Hernández, F. D. (2001). Characterization of a cDNA encoding a cathepsin L-like protein of *Rhodnius prolixus*. *Insect Molecular Biology* **10**, 505–511.

Ni, X., Canuel, M. and Morales, C. R. (2006). The sorting and trafficking of lysosomal proteins. *Histology and Histopathology* **21**, 899–913.

Oliver, E. M., Skuce, P. J., McNair, C. M. and Knox, D. P. (2006). Identification and characterization of an asparaginyl proteinase (legumain) from the parasitic nematode, *Haemonchus contortus*. *Parasitology* **133**, 237–244.

Rao, Y. Z. and Yang, T. B. (2007). cDNA cloning, mRNA expression and recombinant expression of a cathepsin L-like cysteine protease from *Neobenedenia melleni* (Monogenea: Capsalidae). *Aquaculture* **269**, 41–53.

Renard, G., Garcia, J. F., Cardoso, F. C., Richter, M. F., Sakanari, J. A., Ozaki, L. S., Termignoni, C. and Masuda, A. (2000). Cloning and functional expression of a *Boophilus microplus* cathepsin L-like enzyme. *Insect Biochemistry and Molecular Biology* **30**, 1017–1026.

Robinson, M. W., Dalton, J. P. and Donnelly, S. (2008). Helminth pathogen cathepsin proteases: it's a family affair. *Trends in Biochemical Sciences* **3**, 601–608.

Sajid, M. and McKerrow, J. H. (2002). Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology* **120**, 1–21.

Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I., Greenbaum, D., Bogyo, M., Salter, J. P., Lim, K. C., Franklin, C., Kim, J.-H. and Caffrey, C. R. (2003). Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Molecular and Biochemical Parasitology* **131**, 65–75.

Smyth, J. D. and Halton, D. W. (1983). *The Physiology of Trematodes*. Cambridge University Press, Cambridge, UK.

Sojka, D., Hajdušek, O., Dvořák, J., Sajid, M., Franta, Z., Schneider, E. L., Craik, C. S., Vancová, M., Burešová, V., Bogyo, M., Sexton, K. B., McKerrow, J. H., Caffrey, C. R. and Kopáček, P. (2007). IrAE: an asparaginyl endopeptidase (legumain) in the gut of the hard tick *Ixodes ricinus*. *International Journal for Parasitology* **37**, 713–724.

Sojka, D., Franta, Z., Horn, M., Hajdušek, O., Caffrey, C. R., Mareš, M. and Kopáček, P. (2008). Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases. *Parasites & Vectors* **1**, 7.

Sojka, D., Franta, Z., Frantová, H., Bartošová, P., Horn, M., Váchová, J., O'Donoghue, A. J., Eroy-Reveles, A. A., Craik, C. S., Knudsen, G. M., Caffrey, C. R., McKerrow, J. H., Mareš, M. and Kopáček, P. (2012). Characterization of gut-associated cathepsin D hemoglobinase from tick *Ixodes ricinus* (IrCD1). *Journal of Biological Chemistry* **287**, 21152–21163.

Sonenshine, D. E. (1991). *Biology of Ticks Volume 1*. Oxford University Press, New York.

Steverding, D. (2011). The cathepsin B-selective inhibitors CA-074 and CA-074Me inactivate cathepsin L under reducing conditions. *Open Enzyme Inhibition Journal* **4**, 11–16.

Steverding, D., Sexton, D. W., Wang, X., Gehrke, S. S., Wagner, G. K. and Caffrey, C. R. (2012). *Trypanosoma brucei*: chemical evidence that cathepsin L is essential for survival and a relevant drug target. *International Journal for Parasitology* **42**, 481–488.

Tinsley, R. C. (1973). Ultrastructural studies on the form and function of the gastrodermis of *Protopolystoma xenopi* (Monogenoidea: Polyopisthocotylea). *Biological Bulletin* **144**, 541–555.

Turk, B., Turk, D. and Turk, V. (2000). Lysosomal cysteine proteases: more than scavengers. *Biochimica et Biophysica Acta* **1477**, 98–111.

Valigurová, A., Hodová, I., Sonnek, R., Koubková, B. and Gelnar, M. (2011). *Eudiplozoon nipponicum* in focus: monogenean exhibiting a highly specialized adaptation for ectoparasitic lifestyle. *Parasitology Research* **108**, 383–394.

Verity, C. K., Loukas, A., McManus, D. P. and Brindley, P. J. (2001). *Schistosoma japonicum* cathepsin D aspartic protease cleaves human IgG and other serum components. *Parasitology* **122**, 415–421.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M. C., Storer, A. C. and Thomas, D. Y. (1995). Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *Journal of Biological Chemistry* **270**, 10838–10846.

Williamson, A. L., Brindley, P. J., Knox, D. P., Hotez, P. J. and Loukas, A. (2003*a*). Digestive proteases of blood-feeding nematodes. *Trends in Parasitology* **19**, 417–423.

Williamson, A. L., Brindley, P. J., Abbenante, G., Datu, B. J. D., Prociv, P., Berry, C., Girdwood, K., Pritchard, D. I., Fairlie, D. P., Hotez, P. J., Zhan, B. and Loukas, A. (2003*b*). Hookworm aspartic protease, Na-APR-2, cleaves human hemoglobin and serum proteins in a host-specific fashion. *Journal of Infectious Diseases* **187**, 484–494.

Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., Cantley, L. C., Sajid, M., Craik, C. S. and Loukas, A. (2004). A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *Journal of Biological Chemistry* **279**, 35950–35957.

Publication No. 2

Cysteine peptidases of Eudiplozoon nipponicum: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp

Jedličková, L., Dvořáková, H., Dvořák, J., Kašný, M., Ulrychová, L., Vorel, J., Žárský, V., Mikeš, L., 2018.**.** Parasit. Vectors 11, 1–17. https://doi.org/10.1186/s13071-018-2666-2

The first publication indicated that cysteine peptidases of clan CA, family C1 account for a major part of proteolytic activity in the haematophagous monogenean *Eudiplozoon nipponicum.* It was also proposed that cathepsin L-like peptidases, in cooperation with cathepsins D and B, play a crucial role in the digestion of haemoglobin in this sanguinivorous parasite. Therefore, in this study we focused on detailed characterisation of cathepsins L and B from adult *E. nipponicum.* Transcriptomic analysis revealed a set of ten distinct transcripts that encode cathepsin L. Bioinformatic analysis showedthat some of them display unusual structural features that resemble cathepsins B and S. Two heterologously expressed cathepsins L (EnCL1 and EnCL3) had different pH activity profiles with both synthetic and macromolecular substrates, and were able to hydrolyse blood proteinsin vitro (haemoglobin, albumin, fibrinogen, IgG) and type I collagen of the connnective tissue. Both enzymes were localised in the haematin cells of the worm's digestive tract and also in the lumen of the parasite´s gut. These results confirm the putative role of cathepsins L in the digestion of host blood proteins and reveal the potential extracellular phase of digestion.

In contrast to a broad repertoire of cathepsins L, a single cathepsin B is expressed by the adult *E. nipponicum* (EnCB). It is most similar to cathepsin B2 of *Schistosoma mansoni* (SmCB2), which is not involved in digestion of blood proteins. It displays molecular features typical of cathepsins B, including an occluding loop responsible for its exopeptidase activity. It also contains a 'haemoglobinase motif' assigned to helminth cathepsins B with substrate specificity for haemoglobin. In the case of EnCB, this motif is slightly modified. Although the sequence of EnCB contains this motif and the re combinant EnCB was able to hydrolyze haemoglobin in vitro, the enzyme was localised in the vitelline cells of the parasite and not in the digestive tract. According to this result, it seems that EnCB is probably not involved in digestion, but rather in vitellogenesis.

My contribution was the production of recombinant cathepsin B, its biochemical and functional characterisation and immunohistochemistry, participation in the bioinformatic annotation of enzyme structure, data interpretation, and manuscript writing.

Publication No. 2

Cysteine peptidases of Eudiplozoon nipponicum: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp

Jedličková, L., Dvořáková, H., Dvořák, J., Kašný, M., Ulrychová, L., Vorel, J., Žárský, V., Mikeš, L

(2018)

Parasit. Vectors 11, 1–17. https://doi.org/10.1186/s13071-018-2666-2

Jedličková et al. Parasites & Vectors (2018) 11:142
DOI 10.1186/s13071-018-2666-2

Parasites & Vectors

RESEARCH Open Access

Cysteine peptidases of *Eudiplozoon nipponicum*: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp

Lucie Jedličková[1*], Hana Dvořáková[1], Jan Dvořák[2,3], Martin Kašný[1,4], Lenka Ulrychová[1,5], Jiří Vorel[4], Vojtěch Žárský[6] and Libor Mikeš[1]

Abstract

Background: Cysteine peptidases of clan CA, family C1 account for a major part of proteolytic activity in the haematophagous monogenean *Eudiplozoon nipponicum*. The full spectrum of cysteine cathepsins is, however, unknown and their particular biochemical properties, tissue localisation, and involvement in parasite-host relationships are yet to be explored.

Methods: Sequences of cathepsins L and B (*EnCL* and *EnCB*) were mined from *E. nipponicum* transcriptome and analysed bioinformatically. Genes encoding two *EnCLs* and one *EnCB* were cloned and recombinant proteins produced in vitro. The enzymes were purified by chromatography and their activity towards selected substrates was characterised. Antibodies and specific RNA probes were employed for localisation of the enzymes/transcripts in tissues of *E. nipponicum* adults.

Results: Transcriptomic analysis revealed a set of ten distinct transcripts that encode *EnCLs*. The enzymes are significantly variable in their active sites, specifically the S2 subsites responsible for interaction with substrates. Some of them display unusual structural features that resemble cathepsins B and S. Two recombinant *EnCLs* had different pH activity profiles against both synthetic and macromolecular substrates, and were able to hydrolyse blood proteins and collagen I. They were localised in the haematin cells of the worm's digestive tract and in gut lumen. The *EnCB* showed similarity with cathepsin B2 of *Schistosoma mansoni*. It displays molecular features typical of cathepsins B, including an occluding loop responsible for its exopeptidase activity. Although the *EnCB* hydrolysed haemoglobin in vitro, it was localised in the vitelline cells of the parasite and not the digestive tract.

(Continued on next page)

* Correspondence: lucka.jedlickova@seznam.cz
[1]Department of Parasitology, Faculty of Science, Charles University, Viničná 7, 12844 Prague 2, Czech Republic
Full list of author information is available at the end of the article

BioMed Central

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

Conclusions: To our knowledge, this study represents the first complex bioinformatic and biochemical characterisation of cysteine peptidases in a monogenean. *Eudiplozoon nipponicum* adults express a variety of CLs, which are the most abundant peptidases in the worms. The properties and localisation of the two heterologously expressed *EnCLs* indicate a central role in the (partially extracellular?) digestion of host blood proteins. High variability of substrate-binding sites in the set of *EnCLs* suggests specific adaptation to a range of biological processes that require proteolysis. Surprisingly, a single cathepsin B is expressed by the parasite and it is not involved in digestion, but probably in vitellogenesis.

Keywords: Cysteine peptidase, Protease, Cathepsin, S2 subsite, Haematophagy, Blood digestion, Monogenea, Diplozoidae, *Eudiplozoon nipponicum*, Fish parasite

Background

Blood-feeding monogeneans of the family Diplozoidae (Heteronchoinea) are ectoparasites that live on the gills of cyprinid fishes. One member of the family, *Eudiplozoon nipponicum*, is an important invasive species, first recorded in Europe in 1983 on farmed carp in France [1]. It is currently found throughout Europe and is a widespread representative of the helminth fauna of the common carp (*Cyprinus carpio*) in the Czech Republic [2]. The common carp is a fish of high economic importance in many Asian and European countries, with global aquaculture production yielding over 4 million tons in 2014 [3]. Since the parasite affects the health of farmed fish, although precise estimates are unknown, it is assumed it causes economic losses. Pathogenic effects of *E. nipponicum* are associated mainly with inducing hypochromic microcytic anaemia in the fish by a continuous intake of blood by worms attached to the carp gills [4].

Monogenea are a rather neglected group of Neodermata and only a handful of papers on their biochemistry and bioactive molecules have ever been published. Based on previous ultrastructural studies, it has been assumed that in diplozoid monogeneans, the digestion of blood, gathered by combined action of their powerful buccal suckers and muscular pharynx, takes place within the lysosomal cycle in the specialised cells of the intestinal epidermis [5–9], similarly as in blood-sucking mites such as ticks [10, 11]. Our previous study [12] has shown that the processing of blood in *E. nipponicum* relies on an evolutionarily conserved multi-enzyme network of cysteine and aspartic peptidases, similar to the proteolytic cascades of other blood-feeding helminths such as *Schistosoma*, *Fasciola*, *Ancylostoma*, etc. [13–16]. Among the endopeptidases of *E. nipponicum*, clan CA cysteine peptidase activities, including cathepsin L-like and cathepsin B-like activity, are dominant [12].

In helminths, cathepsins L and B play various roles. Due to their histolytic potency, they are involved in host invasion and tissue migration, but they also play a role in various pathological processes, immune evasion, and other parasite-host interactions, as well as in helminth reproduction, nutrient digestion, etc. [17–19]. In general, peptidases encoded in the helminth genomes are mostly temporarily expressed in the various life stages, thus reflecting the parasite's specific needs regarding digestion or other hydrolytic processes. For instance, a characterisation of several cathepsins L of the liver fluke *Fasciola hepatica* had shown that the infective larvae use cathepsin L3 to traverse the host's intestinal wall, while the fluke's migratory stages utilise cathepsin L2 to penetrate host liver tissue, and adults employ cathepsin L1 jointly with cathepsins L2 and L5 to digest host proteins [20]. The use of different peptidases with overlapping substrate specificities helps heteroxenous and tissue-migrating parasites to adapt to various environments and sources of nutrition within the hosts. On the other hand, little is known about the complex functioning of cysteine peptidases in monoxenous blood-feeding monogeneans that spend most of their life attached to a single type of host tissue, such as the gills.

In the present study, we focused on clan CA cysteine peptidases, namely cathepsins L and B, selected due to their abundance in the transcriptome of adult *E. nipponicum* worms. We have employed phylogenetic and bioinformatic analysis to investigate their relationship to other helminth peptidases. We have selected two of the most abundant cathepsins L and the only expressed cathepsin B, and produced these as functionally active recombinant forms using the *Pichia pastoris* expression system. A biochemical and functional characterisation of the recombinant enzymes was performed in order to understand their specificity and substrate preference. Furthermore, we have developed specific antibodies and RNA probes to localise proteins/transcripts within the worms' bodies by immunofluorescence assay and RNA *in situ* hybridisation technique. Our work thus presents the first detailed functional characterisation of monogenean peptidases.

Methods

Parasite material

Adult worms of *E. nipponicum* were collected from the gills of infected carps (*C. carpio*) freshly slaughtered in a commercial facility of the fishery Rybářství Třeboň, Plc., Třeboň basin, South Bohemia, Czech Republic. A soluble worm extract, parasite RNA, and first-strand cDNA were

obtained as previously described [12]. For RNA *in situ* hybridisation, the worms were flat-fixed between microscopic slides in Bouin's solution (Sigma-Aldrich, Darmstadt, Germany) at RT for 1 h, then transferred into fresh fixative and incubated at 4 °C overnight. For immunohistochemistry, 4% paraformaldehyde in PBS was employed as a fixative. Fixed worms were dehydrated with increasing concentrations of ethyl alcohol (70–100%), cleared with xylene, and embedded in paraffin.

Sequence analyses of cathepsin L-like and B-like peptidases of *E. nipponicum*

Transcripts from the transcriptome of adult *E. nipponicum* worms were assembled within a parallel project and deposited in the NCBI GenBank database under accession number GFYM00000000. They were annotated by searching for the closest homologues. We have used BLASTp and BLASTn algorithms (E-value 1e-5) to mine the following public sequence databases: NCBI non-redundant protein database [21], MEROPS database of peptidases and their inhibitors [22], UniProtKB/UniRef100 database and UniProtKB/TrEMBL only used for searching for protein sequences related to phylum Platyhelminthes (Taxon ID: 6157) [23], UniProtKB/Swiss-Prot [24], RCSB PDB [25] and DDBJ [26]. Sequences encoding cathepsin L-like and B-like peptidases of *E. nipponicum* were identified and their amino acid sequences (catalytic domains) aligned with sequences of formerly recognised as *E. nipponicum* cathepsins L, *EnCL1* (GenBank: KP793605) and *EnCL3* (GenBank: KP793606), *F. hepatica* cathepsin L1 (GenBank: AAT76664.1), *F. hepatica* cathepsin L3 (GenBank: CAC12807.1) and with *E. nipponicum* cathepsin B, *EnCB* (GenBank: MF346929), *Schistosoma mansoni* cathepsins B1 (GenBank: CAD44624.1) and B2 (GenBank: XP_018651608.1), respectively, using Clustal Omega tool (http://www.ebi.ac.uk/Tools/msa/clustalo/) [27]. The relative abundance of cathepsin L transcripts, which reflects the rate of transcription of the corresponding genes, was predicted by back mapping of raw Illumina reads to the assembled transcripts using Kallisto v. 0.43.0 [28]. The presence of a signal sequence in amino acid sequences of the enzymes was predicted by SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/) [29]. A theoretical position of the pro-region and composition of the enzymes' S2 subsites of the active site were determined by multiple sequence alignment (see above). Molecular mass and theoretical pI of the deduced proteins were determined by Compute pI/Mw software available at the ExPASy web portal (https://web.expasy.org/compute_pi/) [30]. Potential N-glycosylation and O-glycosylation sites were identified using an online tool at NetNGlyc 1.0 and NetOGlyc 4.0 Servers (http://www.cbs.dtu.dk/services/NetNGlyc [30] and http://www.cbs.dtu.dk/services/NetOGlyc [31]).

Eudiplozoon nipponicum cathepsin L1 and L2 amino acid sequences were used as queries for BLAST searches in the NCBI non-redundant protein database [32] and the Wormbase database of platyhelminths [33]. Recovered sequences were further analysed using CLANS software [34] and a cluster of 515 sequences containing *E. nipponicum* cathepsins was selected. Redundant sequences were filtered at 95% level of identity using the CD-HIT software [35]. Divergent and incomplete sequences were manually removed, which resulted in a dataset of 288 sequences. These were aligned using the MAFFT [36] and the alignment was trimmed using BMGE (239 sites) [37]. We inferred a maximum-likelihood tree using the best-fit model (LG + C60 + G), and tested the topology with 10,000 ultrafast bootstraps in IQ-TREE [38]. Bayesian inference of phylogeny was run under the C60 site-heterogeneous mixture model using Phylobayes.

Expression of recombinant cathepsins L1, L3 and B in *Pichia pastoris* and their purification

Two abundant *EnCLs* (*EnCL1* and *EnCL3*) and the only *EnCB* were selected for a functional expression. Gene-specific primers (Additional file 1) and the first-strand cDNA of adult *E. nipponicum* were used to amplify the *yrEnCL1* (GenBank: KP793605), *yrEnCL3* (GenBank: KP793606), and *yrEnCB* (GenBank: MF346929) genes. The amplified products were ligated into a *P. pastoris* expression vector pPICZα B as described previously [39], and the constructs were verified by Sanger sequencing. Protein expression in *P. pastoris* X-33 was carried out following a protocol described previously [40] and according to the manual of the Easy Select Pichia Expression Kit (Thermo Fisher Scientific, Waltham, Massachusetts). The yeast medium was centrifuged and the supernatant filtered (0.22 μm) and concentrated on AmiconUltra-15 filters 10,000 MWCO (Millipore, Darmstadt, Germany). The enzymes were affinity purified via their His-tag using Ni-chelating chromatography (HisTrap™ FF crude, GE Healthcare, Little Chalfont, United Kingdom) and a stepwise elution by imidazole (0.05–0.5 M). Yeast-origin hyperglycosylation of *EnCB* was removed by endoglycosidase F1 (Calbiochem, Darmstadt, Germany). Purification was completed by cation exchange FPLC on Mono S™ 5/50 GL column (GE Healthcare) using 20 mM MES buffer pH 6.0 and gradient elution by NaCl (0–1 M).

Expression of recombinant cathepsin L1 in *Escherichia coli* and its purification

The *EnCL1* gene (GenBank: KP793605) was amplified using first-strand cDNA of adult *E. nipponicum* and specific primers (Additional file 2). The amplified product was ligated into the *E. coli* expression vector pet28a + (Novagen, Darmstadt, Germany) and verified by DNA sequencing.

EnCL1 was expressed in *E. coli* BL21 Star™ (DE3) (Invitrogen, Carlsbad, California) according to pET System Manual (Novagen). Harvested cells were lysed in 20 mM Tris-HCl pH 8/ 0.3 M NaCl/1% lauryl sarcosine/10 mM imidazole by sonication, and centrifuged at 10,000× *g* for 10 min. Inclusion bodies were solubilised with 6 M guanidine hydrochloride as described elsewhere [41]. The solubilised mixture was filtered (0.65 μm), passed over the column containing Ni-NTA beads (Qiagen, Hilden, Germany), and eluted by imidazole (0.05–0.5 M). Chromatographic fractions were analysed by SDS-PAGE and the identity of suspected protein bands was verified by mass spectrometry as described in [12].

Processing the *yrEnCL3* pro-enzyme

Auto-processing of *pro-yrEnCL3* was tested by incubation of the recombinant protein in 50 mM/100 mM CPB, with 2 mM DTT at pH values of 4, 5 and 6. Aliquots of 20 μl were taken from reactions at specific time intervals (15, 30, 60 and 120 min) at 37 °C, whereby incubations were stopped by adding 10 μM E-64. Processing of the enzyme to its mature form was monitored by SDS-PAGE, Western blotting and Edman degradation of N-terminus and subsequently confirmed by fluorometry with the corresponding synthetic peptide substrates (see below). Processing of *yrEnCL3* was also monitored by affinity labelling using 20 μM fluorescent probe BODIPYgreen-DCG-04 [42] and 2.5 μg of the enzyme (30 min incubation at pH 5) as described previously [40]. Specificity of the probe towards the active site of the peptidase was verified by pre-incubation with the cysteine peptidase-specific inhibitor E-64 (10 μM) for 5 min.

Peptidase activity assays

Peptidase activities of recombinant cathepsins were measured with fluorogenic aminomethylcoumarin peptide substrates (Bachem) as described previously [12]. Substrates tested with *yrEnCL1* and *yrEnCL3* included Z-Phe-Arg-AMC (FR), Z-Leu-Arg-AMC (LR), Z-Arg-Arg-AMC (RR), Z-Pro-Arg-AMC (PR) and Z-Gly-Pro-Arg-AMC (GPR). The *yrEnCB* was tested with FR and RR substrates only. The purified *rEnCLs* (15 nM) and *rEnCB* (30 nM) were dissolved in 100 μl of 50 mM/100 mM CPB supplemented with 2 mM DTT, in pH range of 3–8. The reactions were initiated by adding the individual substrates (final concentration 50 μM) in 100 μl of the same buffer at 28 °C. The release of free AMC was measured at excitation and emission wavelengths of 355 and 460 nm, respectively, in Infinite M200 fluorometer (TECAN, Männedorf, Switzerland).

For inhibition assays, three inhibitors were used (final concentration 10 μM, Sigma-Aldrich): E-64 (L-trans-epoxysuccinyl-leucylamido [4-guanidino] butane), which is a general cysteine peptidase inhibitor, iCL (Arg-Lys-Leu-Leu-Trp-NH2), a reversible inhibitor of cathepsin L; and CA-074 (N-[L-3-trans-propylcarbamoyloxirane-2-carbonyl]-Ile-Pro-OH), which is a selective inhibitor of cathepsin B. Inhibitors were mixed with enzyme samples and incubated at 28 °C for 15 min prior to adding the substrates. Remaining hydrolysis rates were measured with the FR substrate (cathepsins L and B), with LR (cathepsins L), or with RR (cathepsin B) as described above.

The assay of exopeptidase activity of activated *yrEnCB* (75 nM) was carried out with benzoyl-glycinyl-histidinyl-leucine (Bz-Gly-His-Leu; final conc. 2 mm) as a substrate, using a modified protocol [43]. The activity was measured in 50 mM/100 mM CPB pH 4–6 containing 2 mM DTT and 0.05 mg/ml fluorescamine.

Hydrolysis of macromolecular substrates by recombinant cathepsins L and B

Processed *yrEnCL1*, *yrEnCL3* and *yrEnCB* (2.5 μg each) were incubated at 37 °C with 50 μg of macromolecular substrates (bovine haemoglobin and albumin, human collagen I, IgG and fibrinogen; Sigma-Aldrich) dissolved in 100 μl of 50 mM/100 mM CPB with 2 mM DTT in pH 4.5–6. Aliquots (10 μl) for SDS-PAGE analysis in 12% gels were taken at various intervals (0, 30, 60 and 120 min and 16 h).

RNA *in situ* hybridisation

EnCL1/*EnCL3* gene-specific primers (Additional file 3) and first-strand cDNA of adult *E. nipponicum* were used for the PCR. The amplified products were ligated into pGEM®-T Easy vector (Promega, Madison, Wisconsin) and verified by DNA sequencing. The constructs were linearised and used as a template for RNA probes. Both sense and anti-sense RNA probes were synthesised in vitro (Dig RNA Labelling Kit (SP6/T7); Roche, Basel, Switzerland). *In situ* hybridisation was performed using a modified protocol [44]. Briefly, the sections of adult worms on slides were incubated for 19 h at 41 °C (*EnCL1*) and at 37 °C (*EnCL3*) with specific RNA probes diluted 1:100 in a hybridisation mixture (5× SSC, 1× PBS, 50% deionised formamide, 1% Tween-20, 10% dextran sulphate Mw 500×, 1 mg/ml Torula yeast RNA). Detection was achieved with alkaline phosphatase-conjugated anti-digoxigenin antibodies (1:500, Roche) and Fast Red TR substrate (Sigma-Aldrich). Negative controls were incubated under the same conditions but with an anti-sense probe or without a probe. Specific strand RT-PCR was used for the detection of natural anti-sense transcripts occurring in the monogenean cells [45].

Production of peptidase-specific polyclonal antibodies

For the production of monospecific polyclonal *anti-EnCL3* and *anti-EnCB* antibodies, ICR/CD1 mice (ENVIGO,

Huntingdon, United Kingdom) were injected subcutaneously twice, at a 14-day interval, with 30 μg of purified *pro-yrEnCL3* or deglycosylated *yrEnCB* in sterile saline and TitermaxGold adjuvant (Sigma-Aldrich). The mice were boosted by an intramuscular injection of purified antigens (15 μg) in sterile saline after another two weeks. *Anti-EnCL1* antibodies were produced in mice injected intraperitoneally with *c.*50 μg of *brEnCL1* cut as a band from a 12% gel after SDS-PAGE. The strip of the gel containing the antigen was repeatedly and thoroughly washed in 50% methanol/10% acetic acid for several hours to remove SDS, washed several times in sterile saline, and homogenised in sterile saline using a motorised pellet pestle. Two antigen deliveries were performed at an interval of 14 days. All mice were bled under deep ketamine/xylazine anaesthesia 14 days after the last injection and sera were collected by centrifugation. Control sera were taken from the same mice prior to immunisation.

Western and affinity blotting

Specificity of the anti-cathepsin antibodies produced in mice was verified on PVDF membrane immunoblots of soluble worm extracts previously separated by SDS-PAGE in 12% gels (20 μg/well). The experiment was performed according to a modified protocol [40]. The membranes were first incubated for 1 h with mouse *anti-EnCL3/anti-EnCL1/anti-EnCB* /control sera diluted 1:100 in PBS-T and then for 1 h with horseradish peroxidase-labelled goat anti-mouse IgG (Sigma-Aldrich) diluted 1:5000 in PBS-T. Finally, the membranes were developed by the Opti-4CN Substrate Kit (Bio-Rad, Hercules, California).

Purified recombinant proteins (1 μg/well) were resolved by SDS-PAGE in 12% gels and either stained by CBB or trans-blotted onto a PVDF membrane. His-tagged enzymes were detected on blots either with a mouse biotinylated monoclonal Anti-polyhistidine antibody (Sigma-Aldrich), alternatively with 5 nM iBody4, which is a biotinylated copolymer containing nitrilotriacetic acid-bound nickel cations [46]. In both cases, detection was finalized using peroxidase-labelled streptavidin (Sigma-Aldrich) and the Opti-4CN™ Substrate kit.

Immunohistochemistry

Deparaffined and rehydrated 5 μm sections of worms on slides were heated 3 × 3 min in 0.01 M citrate buffer pH 6 containing 0.05% Tween 20 in a microwave oven (500 W) and then allowed to cool for 20 min. Subsequently, the sections were blocked and immunostained according to a modified Immunocytochemistry and Immunofluorescence Protocol (Abcam). After an overnight incubation with sera (1:50), the slides were placed in a dark chamber for 1 h at RT with Goat Anti-Mouse IgG H&L-Alexa Fluor® 568 (Ex: 578 nm/Em: 603 nm, Abcam, Cambridge, United Kingdom) diluted 1:200 in PBS-TX100 with 1% BSA. Finally, the sections were mounted in Vectashield with DAPI (Vector laboratories, Burlingame, California). Immunofluorescence was observed and photographed under fluorescence or confocal microscope.

Results

Bioinformatic and phylogenetic analyses of the primary structures of the enzymes

Ten unique transcripts encoding different cathepsin L-like sequences were discovered in the transcriptome of adult *E. nipponicum*. Of the predicted CL amino acid sequences, seven represented complete sequences of mature (processed) enzymes, with six encoding the complete sequences of zymogens including the pro-sequences. Two of the predicted complete zymogen sequences were identical with the sequence of *EnCL1* (GenBank: KP793605) and *EnCL3* (GenBank: KP793606) obtained previously by reverse transcription of parasite mRNA using degenerate primers, followed by RACE-PCR [12]. The complete pro-enzyme molecules were assigned *EnCL1-EnCL6*. The remaining four incomplete sequences shared a relatively high similarity with *EnCL6*: the five close orthologs were therefore assigned *EnCL6a-EnCL6e*. The alignment of the predicted amino acid sequences of mature *EnCLs* with cathepsins L1 (GenBank: AAT76664.1) and L3 (GenBank: CAC12807.1) of *F. hepatica* is presented in Additional file 4, including a mutual comparison of the amino acids of the S2 pocket, which determines the substrate-binding specificity of the enzymes. The alignment of all complete/incomplete amino acid sequences of *EnCLs* is included in Additional file 5.

Interestingly, *EnCL2* displays a substitution of the active site cysteine (Cys25) to serine (Ser25). In terms of amino acid constitutions of the S2 pocket (Additional files 4, 5 and 6), the *EnCLs* can be divided in 3 groups. One is formed by CL1, CL3, CL4 and CL5, which share the Leu67, Ala133, Leu157 motif (papain numbering used throughout this column), with a slightly hydrophobic Ala being at the bottom of the S2 pocket. The second includes *EnCL6a*, *EnCL6b* and *EnCL6e*, where Ala133 is replaced by the amphiphilic Gly133 (for *EnCL6c* and *EnCL6d* data were not available.). The third group contains only *EnCL2*, in which the S2 pocket differs substantially from that of the other *EnCLs*: Ala133 is replaced by the more hydrophobic Val133 at the bottom, and moreover, the two leucines are replaced by Trp67 and Phe157, which could make the S2 pocket highly hydrophobic. The most variable element in the S2 pocket is the amino acid position 205, where hydrophobic residues occur in CL3, CL4 and CL5, and uncharged hydrophilic residues in CL1 and CL6a. In the case of CL2, residue 205 is a positively charged Arg, whereas in CL6e, it is the negatively charged aspartate.

Some of the *EnCLs* include potential N-glycosylation sites: CL2 has one and CL4 has two in the mature

enzyme, others occur in the pro-sequences of CL2, CL4 and CL5. Several O-glycosylation sites have been predicted in all but one (*EnCL6b*) cathepsin L sequences. Most, however, are located in the pro-domains; CL5 has three, while all the other enzymes except for CL6d have one in the mature part (Table 1, Additional file 5).

The rate of transcription of the particular *EnCLs* was predicted by a reverse mapping of raw reads towards the selected transcripts. *EnCL1* is the most abundant and *EnCL3* is the sixth most abundant cathepsin L in *E. nipponicum* adults (Fig. 1).

Phylogenetic analysis has shown that *E. nipponicum* cathepsins L separate into two clusters: one contains CL1, CL3 and CL5, and the other includes CL2, CL4 and CL6a. The two clusters belong to separate clades, whereby each of the clusters also includes different cathepsins L of free-living rhabditophorans and CLs of digeneans and cestodes (Fig. 2, Additional file 7). It can thus be concluded that *E. nipponicum* cathepsins L are polyphyletic.

The cluster containing *EnCL1* and *EnCL3*, for which the function in blood digestion was confirmed in this study, occurs within the same clade as the CL of another haematophagous monogenean, *Protopolystoma xenopodis*. The various CLs of the mucophagous *Gyrodactylus salaris* fall into both clades where the two clusters of *EnCLs* are located. Interesting is the highly supported position of myxosporean CLs within the neodermatan clade which contains also *EnCL2*, *EnCL4* and *EnCL6a*.

In stark contrast to the numerous *EnCLs*, only one transcript encoding *EnCB* was found in the *E. nipponicum* transcriptome; it contained the whole sequence of the pro-enzyme. Its closest relative found in the databases is cathepsin B2 of *Schistosoma mansoni* (GenBank: XP_018651608.1) (see the alignment in Additional file 8).

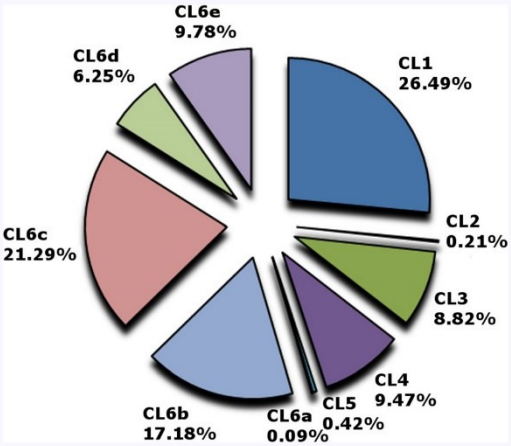


Fig. 1 Relative transcription rate of *E. nipponicum* cathepsins L. The relative transcription rate predicted by a back mapping of raw Illumina reads to the assembled transcripts was calculated as the percentage of all cathepsin L transcripts

Its primary structure contains the typical motif of an occluding loop, which is responsible for peptidyl dipeptidase activity of cathepsins B. One potential N-glycosylation site is positioned on the occluding loop. Seven O-glycosylation motifs were predicted in the molecule, five of them located within the pro-domain. *EnCB* also contains a modified 'haemoglobinase motif' (213)YWLIA**N**SW–EWGD(226) in the asparagine active site region, which is ascribed to haemoglobinolytic cathepsins B of helminth blood-feeders [47]. The predicted/computed basic features of all *EnCLs* and *EnCB* are summarised in Table 1.

Table 1 Predicted/computed sequence-derived features of *E. nipponicum* cathepsins L and B

	Transcript_ID	Bases/AA	Completeness	SP	pI proenzyme/ mature enzyme	MW proenzyme/ mature enzyme (kDa)	N-/O-glyc sites
CL1	E_nip_trans_58808_m.372114	954/317	yes	no	5.86/6.08	34.9/24.36	0/7
CL2	E_nip_trans_70234_m.461805	1029/342	yes	yes	7.23/4.87	37/24.9	3/2
CL3	E_nip_trans_02967_m.11341	1107/368	yes	yes	4.80/4.15	38/24.1	0/5
CL4	E_nip_trans_06099_m.25531	1062/353	yes	yes	5.59/4.94	37.6/24.2	3/8
CL5	E_nip_trans_65378_m.396731	1071/356	yes	no	5.62/5.02	39.5/24	1/8
CL6a	E_nip_trans_15113_m.115989	1059/352	yes	yes	5.55/4.8	37.7/24.4	0/2
CL6b	E_nip_trans_04751_m.20488	519/173	iN	?	?/?	?/?	0/0
CL6c	E_nip_trans_04670_m.20209	642/214	iC	yes	?/?	?/?	0/4
CL6d	E_nip_trans_55822_m.362291	705/235	iC	yes	?/?	?/?	0/1
CL6e	E_nip_trans_60687_m.380367	768/255	iN	?	?/4.58	?/24.47	0/2
CB	E_nip_trans_02724_m.9562	1149/382	yes	yes	7.21/5.55	39.9/28.5	1/7

Abbreviations: *AA* Amino acids, *SP* Signal peptide, *N-/O-glyc sites* Number of potential N-glycosylation/O-glycosylation sites, iN Incomplete N-terminus, *iC* Incomplete C-terminus, *?* data not available due to an incomplete sequence

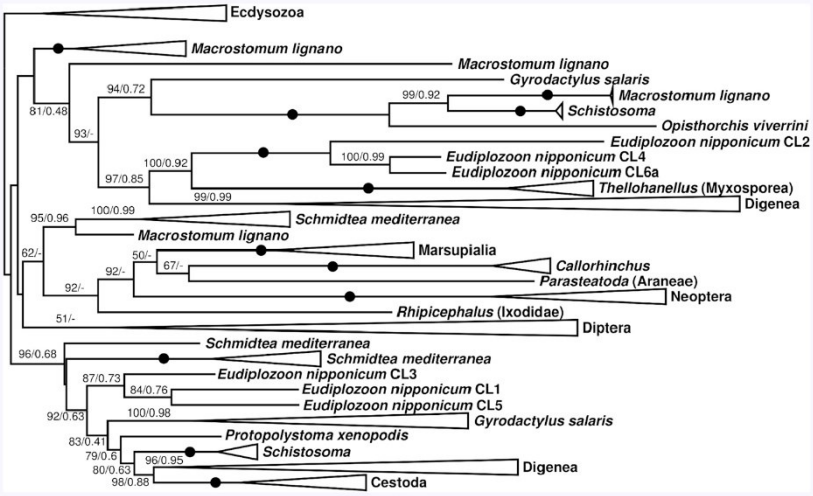


Fig. 2 Collapsed phylogram showing relationships between *E. nipponicum* cathepsins L and cathepsins L of other selected organisms. Collapsed unrooted maximum-likelihood tree of selected *E. nipponicum* cathepsins L inferred using the best-fit model (LG + C60 + G). Leaves of related organisms are collapsed. Ultrafast bootstrap supports and posterior probabilities are shown where ultrafast bootstrap is ≥ 50 or posterior probability ≥ 0.9. Maximum support (100/1.0) is indicated by a black circle

Functional expression and purification of recombinant cathepsins L1, L3 and B

EnCL1, *EnCL3* and *EnCB* were expressed as pro-enzymes in yeasts. Since *yrEnCL1* and *yrEnCB* had undergone self-processing already in yeast cultivation medium, a high activity of both was confirmed in the presence of appropriate fluorogenic peptide substrates. The *yrEnCL1* thus migrated in electrophoretic gels as a ~28 kDa dominant double band and reacted on the blot with anti-His-Tag mouse antibody and iBody4 at the same size (Fig. 3). The enzyme was partially purified by Ni-NTA affinity chromatography, but attempts to purify it to homogeneity by ion-exchange chromatography failed due to its poor stability. Activity and inhibition assays were therefore performed with only partially purified enzyme. Additionally, for the production of anti-*EnCL1* antibodies it was necessary to use the inactive recombinant form produced in *E. coli* (*brEnCL1*).

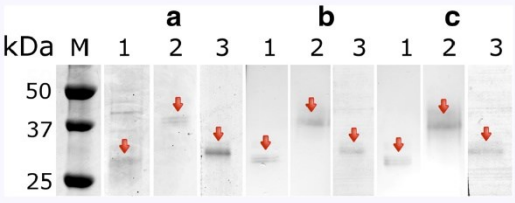


Fig. 3 SDS-PAGE and affinity–/immunoblots of *yrEnCL1*, *yrEnCL3* and *yrEnCB* after purification. **a** 12% gel stained with Coomassie Brilliant Blue. **b** affinity blot labelling by iBody4. **c** immunoblot detection by anti-His-Tag antibody. Lane 1: *yrEnCL1* (28 kDa partially processed form); Lane 2: *yrEnCL3* (38 kDa pro-enzyme); Lane 3: *yrEnCB* (29 kDa deglycosylated processed enzyme). Arrows point to the enzyme bands

The affinity purified unprocessed *pro-yrEnCL3* migrated in the gel as a double band at ~38 kDa, and reacted with the anti-His-Tag antibody and iBody4 at the same size (Fig. 3). Autocatalytic processing of *pro-yrEnCL3* was most effective at pH 4, and resulted in a ~28 kDa band in the gel. Processing of the zymogen was less effective at pH 5, while at pH 6 it did not take place even after prolonged incubation. Affinity probe BODIPYgreen-DCG-04 bound strictly to the mature enzyme, while in the pro-enzyme no labelling was observed (Fig. 4). Edman degradation revealed LPTDVD sequence at the N-terminus of the processed *EnCL3*. Since the processed enzyme had limited stability, the purified pro-enzyme was used for the immunisation of mice.

YrEnCB pre-purified on a Ni-NTA column migrated as a broad band around 50 kDa. After enzymatic deglycosylation and subsequent purification by ion-exchange chromatography, it migrated as a ~29 kDa band in the gels, which correlated with the results of Western and affinity blotting (anti-His-Tag antibody, iBody4) (Fig. 3).

Activity and inhibition assays

Both *EnCLs* efficiently hydrolysed the FR substrate, and to a lesser extent the LR. No activity towards RR, PR and GPR was detected. The optima for hydrolysis of both substrates were pH 5.5 for *yrEnCL1* and pH 6 for *yrEnCL3*. The preferred substrate for *yrEnCB* was the

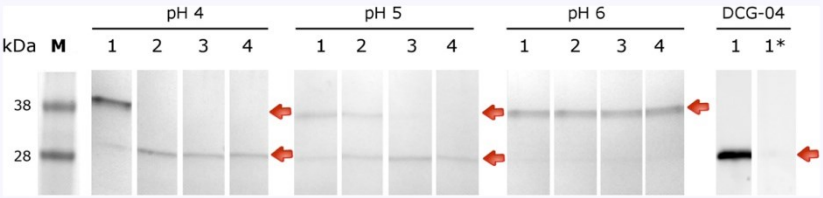


Fig. 4 SDS-PAGE and active site labelling of purified *pro-yrEnCL3* after auto-activation at various pH. The 38 kDa zymogen was autocatalytically processed to a 28 kDa form during incubation in 50 mM/100 mM CPB containing 2 mM DTT at 37 °C. After incubation at various pH values, the protein was run in 12% polyacrylamide gels and stained with Coomassie Brilliant Blue (pH 4, pH 5 and pH 6). Lane 1: 15 min incubation; Lane 2: 30 min; Lane 3: 60 min; Lane 4120 min. DCG-04, a 12% gel showing active site labelling of the processed enzyme with fluorescent affinity probe BODIPY green DCG-04. Lane 1: labelling by the affinity probe after 30 min incubation at pH 5; **1***, binding of the affinity probe was blocked by E-64 inhibitor. The 38 kDa zymogen and the 28 kDa processed enzyme are indicated by arrows

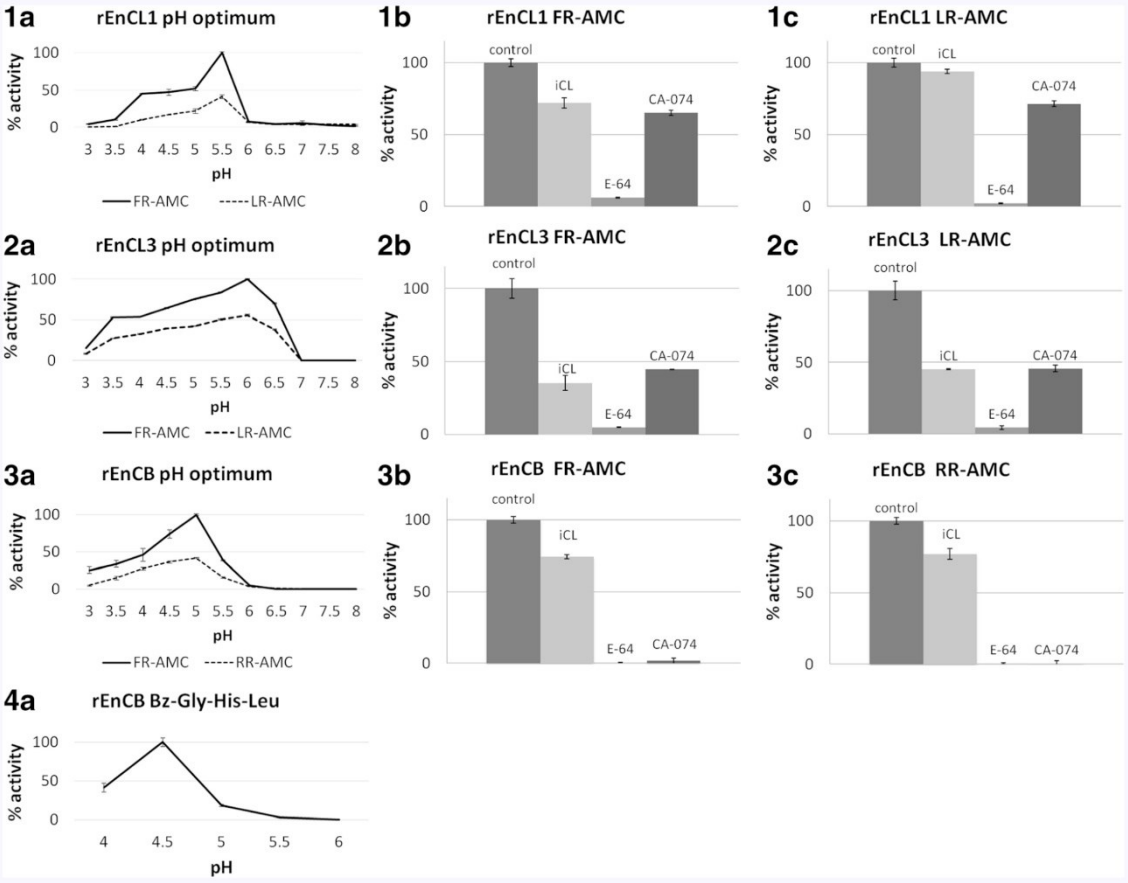


Fig. 5 pH optima of *rEnCLs* and *rEnCB* activities, and the effect of inhibitors on enzyme activities. Activity assays were performed in 50 mM/100 mM CPB with 2 mM DTT at pH 3–8. Inhibition assays were run at the pH optima of the particular enzymes. The values are expressed as the percentage of maximum activity in the sample without inhibitor. **1**, *yrEnCL1*; **2**, *yrEnCL3*; **3**, *yrEnCB*; **4**, *yrEnCB*. **a** pH optima of endopeptidase (1–3) and exopeptidase (4) activities. **b** inhibition assays with FR substrate. **c** inhibition assays with LR substrate (1–2) and RR substrate (3). The inhibitors (10 μM) are indicated in the graphs above the columns: E-64, general irreversible inhibitor of clan CA cysteine peptidases; iCL, reversible peptide inhibitor of cathepsin L; CA-074, irreversible inhibitor of cathepsin B

FR, while RR was cleaved less efficiently at the optimum of pH 5 (Fig. 5).

Activity of all three enzymes was effectively inhibited by 10 μM E-64 (the general inhibitor of clan CA family C1 cysteine peptidases) in the presence of either substrate. The cathepsin B-selective inhibitor CA-074 had completely suppressed the activity of *yrEnCB* at 10 μM and diminished the activity of *yrEnCL1* and *yrEnCL3* to ~70% and 45–65%, respectively, depending on the substrate. The reversible peptide inhibitor of cathepsin L (10 μM, iCL) partially inhibited all three cathepsins but was most effective in the case of *yrEnCL3* with both substrates (*c.*40% remaining activity). The inhibition rate of *yrEnCL1* was comparable with the effect on *yrEnCB* (*c.*90–70% remaining activity) (Fig. 5).

The exopeptidase (dipeptidyl peptidase) activity of *yrEnCB* peaked at pH 4.5, then rapidly dropping to *c.*40% at pH 4, and to less than 20% at pH 5 (Fig. 5).

Hydrolysis of macromolecular substrates by recombinant cathepsins L and B

Both *EnCLs* and *EnCB* efficiently degraded selected protein substrates (bovine haemoglobin and albumin, human collagen type I and fibrinogen) at pH 4.5. *EnCLs* readily hydrolysed also human IgG, which was poorly cleaved by *yrEnCB* at pH range 4.5–6.0 (Fig. 6). While the action of *yrEnCL1* did not vary substantially between pH 4.5–6.0, *yrEnCL3* did not act on the substrates (except fibrinogen) at pH 5.0 or higher (not shown). At pH, hydrolysis of all the macromolecular substrates by *yrEnCB* failed (not shown).

Western blotting and immunohistochemistry

Specificity of mouse *anti-EnCL1*, *anti-EnCL3*, and *anti-EnCB* antibodies was verified on immunoblots of the soluble worm extracts. The detected bands appeared at ~24 kDa (*EnCL1*), ~28 kDa (*EnCL3*) and ~29 kDa (*EnCB*). Reactions of the particular sera with the appropriate enzyme bands were highly specific (Fig. 7), so they could be used for immunohistochemistry. Control pre-immune sera did not exhibit any reaction.

The reactions of specific antibodies with histological sections of the worms showed the presence of both *EnCL1* and *EnCL3* in the lumen of the gut (Fig. 8). No reaction was observed in the haematin digestive cells or outside the digestive system.

Anti-yrEnCB antibodies reacted with vesicular structures in the vitelline cells of the adult parasites (Fig. 8). All control sera displayed no reaction.

RNA *in situ* hybridisation

In situ hybridisation with specific antisense-RNA probes showed a localisation of *EnCL1* and *EnCL3* transcripts specifically in the haematin (digestive) cells of the worms' gastrodermis (Fig. 9). No reactivity was found in other parts of the parasite body, including tegument, vitelline cells and the initial part of the digestive tract. The same pattern of reactivity was obtained when the sense-RNA probes were used. This was not a non-specific reaction, as demonstrated by the strand-specific RT-PCR which confirmed the presence of anti-sense transcripts in *E. nipponicum*. In the negative controls (without any probe), no reaction was observed.

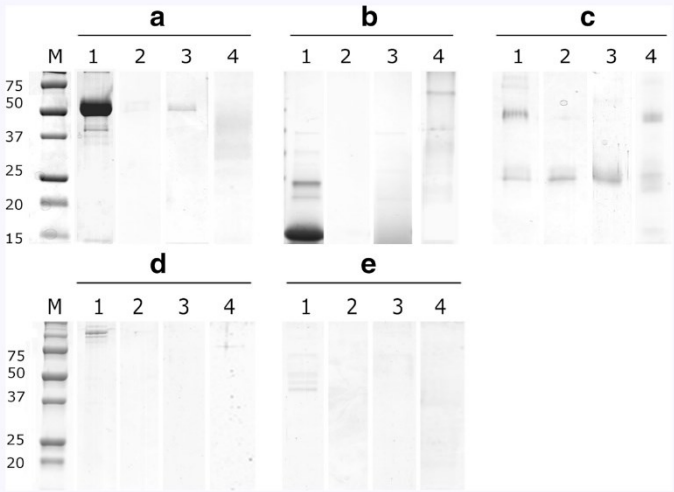


Fig. 6 SDS-PAGE of protein substrates degraded by recombinant *EnCL1*, *EnCL3* and *EnCB*. Prior to electrophoresis in 12% polyacrylamide gel, the enzymes were incubated overnight with various protein substrates in 50 mM/100 mM CPB pH 4.5 containing 2 mM DTT. **a** Bovine serum albumin. **b** Bovine haemoglobin. **c** Human IgG. **d** Human type I collagen. **e** Human fibrinogen. Lane 1: controls (substrates without enzyme); Lane 2: *yrEnCL1*; Lane 3: *yrEnCL3*; 4: *yrEnCB*. Gels were stained with Coomassie Brilliant Blue

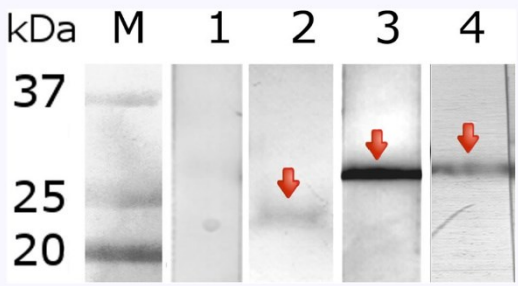


Fig. 7 Reactions of anti-cathepsin antibodies with proteins on blots of *E. nipponicum* soluble extracts. Lane 1: mixed control (pre-immune sera); Lane 2: *anti-brEnCL1* antibodies; Lane 3: *anti-pro-yrEnCL3*; Lane 4: *anti-yrEnCB*. Arrows point to the enzymes detected by specific antibodies

Discussion

In a previous study [12], we proposed that cathepsin L-like peptidases, in cooperation with cathepsins D and B, play a leading role in the digestion of haemoglobin in the sanguinivorous monogenean *E. nipponicum*. Although the presence of cathepsins L in soluble extracts of the worms was suggested by several methods in that study, the large number of different cathepsin L transcripts (10 in total) found in the transcriptome of the adult life stage was not anticipated. In other blood-feeding helminths, such as *Fasciola hepatica*, multiple cathepsins L are involved in feeding and other important tasks, such as tissue migration, immune evasion, and possibly egg production and metacercarial excystment [48, 49]. Small changes (even single amino acid substitutions) in the primary structure, but especially in the S2 subsite of the active site of those cathepsin L types/isoforms, alter the substrate specificity [50–52]. The production of peptidases with overlapping

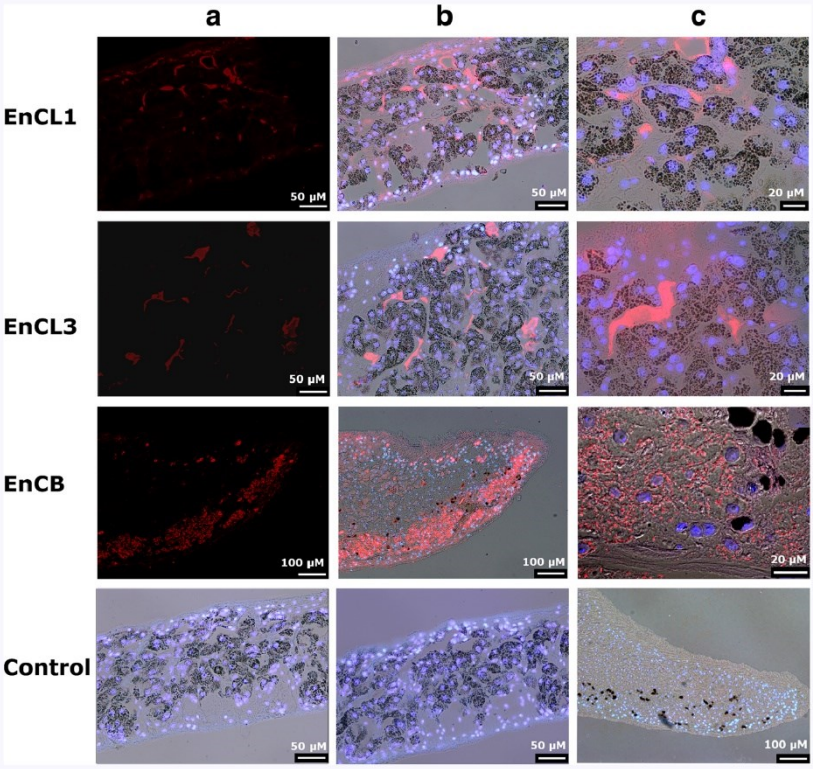


Fig. 8 Immunolocalisation of *EnCL1*, *EnCL3*, and *EnCB* on histological sections of adult worms. Longitudinal sections of *E. nipponicum* adult; anterior part of the worm's body is shown. *EnCL1*, reaction with *anti-brEnCL1* antibodies; *EnCL3*, *anti-pro-yrEnCL3* antibodies; *EnCB*, *anti-yrEnCB* antibodies. **a** Red fluorescence (Alexa Fluor 568 goat anti-mouse IgG) of the labelled secondary antibody in a dark field. **b** Merged picture of bright field view, red antibody fluorescence, and blue DAPI fluorescence. **c** Higher magnification of **b**. Reaction of antibodies raised against *brEnCL1* and *yrEnCL3* takes place in the lumen of the gut. Reaction of antibodies raised against *yrEnCB* takes place in the vesicular structures within vitelline cells. Control, pre-immune mouse sera in the order *EnCL1* (**a**), *EnCL3* (**b**), *EnCB* (**c**); merged picture of bright field view, red and blue fluorescence. No reaction with any of the control sera was detected

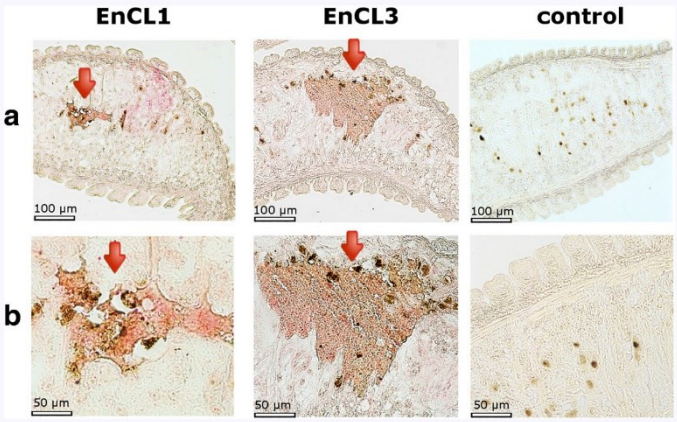


Fig. 9 Localisation of mRNA encoding *EnCL1* and *EnCL3* using RNA *in situ* hybridisation. Longitudinal sections of *E. nipponicum* adult worms incubated with specific *EnCL1*/*EnCL3* antisense-RNA probes. **a** Reaction in haematin cells of the digestive tract. **b** Higher magnification of the reacting areas from the same section. Controls were performed without any probe because the worms also produce anti-sense transcripts

specificities enables the parasite to degrade host-derived macromolecules more efficiently.

All but one of the *EnCLs* characterised in this study possess a typical catalytic triad of cysteine peptidase active site (Cys/His/Asn); the exception is *EnCL2*, in which the active site cysteine (Cys25) is replaced by serine (Ser25). However, as described for congopain (cathepsin L-like peptidase) in *Trypanosoma congolense*, Cys-to-Ser mutation need not necessarily affect the proteolytic ability of an enzyme [53]. It has been also suggested that Cys-to-Ser mutation in the cysteine peptidases of parasites might be a way to moderate oxidation of the active site in case parasites have to deal with an oxygen-rich environment, such as blood or external environment [17]. In addition to this substitution in the active site, *EnCL2* has in its S2 pocket Trp at position 67 (papain numbering used throughout the Discussion), which makes it a potential collagenase/elastase. Presence of the bulky aromatic residues Tyr67/Trp67 is typical of *F. hepatica* cathepsins L2 and L3, and of human cathepsin K, which all efficiently cleave collagen and elastin of the connective tissues due to their ability to accommodate Pro residues at P2 position in the substrates via a stabilising interaction with Tyr67/Trp67 [50–52, 54, 55]. However, the functionality and possible role of *EnCL2* has not yet been studied. Its expression rate in adult worms is very low, which indicates that the enzyme is predominantly employed by another life stage (the larva: oncomiracidium; or the post-oncomiracidial stage: diporpa), or in some less prominent processes unconnected with blood digestion.

Since the composition of the S2 pocket is the key determinant of substrate specificity of papain-like cysteine peptidases [56], the amino acid sequences of *EnCLs* were aligned (Additional files 4, 5 and 6) and their S2 residues (positions 67, 133, 157, 205) compared. The comparison of various *EnCLs* revealed the existence of three groups differing in the polarity and charge of the S2 pocket. In *EnCL2*, the S2 pocket is strongly hydrophobic and carries a positive charge. *EnCL1*, *EnCL3*, *EnCL4* and *EnCL5* have a moderately hydrophobic S2 pocket with a neutral charge. The least hydrophobic S2 pocket is found in the closely related *EnCL6* orthologs; it has a neutral charge in *EnCL6a* and 6b, and a negative charge in *EnCL6e* (for *EnCL6c* and 6d the complete data are not available). The presence of a negatively charged Asp205 in the S2 pocket may endow *EnCL6e* with the ability to accommodate a positively charged Arg at P2 position of the substrate, which would give this cathepsin L ortholog a cathepsin B-like substrate specificity. Additionally, the group of *EnCL6* orthologs is atypical among cathepsins L by having an amphipathic Gly133 in the S2 pocket. This feature is, however, typical of cathepsins S, and Gly133 can also be found in a few other, yet unclassified, clan CA family C1 peptidases, such as the plerocercoid growth factor of the larval stage of the tapeworm *Spirometra* (Genbank: BAB62718.1) (see the alignment in Additional file 6). Nonetheless, specificity of the S2 pocket in papain-like peptidases is largely determined by the residue 205. Cathepsins S contain at this position a large aromatic non-polar Phe, which implies a short and more restricted pocket. Cathepsins L, on the other hand, usually include at this position branched aliphatic residues, which enables the accommodation of larger aromatic residues [57]. This was not observed in *EnCL6a*, which contains a polar Ser205. It thus seems that diversity in the composition of S2 subsites of *E. nipponicum* cathepsins L gives the parasite a wide array of efficient proteolytic tools with diverse affinities to potential P2 positions of protein substrates.

Phylogenetic analysis seems to support a supposition, inferred from the structural data and relative rates of expression, that in *E. nipponicum*, cathepsins L may be involved in various biological processes. The two functionally characterised enzymes, *EnCL1* and *EnCL3*, which seem to play a central role in blood digestion by the parasite, form a monophyletic group with *EnCL5*. The other monophyletic clade formed by *EnCL2*, *EnCL4* and *EnCL6* is relatively distant from the first group, thus suggesting involvement in other processes or, in the case of the least proportionally represented transcripts such as *EnCL2* and *EnCL6a*, a more substantial expression of these orthologs in other life stages. The presence of myxosporean-like cathepsins L within the otherwise purely neodermatan clade may lead to a speculation about a lateral gene transfer, since myxosporeans share their fish hosts with monogeneans. Investigation of this possibility would, however, require a thorough analysis that is beyond the scope of the present study. Moreover, it should also be noted that the evolution of genes need not follow the branching structure of species-level evolutionary trees due to mechanisms such as gene duplication and independent gene loss. Particularly, genes of parasites are prone to phylogenetic artefacts caused by attraction between long branches, an effect we tried to mitigate by using site-heterogeneous mixture models [58].

Surprisingly, only one transcript coding for cathepsin B was found in the transcriptome of adult *E. nipponicum*. Amino acid sequence of *EnCB* had shown that this enzyme is a typical representative of cathepsins B, including its active site, composition of the S2 subsite, and the occluding loop responsible for exopeptidase activity. It also contains a 'haemoglobinase motif' assigned to helminth cathepsins B with substrate specificity for haemoglobin [47]. In the case of *EnCB*, this motif is slightly modified in its C-terminal region, where the original sequence described for helminth haemoglobinases 'DWGE' is replaced by 'EWGD' (Additional file 8). Nevertheless, as we experimentally verified, these purely conservative amino acid substitutions did not affect the ability of *EnCB* to cleave haemoglobin. Similar substitutions in the motif were described for schistosome cathepsins B2 (*S. mansoni* and *Trichobilharzia regenti*) that were found only outside the parasites' digestive system; it is assumed that their functions do not include the digestion of haemoglobin [59, 60]. In contrast to *EnCB*, *T. regenti* CB2 degraded haemoglobin at a negligible rate.

In the process of heterologous functional expression in *P. pastoris*, *YrEnCL1* had undergone auto-activation in the cultivation medium by cleaving-off a significant proportion of its pro-sequence and migrated on SDS-PAGE as a 28 kDa band (the value thus differed from the expected ~35 kDa and ~24 kDa of the zymogen and fully processed enzyme, respectively). Other, less abundant bands also occasionally appeared in the gel; mass spectrometry had confirmed they were fragments of *yrEnCL1*, generated probably by auto-degradation (data not shown). This is consistent with the very limited stability of the partially processed enzyme observed during purification. Monospecific *anti-yrEnCL1* antibodies did, however, react with a ~24 kDa band on blots of soluble worm extracts, indicating a full processing of *EnCL1* under natural conditions. The *yrEnCL3*, on the other hand, was obtained as a relatively stable zymogen. It was capable of autocatalytic activation under acidic conditions, but the size of the resulting product, ~28 kDa, was also somewhat higher than the predicted mass of ~24.1 kDa of the mature enzyme. Nevertheless, Edman degradation confirmed an amino acid sequence at the N-terminus of the processed *yrENCL3* which had been predicted by bioinformatical methods, thus indicating that the processing was complete. Monospecific *anti-yrEnCL3* antibodies reacted similarly with a ~28 kDa band on blots of soluble worm extracts. Given that the *EnCL3* does not possess any N-glycosylation motifs, these differences cannot be explained by either natural glycosylation of the enzyme or by N-hyperglycosylation from *P. pastoris* in the case of the recombinant enzyme. It could be speculated that the observed shift in Mw of the enzyme was caused by disruption of disulphide bonds stabilizing cathepsin L structure, which resulted in mobility change during electrophoresis [60, 61]. Although *EnCL3* possesses one potential O-glycosylation site in the mature part, the prediction software on the NetOglyc server produces predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins, and the situation may be different in invertebrates. Moreover, we are not aware of any descriptions in literature of a cathepsin L with experimentally confirmed O-glycosylation.

S2 subsite specificity of *yrEnCL1* and *yrEnCL3* was tested over a range of pH with a set of oligopeptide synthetic substrates, varying in their Phe, Leu, Arg, and Pro at P2 position. As predicted from the composition of the S2 pockets, the enzymes were not able to cleave substrates containing Pro or Arg in this position. Consistently with most papain-like peptidases, both enzymes preferred at this position the bulky hydrophobic aromatic Phe and the branched aliphatic Leu, in order Phe > Leu [56]. The enzymes did, however, differ in their pH profiles of activity towards substrates, so that whereas *yrEnCL1* was active only in a narrow pH range with a pH optimum at 5.5, *yrEnCL3* had a broader pH activity profile with an optimum at pH 6. The slightly acidic pH optima correspond to those of cathepsin L-like peptidases of other blood-feeding parasites [61–64].

As expected, activity of both of the recombinant *EnCLs* was inhibited by the general papain-like cysteine

peptidase inhibitor E-64. Surprisingly, however, the iCL, a highly potent amphiphilic reversible peptide inhibitor of human cathepsin L [65], had but a little effect on the enzymes' activity, although it has Leu at the P2 position. One can thus hypothesise that a lower affinity to Trp at P1 position, found also in, e.g. cathepsin L of *Leishmania mexicana* [56], might account for this result. Remarkably, CA-074 (a specific inhibitor of cathepsin B) also partially reduced the activity of both cathepsins L. It has been reported elsewhere [66, 67], however, that CA-074 can decrease the activity of cathepsins L under reducing conditions and that could also explains our results.

Activity profiles, pH optimum, specificity to peptide substrates, inhibition by CA-074 and exopeptidase activity clearly show that *EnCB* is a typical member of cathepsins B group. The pro-enzyme had undergone auto-activation either in the yeast medium or during purification, and enzymatic deglycosylation demonstrated hyperglycosylation of the recombinant enzyme by yeast, which corresponds to the presence of an N-glycosylation motif in the *EnCB* molecule. *Anti-yrEnCB* monospecific antibodies detected native *EnCB* on blots of soluble worm extracts at ~29 kDa, which is adequate to the expected size of the processed enzyme which was probably slightly glycosylated naturally.

Both recombinant *EnCL1* and *EnCL3* efficiently degraded various blood proteins and type I collagen of the connective tissue. Unlike synthetic peptide substrates, macromolecular substrates (except for fibrinogen) were cleaved only at pH 4.5 by *yrEnCL3*, while *yrEnCL1* hydrolysed all macromolecular substrates at pH 4.5–6.0. Activity at low pH values is in accordance with the site of expression of both genes encoding the enzymes that was localised by *in situ* hybridisation in the haematin (digestive) cells, and with the presumed intracellular digestion of haemoglobin in the endolysosomal vesicles where pH below 4.5 would be expected [68]. Both enzymes, however, were also immunolocalised in the lumen of the parasite's gut. For other blood-feeding helminths, it has been estimated that the pH in the gut lumen is only slightly acidic (*c.*5.5–6.5) [43, 69, 70], and digestion therefore takes place in a local microenvironment between the lamellae of the digestive cells where the pH is more acidic (~4.5), as documented for *F. hepatica* [71]. Haematin cells of another studied blood-feeding diplozoid monogenean, *Paradiplozoon homoion*, also contain numerous lamellae on the side facing the lumen of the gut. Host erythrocytes are lysed in proximity of the haematin cells and released haemoglobin is taken in by lamellae of the cells via pinocytosis [9]. These facts, together with presence of *EnCL1* and *EnCL3* in excretory/secretory products of the worms [12], and with the results of immunolocalisation, imply that in diplozoid monogeneans extracellular, luminal digestion of blood can take place. Moreover, if the enzymes are released from the gut lumen into host circulation, they could potentially participate in some host-parasite interactions by affecting, e.g. fibrinolysis (they cleaved fibrinogen in vitro) or immune response (degradation of immunoglobulins).

Surprisingly, both *EnCLs* hydrolysed collagen type I. This ability was attributed in human body primarily to family C1 lysosomal cysteine peptidase cathepsin K. Because of acceptance of Pro in the P2 position, cathepsin K can completely degrade collagen by cleaving within the repeated Gly-Pro-Xaa motifs that occur in the triple helix of collagen [54]. The ability to accommodate substrates with P2 proline has also been observed in some cathepsins L of *F. hepatica*, namely *FhCL2*, *FhCL3* and *FhCL5* [50, 55, 72]. On the other hand, neither *yrEnCL1* nor *yrEnCL3* cleaved synthetic substrates GPR and PR. So, though both enzymes degraded native collagen, they probably do not possess intrahelical activity and this likely makes them less effective collagenases. Such a phenomenon has been described in human cathepsin L and *FhCL1* [51, 54].

During the localisation of expression of *EnCL1* and *EnCL3* by RNA *in situ* hybridisation, sense-RNA probes used as controls had shown similar reaction as antisense probes. It is known from the literature that over 30% of genes transcribed in humans also have antisense transcription [73], and antisense transcripts have been described for genes in a diverse group of eukaryotes [74–76]. The use of a specific strand RT-PCR has revealed that the *EnCL1* and *EnCL3* encoding genes also have naturally present antisense transcripts. Since the cathepsin L-like peptidases play important roles during the life of the monogeneans, antisense transcripts may be associated with the regulation of gene transcription [77].

It has originally been thought that *EnCB* is also involved in the proteolytic cascade of blood digestion in *E. nipponicum* [12]. In haematophagous parasites, many cathepsins B, together with other peptidases of the digestive cascades, function as important enzymes with a mixed mode of action, as endopeptidases that cleave proteins into smaller fragments and as exopeptidases (peptidyldipeptidases, carboxypeptidases) that remove dipeptides from the C-termini of those protein fragments due to the presence of the occluding loop [78]. Immunofluorescence with monospecific *anti-yrEnCB* antibodies clearly showed the localisation of *EnCB* in vesicular structures in vitelline cells, but its presence in the digestive tract has not been detected at all. We could thus speculate on its function in vitellogenesis. The presence of cysteine peptidases in the reproductive system has been detected also in other helminths [79–82], but their role in worms' reproduction is still speculative. The lack of exopeptidase activity in the gut due to the absence of cathepsin B may be compensated by the activity of cathepsins C; at least three different transcripts of these

dipeptidyl peptidases (aminopeptidases) were found in *E. nipponicum* transcriptome (not shown).

Conclusions

To our knowledge, this work represents the first comprehensive exploration and functional annotation of a group of peptidases from a monogenean transcriptome. It also describes the very first functional characterisation of recombinant cysteine peptidases of monogenean origin. The lack of experimental data from other monogeneans, particularly those of a high economic importance (such as *Gyrodactylus salaris*, *Benedenia* spp.), unfortunately precludes the possibility of drawing any comparisons. We have shown that the haematophagous monogenean *E. nipponicum* possesses a wide range of endopeptidases of cathepsin L type, most of which are highly expressed in adult worms. Some of these enzymes possess structural features rather unusual in cathepsins L. It holds true especially for the composition of S2 subsite of the active site, which is among the studied enzymes rather variable and includes even substitutions typical more of cathepsins B and cathepsins S. This expands our understanding of the structural diversity of cathepsins L in general. In contrast to the rich world of cathepsins L, the presence of just one cathepsin B in the worms was somewhat surprising. Despite its ability to hydrolyse haemoglobin in vitro, *EnCB*, due to its localisation in the vitelline cells of the parasite, does not seem to be involved in digestion. On the other hand, functional characterisation and localisation has revealed that *EnCL1* and *EnCL3* are important in the digestion of blood. Despite previous claims, it seems that the digestive process takes place not only within the intracellular lysosomal cycle, but at least in part also outside the gastrodermis, in the lumen of the gut. From there, the *EnCLs* may be eventually released into the blood circulation of the host and participate in elicitation of immune response and/or immunomodulation. Genomic data, currently available only for the mucophagous polyonchoinean *G. salaris* and the haematophagous heteronchoinean *Protopolystoma xenopodis*, and transcriptomic data from the mucophagous polyonchoinean *Neobenedenia melleni*, open a possibility for future comparative biochemical research of the two trophic strategies characteristic of the monogeneans. For example, preliminary data explorations show that mucus feeders may employ elastase-like serine peptidases that have not been found in blood-feeders (Mikeš, unpublished). Further research is likely to provide data interesting from an evolutionary point of view and information essential to understanding parasite-host interactions on a molecular level in this group of parasites, which would be of interest especially in context of pathogenesis and fish immunity. Although parasite peptidases have often been seen as promising targets for the development of vaccines or new inhibitor-based remedies, in the case of monogeneans we are somewhat sceptical regarding this possibility. The main reasons why we do not believe this is a promising direction are the likely costs linked to a complicated handling of large numbers of individual fish hosts in aquaculture, difficulties with administering a correct dosage of medication (peptidase inhibitors) in an aquatic environment (assuming that a peroral formula would even be available) and the necessity of repeated administering in order to reach a lethal effect on the parasites. At the moment, most methods applied against monogeneans rely on mass treatment of farmed fish by readily available, low-cost chemical biocides, such as hydrogen peroxide, or antihelminthics, such as praziquantel [83].

Additional files

Additional file 1: Primers for the expression of *yrEnCL1*, *yrEnCL3*, and *yrEnCB* in *P. pastoris*. (PDF 239 kb)

Additional file 2: Primers for the expression of *brEnCL1* in *E. coli*. (PDF 83 kb)

Additional file 3: Specific *EnCL1/EnCL3* primers for the synthesis of probes for RNA *in situ* hybridisation. (PDF 84 kb)

Additional file 4: Amino acid sequence alignment of *E. nipponicum* cathepsins L with cathepsins L of *F. hepatica*. Signal peptides and pro-sequences were omitted. Numbers at the end of the lines show amino acid numbering of particular mature parts of the enzymes. Catalytic triad of the active site (C, H, N) is marked by triangles. Conserved motifs around active site residues are shaded in grey. Residues within the S2 subsite of the active site involved in determining the substrate specificity are shaded in black and indicated with numbers (papain numbering). Tripeptides of potential N-glycosylation sites are boxed. Predicted O-glycosylated residues are marked by grey squares. The alignment was made using the sequences of *F. hepatica* mature cathepsins L1 (GenBank: AAT76664.1) and L3 (GenBank: CAC12807.1). (PDF 1949 kb)

Additional file 5: Multiple alignment of all the complete/incomplete amino acid sequences of *E. nipponicum* cathepsins L inferred from the transcriptome of adult worms. Signal peptides are shaded in dark grey, position of the pro-region cleavage site is marked by arrows. ERFNIN- and GNFD-like motifs are underlined and indicated by underlined headings. The catalytic triad of the active site (C, H, N) is marked by triangles. Conserved motifs around active site residues are shaded in light grey. Residues within the S2 subsite of the active site involved in determining the substrate specificity are shaded in black and indicated with numbers (papain numbering). Tripeptides of potential N-glycosylation sites are boxed. Predicted O-glycosylated residues are marked by grey squares. (PDF 5783 kb)

Additional file 6: Composition of S2 subsites of the active sites shown in the alignment of parts of amino acid sequences of *E. nipponicum* cathepsin L orthologs 6a, 6b and 6e with human cathepsin S and with plerocercoid growth factor of *Spirometra* tapeworm. The catalytic triad of the active site (C, H, N) is marked by triangles. Conserved motifs around active site residues are shaded in grey. Residues within the S2 subsite of the active site involved in determining the substrate specificity are shaded in black and indicated with numbers (papain numbering). *Abbreviations*: PGF, *Spirometra erinaceieuropaei* plerocercoid growth factor (Genbank: BAB62718.1); *HsCS*, human cathepsin S (Genbank: AAC37592). (PDF 1406 kb)

Additional file 7: Phylogram showing relationships between *E. nipponicum* cathepsins L and cathepsins L of other selected organisms. Unrooted maximum-likelihood tree of selected *E. nipponicum* cathepsins L inferred using the best-fit model (LG + C60 + G). Ultrafast bootstrap supports and posterior probabilities are shown. The leaf descriptions contain the sequence ID, genus and taxonomic placement. (PDF 4260 kb)

Additional file 8: Amino acid sequence alignment of *E. nipponicum* cathepsin B with cathepsins B of *S. mansoni*. Whole zymogens including a signal peptide were aligned. Numbers at the end of the lines show amino acid numbering of the mature parts of particular enzymes. Position of the pro-region cleavage site is marked by an arrow. The catalytic triad of the active site (C, H, N) is marked by triangles. Conserved motifs around active site residues are shaded in grey. Tripeptides of potential N-glycosylation sites are boxed. Predicted O-glycosylated residues are marked by grey squares. The occluding loop typical of cathepsins B is underlined. The 'haemoglobinase motif' ascribed to cathepsins B with an assumed function in haemoglobinolysis in blood-feeding helminths, is shaded in black (it is modified in *SmCB2* and *EnCB*). Residues within the S2 subsite of the active site involved in determining the substrate specificity are marked by a black dot and indicated with number (papain numbering). The alignment was made using sequences of *Schistosoma mansoni* cathepsins B1 (GenBank: CAD44624.1) and B2 (GenBank: XP_018651608.1). (PDF 1923 kb)

Abbreviations

AMC: 7-amino-4-methylcoumarin; BSA: Bovine serum albumin; Bz-Gly-His-Leu: Benzoyl-glycinyl-histidinyl-leucine; CA-074: N-[L-3-trans-propylcarbamoyloxirane-2-carbonyl]-Ile-Pro-OH; CBB: Coomassie Brilliant Blue; CPB: Citrate/phosphate buffer; DCG-04: fluorescent labelled probe, derivate of inhibitor E-64; DTT: Dithiothreitol; E-64: (L-trans-epoxysuccinyl–leucylamido [4-guanidino] butane); *EnCB*: *Eudiplozoon nipponicum* cathepsin B; *EnCL*: *Eudiplozoon nipponicum* cathepsin L; ESP: Excretory secretory products; FPLC: Fast protein liquid chromatography; FR: Substrate for serine and cysteine peptidases Z-Phe-Arg-AMC; GPR: Substrate for serine peptidases Z-Gly-Pro-Arg-AMC; His-tag: polyhistidine-tag (6xHis); ICL: Reversible inhibitor of cathepsin L(Arg-Lys-Leu-Leu-Trp-NH2); LR: Substrate for serine and cysteine peptidases Z-Leu-Arg-AMC; MES: (2-N-morpholino) ethanesulfonic acid; Ni-NTA: Nickel-nitrilotriacetic acid; PBS- TX100: Phosphate buffered saline containing 0.3% Triton X100; PBS: Phosphate buffered saline; PBS-T: Phosphate-buffered saline containing 0.05% Tween 20; PR: Substrate for serine peptidases Z-Pro-Arg-AMC; PVDF: Polyvinylidene difluoride; RACE-PCR: Rapid amplification of cDNA ends; RR: Substrate for cathepsin B Z-Arg-Arg-AMC; RT: Room temperature; RT-PCR: Reverse transcription-PCR; SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis; SSC: Saline-sodium citrate buffer

Acknowledgements

We would like to express our sincere gratitude to Professor John P. Dalton, Medical Biology Centre, School of Biological Sciences, Queen's University Belfast, for enabling LJ to make use of his institute's experimental facilities, for valuable advice offered to LJ during her scientific stay, and for critical comments on the final version of the manuscript. We wish to thank Professor Jan Konvalinka and Dr Tomáš Knedlík, Institute of Organic Chemistry and Biochemistry, ASCR, Praha, for providing us with iBody4. We would also like to thank Dr Conor R. Caffrey and Dr Doron C. Greenbaum, UCSD, Skaggs School of Pharmacy & Pharmaceutical Sciences, CA, USA, who kindly provided the Green BODIPY-DCG-04 probe. Our gratitude also extends to Mr Stanislav Sojka, Rybářství Třeboň, Plc., for arrangements concerning the collection of parasites. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, as part of the 'Projects of Large Research, Development, and Innovations Infrastructures' programme.

Funding

This work was supported by the project "Centre for Research of Pathogenicity and Virulence of Parasites" (No. CZ.02.1.01/0.0/0.0/16_019/0000759) funded by European Regional Development Fund and Ministry of Education, Youth and Sports of the Czech Republic. The research was also funded by Czech Science Foundation grants (P506/12/1258, GBP505/12/G112), Charles University institutional support (PROGRES Q43, UNCE 204017, and SVV 244-260432/2017), Masaryk University (MUNI/A/1362/2016), Ministry of Education, Youth, and Sports of the Czech Republic (InterBioMed LO1302), and by the IOCB institutional project (RVO 61388963).

Availability of data and materials

Transcripts from the assembled transcriptome of adult *E. nipponicum* worms have been deposited at DDBJ/EMBL/GenBank under the accession GFYM00000000. The version described in this paper is the first version, GFYM01000000.

Authors' contributions

LJ produced recombinant cathepsins L, performed their functional characterisation, *in situ* hybridisation and immunohistochemistry, participated in the bioinformatic annotation of enzyme structures, data interpretation, and in manuscript writing. HD produced the recombinant cathepsin B, performed its functional characterisation and immunohistochemistry, participated in the bioinformatic annotation of enzyme structure, data interpretation, and in manuscript writing. JD participated in the expression of recombinant cathepsins L and their biochemical characterisation. MK took part in designing the study and in data interpretation. LU was involved in designing of RNA probes and in the process of *in situ* hybridization. JV assembled the transcriptome, mined the sequence data and evaluated expression rates of the enzymes. VZ performed the phylogenetic analyses. LM conducted the work, prepared the study design, and participated in the purification of recombinant enzymes, bioinformatic annotation, interpretation of data, and in the production of the final draft of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The maintenance and care of experimental animals was carried out in accordance with European Directive 2010/63/EU, and according to Czech law (246/1992 and 359/2012) for biomedical research involving animals. Experiments were performed with a legal consent of the relevant institutional animal research ethics committees, namely the Professional Ethics Committee of the Faculty of Science, Charles University in Prague and the Branch for Research and Development of the Ministry of Education, Youth and Sports of the Czech Republic. The animal facility, its equipment, animal welfare and accompanying services, including maintenance of experimental animals, have been approved by the Branch of Animal Commodities of the Ministry of Agriculture of the Czech Republic (approval no. 13060/2014-MZE-17214).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Department of Parasitology, Faculty of Science, Charles University, Viničná 7, 12844 Prague 2, Czech Republic. [2]Medical Biology Centre, School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK. [3]Department of Zoology and Fisheries, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 16500 Prague 6, Czech Republic. [4]Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic. [5]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16000 Prague 6, Czech Republic. [6]Department of Parasitology, Faculty of Science, Charles University, Průmyslová 595, Vestec 25250, Czech Republic.

Received: 14 September 2017 Accepted: 22 January 2018
Published online: 06 March 2018

References

1. Denis A, Gabron C, Lambert A. Presence en France du deux parasites d'origine Est-asiatique: *Diplozoon nipponicum* Goto 1891 et *Bothriocephalus acheilognathi* Yamaguti, 1934 (Cestoda) chez *Cyprinus carpio* (Teleostei, Cyprinidae). Bull Fr Piscic. 1983;289:128–34.
2. Matějusová I, Koubková B, D' Amelio S, Cunningham CO. Genetic characterization of six species of diplozoids (Monogenea; Diplozoidae). Parasitology. 2001;123:465–74.
3. Fisheries and Aquaculture Department. About us - Fisheries and Aquaculture Department.. http://www.fao.org/fishery/about/en Updated 17 March 2017.
4. Kawatsu H. Studies on the anemia of fish-IX, hypochromic microcytic anemia of crucian carp caused by infestation with a trematode, *Diplozoon nipponicum*. Bull Japan Soc Sci fish. 1978;44:1315–9.

5. Smyth JD, Halton DW. The Monogenean - functional morphology. In: Smyth JD, Halton DW, editors. The Physiology of Trematodes. Cambridge: Cambridge University Press; 1983. p. 250–88.
6. Smyth JD, Halton DW. The Monogenean – physiology. In: Smyth JD, Halton DW, editors. The Physiology of Trematodes. Cambridge: Cambridge University Press; 1983. p. 289–320.
7. Hodová I, Matějusová I, Gelnar M. The surface topography of *Eudiplozoon nipponicum* (Monogenea) developmental stages parasitizing carp (*Cyprinus carpio* L.). Cent Eur J Biol. 2010;5:702–9.
8. Valigurová A, Hodová I, Sonnek R, Koubková B, Gelnar M. *Eudiplozoon nipponicum* in focus: monogenean exhibiting a highly specialized adaptation for ectoparasitic lifestyle. Parasitol Res. 2011;108:383–94.
9. Konstanzová V, Koubková B, Kašný M, Ilgová J, Dzika E, Gelnar M. Ultrastructure of the digestive tract of *Paradiplozoon homoion* (Monogenea). Parasitol Res. 2015;114:1485–94.
10. Sonenshine D. Biology of ticks. Volume 1. New York: Oxford University Press; 1991.
11. Sojka D, Franta Z, Horn M, Caffrey CR, Mareš M, Kopáček P. New insights into the machinery of blood digestion by ticks. Trends Parasitol. 2013;29:276–85.
12. Jedličková L, Dvořáková H, Kašný M, Ilgová J, Potěšil D, Zdráhal Z, et al. Major acid endopeptidases of the blood-feeding monogenean *Eudiplozoon nipponicum* (Heteronchoinea: Diplozoidae). Parasitology. 2016;143:494–506.
13. Sajid M, JH MK. Cysteine proteases of parasitic organisms. Mol Biochem Parasitol. 2002;120:1–21.
14. Williamson AL, Brindley PJ, Knox DP, Hotez PJ, Loukas A. Digestive proteases of blood-feeding nematodes. Trends Parasitol. 2003;19:417–23.
15. Williamson AL, Lecchi P, Turk BE, Choe Y, Hotez PJ, JH MK, et al. A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. J Biol Chem. 2004;279:35950–7.
16. Delcroix M, Sajid M, Caffrey CR, Lim KC, Dvořák J, Hsieh I, et al. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. J Biol Chem. 2006;281:39316–29.
17. Atkinson H, Babbitt P, Sajid M. The global cysteine peptidase landscape in parasites. Trends Parasitol. 2009;25:573–81.
18. Kašný M, Mikeš L, Hampl V, Dvořák J, Caffrey CR, Dalton JP, et al. Peptidases of trematodes. Adv Parasitol. 2009;69:205–97.
19. Robinson MW, Dalton JP. Cysteine proteases of pathogenic organisms (Vol. 712). Boston: Springer; 2011.
20. McVeigh P, Maule AG, Dalton JP, Robinson MW. *Fasciola hepatica* virulence-associated cysteine peptidases: a systems biology perspective. Microbes Infect. 2012;14:301–10.
21. NCBI Resource Coordinators NR. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016;44:D7–19.
22. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2016;44:D343–50.
23. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45:D158–69.
24. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. Methods Mol Biol. 2016;1374:23–54.
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000;28:235–42.
26. Mashima J, Kodama Y, Kosuge T, Fujisawa T, Katayama T, Nagasaki H, et al. DNA data bank of Japan (DDBJ) progress report. Nucleic Acids Res. 2016;44:D51–7.
27. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol. 2011;7:539.
28. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.
29. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
30. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40:W597–603.
31. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, KT-BG S, et al. Precision mapping of the human O-GalNAc glycoproteome through Simple Cell technology. EMBO J. 2013;32:1478–88.
32. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35:D61–5.
33. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2–10.
34. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004;20:3702–4.
35. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
37. Criscuolo A, Gribaldo S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:210.
38. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
39. Caffrey CR, Salter JP, Lucas KD, Khiem D, Hsieh I, Lim K-C, et al. SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. Mol Biochem Parasitol. 2002;121:49–61.
40. Dvořák J, Delcroix M, Rossi A, Vopálenský V, Pospíšek M, Šedinová M, et al. 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.
41. Wang B-L, Xu Y, Wu C-Q, Xu Y-M, Wang H-H. Cloning, expression, and refolding of a secretory protein ESAT-6 of *Mycobacterium tuberculosis*. Protein Expr Purif. 2005;39:184–8.
42. Greenbaum D, Baruch A, Hayrapetian L, Darula Z, Burlingame A, Medzihradszky KF, et al. Chemical approaches for functionally probing the proteome. Mol Cell Proteomics. 2002;1:60–8.
43. Sajid M, JH MK, Hansell E, Mathieu MA, Lucas KD, Hsieh I, et al. 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.
44. Quack T, Knobloch J, Beckmann S, Vicogne J, Dissous C, Grevelding CG. The formin-homology protein SmDia interacts with the Src Kinase SmTK and the GTPase SmRho1 in the gonads of *Schistosoma mansoni*. PLoS One. 2009;4:e6998.
45. ECH H, Donaldson ME, Saville BJ. Detection of antisense RNA transcripts by strand-specific RT-PCR. Methods Mol Biol. 2010;630:125–38.
46. Šácha P, Knedlík T, Schimer J, Tykvart J, Parolek J, Navrátil V, et al. iBodies: modular synthetic antibody mimetics based on hydrophilic polymers decorated with functional moieties. Angew Chem Int Ed Engl. 2016;55:2356–60.
47. Baig S, Damian RT, Peterson DS. A novel cathepsin B active site motif is shared by helminth bloodfeeders. Exp Parasitol. 2002;101:83–9.
48. Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP. Proteinases and associated genes of parasitic helminths. Adv Parasitol. 1999;43:161–266.
49. Dalton JP, Caffrey CR, Sajid M, Stack C, Donnelly S, Loukas A, et al. Proteases in trematode biology. In: Maule A, Marks N, editors. Parasitic flatworms: molecular biology, biochemistry, immunology and physiology. Cambridge: CABI; 2006. p. 348–68.
50. Smooker PM, Whisstock JC, Irving JA, Siyaguna S, Spithill TW, Pike RN. A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. Protein Sci. 2000;9:2567–72.
51. Stack CM, Caffrey CR, Donnelly SM, Seshaadri A, Lowther J, Tort JF, et al. Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. J Biol Chem. 2008;283:9896–908.
52. Norbury LJ, Hung A, Beckham S, Pike RN, Spithill TW, Craik CS, et al. Analysis of *Fasciola* cathepsin L5 by S2 subsite substitutions and determination of the P1-P4 specificity reveals an unusual preference. Biochimie. 2012;94:1119–27.
53. Pillay D, Boulangé AF, THT C. Expression, purification and characterisation of two variant cysteine peptidases from *Trypanosoma congolense* with active site substitutions. Protein Expr Purif. 2010;74:264–71.
54. Lecaille F, Choe Y, Brandt W, Li Z, Craik CS, Brömme D. Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. Biochemistry. 2002;41:8447–54.
55. Corvo I, Cancela M, Cappetta M, Pi-Denis N, Tort JF, Roche L. The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen. Mol Biochem Parasitol. 2009;167:41–7.

56. Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Brömme D, et al. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J Biol Chem. 2006;281:12824–32.
57. Brömmes D, Bonneaug PR, Lachance P, Storer AC. Engineering the S2 subsite specificity of human cathepsin S to a cathepsin L-and cathepsin B-like specificity. J Biol Chem. 1994;269:30238–42.
58. Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 2007;7:S4.
59. Dolečková K, Kašný M, Mikeš L, Cartwright J, Jedelský P, Schneider EL, et al. The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome *Trichobilharzia regenti*. Int J Parasitol. 2009;39:201–11.
60. McGrath ME. The lysosomal cysteine proteases. Annu Rev Biophys Biomol Struct. 1999;28:181–204.
61. Mikeš L, Man P. Purification and characterization of a saccharide-binding protein from penetration glands of *Diplostomum pseudospathaceum* - a bifunctional molecule with cysteine protease activity. Parasitology. 2003;127:69–77.
62. Brady CP, Brindley PJ, Dowd AJ, Dalton JP. *Schistosoma mansoni*: differential expression of cathepsins L1 and L2 suggests discrete biological functions for each enzyme. Exp Parasitol. 2000;94:75–83.
63. Dvořák J, Mashiyama ST, Sajid M, Braschi S, Delcroix M, Schneider EL, et al. SmCL3, a gastrodermal cysteine protease of the human blood fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 2009;3:e449.
64. Lowther J, Robinson MW, Donnelly SM, Xu W, Stack CM, Matthews JM, et al. The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease. PLoS Negl Trop Dis. 2009;3:e369.
65. Franta Z, Sojka D, Frantová H, Dvořák J, Horn M, Srba J, et al. IrCL1 - the haemoglobinolytic cathepsin L of the hard tick, *Ixodes ricinus*. Int J Parasitol. 2011;41:1253–62.
66. Brinker A, Weber E, Stoll D, Voigt J, Müller A, Sewald N, et al. Highly potent inhibitors of human cathepsin L identified by screening combinatorial pentapeptide amide collections. Eur J Biochem. 2000;267:5085–92.
67. Steverding D. The cathepsin B-selective inhibitors CA-074 and CA-074Me inactivate cathepsin L under reducing conditions. Open Enzym Inhib J. 2011;4:11–6.
68. Steverding D, Sexton DW, Wang X, Gehrke SS, Wagner GK, Caffrey CR. *Trypanosoma brucei*: chemical evidence that cathepsin L is essential for survival and a relevant drug target. Int J Parasitol. 2012;42:481–8.
69. Mindell JA. Lysosomal acidification mechanisms. Annu Rev Physiol. 2012;74:69–86.
70. Chappell CL, Dresden MH. *Schistosoma mansoni*: Proteinase activity of "hemoglobinase" from the digestive tract of adult worms. Exp Parasitol. 1986;61:160–7.
71. Halton DW. Nutritional adaptations to parasitism within the Platyhelminthes. Int J Parasitol. 1997;27:693–704.
72. Dalton JP, Skelly P, Halton DW. Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. Can J Zool. 2004;82:211–32.
73. Robinson MW, Corvo I, Jones PM, George AM, Padula MP, To J, et al. Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*. PLoS Negl Trop Dis. 2011;5:e1012.
74. Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell. 2010;143:1018–29.
75. Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, Suzuki K, et al. Antisense transcripts with rice full-length cDNAs. Genome Biol. 2003;5:R5.
76. Steigele S, Nieselt K. Open reading frames provide a rich pool of potential natural antisense transcripts in fungal genomes. Nucleic Acids Res. 2005;33:5034–44.
77. Henke C, Strissel PL, Schubert M-T, Mitchell M, Stolt CC, Faschingbauer F, et al. Selective expression of sense and antisense transcripts of the sushi-ichi-related retrotransposon-derived family during mouse placentogenesis. Retrovirology. 2015;12:9.
78. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet. 2013;14:880–93.
79. Illy C, Quraishi O, Wang J, Purisima E, Vernet T, Mort JS. Role of the occluding loop in cathepsin B activity. J Biol Chem. 1997;272:1197–202.
80. Michel A, Ghoneim H, Resto M, Klinkert MQ, Kunz W. Sequence, characterization and localization of a cysteine proteinase cathepsin L in *Schistosoma mansoni*. Mol Biochem Parasitol. 1995;73:7–18.
81. Britton C, Murray L. A cathepsin L protease essential for *Caenorhabditis elegans* embryogenesis is functionally conserved in parasitic nematodes. Mol Biochem Parasitol. 2002;122:21–33.
82. Pinlaor P, Kaewpitoon N, Laha T, Sripa B, Kaewkes S, Morales ME, et al. Cathepsin F cysteine protease of the human liver fluke. PLoS Negl Trop Dis. 2009;3:e398.
83. Ogawa K. Diseases of cultured marine fishes caused by Platyhelminthes (Monogenea, Digenea, Cestoda). Parasitology. 2015;142:178–95.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

Manuscript in preparation No. 3

Cathepsins D of *Eudiplozoon nipponicum*
Dvořáková H. and Mikeš L

In various blood-feeding helminths, cathepsin D-like enzymes have been proved to be among the most crucial enzymes in the processing of host proteins. In this study, we aimed to reveal the localisation and possible function of cathepsin D expressed by the monogenean *Eudiplozoon nipponicum.* Mining transcriptomic data of the adult *E. nipponicum*, we found four different cathepsin D transcripts. Moreover, a pulldown technique employing iBody3 (synthetic polymer conjugate containing the class-specific aspartic peptidase inhibitor pepstatin A) was used for a direct proof of cathepsins D in soluble protein extracts from adult worms. One type of cathepsin D (EnCD2) was selected for heterologous expression and specific antibody was produced. EnCD2was localised in haematin cells of the worm's digestive tract. We conclude from our findings that cathepsin D is involved in proteolysis of haemoglobin in the gut of the parasite. This study completes our analysis of the endopeptidases involved in the haemoglobinolytic network inside the gut of *E. nipponicum.*

My contribution was searching transcriptome database for cathepsin D transcripts, bioinformatic sequence analyses, pulldown technique with iBody3, immunolocalisation and preparation of the manuscript.

Manuscript in preparation No. 3

Cathepsins D of *Eudiplozoon nipponicum*

Dvořáková H. and Mikeš L

Cathepsins D of *Eudiplozoon nipponicum*

Dvořáková H. and Mikeš L.

Abstract

Transcriptome of the adult blood-feeding monogenean *Eudiplozoon nipponicum* includes transcripts encoding four different cathepsins D (EnCD1-4). All but one of them possess a typical catalytic dyad of the aspartic peptidase active site (Asp/Asp); an exception is represented by EnCD1, in which one aspartate is replaced by serine. EnCD2 was purified from parasite protein extracts by a pull-down method employing pepstatin A – based affinity chromatography. The EnCD2-coding gene was cloned, and a recombinant form of the enzyme was produced in *Escherichia coli*. Using mouse anti-EnCD2-specific antibody, the enzyme was immunolocalized in haematin cells of the digestive tract in adult worms. This finding strongly boosts the suggestion that aspartic peptidases play a key role in blood digestion in haematophagous monogeneans, supported by their ability to efficiently degrade haemoglobin in vitro. In context of previously characterized cysteine cathepsins from parasite's digestive cells, this study completes our view on the endopeptidase part of the haemoglobinolytic network in the gut of *E. nipponicum*.

Introduction

Cathepsin D (CD; Enzyme Commission number EC 3.4.23.5) belongs to the MEROPS clan AA, family A1 aspartic endopeptidases (known as pepsin-like enzymes) which includes pepsin, renin, chymosin, cathepsin E, HIV1 peptidase, and several other enzymes (Fusek et al., 2013; Rawlings et al., 2018). Peptidases of the family A1 are characterized by the presence of two Asp residues that are required for their catalytic activity. Cathepsins D are expressed in a diverse range of organisms, cells, and tissues and are located predominantly in lysosomes. Parasite CDs often operate within multi-enzyme digestive cascades together with cysteine peptidases. Examples can be found, e.g., among protists including malarial parasites, helminths including trematodes and nematodes, and ticks, where they usually ensure the initial or early cleavage of ingested blood proteins (Brindley et al., 2001; Brinkworth et al., 2001; Goldberg et al., 1991; Kang et al., 2019; Sojka et al., 2012; 2016; Williamson et al., 2002). It was proposed in our previous study (Jedličková et al., 2016) that similar peptidase tools seem to be employed inside the digestive tract of blood-feeding monogeneans. Aspartic cathepsin D-like activity was significant in both the soluble protein extracts and excretory-secretory (E/S) products of adult *Eudiplozoon. nipponicum* (Heteronchoinea: Diplozoidae), an invasive monogenean species parasitizing

common carp. The proteolysis of haemoglobin in these samples could be completely inhibited by a mixture of cysteine and aspartic peptidase inhibitors, and diminished by individual inhibitors of cathepsins L, B and D. Furthermore, we characterized in detail two gut-associated haemoglobinolytic cathepsins L (EnCL1 and EnCL3) of the parasite (Jedličková et al., 2018) and three types of peptidase inhibitors which are likely involved in regulation of blood processing and digestion and host-parasite interactions (Ilgová et al., 2017; Jedličková et al., 2019; Roudnický et al., 2018). In the assembled transcriptome of *E. nipponicum* adults (Vorel et al., 2021) we found four transcripts of genes coding for different cathepsin D-like aspartic peptidases termed EnCD1 – EnCD4. While three of the enzyme's amino acid sequences possess both Asp residues of the active site, in another paralog, one of the aspartic acid is substituted for serine, likely making the peptidase proteolytically inactive. In this study we aimed to verify and annotate the primary structures of *E. nipponicum* cathepsins D and confirm and localize in worm's body the most abundant enzyme on protein level employing pull-down affinity purification, mass spectrometry, and immunohistochemistry.

Materials and methods

Sample preparation

Adult *E. nipponicum* were collected from *Cyprinus carpio* freshly slaughtered in a commercial facility of the fishery Rybářství Třeboň Plc., Czech Republic. Soluble worm protein extract (solPE), parasite RNA, and first-strand cDNA were prepared as described previously (Jedličková et al., 2016). Preparation of samples for immunohistochemistry was carried out according to a protocol of Jedličková et al. (2018).

Ethics statement

The animals were treated in concordance with the legislation of the Czech Republic (246/1992 and 359/2012), the European Directive (2010/63/EU), and ARRIVE guidelines (arriveguidelines.org). Our experimental procedures conform with the legal consent of the Professional Ethics Committee of the Faculty of Science, Charles University, Prague, Czech Republic and the Branch for Research and Development of the Ministry of Education, Youth and Sports of the Czech Republic. The animal facility, its equipment, animal welfare, and accompanying services including the maintenance of experimental animals, have been approved by the Section of Animal Commodities of the Ministry of Agriculture of the Czech Republic (approval no. 13060/2014-MZE-17214).

Sequence analyses of cathepsin D-like peptidases of E. nipponicum

Completely assembled transcripts from adult *E. nipponicum* transcriptome deposited in the NCBI GenBank database under accession number GFYM00000000 (sequences and annotation also available on https://github.com/jirivorel/Eudiplozoon-nipponicum-transcriptome-secretome) was used for searching of transcripts encoding cathepsin D-like peptidases. BLASTp and BLASTn algorithms (E-value 1e-5) were employed to carry out a similarity search using helminth cathepsins D of *Schistosoma mansoni* [XP_018651076.1], *Clonorchis sinensis* [GAA56870.1], *Echinococcus granulosus* [CDS16987.1] and *Ancylostoma ceylanicum* [AAO22152.1] as query sequences (GenBank accession numbers in square brackets). Subsequently, PCR amplifications were performed with the cDNA of adult *E. nipponicum* and gene specific primers (Table 1) designed on the basis of the cathepsin D-coding sequences identified in *E. nipponicum* transcriptome. Amplified products were cloned into pGEM®-T Vector System (Promega) and *E. coli* strain XL-1 blue cells. Clones were sequenced with M13 primers (DNA Sequencing Laboratory, Faculty of Science, Charles University). The amino acid sequences encoding EnCDs were aligned with human pro-cathepsin D [GenBank: AAA51922.1] using Clustal Omega tool at https://www.ebi.ac.uk/Tools/msa/clustalo/ (Sievers et al., 2014). The protein parameters were predicted in silico using proteomic tools ProtParam at https://web.expasy.org/protparam/ (Gasteiger et al., 2005), Signal P at http://www.cbs.dtu.dk/services/SignalP-4.1/ (Nielsen, 2017), and NetNGlyc at http://www.cbs.dtu.dk/services/NetNGlyc/ (Gupta et al., 2004).

Table 1. Primers used for amplification of *Eudiplozoon nipponicum* cathepsin D genes

Primer name	Primer sequence
EnCD1F	5′ ATG TAT TCT CGA AAT ACA AGG TTG 3′
EnCD1R	5′ TCA TCT CGC AAT CGC AAT TC 3′
EnCD2F	5′ AAA CCC GGT AAA TTG ATA AGC ATT 3′
EnCD2R	5′ TTA ATT TCC TTT TGG ATG ACG TG 3′
EnCD3F	5′ TCA ATA ATA AGA GTT CCA CTG CGT 3′
EnCD3R	5′ TTA ATT TCT AAT AGC CGG TT 3′
EnCD4F	5′ ATG TGC CAT TAG GTC GGC AAT CT 3′
EnCD4R	5′ TCA GTT CTT AGC GAG AGC AAA 3′

Purification of EnCD from soluble worm extract of E. nipponicum

A synthetic polymer conjugate containing biotin and the class-specific aspartic peptidase inhibitor pepstatin A (iBody3) was used for pull-down affinity purification of EnCDs (substance kindly provided by Prof. Jan Konvalinka and Dr. Tomáš Knedlík, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic). The experiment was performed according to a modified protocol of Šácha et al. (2016). Briefly, 200 nM iBody3 was bound to 30 µl Streptavidin Mag Sepharose (GE Healthcare) in 100 µl binding buffer (20 mM Tris-HCl, 150 mM NaCl, 0.1 % Tween 20, pH 7.8) at RT for 1h. A negative control was based on iBody6 lacking the targeting ligand (pepstatin A). Biotin binding sites were blocked for 10 min with 2 mM biotin in binding buffer at RT. After 3x washing with 100 mM sodium acetate, 300 mM NaCl, 0.1 % Tween 20, pH 4 (washing buffer), the resin with bound conjugate was mixed with 100 µl worm protein extract (1.2 µg/µl) in washing buffer and incubated for 30 min at RT. The resin was washed with acetate buffer and affinity-bound EnCD was eluted by adding 25 µl reducing SDS-PAGE sample buffer and heating at 98 °C for 10 min. The samples were run on SDS-PAGE. Bands were manually excised and analysed by mass spectrometry as described previously (Jedličková et al., 2016). MS/MS ion searches were done against an in-house database containing cathepsin D sequences found in the transcriptome of *E. nipponicum.*

Expression of EnCD2 proenzyme and production of anti-EnCD2 antibodies

The *E.coli* bacterial pET expression system was used for heterologous expression of EnCD2 as described previously for *E. nipponicum* cysteine cathepsins (Jedličková et al., 2018). N-terminal 6x His-tagged fusion EnCD2 was prepared using the pET28a+ vector (Novagen) and gene-specific primers (Table 2). The recombinant protein was purified by Immobilized Metal Affinity Chromatography (HisTrap FF Crude column, GE Healthcare). Two ICR mice were immunized by recombinant EnCD2 according to a protocol of Jedličková et al. (2018). Specificity of the anti-EnCD2 antibody was verified using PVDF membrane immunoblotting with soluble worm extract separated by SDS-PAGE in a 12% gel (20 µg/well). After blocking with 5% non-fat milk (Bio-Rad), the membrane was incubated for 1 h with mouse anti-EnCD2 or control pre-immune sera (diluted 1:100) and then for 1 h with horse radish peroxidase-labelled goat anti-mouse IgG (Sigma-Aldrich, diluted 1:5000). All washing steps were performed with 50 mM PBS pH 7.2 containing 0.05% Tween-20. Finally, the membrane was developed by the Opti-4CN Substrate Kit (Bio-Rad).

Table 2. Primers used for recombinant expression of EnCD2 from cDNA.

Primer name	Primer sequence
EnCD2E.coliF	5´ GAT Agc tag cAA ACC CGG TAA ATT GAT AAG C 3´
EnCD2E.coliR	5´ ATA gag ctc TTA ATT TCC TTT TGG ATG ACG TG 3´

Restriction sites for NHE-I/SAC-I are in lower case letters

Localisation of EnCD2 by indirect immunofluorescence histochemistry

The sections of worms were prepared, blocked and immuno-labelled according to a protocol described by Jedličková et al. (2018). Briefly, after an overnight incubation with anti-EnCD2 or pre-immune serum (diluted 1:50), the slides were placed in a dark chamber for 1 h at RT with Goat Anti-Mouse IgG H&L-Alexa Fluor® 568 diluted 1:200 (Abcam). Sections were mounted in Vectashield with DAPI (Vector Laboratories). Immunofluorescence was observed and photographed under an Olympus BX51 microscope equipped with an Olympus DP72 camera.

Results

Bioinformatic analysis of cathepsin D-like transcripts

Data mining of *E. nipponicum* transcriptome identified four different cathepsin D transcripts tagged as EnCD1 – EnCD4 (see transcript IDs in Table 3). The presence of all paralogs in cDNA of *E. nipponicum* were verified by PCR amplification and Sanger sequencing. The predicted/computed basic parameters of EnCDs including mRNA length, protein length (zymogen), signal sequence prediction, theoretical pI and molecular weight, and N-glycosylation sites are summarised in Table 3. EnCD1, EnCD2 and EnCD4 have two potential N-glycosylation sites and EnCD3 has only one (all occur in the mature enzyme). The N-terminal region of EnCD2 and EnCD3 contains a signal peptide (21 amino acids long); no signal peptides were identified in the sequences of EnCD1 and EnCD4. EnCD1.

An amino acid sequence alignment of the four EnCDs with human pro-cathepsin D is displayed in Fig.1A. Except for EnCD1, the mature EnCDs have two catalytic Asp residues in the evolutionary conserved “DTG” motif and do not incorporate a processing loop known from mammalian lysosomal cathepsins D. Interestingly, EnCD1 displays a substitution of one active site aspartate for serine and contains a modification in the conserved motif. The amino acid sequence identity between the EnCDs is 25.5 – 40 % (Fig.1B).

Table 3. Basic parameters of the identified *E. nipponicum* cathepsin D paralogs

Name	transcript ID	base/AA	SP	pI proenzyme/ mature enzyme	Mw proenzyme/ mature (kDa)	N-gly sites
EnCD1	E_nip_trans_02749|m.9764	1252/416	-	7.11/6.01	47.7/37.9	2
EnCD2	E_nip_trans_02773|m.9930	1197/377	21	*5.88/5.01	*41.9/37.5	2
EnCD3	E_nip_trans_50085|m.338414	1356/430	21	*8.79/7.67	*47.9/43.3	1
EnCD4	E_nip_trans_14169|m.106900	1239/412	-	7.56/6.41	45.7/36.7	2

Abbreviations: AA - amino acids, Sp - signal peptide, * proenzyme without signal peptide, N-gly sites - number of potential N-glycosylation sites

Purification of EnCDs from soluble worm extracts of E. nipponicum

One protein band (~38 kDa) was observed in SDS-PAGE gel after pepstatin A inhibitor-based affinity purification using worm extracts and iBody3. No protein band occurred in the negative control (Fig.2A). Ten peptide sequences identified by MS analysis matched to the sequence of EnCD2 only (Fig.2B).

Anti-EnCD2 antibody specificity and immunohistochemical localisation of EnCD2

The specificity of mouse anti-EnCD2 sera was verified on immunoblot with soluble worm extracts. The antibodies reacted specifically with a band appearing at ~38 kDa. The control serum did not exhibit any reaction (Fig.3A). The reaction of the antibodies with histological sections of the worms was obvious inside digestive haematin cells of the gut. No reaction was observed outside the digestive system (Fig.3B). The control serum displayed no reaction at all (Fig.3B).

Discussion

The role of cathepsin D in parasite digestive networks has been relatively well characterized in blood-feeding nematodes (Williamson et al., 2004, 2003), trematodes (Caffrey et al., 2004; Delcroix et al., 2006), and ticks (Hánová et al., 2018; Horn et al., 2009; Sojka et al., 2012) However, to our knowledge, there are no reports of aspartic peptidases from any species of monogeneans, except for our previous works on *E. nipponicum* (Jedličková et al., 2016; Roudnicky et al., 2020; Vorel et al., 2021) and a few genomic projects dealing with other monogenean species (Coghlan et al., 2018; Hahn et al., 2014; Konczal et al., 2020).

Here we got back to aspartic peptidases of *E. nipponicum* intending to support our previous findings and assumptions about their biological function in the parasite. We identified four transcripts coding for cathepsins D in the transcriptome of the adult worms. EnCD1 has a second Asp residue of the active site substituted by Ser. Aspartic peptidases are

characterized by the presence of two Asp residues that are required for their catalytic activity. The side chains of those Asp residues hold activated water molecule in a proper position for attacking the scissile bond of the polypeptide substrate (Rawlings and Salvesen, 2013). However, an aspartic peptidase isolated from eggs of the hard tick *Boophilus microplus* (BYC, *Boophilus* Yolk Cathepsin) was able to process haemoglobin and vitellin despite the lack of a second aspartic catalytic residue (Logullo et al., 1998). In addition to the Asp-to-Ser substitution in the active site, other substitutions occur in both aspartate-threonine-glycine ("DTG") active site motifs of EnCD1, which are highly conserved in members of the A1 family of pepsin-like peptidases, where each catalytic Asp residue occurs in a very similar motif within the sequence Xaa-Xaa-Asp-Xbb-Gly-Xbb (Xaa is a hydrophobic residue and Xbb is either Thr or Ser) (Rawlings and Salvesen, 2013). The substitutions in EnCD1 are Gly-to-Ser in the first motif (closer to the N-terminus) and Gly-to-Ala in the second one; thus, their mutated sequences are Asp-Thr-Ser (DTS) and Ser-Ser-Ala (SSA). Also, the usual Thr/Ser residue at the C-terminus of the second "DTG" motif (SSA in EnCD1) is replaced by Leu (SSAL) in EnCD1. Additionally, EnCD1 has aliphatic Ile^{78} instead of the bulky aromatic Tyr^{78} (mature human cathepsin D numbering) in the N-terminal domain in β-hairpin loop (termed "Y flap"). In pepsin-like peptidases, this loop partially occludes the active site and carries residues important for specificity, which can interact with the substrate and are a part of the S1 subsite of the peptidase (Rawlings and Salvesen, 2013). The flap is a highly flexible part of the enzyme which plays a critical role in substrate binding by displaying a scissor-like motion necessary for substrate uptake. It has been suggested that rotation of Tyr^{78} side-chain plays an important role in the flap dynamics (Bhakat and Söderhjelm, 2020). So, the non-conservative replacement of Tyr for Ile in EnCD1 could alter the preference for protein substrates or cause the decrease or loss of peptidolytic activity of EnCD1. Although the impact of the abovementioned mutations on the functionality and possible role of EnCD1 has not yet been studied, is seems likely that the vast changes in the second active site motif led to the loss of peptidolytic activity.

All EnCDs possess potential N-glycosylation sites in their mature regions (holoenzyme parts). No signal sequences were identified in the sequences of two paralogs, EnCD1 and EnCD4.

Since proteins without signal sequence are probably not to be exposed to the N-glycosylation pathway, they may not be in vivo glycosylated even though they contain potential motifs. Most likely, these two paralogs are intracellular.

A major difference among cathepsins D is a structural part involved in the processing of a single chain form to a two-chain form of the enzyme. Processing of vertebrate cathepsins D involves usually two steps of the maturation process: in the first step the N-terminal pro-peptide is removed to yield an active intermediate single-chain molecule, the intermediate single-chain species is then cleaved into a mature two-chain enzyme (Gieselmann et al., 1985). The processing site at which the intermediate single-chain form is converted into the mature two-chain form was determined as the β-hairpin loop (Kobayashl et al., 1992). EnCDs, similarly to CDs of other invertebrates such as the blood fluke *Schistosoma japonicum* (Becker et al., 1995) or the hard tick *Ixodes ricinus* (Sojka et al., 2012), do not incorporate a processing loop (β-hairpin) characteristic of mammalian lysosomal cathepsins D. Cathepsins D that lack the β-hairpin are probably processed only in one step resulting in single-chain molecules, as described, e,g., for cathepsin D of *I. ricinus* (Sojka et al., 2012).

Interestingly, EnCD3 contained a C-terminal extension that was found in some other family A1 peptidases of helminths (Becker et al., 1995; Suttiprapa et al., 2009) but occurs neither in human pepsin and cathepsin D nor in other EnCD paralogs. The function of this extension is unknown.

The protein band of ~38 kDa affinity-isolated from the homogenate of adult *E. niponicum* was confirmed by mass spectrometry as EnCD2, providing a direct proof of cathepsin D presence in the parasite. Regarding its MW, it was present as a processed mature enzyme in worm extracts. It seems from this result that EnCD2 is the most abundant (on a protein level) of four cathepsins D found in *E. nipponicum* transcriptome or at least of those three with the complete active site, which are likely bind to pepstatin.

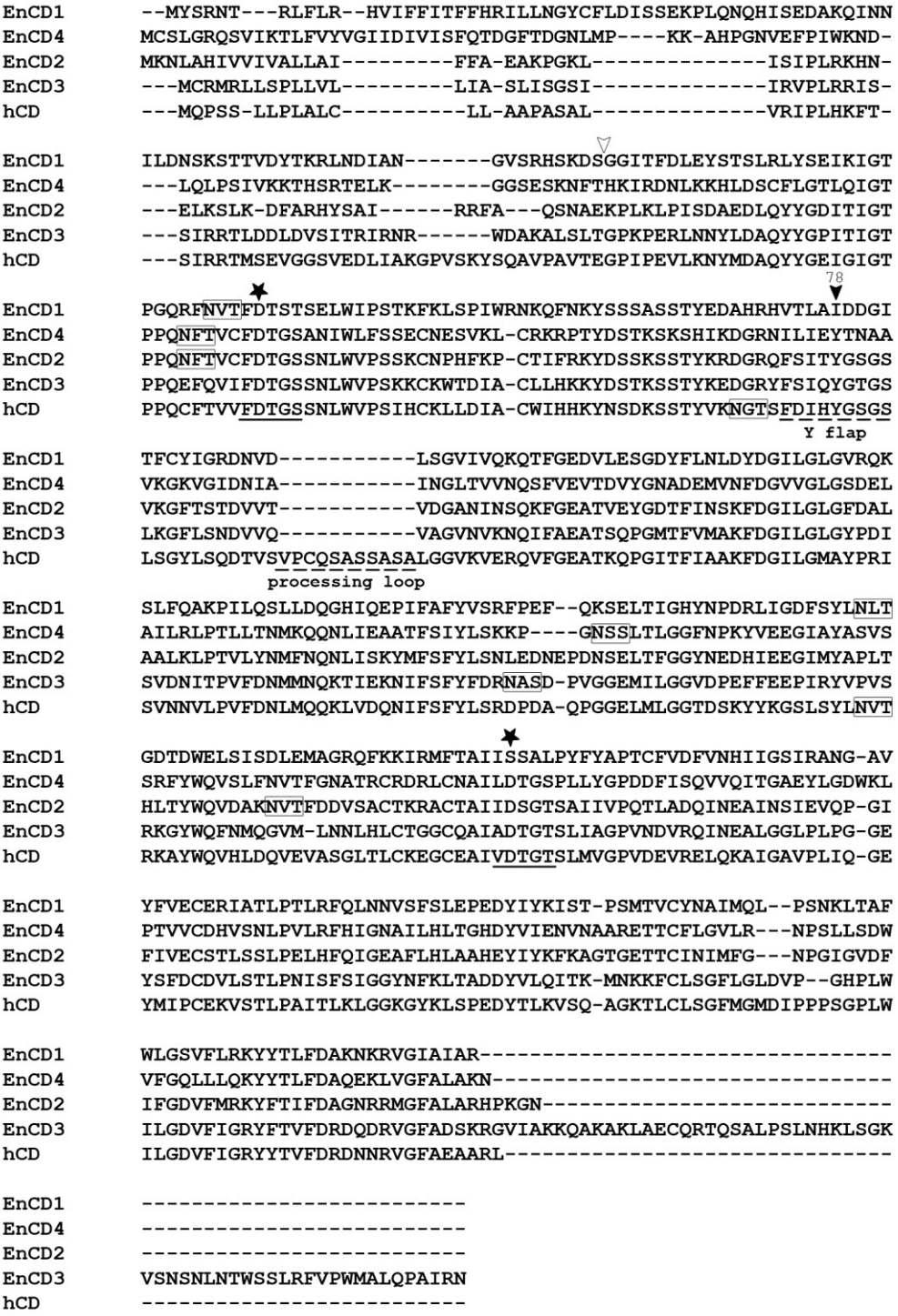
According to pioneering ultrastructural and histochemical studies, the blood digestion in heteronchoinean monogeneans has been considered intracellular, taking place in the acidic lysosomes within specialized types of gut cells (Jennings, 1959; Konstanzová et al., 2015; Tinsley, 1973). However, subsequent detection of high cysteine peptidase activities and finding of digestive cathepsins L in E/S products of *E. nipponicum* (Jedličková et al., 2016; Vorel et al., 2021), and the localisation of two gastrodermal cathepsins L in the lumen of parasite's gut (Jedličková et al., 2018) created support for the assumption that the digestive process takes place, at least in part, extracellularly, outside the gastrodermis. On the other hand, EnCD2 was immunolocalized only inside the gastrodermis in this study. Previously, significant cathepsin D-like activity was observed in worm extracts, whereas only a minute activity was usually recorded in the samples of E/S products (Jedličková et al., 2016). This may imply that EnCD2 would function in a later, intracellular phase of haemoglobin

digestion rather than being responsible for the initial scission typical for most cathepsins D of blood-feeding helminths. Localization of the enzyme on ultrastructural level might help to resolve this issue.

In conclusion, the annotation of amino acid sequences of EnCDs revealed that three of four paralogs are typical pepsin-like enzymes sharing active site motifs that are conserved in the cathepsin D superfamily of peptidases. One paralog possesses significant non-conservative mutations in the active site motifs which are not likely to retain their function in terms of catalytic activity. The most abundant paralog EnCD2 was isolated from worm protein extracts and its identity was confirmed by mass spectrometry. Its localization in the cells of parasite's gastrodermis, together with previous results demonstrating the ability of *E. nipponicum* aspartic peptidases to degrade haemoglobin in vitro, supports its role in blood digestion. Unfortunately, the recombinant EnCD2 was produced in *E. coli* as an insoluble protein which could be solubilized in high concentrations of chaotropic denaturating solvents and our attempts to transfer it to non-denaturating buffers for refolding under diverse conditions failed. Therefore, detailed biochemical characterisation of the enzyme was not performed. Recombinant expression in other system(s), functional characterisation and localisation of other CDs are therefore desirable to fully elucidate the roles of these peptidases in monogeneans.

Figures

A



B

	EnCD1	EnCD4	EnCD2	EnCD3	hCD
EnCD1	100.00	24.69	33.07	27.27	29.97
EnCD4	24.69	100.00	38.24	31.43	30.05
EnCD2	33.07	38.24	100.00	39.79	36.08
EnCD3	27.27	31.43	39.79	100.00	52.30
hCD	29.97	30.05	36.08	52.30	100.00

Fig. 1. Sequence alignment and similarities of *Eudiplozoon nipponicum* and human cathepsins D

A. Lysosomal human pro-cathepsin D [GenBank: AAA51922.1] was aligned with pro-EnCD1, pro-EnCD2, pro-EnCD3 and pro-EnCD4. Position of the pro-region cleavage site is marked by white arrow. The catalytic aspartic acid residues are marked by asterisks. Conserved motifs around active site residues are underline. Tripeptides of potential N-glycosylation sites are horizontal boxed. Y flap and processing loop are underlined by a dashed line. Tyr^{78} is marked by black arrow (mature human cathepsin D numbering). EnCD, *E*. *nipponicumm* cathepsin D; hCD, human cathepsin D.

B. Amino acid identities between EnCD pro-enzymes and human pro-cathepsin D created by Clustal2.1 (expressed in %).

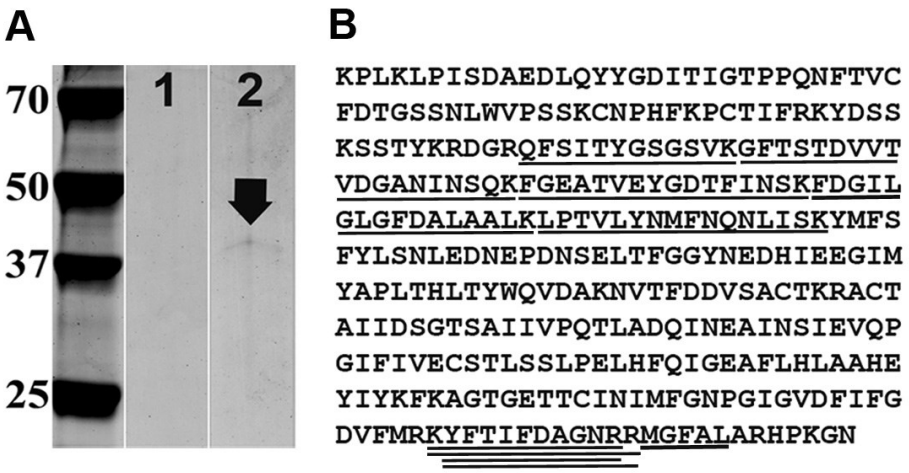


Fig. 2. Affinity isolation and MS identification of *Eudiplozoon nipponicum* cathepsin D2 from soluble worm extracts of *E. nipponicum*

A. SDS-PAGE of worm extracts incubated with **1.** iBody 6 lacking the targeting ligand (negative control); **2.** iBody 3 containing pepstatin A (affinity probe). The ~38 kDa band indicated by arrow was excised from the gel and prepared for mass spectrometry analysis.

B. Identification of EnCD2 by mass spectrometry. All 10 tryptic peptides identified in the band of affinity-isolated protein matched to the sequence of EnCD2. Each line corresponds to a single identified peptide; the sequences of some peptides partially overlap with the others.

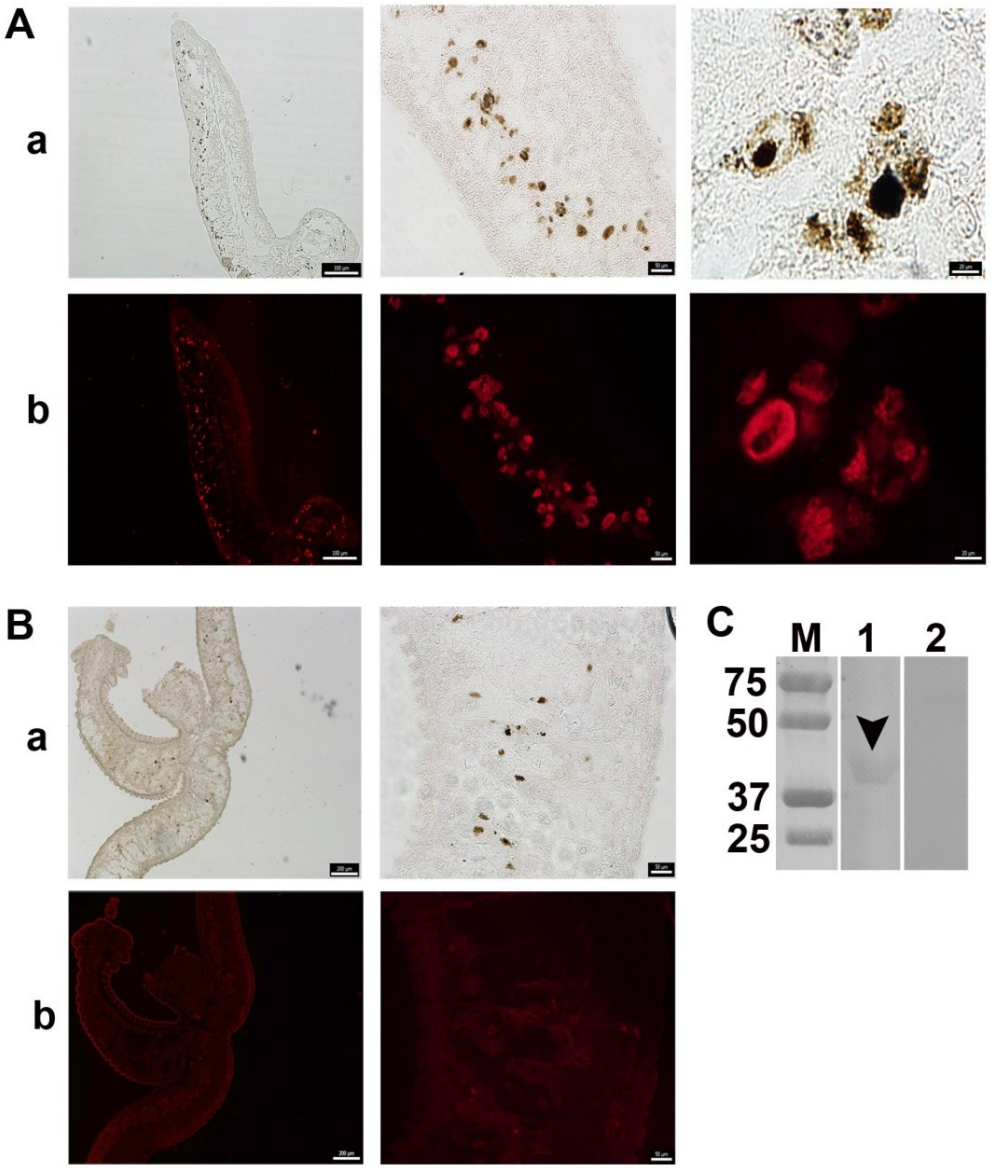


Fig. 3. Immunolocalisation of *Eudiplozoon nipponicum* cathepsin D2 on histological sections of adult worms and on Western blot of worm protein extract.

A. Longitudinal section of *E. nipponicum* was labelled with specific anti-EnCD2 antibody (scale bars 100 µm, 50 µm, 20 µm). B. Negative control, pre-immune mouse serum (scale bars 200 µm, 50 µm). a) Bright field view. b) Reaction with sera. Reaction of antibody raised against EnCD2 takes place inside the haematin digestive cells of the gastrodermis. No reaction with the control sera was detected.

C) Reaction of anti-EnCD2 antibody with proteins on blots of *E. nipponicum* soluble worm extracts. Lane 1, anti-EnCD2 antibody (arrow points to the enzyme band detected by antibody); Lane 2, pre-immune mouse sera.

References

Becker, M.M., Harrop, S.A., Dalton, J.P., Kalinna, B.H., Mcmanus, D.P., Brindley, P.J., 1995. Cloning and Characterization of the Schistosoma japonicum Aspartic Proteinase Involved in Hemoglobin Degradation, Journal of Biological Chemistry. https://doi.org/10.1074/jbc.270.41.24496

Bhakat, S., Söderhjelm, P., 2020. Flap dynamics in pepsin-like aspartic proteases: A computational perspective using Plasmepsin-II and BACE-1 as model systems. bioRxiv. https://doi.org/10.1101/2020.04.27.062539

Brindley, P.J., Smyth, Danielle Julianna, Fairlie, D., Milburn, P., Kalinna, B.H., Wong, J.Y.M., Bogitsh, B.J., King, L.T., Smyth, Danielle J, Verity, C.K., Abbenante, G., Brinkworth, R.I., Fairlie, D.P., Smythe, M.L., Milburn, P.J., Bielefeldt-Ohmann, H., Zheng, Y., Mcmanus, D.P., 2001. Proteolysis of human hemoglobin by schistosome cathepsin D. Artic. Mol. Biochem. Parasitol. 112, 103–112. https://doi.org/10.1016/S0166-6851(00)00351-0

Brinkworth, R.I., Prociv, P., Loukas, A., Brindley, P.J., 2001. Hemoglobin-degrading, aspartic proteases of blood-feeding parasites. Substrate specificity revealed by homology models. J. Biol. Chem. 276, 38844–38851. https://doi.org/10.1074/jbc.M101934200

Caffrey, C.R., McKerrow, J.H., Salter, J.P., Sajid, M., 2004. Blood "n" guts: An update on schistosome digestive peptidases. Trends Parasitol. 20, 241–248. https://doi.org/10.1016/j.pt.2004.03.004

Coghlan, A., Tyagi, R., Cotton, J.A., Holroyd, N., Rosa, B.A., Tsai, I.J., Laetsch, D.R., Beech, R.N., Day, T.A., Hallsworth-Pepin, K., Ke, H.M., Kuo, T.H., Lee, T.J., Martin, J., Maizels, R.M., Mutowo, P., Ozersky, P., Parkinson, J., Reid, A.J., Rawlings, N.D., Ribeiro, D.M., Swapna, L.S., Stanley, E., Taylor, D.W., Wheeler, N.J., Zamanian, M., Zhang, X., Allan, F., Allen, J.E., Asano, K., Babayan, S.A., Bah, G., Beasley, H., Bennett, H.M., Bisset, S.A., Castillo, E., Cook, J., Cooper, P.J., Cruz-Bustos, T., Cuéllar, C., Devaney, E., Doyle, S.R., Eberhard, M.L., Emery, A., Eom, K.S., Gilleard, J.S., Gordon, D., Harcus, Y., Harsha, B., Hawdon, J.M., Hill, D.E., Hodgkinson, J., Horák, P., Howe, K.L., Huckvale, T., Kalbe, M., Kaur, G., Kikuchi, T., Koutsovoulos, G., Kumar, S., Leach, A.R., Lomax, J., Makepeace, B., Matthews, J.B., Muro, A., O'Boyle, N.M., Olson, P.D., Osuna, A., Partono, F., Pfarr, K., Rinaldi, G., Foronda, P., Rollinson, D., Samblas, M.G., Sato, H., Schnyder, M., Scholz, T., Shafie, M., Tanya, V.N., Toledo, R., Tracey, A., Urban, J.F., Wang, L.C., Zarlenga, D., Blaxter, M.L., Mitreva, M., Berriman, M., 2018. Comparative genomics of the major parasitic worms. Nat. Genet. 2018 511 51, 163–174. https://doi.org/10.1038/s41588-018-0262-1

Delcroix, M., Sajid, M., Caffrey, C.R., Lim, K.C., Dvořák, J., Hsieh, I., Bahgat, M., Dissous, C., McKerrow, J.H., 2006. A Multienzyme Network Functions in Intestinal Protein Digestion by a Platyhelminth Parasite. J. Biol. Chem. 281, 39316–39329. https://doi.org/10.1074/jbc.M607128200

Fusek, M., Mares, M., Vetvička, V., 2013. Cathepsin D, in: Rawling, N., Salvesen, G. (Eds.), Handbook of Proteolytic Enzymes. Academic Press, Amsterdam, pp. 54–63.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server,

in: Walker, J.M. (Ed.), The Proteomics Protocols Handbook. Humana Press, pp. 571–607.

Gieselmann, V., Hasilik, A., Von Figura, K., 1985. Processing of Human Cathepsin D in Lysosomes in Vitro. J. Biol. Chem. 260, 3215–3220. https://doi.org/10.1016/S0021-9258(18)89493-5

Goldberg, D.E., Slater, A.F.G., Beavis, R., Chait, B., Cerami, A., Henderson, G.B., 1991. Hemoglobin Degradation in the Human Malaria Pathogen Plasmodium falciparum : A Catabolic Pathway Initiated by a Specific Aspartic Protease. J. Exp. Med. 173, 961–969.

Gupta, R., Jung, E., Brunak, S., 2004. Prediction of N-glycosylation sites in human proteins (in preparation).

Hahn, C., Fromm, B., Bachmann, L., 2014. Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparastic neodermata. Genome Biol. Evol. 6, 1105–1117. https://doi.org/10.1093/gbe/evu078

Hánová, I., Brynda, J., Houštecká, R., Alam, N., Sojka, D., Kopáček, P., Marešová, L., Vondrášek, J., Horn, M., Schueler-Furman, O., Mareš, M., 2018. Novel Structural Mechanism of Allosteric Regulation of Aspartic Peptidases via an Evolutionarily Conserved Exosite. Cell Chem. Biol. 25, 318-329.e4. https://doi.org/10.1016/J.CHEMBIOL.2018.01.001

Horn, M., Nussbaumerová, M., Šanda, M., Kovářová, Z., Srba, J., Franta, Z., Sojka, D., Bogyo, M., Caffrey, C.R., Kopáček, P., Mareš, M., 2009. Hemoglobin Digestion in Blood-Feeding Ticks: Mapping a Multipeptidase Pathway by Functional Proteomics. Chem. Biol. 16, 1053–1063. https://doi.org/10.1016/j.chembiol.2009.09.009

Ilgová, J., Jedličková, L., Dvořáková, H., Benovics, M., Mikeš, L., Janda, L., Vorel, J., Roudnický, P., Potěšil, D., Zdráhal, Z., Gelnar, M., Kašný, M., 2017. A novel type I cystatin of parasite origin with atypical legumain- binding domain. Sci. Rep. 7, 1–12. https://doi.org/10.1038/s41598-017-17598-2

Jedličková, L., Dvořák, J., Hrachovinová, I., Ulrychová, L., Kašný, M., Mikeš, L., 2019. A novel Kunitz protein with proposed dual function from Eudiplozoon nipponicum (Monogenea) impairs haemostasis and action of complement in vitro. Int. J. Parasitol. 49, 337–346. https://doi.org/10.1016/J.IJPARA.2018.11.010

Jedličková, L., Dvořáková, H., Dvořák, J., Kašný, M., Ulrychová, L., Vorel, J., Žárský, V., Mikeš, L., 2018. Cysteine peptidases of Eudiplozoon nipponicum: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp. Parasit. Vectors 11, 1–17. https://doi.org/10.1186/s13071-018-2666-2

Jedličková, L., Dvořáková, H., Kašný, M., Ilgová, J., Potěšil, D., Zdráhal, Z., Mikeš, L., 2016. Major acid endopeptidases of the blood-feeding monogenean Eudiplozoon nipponicum (Heteronchoinea: Diplozoidae). Parasitology 143, 494–506. https://doi.org/10.1017/S0031182015001808

Jennings, J.B., 1959. Studies on digestion in the monogenetic trematode Polystoma integerrimum. J. Helminthol. 33, 197–204.

Kang, J.-M., Yoo, W.-G., Le, H.G., Thai, T.L., Hong, S.-J., Sohn, W.-M., Na, B.-K., 2019. Partial Characterization of Two Cathepsin D Family Aspartic Peptidases of Clonorchis sinensis. Korean J Parasitol 57, 671–680.

https://doi.org/10.3347/kjp.2019.57.6.671
Kobayashl, K., Honke, T., Gasa, S., Fujii, T., Maguchi, S., Miyazaki', T., Makita, A., 1992. Proteolytic processing sites producing the mature form of human cathepsin D. Int. J. Biochem. 24, 1487–1491.
Konczal, M., Przesmycka, K.J., Mohammed, R.S., Phillips, K.P., Camara, F., Chmielewski, S., Hahn, C., Guigo, R., Cable, J., Radwan, J., 2020. Gene duplications, divergence and recombination shape adaptive evolution of the fish ectoparasite Gyrodactylus bullatarudis. Mol. Ecol. 29, 1494–1507. https://doi.org/10.1111/MEC.15421
Konstanzová, V., Koubková, B., Kašný, M., Ilgová, J., Dzika, E., Gelnar, M., 2015. Ultrastructure of the digestive tract of Paradiplozoon homoion (Monogenea). Parasitol. Res. 114, 1485–1494. https://doi.org/10.1007/s00436-015-4331-4
Logullo, C., Da Silva Vaz, I., Sorgine, M.H.F., Paiva-Silva, G.O., Faria, F.S., Zingali, R.B., De Lima, M.F.R., Abreu, L., Fialho Oliveira, E., Alves, E.W., Masuda, H., Gonzales, J.C., Masuda, A., Oliveira, P.L., 1998. Isolation of an aspartic proteinase precursor from the egg of a hard tick, Boophilus microplus. Parasitology 116, 525–532.
Nielsen, H., 2017. Predicting secretory proteins with signaIP, in: Methods in Molecular Biology. Humana Press Inc., pp. 59–73. https://doi.org/10.1007/978-1-4939-7015-5_6
Rawlings, N., Salvesen, G. (Eds.), 2013. Handbook of Proteolytic enzymes, 3rd ed. Elsavier, Amsterdam.
Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46, D624–D632. https://doi.org/10.1093/nar/gkx1134
Roudnicky, P., Potěšil, D., Zdráhal, Z., Gelnar, M., Kašný, M., 2020. Laser capture microdissection in combination with mass spectrometry: Approach to characterization of tissue-specific proteomes of Eudiplozoon nipponicum (Monogenea, Polyopisthocotylea). PLoS One 15, 1–15. https://doi.org/10.1371/journal.pone.0231681
Roudnický, P., Vorel, J., Ilgová, J., Benovics, M., Norek, A., Jedličková, L., Mikeš, L., Potěšil, D., Zdráhal, Z., Dvořák, J., Gelnar, M., Kašný, M., 2018. Identification and partial characterization of a novel serpin from Eudiplozoon nipponicum (Monogenea, Polyopisthocotylea). Parasite 25. https://doi.org/10.1051/PARASITE/2018062
Šácha, P., Knedlík, T., Schimer, J., Tykvart, J., Parolek, J., Navrátil, V., Dvořákovμ, P., Sedlák, F., Ulbrich, K., Strohalm, J., Majer, P., Ubr, V., Konvalinka, J., 2016. iBodies: Modular Synthetic Antibody Mimetics Based on Hydrophilic Polymers Decorated with Functional Moieties. Angew. Chemie Int. Ed. 55, 2356 –2360.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539–539. https://doi.org/10.1038/msb.2011.75
Sojka, D., Franta, Z., Frantová, H., Bartošová, P., Horn, M., Váchová, J., O'Donoghue, A.J., Eroy-Reveles, A. a., Craik, C.S., Knudsen, G.M., Caffrey, C.R., McKerrow, J.H., Mareš, M., Kopáček, P., 2012. Characterization of gut-associated cathepsin D

hemoglobinase from tick Ixodes ricinus (IrCD1). J. Biol. Chem. 287, 21152–21163. https://doi.org/10.1074/jbc.M112.347922

Sojka, D., Hartmann, D., Bartoš Ová-Sojková, P., Dvor ̌ák, J., 2016. Parasite Cathepsin D-Like Peptidases and Their Relevance as Therapeutic Targets. https://doi.org/10.1016/j.pt.2016.05.015

Suttiprapa, S., Mulvenna, J., Huong, N.T., Pearson, M.S., Brindley, P.J., Laha, T., Wongkham, S., Kaewkes, S., Sripa, B., Loukas, A., 2009. Ov-APR-1, an aspartic protease from the carcinogenic liver fluke, Opisthorchis viverrini: functional expression, immunolocalization and subsite specificity. Int J Biochem Cell Biol 41, 1148–1156. https://doi.org/10.1016/j.biocel.2008.10.013

Tinsley, R.C., 1973. Ultrastructural studies on the form and function of the gastrodermis of Protopolystoma xenopi (Monogenoidea: Polyopisthocotylea) Title. Biol. Bull. 144, 541–555.

Vorel, J., Cwiklinski, K., Roudnický, P., Ilgová, J., Jedličková, L., Dalton, J.P., Mikeš, L., Gelnar, M., Kašný, M., 2021. Eudiplozoon nipponicum (Monogenea, Diplozoidae) and its adaptation to haematophagy as revealed by transcriptome and secretome profiling. BMC Genomics 22. https://doi.org/10.1186/s12864-021-07589-z

Williamson, A.L., Brindley, P.J., Abbenante, G., Prociv, P., Berry, C., Girdwood, K., Pritchard, D.I., Fairlie, D.P., Hotez, P.J., Dalton, J.P., Loukas, A., 2002. Cleavage of hemoglobin by hookworm cathepsin D aspartic proteases and its potential contribution to host specificity. FASEB J. 16 (11), 1458–1460. https://doi.org/10.1096/fj.02-0181fje

Williamson, A.L., Brindley, P.J., Knox, D.P., Hotez, P.J., Loukas, A., 2003. Digestive proteases of blood-feeding nematodes. Trends Parasitol. 19, 417–423. https://doi.org/10.1016/S1471-4922(03)00189-2

Williamson, A.L., Lecchi, P., Turk, B.E., Choe, Y., Hotez, P.J., McKerrow, J.H., Cantley, L.C., Sajid, M., Craik, C.S., Loukas, A., 2004. A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. J. Biol. Chem. 279, 35950–35957. https://doi.org/10.1074/jbc.M405842200

Publication No.4

Isoforms of Cathepsin B1 in Neurotropic Schistosomula of Trichobilharzia regenti Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin

Dvořáková, H., Leontovyč, R., Macháček, T., O'Donoghue, A.J., Šedo, O., Zdráhal, Z., Craik, C.S., Caffrey, C.R., Horák, P., Mikeš, L., 2020.. Front. Cell. Infect. Microbiol. 10, 1–17. https://doi.org/10.3389/fcimb.2020.00066

Cathepsin B-like peptidases have been shown to be essential for parasite survival. They also contribute to pathogenic processes in many parasites. The neuropathogenic bird schistosome *Trichobilharzia regenti* possesses multiple forms of cathepsin B1 (TrCB1.1-TrCB1.6) described from the gut of schistosomula with probable involvement in nutrient digestion. Two of them (namely TrCB1.5, TrCB1.6) have a substitution of the catalytic cysteine in the active site (Cys29-to-Gly substitution) that renders the enzymes inactive. In this study, the comparison of substrate preferences of the two most divergent active homologs TrCB1.1 and TrCB1.4 showed that TrCB1.4 had very narrow substrate specificity and after processing it was less effective toward protein substrates when compared to TrCB1.1. Auto-activation of both active isoforms could be facilitated by sulphated polysaccharides due to a specific "heparin-binding motif" in the pro-sequence. Our work also demonstrated that the catalycally inactive TrCB1.6wt could be processed by the active TrCB1 forms and by an asparaginyl endopeptidase. Even trans-processed TrCB1.6wt lacked hydrolytic activity. However, endopeptidase, but not dipeptidase, activity could be restored by retrograde Gly29-to-Cys mutation. The effects of TrCB1.6wt paralog on cell viability and nitric oxide production using murine astrocytes, microglia and macrophages observed for TrCB1.1 in a previous study (Macháček et al. 2016) was not confirmed for TrCB1.6. Detailed analysis of TrCB1gene expression performed for all isoforms surprisingly revealed a high rate of expression of the enzymatically inactive paralogs TrCB1.5 and TrCB1.6. All recombinant TrCB1 isoforms were able to form SDS-stable enzyme/inhibitor complexes after incubation with chicken cystatin. So, the ability of the TrCB1.6 paralog with non-functional, Cys-to-Gly mutated active site to bind cystatin and its unexpectedly high expression support the hypothesis of inhibitor sequestration and thus protection of active cysteine peptidases.

My contribution was the production of the recombinant wild type TrCB1.6wt and its site-directed mutant TrCB1.6G/C, expression and purification of recombinant TrCB1 forms, activation/processing of recombinant enzymes, comparison of substrate preferences of the two active homologs, determination of cleavage sites in haemoglobin and myelin basic protein and interpretation of results, measurement of peptidase activities and inhibition

assays, testing formation of peptidase-cystatin complex and preparation of the manuscript**.**

Publication No.4

Isoforms of Cathepsin B1 in Neurotropic Schistosomula of Trichobilharzia regenti Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin

Dvořáková, H., Leontovyč, R., Macháček, T., O'Donoghue, A.J., Šedo, O., Zdráhal, Z., Craik, C.S., Caffrey, C.R., Horák, P., Mikeš, L.

(2020)

Front. Cell. Infect. Microbiol. 10, 1–17. https://doi.org/10.3389/fcimb.2020.00066

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 26 February 2020
doi: 10.3389/fcimb.2020.00066

Isoforms of Cathepsin B1 in Neurotropic Schistosomula of *Trichobilharzia regenti* Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin

Hana Dvořáková[1], Roman Leontovyč[1], Tomáš Macháček[1], Anthony J. O'Donoghue[2†], Ondřej Šedo[3], Zbyněk Zdráhal[3], Charles S. Craik[2], Conor R. Caffrey[4†], Petr Horák[1] and Libor Mikeš[1]*

[1] Department of Parasitology, Faculty of Science, Charles University, Prague, Czechia, [2] Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, San Francisco, CA, United States, [3] Central European Institute of Technology, Masaryk University, Brno, Czechia, [4] Center for Discovery and Innovation in Parasitic Diseases, Department of Pathology, University of California, San Francisco, San Francisco, CA, United States

Schistosomula (the post-infective stages) of the neurotropic schistosome *Trichobilharzia regenti* possess multiple isoforms of cathepsin B1 peptidase (TrCB1.1-TrCB1.6) with involvement in nutrient digestion. The comparison of substrate preferences of TrCB1.1 and TrCB1.4 showed that TrCB1.4 had a very narrow substrate specificity and after processing it was less effective toward protein substrates when compared to TrCB1.1. Self-processing of both isoforms could be facilitated by sulfated polysaccharides due to a specific binding motif in the pro-sequence. Trans-activation by heterologous enzymes was also successfully employed. Expression profiling revealed a high level of transcription of genes encoding the enzymatically inactive paralogs TrCB1.5 and TrCB1.6. The transcription level of TrCB1.6 was comparable with that of TrCB1.1 and TrCB1.2, the most abundant active isoforms. Recombinant TrCB1.6wt, a wild type paralog with a Cys^{29}-to-Gly substitution in the active site that renders the enzyme inactive, was processed by the active TrCB1 forms and by an asparaginyl endopeptidase. Although TrCB1.6wt lacked hydrolytic activity, endopeptidase, but not dipeptidase, activity could be restored by mutating Gly^{29} to Cys^{29}. The lack of exopeptidase activity may be due to other mutations, such as His^{110}-to-Asn in the occluding loop and Asp^{224}-to-Gly in the main body of the mature TrCB1.6, which do not occur in the active isoforms TrCB1.1 and TrCB1.4 with exopeptidase activity. The catalytically active enzymes and the inactive TrCB1.6 paralog formed complexes with chicken cystatin, thus supporting experimentally the hypothesis that inactive paralogs could potentially regulate the activity of the active forms or protect them from being inhibited by host inhibitors. The effect on cell viability and nitric oxide production by selected immune cells observed for TrCB1.1 was not confirmed for TrCB1.6. We show here that the active isoforms of TrCB1 have different affinities for peptide substrates thereby facilitating diversity in protein-derived nutrition for

OPEN ACCESS

Edited by:
Mario Alberto Rodriguez, National Polytechnic Institute, Mexico

Reviewed by:
César López-Camarillo, Universidad Autónoma de la Ciudad de México, Mexico
Anahid Jewett, University of California, Los Angeles, United States

***Correspondence:**
Libor Mikeš
mikes@natur.cuni.cz

†Present address:
Anthony J. O'Donoghue and Conor R. Caffrey, Center for Discovery and Innovation in Parasitic Diseases, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, United States

Specialty section:
This article was submitted to Parasite and Host, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *30 September 2019*
Accepted: *07 February 2020*
Published: *26 February 2020*

Citation:
Dvořáková H, Leontovyč R, Macháček T, O'Donoghue AJ, Šedo O, Zdráhal Z, Craik CS, Caffrey CR, Horák P and Mikeš L (2020) Isoforms of Cathepsin B1 in Neurotropic Schistosomula of Trichobilharzia regenti Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin. Front. Cell. Infect. Microbiol. 10:66. doi: 10.3389/fcimb.2020.00066

the parasite. The inactive paralogs are unexpectedly highly expressed and one of them retains the ability to bind cystatins, likely due to specific mutations in the occluding loop and the enzyme body. This suggests a role in sequestration of inhibitors and protection of active cysteine peptidases.

Keywords: peptidase, cathepsin B, processing, substrate specificity, occluding loop, cystatin, helminth, schistosome

INTRODUCTION

Flukes (Trematoda: Digenea) of the family Schistosomatidae include blood-dwelling flatworms that are pathogenic in birds and mammals, including man. Their infectious juveniles (cercariae) developing in intermediate snail hosts penetrate the skin of the definitive host in the water and enter the blood circulation while transforming to post-infective stages—the schistosomula (Kašný et al., 2016; Řimnáčová et al., 2017). In humans, the invasion of the skin can be accompanied by an allergic reaction called cercarial dermatitis (Horák et al., 2015; Macháček et al., 2018). However, schistosomula of the avian schistosome *Trichobilharzia regenti* use peripheral nerves and the central nervous system (CNS) as a migratory route to the final localization in the nasal mucosa of the duck host (Horák et al., 1998, 1999). The migration through the CNS can be manifested by neuromotor symptoms (leg paralysis and balance disorders) in birds and experimentally infected mice, occasionally leading to death of the host (Horák et al., 1999, 2008).

The migrating schistosomula actively feed on nervous tissue (Lichtenbergová et al., 2011; Leontovyč et al., 2019) and the adults in the nasal cavity ingest blood (Chanová and Horák, 2007). Indeed, both stages require gut-associated peptidases to digest host proteins. In *T. regenti*, isoforms of the cysteine peptidase cathepsin B1 (TrCB1) are expressed in the gut and represent a subset of a larger group of peptidases that are responsible for digestion (Dvořák et al., 2005; Dolečková et al., 2010; Leontovyč et al., 2016, 2019), much like in human *Schistosoma* species (Caffrey et al., 2004) and other helminths (Williamson et al., 2003, 2004; Delcroix et al., 2006; Caffrey et al., 2018). Cathepsin B (IUPAC: EC 3.4.22.1; MEROPS: Clan CA, Family C1) is unique among other papain-like cysteine peptidases due to its ability to act as both an endopeptidase and carboxy-exopeptidase (peptidyl dipeptidase) (Barrett and Kirschke, 1981). The latter activity is enabled due to the presence of an extra structural element termed "occluding loop," which occupies S′ subsites of the enzyme (Musil et al., 1991; Illy et al., 1997). Cathepsin B is synthesized as an inative precursor and a mature enzyme arises as a consequence of proteolytic removal of the N-terminal pro-peptide that sterically blocks the active site (Mort and Buttle, 1997). Six forms of TrCB1 (TrCB1.1 – TrCB1.6, GenBank: AY648119 – 24) were previously identified (Dvořák et al., 2005). Two of these, namely TrCB1.5 and TrCB1.6, are paralogs possessing a substitution of the catalytic cysteine Cys^{29} for a glycine Gly^{29} (mature TrCB1 numbering). Paralogs with similar single-amino acid mutations in the catalytic site are known for many parasite peptidases (Merckelbach et al., 1994; Caffrey et al., 2000; Holt et al., 2004; He et al., 2005; McCoubrie et al., 2007; Mendoza-Palomares et al., 2008; Pillay et al., 2010; Jedličková et al., 2018). Substitution of the active site cysteine need not always necessarily abolish the enzyme's activity, as shown in cases of Cys-to-Ser nucleophile substitutions (McCoubrie et al., 2007; Pillay et al., 2010), whereas other substitutions are expected to produce loss of peptidase activity, which is the case of TrCB1.5 and 1.6. It has been proposed that inactive paralogs of peptidases might regulate the activity of active forms by competing for substrates or inhibitors (Merckelbach et al., 1994; Holt et al., 2003, 2004; Dvořák et al., 2005). Moreover, they may alter host immune response through the inhibition of host complement pathways by binding the complement components C1q, mannose-binding lectin or properdin (Holt et al., 2004; Bergström et al., 2009; Reynolds et al., 2014). Active cathepsin B1 of *Schistosoma mansoni* (SmCB1) diminished the proinflamatory reaction of activated macrophages (Donnelly et al., 2009), provoked mixed Th1/Th2/Th17 immune response in mice and induced a transient T-helper 17 response in acute schistosome infection (Soloviova et al., 2019), suggesting it has a considerable immunogenic capacity. As for *T. regenti*, an inflammation-promoting activity was noticed in case of active TrCB1.1 which triggered the production of nitric oxide and proinflammatory cytokines interleukin-6 and tumor necrosis factor α by astrocytes and/or microglia (Macháček et al., 2016), i.e., the cells activated in mice infected experimentally by *T. regenti* (Lichtenbergová et al., 2011). However, no specific immunomodulatory mechanisms were tested so far with inactive peptidase paralogs of helminths.

Although the amino acid sequence identity among the six forms of TrCB1 is high (98 – 99% between the active forms TrCB1.1–TrCB1.4) (Dvořák et al., 2005), it seems that even a small change (a single amino acid substitution) in the sequences, especially in the S2 subsite, is likely to alter substrate specificity. Thus, the presence of multiple forms of a particular enzyme may indicate an adaptation of the parasite to physiological changes in the course of migration through the nervous tissue and final settlement in the nasal mucosa, enabling digestion of various host components.

In this study, we performed an in-depth comparison of substrate preferences of the two most divergent active homologs, TrCB1.1 and TrCB1.4, using positional scanning synthetic combinatorial peptide library and mass spectrometry. We also evaluated the ratios of relative expression levels of all six TrCB1 homologs using transcript-specific Illumina reads and stage-specific transcriptomes of *T. regenti* cercariae and schistosomula.We attempted to identify the possible biological roles of the cathepsin B paralogs employing heterologously

expressed wild type TrCB1.6wt and its site-directed mutant TrCB1.6G/C (retrograde Gly^{29}-to-Cys^{29} mutation). Various methods of enzymatic processing (pro-peptide sequence removal) were applied to all four zymogens included in this study. Finally, we assessed the effects of TrCB1.6wt paralog on cell viability and nitric oxide production using murine astrocytes, microglia and macrophages.

RESULTS

Expression of TrCB1 Forms

All recombinant pro-TrCB1 forms produced in *Pichia pastoris* X-33 were hyperglycosylated by the yeast. Each pro-enzyme migrated as a fuzzy band around 50 kDa on SDS-PAGE. After enzymatic deglycosylation and subsequent purification, they occurred as prominent ≈36 kDa bands (**Supplementary Figure 1**), which corresponds with their theoretical molecular weight (MW) (**Table 1**). N-terminal amino acid sequences NEMQF of pro-TrCB1.6wt and pro-TrCB1.6G/C, and ENEIQ of pro-TrCB1.1 and pro-TrCB1.4 corresponded to the beginning of the enzymes' predicted pro-sequences. In addition, a few minor bands of lower MW appeared in fractions of deglycosylated purified pro-TrCB1.6G/C, pro-TrCB1.1, and pro-TrCB1.4 (**Supplementary Figure 1**).

Processing of Recombinant *Pichia*-Derived Pro-enzymes

(A) Incubation of pro-TrCB1.1 and pro-TrCB1.4 at pH 4 or pH 5.5 for up to 16 h resulted in a decrease in abundance of the pro-protein but no detectable appearance of the lower MW mature enzymes (an example is presented in **Supplementary Figure 2**). However, addition of 20 μg/ml sulfated polysaccharides (SP)—dextran sulfate (DS) in the case of TrCB1.1 and TrCB1.4 or heparin sodium salt (HSS) with TrCB1.4—resulted in reliable production of ≈30 kDa processed enzymes at pH 4.5 (**Figures 1C,D**). N-terminal sequencing revealed a L^{88}EIPS sequence (pro-TrCB1 numbering) in TrCB1.4 that was processed in the presence of HSS (**Figure 2A**). The detected cleavage site was two amino acid residues upstream from the predicted beginning of the mature enzyme (Dvořák et al., 2005). Further efforts to determine N-terminal residues in ≈30 kDa SP-treated TrCB1.1 and DS-treated dsTrCB1.4 failed. Incubation of pro-enzymes with both SPs resulted in higher activity against the fluorogenic substrate Z-Phe-Arg-AMC, especially at pH 4.5 (**Figure 1A**). However, a rapid decline of the activity due to enzyme autodegradation could be observed between ~1 and 4 h (**Figure 1B**), even when stored on ice.

(B) SDS-PAGE demonstrated that agarose-bound pepsin succesfully processed all pro-enzymes, since new single/double bands of ≈30–32 kDa appeared in the gel (**Figure 3A**). N-terminal sequencing of pepsin-treated pepTrCB1.1 and pepTrCB1.4 after 1-h incubation revealed that at least 3 products arose due to different cleavage sites within the pro-sequences (**Figure 2A**), but none of the products started with predicted N-terminal residues of mature enzymes. The pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6G/C had increased activity toward Z-Phe-Arg-AMC compared to corresponding pro-enzymes. Maximum activities were observed within 30 min of incubation in samples of TrCB1.1 and 1.4 and within 5 min in the case of TrCB1.6G/G (**Figure 3B**). The pepTrCB1.1 and pepTrCB1.4 were stable even after longer (weeks) storage at −25°C. On the other hand, the pepTrCB1.6G/C had very limited stability. The fluorescent active-site probe DCG-04 bound to pepTrCB1.1 and pepTrCB1.4 (**Figure 4**). This reaction was blocked by the irreversible inhibitor of papain-like cysteine peptidases E-64. No binding of the probe to pepTrCB1.6wt and pepTrCB1.6G/C was observed (**Figure 4**). Unless otherwise noted, the pepTrCB1 forms (after 30 min incubation with pepsin) were used in subsequent experiments.

(C) Trans-processing of pro-TrCB.6wt by recombinant asparaginyl endopeptidase from the tick *Ixodes ricinus* (IrAE) led to a production of variously processed forms (**Figure 5A**). All tested pH values gave the same results (**Supplementary Figure 3**). Three cleavage sites identified by N-terminal sequencing were N^{23}/E^{24}, D^{59}/A^{60}, N^{85}/V^{86} (pro-TrCB1.6 numbering) (**Figure 2B**). Prolonged incubation at pH 5.5 (up to 36 h) resulted in the formation of a single ≈30 kDa band, but N-terminal sequencing failed in this case. The intensity of the upper band (TrCB1.6 pro-enzyme) present in the lane at 16 h decreased to undetectable levels compared to time 0 (**Figure 5A**).

(D) Overnight incubation of pro-TrCB1.6wt with pepTrCB1.1 or pepTrCB1.4 in an activation buffer of pH 4.5 supplemented with DS resulted in the conversion of the zymogens to a major band of ≈32 kDa, and a minor band of ca 30 kDa (**Figure 5B**). Processing was not observed when E-64 (10 μM) was added to the reactions (**Figure 5B**). The N-terminal sequence G^{65}VMRE (pro-TrCB1.6 numbering; **Figure 2B**) was detected in the more intense ≈32 kDa band after overnight incubation of pro-TrCB1.6wt with pepTrCB1.4, while N-terminal sequencing of the other products was unsuccessful.

TABLE 1 | Theoretical molecular weight and number of potential N-glycosylation sites of pro-enzymes/mature TrCB1 forms.

	Pro-enzyme (kDa)	Mature enzyme (kDa)	N-gly sites (pro-sequence/mature enzyme)
TrCB1.1	36.4	28.5	1/1
TrCB1.4	36.5	28.6	1/1
TrCB1.6wt	36.0	28.2	1/2
TrCB1.6G/C	36.1	28.3	1/2

Peptidase Activity and Inhibition Assays of TrCB1.6, TrCB1.1, and TrCB1.4

Assays were performed with two synthetic fluorogenic substrates: Z-Phe-Arg-AMC and Z-Arg-Arg-AMC. The pepTrCB1.6wt had no activity with both substrates (not shown). The activity optimum of pepTrCB1.6G/C with Z-Phe-Arg-AMC occurred at pH 6, i.e., it was shifted more to the neutral value than that of pepTrCB1.1 (pH 5.5) and pepTrCB1.4 (pH 5) (**Figure 6A**).

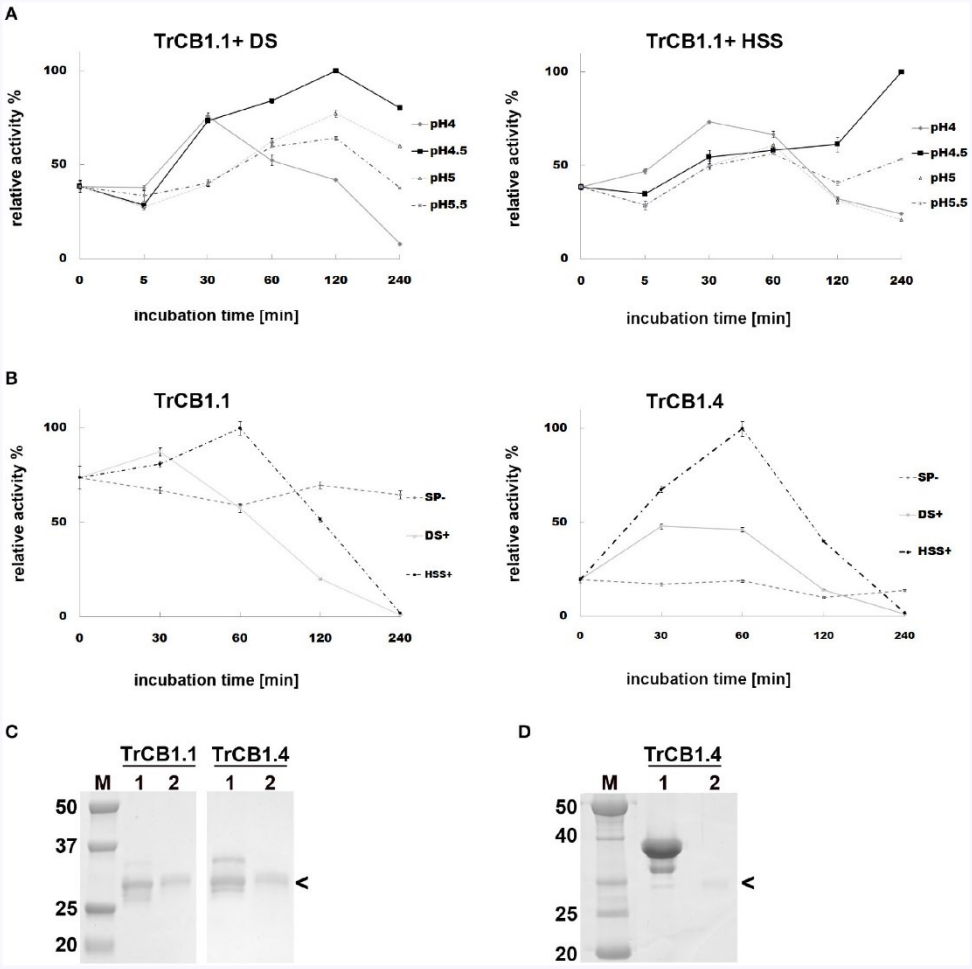


FIGURE 1 | Effect of sulfated polysaccharides on the autocatalytic activation of pro-TrCB1.1 and pro-TrCB1.4. Peptidolytic activities of TrCB1 forms generated after activation were determined in a kinetic assay with Z-Phe-Arg-AMC. The mean values ± standard deviation (SD) of two triplicate assays are expressed as percentage of relative activity in the sample at specified time points. **(A)** pro-TrCB1.1 was incubated in the presence of 10 μg/ml dextran sulfate (DS) or heparin sodium salt (HSS) at various pH values. **(B)** pro-TrCB1.1 and pro-TrCB1.4 were incubated at pH 4.5 in the presence or absence of 20 μg/ml sulfated polysaccharides (SP). **(C)** SDS-PAGE of (pro-)TrCB1.1 and (pro-)TrCB1.4 incubated in the presence of 10 μg/ml DS at pH 4.5 for 30 min (lanes 2). Lanes 1 contain the pro-enzyme without DS. **(D)** SDS-PAGE of (pro-)TrCB1.4 incubated in the presence of 10 μg/ml HSS at pH 4.5 for 30 min (lane 2). Lane 1 contains the pro-enzyme without HSS. M, markers of molecular size (kDa). Arrowheads point to processed enzymes, which were rapidly autodegraded under the experimental conditions. A gel for TrCB1.1 with HSS is not available.

The pepTrCB1.1 was also active toward Z-Arg-Arg-AMC with a similar optimum at pH 5.5 (not shown). No activity toward this substrate was detected for pepTrCB1.6G/C as well as for pepTrCB1.4.

The activities of pepTrCB1.1 and pepTrCB1.4 were inhibited by E-64 and the cathepsin B-specific irreversible inhibitor CA-074 (**Figure 6B**). In the case of pepTrCB1.1, <1% of the activity remained after inhibition by both inhibitors at 0.1 μM concentration. Approximately 1% activity of pepTrCB1.4 remained after treatment by 10 μM E-64, while 0.1 μM CA-074 completely suppressed the activity of TrCB1.4. On the other hand, E-64 and CA-074 had little and no effect, respectively, on the activity of pepTrCB1.6G/C (**Figure 6B**). Even a high concentration (100 μM) of E-64 reduced the activity of pepTrCB1.6G/C only to 34%.

The peptidyl dipeptidase activity of pepTrCB1 was investigated using Bz-Gly-His-Leu substrate (**Figure 7**). Both pepTrCB1.1 and pepTrCB1.4 exhibited carboxypeptidase activity

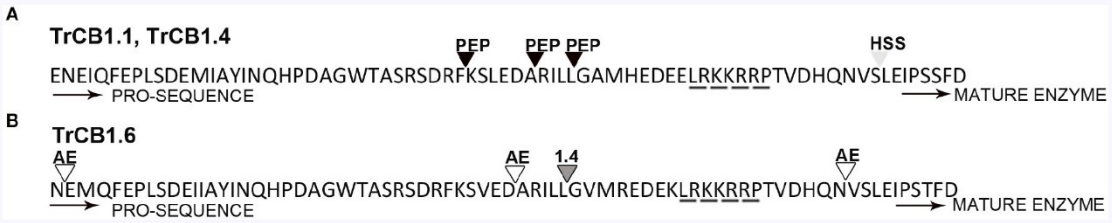


FIGURE 2 | Cleavage sites in the processed pro-sequences of TrCB1 forms. **(A)** cleavage sites determined by Edman degradation in pepsin-processed pepTrCB1.1 and pepTrCB1.4 (black triangles); cleavage site in TrCB1.4 auto-processing product resulted after 1-h incubation in the presence of heparin sodium salt (light gray triangle). **(B)** cleavage sites in rIrAE-activated TrCB1.6wt resulted after 1-h incubation at pH 5 (empty triangles); cleavage site in TrCB1.6wt after processing with pepTrCB1.4 (dark gray triangle). "Heparin-binding" (Horn et al., 2011) motif responsible for glycosaminoglycan binding to the pro-sequences is underlined by a dashed line.

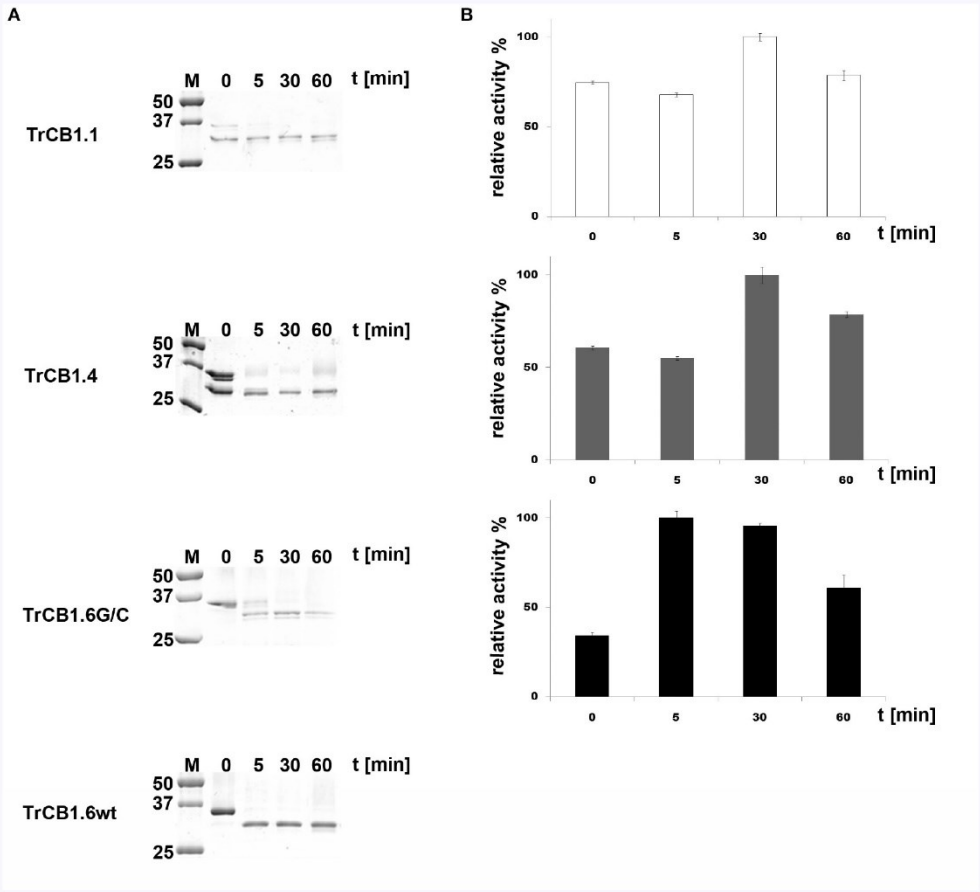


FIGURE 3 | Trans-processing of the TrCB1 zymogens by pepsin. TrCB1 forms were treated with pepsin immobilized on agarose beads at 37°C (TrCB1.6G/C at RT), and the reaction mixtures were analyzed at the indicated time points. **(A)** The processed forms were resolved by SDS-PAGE. M, markers of molecular size. **(B)** Peptidolytic activities of TrCB1.1, TrCB1.4, TrCB1.6G/C generated after processing were measured in a kinetic assay with the substrate Z-Phe-Arg-AMC (25 μM). The results are means ± SD of two triplicate assays and are expressed as percentage of relative activity in the sample.

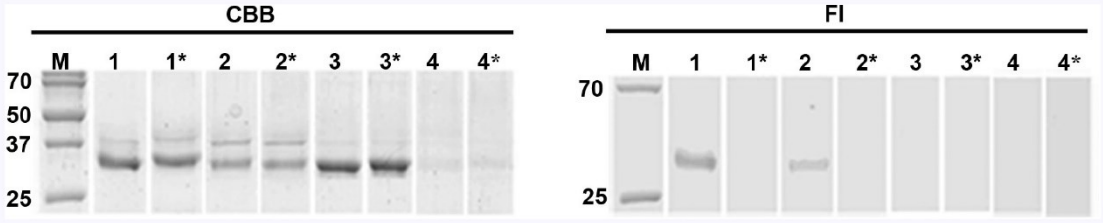


FIGURE 4 | Active site labeling of the pepTrCB1 forms by the cysteine peptidase-specific probe DCG-04. The pepTrCB1 forms were treated with the probe BODIPY green-DCG-04, resolved by SDS-PAGE, and stained with Coomassie Brilliant Blue (CBB) or visualized with a fluorescence imager (Fl). (1) TrCB1.1, (2) TrCB1.4, (3) TrCB1.6wt, (4) TrCB1.6G/C. Samples marked by asterisk were pre-incubated with E-64 inhibitor prior to labeling. M, markers of molecular size (kDa). Note the autodegradation of TrCB1.6G/C (lanes 4 and 4*) during incubation with the probe/inhibitor, resulting in weak protein bands.

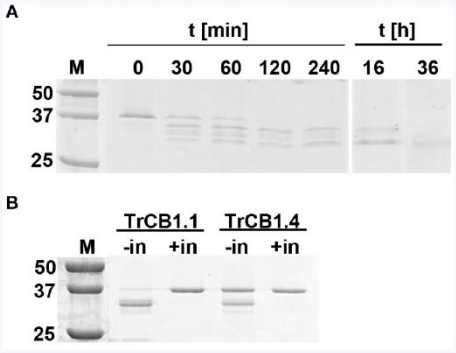


FIGURE 5 | Trans-processing of pro-TrCB1.6wt by IrAE, TrCB1.1, or TrCB1.4. **(A)** The pro-TrCB1.6wt was incubated at pH 5 with purified yeast medium containing activated IrAE for various time periods at 37°C. The TrCB1.6wt processing products were resolved by SDS-PAGE. **(B)** The TrCB1.6wt zymogen was incubated overnight at 37°C with pepTrCB1.1 or pepTrCB1.4 in the presence of dextran sulfate (10 μg/ml). Processing of pro-TrCB1.6wt was visualized by SDS-PAGE, gel stained with Coomassie Brilliant Blue. Processing was inhibited with 10 μM inhibitor of cysteine peptidases E-64. −/+ in denotes absence/presence of the inhibitor. M, markers of molecular size (kDa).

against this substrate with an optimum at pH 4.5 (**Figure 7A**). The activities were efficiently inhibited by 10 μM E-64 and CA-074 (**Figure 7B**). On the other hand, pepTrCB1.6G/C had very low activity against the exo-substrate at pH 5 and 5.5, and no activity occurred at pH ≤ 5. The pepTrCB1.6wt had no exopeptidase activity at any tested pH (not shown).

Repeated attempts to inhibit endopeptidase activity of activated TrCB1 forms (ds/pepTrCB1.1, ds/pepTrCB1.4, ds/pepTrCB1.6G/C) by chicken egg white cystatin and human cystatins B and C in the presence of Z-Phe-Arg-AMC substrate failed under the conditions used (not shown).

Formation of Peptidase Complex With Chicken Egg Cystatin

Overnight incubation at 4°C, but not a 15 min treatment at room temperature (RT) of auto-activated TrCB1 forms (dsTrCB1.1, dsTrCB1.4, dsTrCB1.6G/C) or pepTrCB1.6wt with chicken egg white cystatin resulted in the formation of new weak bands around ≈40 kDa that roughly matched the predicted size of the enzyme/cystatin complex (**Figure 8**). Only a small percentage of the enzymes bound to cystatin. On the other hand, chicken cystatin seemed to protect TrCB1.6G/C from autodegradation. Binding of human cystatins B and C to the pepsin-processed enzymes was not confirmed within 2 h of incubation at RT (not shown).

P1–P4 Specificity Profile of TrCB1.1 and TrCB1.4

The preference of TrCB1.1 and TrCB1.4 for amino acid residues at P1 – P4 positions of the substrates was determined using positional scanning synthetic combinatorial library (PS-SCL) (**Figure 9**). At P1 position, both enzymes preferred small basic amino acids Lys and Arg, but TrCB1.4 showed more distinct inclination to Arg. A stronger preference for Met, the non-natural norleucine (similar unbranched chain structure as Lys), and Lys at this position was also noted for TrCB1.4.

At P2 position, TrCB1.4 exhibited preference for non-polar aliphatic amino acids (Ala, Val, Ile), with overriding preference for Ala. In addition, TrCB1.4 was also able to accommodate polar residues (Ser, Thr and norLeu). TrCB1.1 showed greater promiscuity at P2, with a low preference for large non-polar residues, and could not accept Gly or Pro. However, it was able to hydrolyze substrates with Arg at P2.

Both enzymes displayed similar preferences for residues at P3. The highest preference was recorded for Leu, but the enzymes also well-accepted other amino acid residues (norLeu, Tyr, and Met) at this position. Screening at P4 position revealed a broader substrate specificity, with all amino acid residues accepted by both recombinant enzymes.

Hydrolysis of Protein Substrates by TrCB1 Forms

The pepTrCB1.1 and pepTrCB1.4 were able to degrade all given macromolecular substrates [albumin, IgG, fibrinogen, collagen type I and type IV, myosin, myelin basic protein (MBP) and hemoglobin (Hb)] at pH 5.5 (**Figure 10A**). While MBP was efficiently hydrolyzed between pH 4.5 and 6.5 with

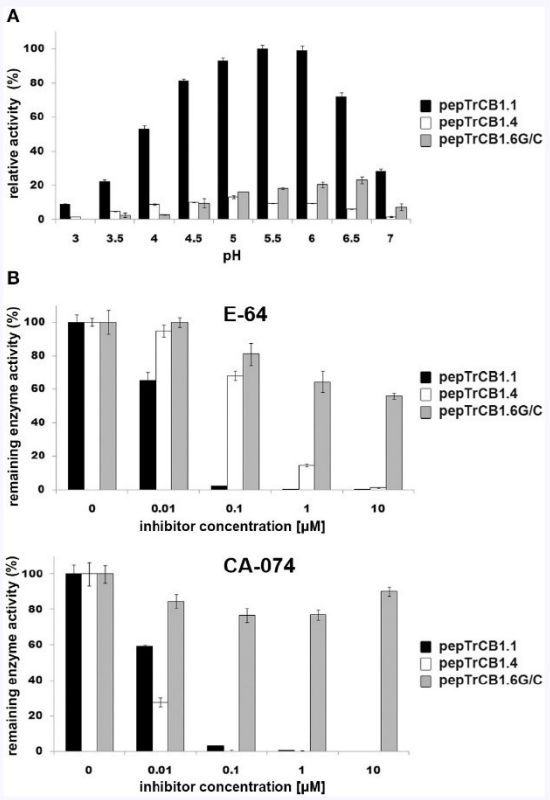


FIGURE 6 | pH optima of the pepTrCB1 forms and effect of inhibitors on enzyme endopeptidase activity. **(A)** The effect of pH on peptidolytic activity of pepTrCB1.1, pepTrCB1.4, pepTrCB1.6G/C toward the fluorogenic peptidyl substrate Z-Phe-Arg-AMC was measured in 50/100 mM CPB (pH range 3–7) with 2 mM DTT. Data are expressed as percentage of relative activity in the sample. **(B)** Effect of cysteine peptidase inhibitors (E-64, CA-074) on the activity of pepTrCB1.1, pepTrCB1.4, pepTrCB1.6G/C toward Z-Phe-Arg-AMC substrate. Inhibition assays were carried out at pH optimum of the particular enzymes. The values are mean ± SD of two triplicate assays.

the best result at pH up to 5.5, Hb was poorly degraded only at lower pH by both recombinant enzymes (**Figure 10B**). The pepTrCB1.6G/C could degrade most of the tested protein substrates (except for IgG) at pH 5.5 (**Figures 10A,B**), while pepTrCB1.6wt completely lacked such activity (not shown). Significant autodegradation of TrCB1.6G/C was noticed under the conditions of the experiment, which could to a certain degree distort the quantitative interpretation of the results.

Identification of Hemoglobin and MBP Fragments by Mass Spectrometry

The MS sequence analysis of peptides resulting from action of TrCB1 forms (pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6G/C) on polypeptides of turkey Hb and human MBP was performed to determine enzymes' cleavage site specificities. Detected peptide fragments are listed in **Supplementary Table 1** (only the fragments which were not detected at time zero are included). The identified cleavage sites in the sequence of α- and β-subunit of Hb and in MBP are shown in **Figures 11A,B**, respectively. Several fragments were obtained with pepTrCB1.1, which did not occur in controls. After 5 min of reaction, eight cleavage sites (i.e., "initial cleavage sites") were identified. The cleavage map shows that TrCB1.1 obviously preferred Lys at P1 position in the two macromolecular substrates, followed by Arg, which corresponds with the results of PS-SCL (see above). Other amino acids could also be accommodated in this position (Leu>Thr/Glu>Phe/Ser>Gln/Ser/Gly/Met/Ala/Val/Phe). At the P2 position, the preference was not that distinctive: pepTrCB1.1 slightly preferred polar uncharged amino acids (Ser/Thr/Glu) and the positively charged Lys, closely followed by Phe/Ala/Leu/Pro/His/Asp, and less Arg/Thr/Ile/Glu. Although PS-SCL results have shown that Pro is unfavorable at P2, two of the fragments generated by pepTrCB1.1 contained Pro at this position. Only one fragment [amino acids 1 – 16 from Hbα (A) subunit] could be reliably detected after hydrolysis of Hb by pepTrCB1.4. While the same fragment generated by pepTrCB1.1 appeared after 5 min of incubation, the fragment produced by pepTrCB1.4 appeared as late as in 2 h. The preference for amino acids in both P1 and P2 positions of macromolecular substrates was not clearly defined and no obvious pattern was observed. Also, no cleavage sites distinct from those produced by TrCB1.1 were detected. Finally, no fragment was detected after incubation of Hb with TrCB1.6G/C and less fragments were also identified after cleavage of MBP comparing to the action of the other enzymes. Moreover, all the cleavage sites were identical with those produced by pepTrCB1.1 (**Figure 11**).

Differential Expression of TrCB1 Forms

Based on the differential transcriptional analysis, TrCB1.1, TrCB1.2, TrCB1.3, TrCB1.5, and TrCB1.6 were stated as being significantly upregulated in schistosomula, when compared to cercariae. TrCB1.4 was not detected in *T. regenti* juvenile transcriptomes at significant levels. Transcription profile of cercariae was limited only to forms TrCB1.1, TrCB1.2, and TrCB1.6 with low counts per million (CPM) rate. In schistosomulum stage, TrCB1.1, TrCB1.2, and TrCB1.6 genes were transcribed at a similar level, followed by TrCB1.3 and TrCB1.5 (**Figure 12**).

In vitro Exposure of Murine Astrocytes, Microglia and RAW 264.7 Macrophages to Pro-TrCB1.6wt

Production of NO by murine astrocytes, microglia and RAW 264.7 macrophages was examined after treatment with pro-TrCB1.6wt for 48 h. As revealed by Griess assay, pro-TrCB1.6wt neither triggered NO production in any of the cells nor reduced NO production in cells concurrently treated by LPS (**Supplementary Figure 4A**). Similarly, no effect on changes in cell viability was observed (**Supplementary Figure 4B**).

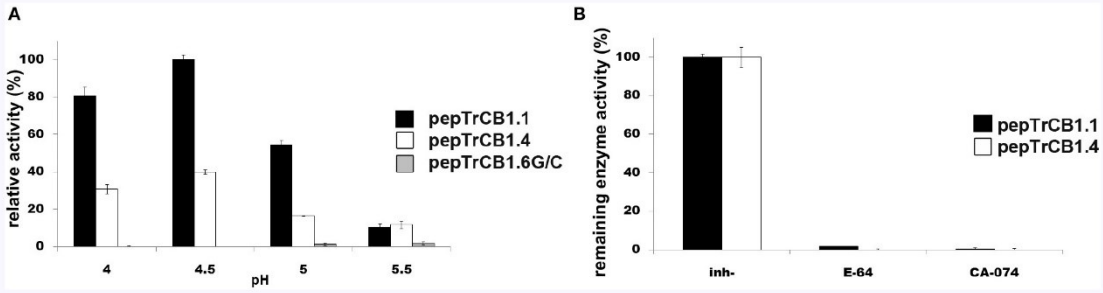


FIGURE 7 | Peptidyl-dipeptidase activity of the pepTrCB1 forms and the effect of inhibitors. **(A)** pepTrCB1 forms were incubated with Bz-Gly-His-Leu substrate in 50/100 mM CPB (pH 4–5.5) containing 2 mM DTT. The appearance of arising free amino groups was monitored by reaction with fluorescamine (0.05 mg/ml). **(B)** The effect of 10μM E-64 and CA-074 on carboxypeptidase activity of pepTrCB1.1 and pepTrCB1.4 assayed at pH optima of the particular enzymes. Values are means of two triplicate assays ± SD.

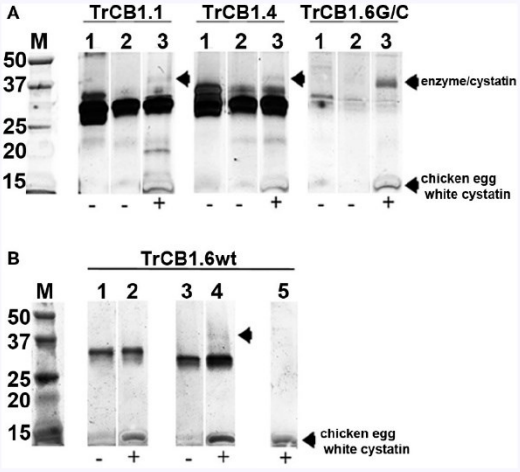


FIGURE 8 | Complex formation between TrCB1 forms and chicken egg white cystatin monitored by SDS-PAGE. **(A)** Auto-activated dsTrCB1.1, dsTrCB1.4, dsTrCB1.6G/C after overnight incubation at 4°C in the presence/absence of chicken egg white cystatin. After incubation, the samples were mixed with non-reducing sample buffer and resolved by SDS-PAGE. Proteins were visualized by silver staining. (1) pro-TrCB1, (2) dsTrCB1 without cystatin, (3) dsTrCB1 with cystatin. Note the autodegradation of TrCB1.6G/C (lanes 1 to 3) happening during incubation, resulting in weak protein bands. **(B)** pepTrCB1.6wt after overnight incubation at 4°C in the presence/absence of cystatin. (1) pro-TrCB1.6wt, (2) pro-TrCB1.6wt with cystatin, (3) pepTrCB1.6wt, (4) pepTrCB1.6wt with cystatin, (5) control lane with cystatin only. Arrows point to peptidase/cystatin complex and/or to cystatin (in section B, lane 5). M, markers of molecular size (kDa).

DISCUSSION

Expression of TrCB1.1 and TrCB1.4 in *Pichia* led to production of zymogens which were partly auto-processed in the yeast media, producing a small proportion of partially activated intermediates. Further processing of the recombinant (pro-) enzymes was therefore necessary. Although Dvořák et al. (2005) obtained auto-processed TrCB1.4 with an eight-fold increase in activity after overnight incubation at acidic pH, we found an insufficient increase of activity in both TrCB1.4 and TrCB1.1 using this approach. Auto-activation of recombinant cathepsin B1 from *Schistosoma mansoni* (SmCB1) is forced by negatively charged polysaccharides, such as sulfated glycosaminoglycans. Their binding to the SmCB1 pro-sequence destabilizes the pro-sequence/enzyme core junction (Jílková et al., 2014) and is mediated by a specific α-helix that contains a "heparin-binding motif"-XBBXBBX- (B is a basic amino acid and X refers to a hydropathic residue) (Horn et al., 2011; Jílková et al., 2014). A sequence alignment revealed that this motif is specific for trematode cathepsins B involved in the digestion of host proteins (Horn et al., 2011). As this motif occurs in all TrCB1 pro-sequences (**Figure 2**), sulfated polysaccharides facilitated auto-activation of both TrCB1.1 and TrCB1.4 (**Figure 1**). However, unlike in SmCB1 (Jílková et al., 2014), the products were relatively unstable at all tested pH values. Therefore, we did not use them in most subsequent experiments. However, there seems to be similarity between SmCB1 and TrCB1 in terms of glycosaminoglycan-mediated acceleration of auto-activation *in vivo*.

In vitro trans-processing of pro-TrCB1 by pepsin was applied according to previously published experience (Lipps et al., 1996; Law et al., 2003). N-terminal sequencing showed that the pro-sequence of TrCB1 is cleaved at carboxyterminal side of Phe and Leu residues, which is typical for pepsin (Hamuro et al., 2008). Unfortunately, none of the cleavage sites resulted in fully mature enzymes. Even prolonged incubations or higher concentrations of pepsin did not give better results. On the other hand, partial removal of the TrCB1 pro-sequences was sufficient for hydrolytic activity toward macromolecular substrates. The advantage of this method was that agarose beads-immobilized pepsin could be easily removed by centrifugation. Therefore, pepsin-processed enzymes were used in most experiments. The asparaginyl endopeptidase (AE, legumain) was confirmed as an

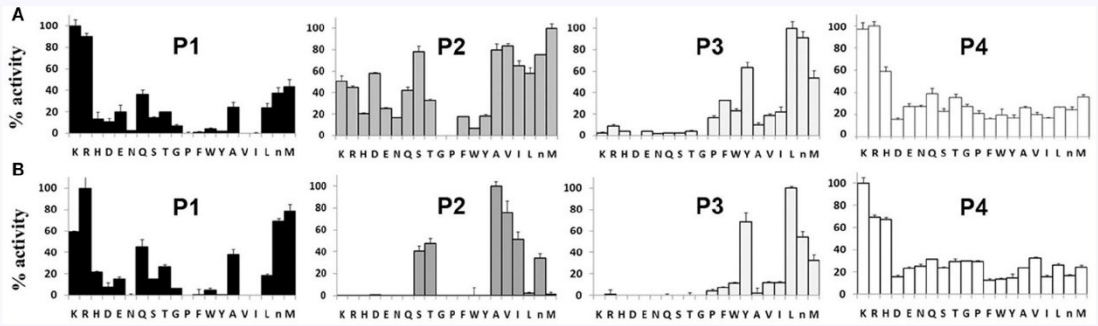


FIGURE 9 | P1–P4 amino acid preferences of pro-TrCB1.1 and pro-TrCB1.4 determined by positional scanning of a synthetic combinatorial library. **(A)** Substrate library of pro-TrCB1.1. **(B)** Substrate library of pro-TrCB1.4. P4-P1 represent the four amino acid side chains recognized by peptidase on the N-terminal side of the cleaved peptide bond. The X-axis indicates the 20 amino acids held constant at each position, designated by the single-letter code (n, norleucine). The Y-axis shows relative preference for the particular amino acid when 100% represents the most preferred amino acid residue.

endogenous *trans*-processing enzyme activating other digestive peptidases in human schistosomes (Sajid et al., 2003) and three transcripts coding this enzyme were also found in *T. regenti* transcriptome (Leontovyč et al., 2016, 2019). We were able to *trans*-process the pro-TrCB1 using tick IrAE, obtaining several intermediates during short-time incubations, but this method was not widely used in our experiments, since it provided low yields.

The substrate specificity of papain-like peptidases seems to be largely determined by the composition of the S2 pocket (Choe et al., 2006), in which the amino acid residues interact with an amino acid residue at P2 position of the peptide substrate (Schechter and Berger, 1967). From this point of view, it was interesting that amino acid preferences of TrCB1.1 and TrCB1.4 differed at the P2 position, while the preference for other positions was quite similar. Using PS-SCL, TrCB1.1 showed broader preferences at P2, having a relative preference for aliphatic amino acids (excepting Gly) over aromatic residues (Phe, Trp, and Tyr). Interestingly, TrCB1.1 had a P2 preference more similar to *S. mansoni* cathepsin B2 than to SmCB1 (Choe et al., 2006), although it has higher overall sequence identity and function similar with SmCB1. By contrast, TrCB1.4 displayed relatively narrow P2 specificity with an overriding preference for aliphatic Ala. In agreement with a previous study (Dvořák et al., 2005), TrCB1.4 did not show any noticeable activity toward substrates with Arg at P2. Surprisingly, the library assay also indicated that even Met is not a favored amino acid for TrCB1.4 at P2, which is not typical of cathepsins B (Choe et al., 2006). On the contrary, Met is the most favored amino acid accepted by TrCB1.1 at P2. Also, Z-Phe-Arg-AMC was a poor substrate for TrCB1.4, compared to TrCB1.1, wich corresponds with the results of PS-SCL.

Both TrCB1.1 and TrCB1.4 degraded a range of macromolecular substrates, such as albumin, IgG, fibrinogen, collagen type I, collagen type IV, myosin, myelin basic protein (MBP), and hemoglobin (Hb). Divergent specificities for the selected protein substrates were not observed between the two enzymes, except for MBP, which was obviously more efficiently hydrolysed by TrCB1.1. In line with previous work (Dvořák et al., 2005), both enzymes poorly degraded Hb. The authors assumed that this may indicate an adaptation to the unique migratory route of *T. regenti* through the CNS. On the other hand, Hb becomes an important source of nutrition for adult worms located in the nasal mucosa of ducks (Horák et al., 1999; Blažová and Horák, 2005; Chanová and Horák, 2007) and the level of TrCB1.1 gene transcription was comparable in *T. regenti* schistosomula and adults (Dolečková et al., 2010).

It is likely that active forms of TrCB1 can work in a cooperative network with other peptidases (e.g., cathepsins L and cathepsins D) in the parasite's gut, which could preferentially initiate hydrolysis of ingested host proteins (including Hb). Such a network was characterized in *S. mansoni*, where the cysteine peptidases (cathepsins L and B) initiate digestion of albumin, whereas cathepsin D is responsible for the primary cleavage of Hb, although cysteine peptidases may aid in this. Cathepsins B also act as potent (carboxy-) exopeptidases (Illy et al., 1997). Several transcripts of cathepsin L and D genes were found also in the transcriptome of *T. regenti* schistosomula (Leontovyč et al., 2016).

MS/MS analysis of protein substrate digests revealed that TrCB1.1 was more efficient in the cleavage of Hb than TrCB1.4 in terms of the number of produced cleavage sites. Nevertheless, intensive signals of undigested Hb were observed even after 2 h of incubation with the enzymes under given conditions (substrate:enzyme ratios). The cleavage sites produced by TrCB1.1 exhibited faintly defined specificity. Thus, it seems that TrCB1 enzymes may contribute to the degradation of Hb, likely after initial cleavage by other endopeptidase(s). Surprisingly, in spite of the differences in amino acid preferences at P2 of the substrates seen in PS-SCL, there was a 100% overlap in cleavage sites produced by TrCB1.4 in the molecule of MBP, with those made by TrCB1.1, which only produced a few more cuts. Unexpectedly, we were not able to detect more fragments of Hb

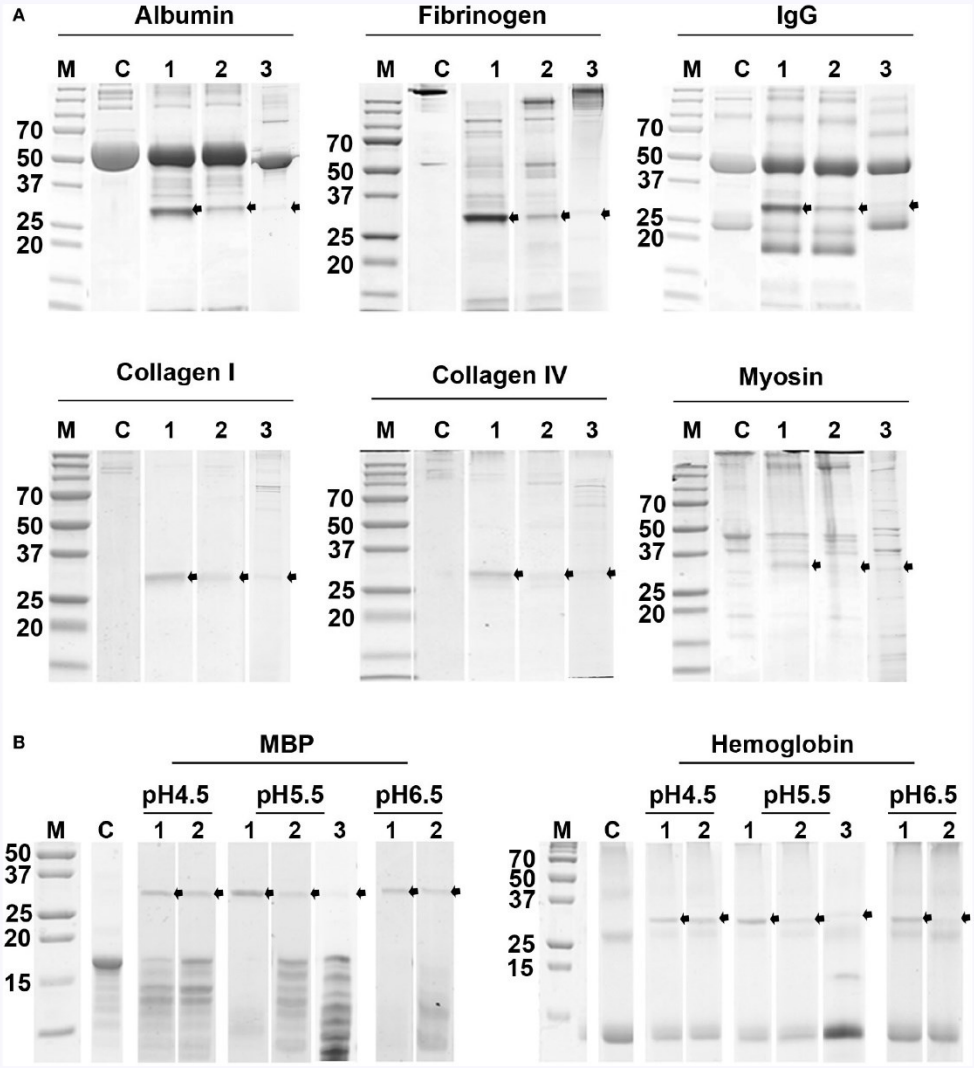


FIGURE 10 | Digestion of selected protein substrates by recombinant TrCB1 forms. The substrates (0.5 mg/ml) were incubated for 6 h with pepTrCB1 forms (1 μg) at 37°C. Aliquots (20 μl) of the resulting hydrolysates were separated by SDS-PAGE and stained with Coomassie Brilliant Blue. **(A)** Incubation of pepTrCB1.1, pepTrCB1.4 and pepTrCB1.6G/C (lanes 1–3, respectively) at pH 5.5 with albumin, fibrinogen, collagen type I, collagen type IV and myosin. **(B)** Incubation of pepTrCB1.1, pepTrCB1.4 and pepTrCB1.6G/C (lanes 1–3, respectively) at pH 4.5–6.5 with myelin basic protein (MBP) and hemoglobin. C, controls (substrates without enzyme). The starting amount of the peptidases was equal in all samples; note the autodegradation of the enzymes under given conditions, resulting in weaker bands especially in TrCB1.6G/C. Arrows indicate the visible peptidase bands in the gels. M, markers of molecular size (kDa).

produced by TrCB1.4, probably due to the low activity of the enzyme toward this substrate.

Although recombinant TrCB1.4 was previously obtained by employing mRNA from 5 to 6 days old schistosomula (Dvořák et al., 2005), our differential expression analysis revealed that transcripts of TrCB1.4 were not present at detectable levels in the transcriptome of cercariae or 7 days old schistosomula. This suggests that TrCB1.4 is not stably expressed in schistosomula and it is likely not essential for migrating worms in the definitive host. However, this, together with a distinct specificity toward amino acid residues of protein substrates, points to "a gene in waiting hypothesis." The redundancy of enzyme isoforms arisen by duplication and subsequent divergence of coding genes "provides a potential mechanism to alter the substrate

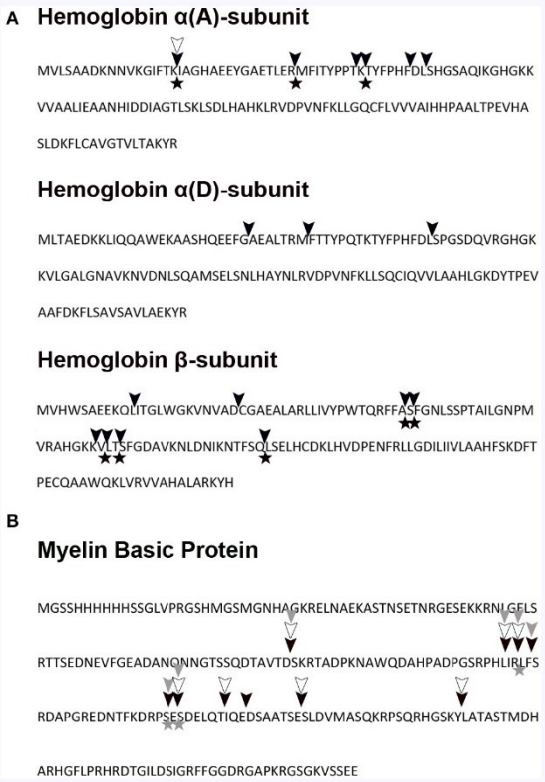


FIGURE 11 | Hemoglobin and myelin basic protein pepTrCB1 cleavage map. **(A)** Turkey hemoglobin and **(B)** recombinant human myelin basic protein were digested *in vitro* with pepTrCB forms at pH 5.5. The fragments were identified by mass spectrometry, and corresponding cleavage sites are indicated in hemoglobin and myelin basic protein sequences: cleavage by pepTrCB1.1 (black triangles), cleavage by pepTrCB1.4 (empty triangles), cleavage by pepTrCB1.6G/C (gray triangles). The cleavage map combines data from digests at three time points. The initial cleavage sites (after 5 min of reaction) are marked with asterisks. Only unique fragments which were not detected at time zero are included.

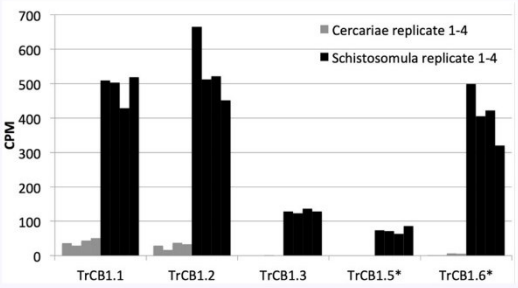


FIGURE 12 | Expression profiles of TrCB1 forms. Bar charts show the level of expression of different forms of cathepsin B1 in cercariae (gray) and schistosomula (black) of *T. regenti* based on mapping of RNA-seq data obtained from cercariae and 7-days old schistosomula. The level of expression is expressed in counts per million (CPM) of mapped reads. Each column is a cluster of four biological replicates. Inactive paralogs are indicated by asterisks.

specificity of the protease. Such a reservoir of genes would provide protection for the parasite against host adaptation or enhancement and expansion of host range" (Salter et al., 2002). Unfortunately, data on the level of TrCB1.4 expression in adult *T. regenti* is not available, thus currently it can be neither excluded nor confirmed if TrCB1.4 is employed by the mature worms in order to increase the total amount or broaden the specificity of peptidases produced in the gut.

Surprising and unexpected was the relatively high expression of TrCB1 inactive paralogs, especially TrCB1.6, which was comparable to that of the active enzymes such as TrCB1.1 in the stage of schistosomulum. This raises a serious question about the importance of the catalytically inactive forms for the biology of the parasite. As expected, recombinant pro-TrCB1.6wt was incapable of auto-processing. However, the presence of an Asp residue in the pro-sequence five amino acids upstream the beginning of the mature domain indicated a potential for processing by an asparaginyl endopeptidase (AE), similarly to SmCB1 (Sajid et al., 2003) and also TrCB1.1 (Dvořák et al., 2005). This was confirmed by incubation with recombinant IrAE (Sojka et al., 2007), which resulted in a single band of ≈30 kDa formed through several intermediates arisen by specific cleavage at sites with Asn and Asp at P1 position. This suggests that the transcripts coding for AE found in the transcriptome of *T. regenti* (Leontovyč et al., 2016, 2019) may be involved in the removal of the pro-sequence from pro-TrCB1.6 *in vivo*. Also, trans-processing of pro-TrCB1.6wt with pepsin-processed active forms pepTrCB1.1 and pepTrCB1.4 resulted in the formation of a major (≈32 kDa) and a minor (≈30 kDa) product in both cases. The 32-kDa product was generated by cleavage between Leu^{64} and Gly^{65}(pro-TrCB1.6 numbering). Unfortunately, the N-terminal sequence of the 30 kDa product could not be determined, but the size almost corresponded to that of the product obtained by IrAE processing. This data indicated that *in vivo* trans-processing of inactive forms of TrCB1 by the active ones is also possible.

For a more detailed view on the inactive TrCB1.6, a site-directed mutant TrCB1.6G/C (reverse $G^{29}C$ mutation) was expressed in *P. pastoris* system and partially characterized in comparison to the recombinant TrCB1.1 and TrCB1.4. The reverse mutation (Gly^{29} > Cys^{29}) of TrCB1.6wt resulted in the restoration of hydrolytic activity toward Z-Phe-Arg-AMC, but the mutant enzyme TrCB1.6G/C was not able to cleave cathepsin B-specific substrate Z-Arg-Arg-AMC, much like TrCB1.4. This feature is caused by uncharged Ala^{245} (TrCB1.4) or Leu^{245} (TrCB1.6) at the position, where negatively charged Glu^{245} typically occurs in the S2 pocket of cathepsins B (Dvořák et al., 2005).

Moreover, TrCB16G/C activity was not inhibited by CA-074 inhibitor (mostly specific for cathepsins B). Because the occluding loop of cathepsins B is important for the interaction with CA-074 (Yamamoto et al., 1997), in which the prolyl component of CA-074 forms hydrogen bonds with the conserved

His^{110} and His^{111} of the loop (bovine cathepsin numbering), the explanation may be that the latter histidine is substituted by an Asn residue in TrCB1.6, which causes the lack of susceptibility to CA-074, and also the virtual lack of exopeptidase activity. In human cathepsin B, two salt-bridge interactions between the occluding loop and the main body of the mature enzyme (Asp^{22}-His^{110} and Arg^{116}-Asp^{224}) anchor the loop down in the active site, and unpaired His^{111} forms a salt bridge with the C-terminus of an exo-substrate (Musil et al., 1991; Illy et al., 1997). According to another study (Krupa et al., 2002), His^{111} is not critical for exopeptidase activity, although the level of this activity might be somewhat impaired by the absence of this residue. This means that also other substitutions in the sequence of TrCB1.6 are probably responsible for the disruption of exopeptidase activity in the reverse Gly^{29} > Cys^{29} mutant. The substitution Asp^{224} > Gly^{224} in the sequence of TrCB1.6, when compared to the active TrCB1 forms or other cathepsins B, seems a likely candidate, since the removal of the salt bridges between the loop and the enzyme core effectively eliminates exopeptidase activity (Illy et al., 1997).

The presence of the occluding loop was proposed to be responsible for the generally lower affinities of cathepsins B to cystatins, compared to other endopeptidases of papain-like family with open active sites (Pavlova et al., 2000). Reduced electrostatic interaction between the loop and the body of the mature TrCB1.6 might cause greater affinity to cystatin. Based on this, we hypothesized that the inactive TrCB1.6 might sequester cystatins and thus protect active TrCB1 forms or other cysteine endopeptidases present in the vicinity of this paralog from inactivation by host cystatin(s). However, repeated attempts to inhibit the peptidolytic activity of TrCB1.1 and 1.4 toward Z-Phe-Arg-AMC with chicken and human cystatins failed. Nevertheless, SDS-stable enzyme/inhibitor complexes were formed after overnight incubation of all processed TrCB1 forms with chicken cystatin, including the catalytically inactive pepTrCB1.6wt. No complex was formed with the non-processed pro-TrCB1.6wt, which suggests that the pro-sequence blocked cystatin binding to the "active site" of the paralog. Binding of chicken cystatin to TrCB1.6G/C resulted even in larger proportion of the complex and prevented the enzyme from autodegradation. Surprisingly, human cystatins B and C did not form a complex with any of the TrCB1 forms after 2 h incubation.

The TrCB1.6G/C mutant degraded most of the protein substrates similarly with TrCB1.4. It seems that the binding pockets (except for the substitution in the active site) of TrCB1.6 are intact; thereby TrCB1.6 might still preserve substrate-binding properties. Thus, the possibility that the TrCB1.6 paralog competes with the active homologs for substrates and participates in this way in post-translational regulation of the activity of these enzymes cannot be omitted.

In a previous study, active TrCB1.1 and TrCB2 were shown to trigger nitric oxide production in murine astrocyte cultures, which could be harmful to both the parasite and the host (Macháček et al., 2016). Therefore, we wanted to test whether also the TrCB1.6 paralog has any effect on the immune response of the host. We exposed immune cells naturally occurring around *T. regenti* schistosomula migrating in murine spinal cord (astrocytes, microglia, and macrophages) (Lichtenbergová et al., 2011) to recombinant TrCB1.6wt *in vitro*. Surprisingly, it neither stimulated nor inhibited nitric oxide production in the naïve or LPS-treated cells, respectively. TrCB1.6wt also did not alter the viability of these cells. Perhaps, the hydrolytic activity of TrCB1.1 and TrCB2 was somehow responsible for the above stated effect.

In conclusion, the characterization of the two active forms TrCB1.1 and TrCB1.4 showed that the enzymes differ from each other in various properties. The levels of relative expression of TrCB1 paralogs in cercariae and schistosomula indicate that both the active forms and the inactive mutants are important for schistosomula migrating and developing in the definitive host, except for TrCB1.4. This enzyme form may be expressed even later in adult worms and this supposition needs to be further verified. The ability of the TrCB1.6 paralog with non-functional, Cys-to-Gly mutated active site to bind cystatin supports the hypothesis of inhibitor sequestration, although the efficacy appeared to be low in our experimental setup. Restoration of the active site of TrCB1.6 resulted in endopeptidase but not exopeptidase activity of the paralog. This elucidated important structural features in TrCB1.6 caused by other mutations in the occluding loop and the main body of the mature enzyme, which either regulate the exopeptidase activity of cathepsin B or may also adjust the affinity of cathepsin B to cystatins. Resolving of 3-D stucture of inactive TrCB1.6 + cystatin complex could help to elucidate the nature of the binding interactions, which is also the case of TrCB1.6G/C with restored active site. Finally, the inability of TrCB1.6 paralog to affect viability and production of nitric oxide by selected immune cells does not exclude other possible interactions with host physiology/immunity, which are a subject to further investigation in this unique and fascinating neurotropic schistosome.

MATERIALS AND METHODS

Parasite Material

Trichobilharzia regenti is routinely maintained in the Department of Parasitology, Charles University through domestic ducks (*Anas platyrhynchos* f. *dom.*) and lymnaeid snails (*Radix lagotis*) as definitive and intermediate hosts, respectively. Schistosomula were obtained from the spinal cord of infected ducks 9 days post-infection, washed in sterile phosphate-buffered saline (PBS) and used for RNA isolation and cDNA preparation as described previously (Dvořák et al., 2005).

Ethics statement: All procedures including animals were performed in concordance with the legislation of the Czech Republic (246/1992 and 359/2012) and the European Directive (2010/63/EU). All experiments were approved with the legal consent of the Professional Ethics Committee of the Faculty of Science, Charles University, i.e., the relevant institutional ethics committee for animal research, and of the Research and Development Section of the Ministry of Education, Youth, and Sports of the Czech Republic (approval no. MSMT-31114/2013-9). The animal facility, its equipment, animal welfare, and accompanying services, including the maintenance of experimental animals, have been approved by the Section of

Animal Commodities of the Ministry of Agriculture of the Czech Republic (approval no. 13060/2014-MZE-17214).

Expression and Purification of Recombinant TrCB1 Forms

Recombinant TrCB1.1 and TrCB1.4 were acquired from frozen glycerol stocks of *P. pastoris* transformants and expressed according to a previously published protocol (Dvořák et al., 2005). A column packed with Macro-Prep High S Support medium (Bio-Rad) was used for pre-purification from yeast media by cation exchange chromatography. TrCB1-containing fractions were treated with Endoglycosidase F1 (Calbiochem) following manufacturer's instructions and the pro-enzymes were further purified to homogeneity on a Mono S 5/50 GL column pre-equilibrated with 20 mM sodium acetate buffer pH 5.5 and employing 0–1 M linear NaCl gradient.

For the expression of recombinant wild-type TrCB1.6 (TrCB1.6wt), purified first-strand cDNA obtained from *T. regenti* schistosomula was used as a template for PCR with gene specific primers (**Supplementary Table 2**). PCR products were inserted into the expression vector pPICZα B (Thermo Fisher Scientific) and constructs were sequenced for verification (DNA Sequencing Laboratory, Faculty of Science, Charles University, Prague). To generate the Gly^{29}-to-Cys^{29} mutation (TrCB1.6G/C), the pPICZα B vector containing the sequence of TrCB1.6wt was mutagenized according to the instruction in QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). The mutated vector was generated by PCR with oligonucleotide primers (each complementary to the opposite strand of the vector) containing the desired mutation (**Supplementary Table 3**). Following PCR, the product was treated with Dpn I endonuclease to digest methylated parental DNA template. Constructs were sequenced to verify the mutations.

The vectors containing sequences of TrCB1.6wt or TrCB1.6G/C were then electroporated into *P. pastoris* X-33 cells and recombinant proteins were produced as described above. Both TrCB1.6wt and TrCB1.6G/C were expressed with a polyhistidine affinity tag (6x His-tag) situated on the C-terminus. The proteins were purified by Ni^{2+} chelating chromatography (HisTrap^{TM}FF crude, GE Healthcare Life Science) equilibrated with 100 mM phosphate buffer pH 7 with 300 mM NaCl and 40 mM imidazole. A linear gradient of 100–500 mM imidazole in the same buffer was used for elution. The eluted proteins were deglycosylated (see above) and then purified to homogeneity by cation exchange FPLC (BioLogic, Bio-Rad) on a Mono S 5/50 GL column (GE Healthcare Life Science) equilibrated with 20 mM phosphate buffer pH 7.2 and using a linear gradient of 0–1 M NaCl.

Protein concentrations were measured by Quant-iT Protein Assay Kit (ThermoFisher Scientific). Molecular size of purified enzymes was estimated by SDS-PAGE and peptidolytic activity was measured with the peptidyl substrate Z-Phe-Arg-AMC (see section Peptidase Activity). Fractions after chromatography were stored at −80°C or lyophilized.

Theoretical MW/pI values and potential N-glycosylation sites were determined by the Compute Mw software and NetNGlyc 1.0 server available at the ExPASy web portal (https://www.expasy.org/tools) (Artimo et al., 2012).

Activation/Processing of Recombinant Enzymes

The possibilities of auto- and heterologous processing of pro-TrCB1 zymogens were investigated in order to produce mature (pro-sequence-free) fully active enzymes for subsequent analyses.

(A) The pro-TrCB1.1 and 1.4 zymogens (0.2 – 0.3 μg/μl) were incubated separately at 37°C for 5, 30, 60, 120, and 240 min in 50 mM sodium acetate buffer (pH 4, 4.5, 5, and 5.5) containing 2 mM DTT (Sigma Aldrich) either in the presence or absence of dextran sulfate (DS, MW 6.5–10 kDa; Sigma Aldrich) or heparin sodium salt (HSS, MW 6–30 kDa; Sigma Aldrich) at final concentrations of 10 or 20 μg/ml.

(B) The *trans*-processing of all recombinant pro-TrCB1 forms (0.2 – 0.3μg/μl) was catalyzed by pepsin immobilized on agarose beads (1U pepsin per 3.3 μg of recombinant pro-TrCB1; P0609 Sigma Aldrich) incubated at 37°C (pro-TrCB1.1, pro-TrCB1.4, and pro-TrCB1.6wt) or at RT (pro-TrCB1.6G/C) for 5, 30, 30, and 60 min in 50 mM sodium acetate buffer pH 4. The reaction was stopped by addition of 10 μM pepstatin A and the samples were centrifuged (1,000 × g, 5 min) to remove agarose-bound pepsin.

(C) The inactive pro-TrCB1.6wt was processed by recombinant asparaginyl endopeptidase of the hard tick *Ixodes ricinus*—IrAE (Sojka et al., 2007; kindly provided by Dr. Daniel Sojka, Institute of Parasitology, ASCR, Ceské Budějovice, Czech Rep.). One gram of the lyophilized yeast medium containing IrAE was dissolved in 7 ml of distilled water, concentrated and transferred to 500 μl of 50 mM sodium acetate buffer pH 4.5 containing 2 mM DTT using Amicon® Ultra 4 mL filters (MWCO 30 kDa; Millipore). IrAE was then incubated for 30 h at 37°C to allow self-processing, which was monitored fluorometrically with Z-Ala-Ala-Asn-AMC substrate (25 μM; Bachem). Subsequently, the pro-TrCB1.6wt (16.5 μg in 55 μl) was incubated with activated IrAE (5 μl, unknown concentration—see above) at 37°C in 50 mM sodium acetate buffer (pH4, 4.5, or 5) containing 2 mM DTT for 30,60, 120, and 240 min or 16 and 36 h.

(D) Alternatively, pro-TrCB1.6wt (0.2–0.3 μg/μl) was dissolved in 50 mM sodium acetate buffer pH 4.5 containing 2 mM DTT, 10 μM pepstatin A, and 10 μg/ml DS. Then the pepsin-activated pepTrCB1.1 or pepTrCB1.4 (2 μg/ml) was added overnight (up to 16 h) at 37°C. Controls contained 10 μM E-64.

Aliquots were taken from reaction mixtures at specified time intervals during incubations and catalytic activity of TrCB1 forms was measured with Z-Phe-Arg-AMC substrate (see section Peptidase Activity). Additionally, a shift in molecular weight was monitored by SDS-PAGE. N-terminal amino acid sequences of the proteins electroblotted onto PVDF membranes were identified by Edman degradation (Dr. Zdeněk Voburka, Laboratory of Medicinal Chemistry, IOCB, ASCR, Prague, Czech Rep.).

Active-site labeling of processed TrCB1 forms with a fluorescent probe BODIPY green-DCG-04 (Greenbaum et al., 2002; obtained from Dr. Doron C. Greenbaum, Tropical Disease Research Unit, Sandler Center for Basic Research in Parasitic Diseases, UCSF, USA—former affiliation) was carried out as described previously (Dvořák et al., 2005). The specificity of probe binding was verified by pre-incubation of controls with 10 μM irreversible cysteine peptidase inhibitor E-64 for 5 min.

Peptidase Activity

The peptidase activity of TrCB1 forms processed as described above (0.45 μg TrCB1 per well, i.e., 70 nM) was assayed with two synthetic fluorogenic substrates (25 μM, Bachem): Z-Phe-Arg-AMC, (cathepsin L/B substrate) and Z-Arg-Arg-AMC (cathepsin B-selective substrate) in 50/100 mM citrate/phosphate buffer (CPB), 2 mM DTT, pH 3–8 (final volume 200 μl). The reactions were performed in 96-well black flat bottom plates (Nunc) using Infinite M200 fluorometer (TECAN) with excitation and emission wavelengths set to 360 and 465 nm, respectively. The release of AMC was measured at RT in a 30 min kinetic cycle at 2 min intervals.

An exopeptidase (peptidyl dipeptidase) activity assay using Bz-Gly-His-Leu as a substrate was performed with pepTrCB1 forms employing a modified protocol (Sajid et al., 2003). The pepTrCB1 forms (1 μg, 160 nM) were incubated for 15 min with the substrate in 50/100 mM CPB pH 4–5.5 containing 2 mM DTT (final volume 100 μl). Spontaneous reaction of the emerging free amino groups of His-Leu with fluorescamine (0.05 mg/ml) was monitored in a fluorometer set to 390/475 nm excitation/emission wavelengths during a 30 min kinetic cycle at 2 min intervals.

Inhibition Assays

The effect of two inhibitors, E-64 and CA-074 (Sigma Aldrich) on endopeptidase activity of the pepTrCB1 forms (0.2 μg, 30 nM) was tested with Z-Phe-Arg-AMC. The concentrations of inhibitors ranged from 0.01 to 10 μM in the case of pepTrCB1.1 and pepTrCB1.4, and from 0.01 to 100 μM with pepTrCB1.6G/C. The samples were pre-incubated with inhibitors for 15 min prior to the measurement. Inhibition assays were carried out at the pH optima of the particular enzymes (pH 5.5 for TrCB1.1, pH 5 for TrCB1.4, and pH 6.5 for pepTrCB1.6G/C). The remaining relative enzyme activity was expressed in relation to control reactions performed in the absence of the inhibitors.

The inhibition assays of exopeptidase activity of the enzymes were performed with E-64 and CA-074 (10 μM) added prior to the substrate.

The effect of chicken egg white cystatin, human cystatin B, and human cystatin C (C8917, C4249, and H5041, Sigma Aldrich) on TrCB1 forms by was tested using dextran sulfate-activated and pepsin-activated enzymes ds/pepTrCB1.1, ds/pepTrCB1.4, ds/pepTrCB1.6G/C. Processed TrCB1 forms (0.2 μg, 30 nM) were pre-incubated with various concentrations of cystatins (0–1 μM) for 15 min prior to the measurement with 25 μM Z-Phe-Arg-AMC in the presence or absence of 2 mM DTT. All samples contained 0.01% non-ionic detergent SP BRIJ L23 (Croda).

Formation of Peptidase-Cystatin Complex

The dsTrCB1.1, dsTrCB1.4, dsTrCB1.6G/C, and pepTrCB1.6wt (3 μg, 3.75 μM) were incubated with chicken egg white cystatin (1 μg, 3 μM) in 50/100 mM CPB pH 5.5 containing 0.01 % BRIJ L23 for 15 min at RT or overnight at 4°C. The pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6 were incubated with human cystatin B and C for 2 h at RT (concentrations as above). Controls did not contain cystatins. Then the samples were mixed with non-reducing sample buffer and allowed to stand for 15 min at RT. The samples were separated by SDS-PAGE under mild conditions (80 V).

Subsite Specificity Exploration of TrCB1.1 and TrCB1.4 by a Positional Scanning Synthetic Combinatorial Library (PS-SCL)

Purified non-processed pro-TrCB1.1 and pro-TrCB1.4 were used for the analysis of S1 – S4 subsite preferences toward P1–P4 residues of substrates. Since both pro-enzymes were able to cleave small peptide substrates, we had chosen this possibility to eliminate a potential risk of residual pepsin activity in pepsin-processed TrCB1 samples, which could bias the results. PS-SCL was applied as described previously (Choe et al., 2006). The peptidolytic activity of purified pro-enzymes (0.5 μg, 70 nM) was measured at RT in 50/100 mM CPB pH 5 containing 100 mM NaCl and 2 mM DTT.

Hydrolysis of Protein Substrates by TrCB1 Forms

The substrates (all Sigma-Aldrich) included turkey hemoglobin (Hb, H0142), bovine myelin basic protein (MBP, M1891), myosin (M7266), human albumin (A9511), collagen I (C7774), collagen IV (C8374), IgG (I4506), and fibrinogen (F3879) at final concentrations of 0.5 mg/ml, which were incubated for 6 h with pepTrCB1 forms (1 μg) in 50 μl of 50/100mM CPB pH 5.5 containing 2 mM DTT and 10 μM pepstatin A. Control reactions did not contain the enzymes. Aliquots of 20 μl of the reaction mixtures were separated by SDS-PAGE. Optimal pH for hydrolysis was ascertained using Hb and MBP as substrates in pH range 4.5–6.5.

Identification of Hb and MBP Fragments by Mass Spectrometry

For MALDI MS/MS analysis, turkey Hb (0.5 mg/ml) or recombinant human MBP (0.2 mg/ml, ab171675, Abcam) were incubated with 1 μg of pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6G/C at 37°C in 25/50 mM CPB pH 5.5 containing 2 mM DTT and 10 μM pepstatin A, total reaction volume was 60 μl. Aliquots of 10 μl were taken after 5, 30, and 120 min and immediately treated with 0.5% trifluoroacetic acid. Control reactions did not contain the enzymes. Buffer employed for TrCB1 processing reactions that was incubated with pepsin-agarose beads was used as another control to assess possible residual activity of pepsin in the pepTrCB1-substrate samples.

MALDI-MS and MALDI-MS/MS were performed using Ultraflextreme instrument (Bruker Daltonics, Germany) operated in the linear or reflectron mode with detection of positive ions. Ferulic acid (12.5 mg/ml in water: acetonitrile:

formic acid, 50:33:17 v/v mixture) or alpha-cyano-4-hydroxycinnamic acid (saturated solution in water: acetonitrile: trifluoroacetic acid, 47.5:50.0:2.5 v/v mixture), respectively, were used as the MALDI matrices. Detected peptide ions were subjected to MS/MS analysis (in a LIFT arrangement) and the resulting fragments were then searched against the primary amino acid sequences of TrCB1 forms [GenBank: AY648119 – 24], porcine pepsinogen [AAA31095.1], turkey α-A-globin [AAB35442.1], α-D-globin [AAB35443.1], β-globin [P02113.1], and human MBP (sequence available on http://www.abcam.com/recombinant-human-myelin-basic-protein-ab171675.html) using MASCOT 2.4 search engine (MatrixScience). No enzyme specificity and oxidation (M) and deamidation (N, Q) as variable modifications were set for all searches.

Differential Expression of TrCB1 Forms

Full-length sequences of TrCB1 genes [GenBank: AY648119 – AY648124] (Dvořák et al., 2005) combined with *T. regenti* juvenile transcriptomes (cercariae and 7 days old schistosomula) (Leontovyč et al., 2016) have been used as the reference for the differential expression analysis (the data can be accessed via the NCBI website https://www.ncbi.nlm.nih.gov/ under BioProject database ID: PRJNA292737). All transcripts homologous to cathepsin B ($n = 14$) (BLASTp; E-value cut off: $< 10^{-5}$) have been removed from *T. regenti* transcriptome and replaced by all six TrCB1 sequences from NCBI database. Trimmed and corrected Illumina reads (HiSeq 2500 platform) from *T. regenti* cercariae and schistosomula were mapped to the reference using RSEM (Li and Dewey, 2011), with zero mismatch rate (i.e., only 100% identical reads were mapped to the reference). Predicted expected counts were rounded to the highest whole number and used for differential gene transcription analysis in edgeR v.3.6.7 (Robinson et al., 2010) and R v.3.2.3 (R Development Core Team, 2015) software packages. Cathepsin B sequences with more than $\log_2$ fold change and false discovery rate of ≤ 0.01 were stated as differentially expressed between cercariae and schistosomula.

In vitro Effect of Pro-TrCB1.6wt on Murine Astrocytes, Microglia and RAW 264.7 Macrophages

Primary murine astrocyte and microglia cultures were obtained from mixed glial cultures as described elsewhere (Macháček et al., 2016). These cells were used because they naturally occur and are activated in the vicinity of *T. regenti* schistosomula migrating in murine spinal cord (Lichtenbergová et al., 2011). Murine RAW 264.7 macrophages were received from Dr. Tereza Leštinová (Charles University, Prague). The cells were seeded in 96-well plates (Nunclon Delta Surface, Thermo Fisher Scientific) and exposed to 1 μg/ml of purified and deglycosylated pro-TrCB1.6wt with or without 0.5 μg/ml of lipopolysaccharide from *E. coli* 0127:B8 (LPS, Sigma Aldrich) for 48 h at 37°C. The medium itself or medium supplemented with 0.5 μg/ml of LPS was used as negative or positive control, respectively. Griess assay was used to analyze NO production by stimulated cells (Macháček et al., 2016). Cell viability after stimulation was assessed by fluorescein diacetate (Sigma Aldrich) according to manufacturer's instructions. The results were evaluated by ordinary one-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons test, performed in GraphPad Prism, version 6.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI BioProject ID: PRJNA292737.

ETHICS STATEMENT

The animal study was reviewed and approved by Professional Ethics Committee of the Faculty of Science, Charles University.

AUTHOR CONTRIBUTIONS

HD performed expression of the enzymes, their processing, activity and inhibition assays, participated in protein purification, prepared samples for mass spectrometry, participated in interpretation of experimental data and preparation of figures, and wrote a larger part of the manuscript. RL obtained transcriptomic data, processed and interpreted the data on differential enzyme expression. TM performed immunological experiments and interpreted the data. AO'D and HD obtained and interpreted PS-SCL data. OŠ and ZZ obtained and interpreted mass spectrometry data. CSC provided know-how, facility and material for PS-SCL experiments, and performed critical reading of the whole manuscript. CRC and HD performed site-directed mutagenesis in TrCB1.6. AO'D and CRC provided critical comments on technical performance and interpretation of the results. PH revised the manuscript critically for important intellectual content. LM prepared conception, coordinated experiments, participated in protein purification, and cystatin-binding assays, contributed to interpretation of experimental data and finalized the manuscript. All authors contributed to writing or finalization of the manuscript.

FUNDING

The work on the project was funded by Czech Science Foundation (Grant Nos. 13-29577S and 18-11140S) and by the project Center for Research of Pathogenicity and Virulence of Parasites (No. CZ.02.1.01/0.0/0.0/16_019/0000759) funded by European Regional Development Fund and Ministry of Education, Youth and Sports of the Czech Republic. Charles University institutional support (PROGRES Q43, UNCE/SCI/012 - 204072/2018, and SVV 244-260432/2017) applied to RL, TM, PH, and LM. CIISB research infrastructure projects LM2015043 and LM2018127 funded by Ministry of Education, Youth and Sports of the Czech Republic are gratefully acknowledged for financially supporting the MALDI-MS measurements at the CEITEC Proteomics Core Facility. OS and ZZ thank for financial support of the project CEITEC 2020 (LQ1601) funded by Ministry of Education, Youth and Sports of the Czech Republic. Computational resources were

provided via the Melbourne Bioinformatics Platform and partly by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme Projects of Large Research, Development, and Innovations Infrastructures. CSC and AO'D were funded by P50 AI150476 and R01 GM104659 from the National Institute of Health.

ACKNOWLEDGMENTS

We thank Dr. Daniel Sojka (Institute of Parasitology, Biology Center of the Czech Academy of Sciences, Ceské Budějovice, Czech Republic) for providing recombinant tick legumain. We are also grateful to Dr. Tereza Leštinová (Department of Parasitology, Charles University, Prague, Czech Republic), who kindly provided murine RAW 264.7 macrophages, and Dr. Zdeněk Voburka (Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic), who performed N-terminal protein sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.00066/full#supplementary-material

REFERENCES

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi: 10.1093/nar/gks400

Barrett, A. J., and Kirschke, H. (1981). Cathepsin-B, cathepsin-H, and cathepsin-L. *Meth. Enzymol.* 80, 535–561. doi: 10.1016/S0076-6879(81)80043-2

Bergström, F. C., Reynolds, S., Johnstone, M., Pike, R. N., Buckle, A. M., Kemp, D. J., et al. (2009). Scabies mite inactivated serine protease paralogs inhibit the human complement system. *J. Immunol.* 182, 7809–7817. doi: 10.4049/jimmunol.0804205

Blažová, K., and Horák, P. (2005). *Trichobilharzia regenti*: the developmental differences in natural and abnormal hosts. *Parasitol. Int.* 54, 167–172. doi: 10.1016/j.parint.2005.03.003

Caffrey, C. R., Goupil, L., Rebello, K. M., Dalton, J. P., and Smith, D. (2018). Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS Neglect. Trop. Dis.* 12:e0005840. doi: 10.1371/journal.pntd.0005840

Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Salter, J. P., Sajid, M., Lucas, K. D., et al. (2000). Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Lett.* 466, 244–248. doi: 10.1016/S0014-5793(99)01798-6

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. doi: 10.1016/j.pt.2004.03.004

Chanová, M., and Horák, P. (2007). Terminal phase of bird schistosomiasis caused by *Trichobilharzia regenti* (Schistosomatidae) in ducks (*Anas platyrhynchos* f. *domestica*). *Folia Parasit.* 54, 105–107. doi: 10.14411/fp.2007.014

Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Brömme, D., et al. (2006). Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J. Biol. Chem.* 281, 12824–12832. doi: 10.1074/jbc.M513331200

Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvořák, J., Hsieh, I., et al. (2006). A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* 281, 39316–39329. doi: 10.1074/jbc.M607128200

Dolečková, K., Albrecht, T., Mikeš, L., and Horák, P. (2010). Cathepsins B1 and B2 in the neuropathogenic schistosome *Trichobilharzia regenti*: distinct gene expression profiles and presumptive roles throughout the life cycle. *Parasitol. Res.* 107, 751–755. doi: 10.1007/s00436-010-1943-6

Donnelly, S., O'Neil, S. M., Stack, C. M., Robinson, M. W., Turnbull, L., Whitchurch, C., et al. (2009). Helminth cysteine proteases inhibit TRIF-dependent activation of macrophages via degradation of TLR3. *J. Biol. Chem.* 285, 3383–3392. doi: 10.1074/jbc.M109.060368

Dvořák, J., Delcroix, M., Rossi, A., Vopálenský, V., Pospíšek, M., Šedinová, M., et al. (2005). Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. *Int. J. Parasitol.* 35, 895–910. doi: 10.1016/j.ijpara.2005.02.018

Greenbaum, D. C., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K., et al. (2002). Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* 1, 60–68. doi: 10.1074/mcp.T100003-MCP200

Hamuro, Y., Coales, S. J., Molnar, K. S., Tuske, S. J., and Morrow, J. A. (2008). Specificity of immobilized porcine pepsin in H/D exchange compatible conditions. *Rapid Commun. Mass Sp.* 22, 1041–1046. doi: 10.1002/rcm.3467

He, W., Ohashi, K., Sugimoto, C., and Onuma, M. (2005). *Theileria orientalis*: Cloning a cDNA encoding a protein similar to thiol protease with haemoglobin-binding activity. *Exp. Parasitol.* 111, 143–153. doi: 10.1016/j.exppara.2005.06.003

Holt, D. C., Fischer, K., Allen, G. E., Wilson, D., Wilson, P., Slade, R., et al. (2003). Mechanisms for a novel immune evasion strategy in the scabies mite *Sarcoptes Scabiei*: a multigene family of inactivated serine proteases. *J. Invest. Dermatol.* 121, 1419–1424. doi: 10.1046/j.1523-1747.2003.12621.x

Holt, D. C., Fischer, K., Pizzutto, S. J., Currie, B. J., Walton, S. F., and Kemp, D. J. (2004). A multigene family of inactivated cysteine proteases in *Sarcoptes scabiei*. *J. Invest. Dermatol.* 123, 240–241. doi: 10.1111/j.0022-202X.2004.22716.x

Horák, P., Dvořák, J., Kolářová, L., and Trefil, L. (1999). *Trichobilharzia regenti*, a pathogen of the avian and mammalian central nervous systems. *Parasitology* 119, 577–581. doi: 10.1017/S0031182099005132

Horák, P., Kolářová, L., and Dvořák, J. (1998). *Trichobilharzia regenti* n. sp. (Schistosomatidae, Bilharziellinae), a new nasal schistosome from Europe. *Parasite* 5, 349–357. doi: 10.1051/parasite/1998054349

Horák, P., Mikeš, L., Lichtenbergová, L., Skála, V., Soldánová, M., and Brant, S. V. (2015): Avian schistosomes and outbreaks of cercarial dermatitis. *Clin. Microbiol. Rev.* 28, 165–190. doi: 10.1128/CMR.00043-14

Horák, P., Mikeš, L., Rudolfová, J., and Kolářová, L. (2008). Penetration of *Trichobilharzia* cercariae into mammals: dangerous or negligible event? *Parasite* 15, 299–303. doi: 10.1051/parasite/2008153299

Horn, M., Jílková, A.,Vondrášek, J., Marešová, L., Caffrey, C. R., and Mareš, M. (2011). Mapping the pro-peptide of the *Schistosoma mansoni* cathepsin B1 drug target: modulation of inhibition by heparin and design of mimetic inhibitors. *ACS Chem. Biol.* 6, 609–617. doi: 10.1021/cb100411v

Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. S. (1997). Role of the occluding loop in cathepsin B activity. *J. Biol. Chem.* 272, 1197–1202. doi: 10.1074/jbc.272.2.1197

Jedličková, L., Dvořáková, H., Dvořák, J., Kašný, M., Ulrychová, L., Vorel, J., et al. (2018). Cysteine peptidases of *Eudiplozoon nipponicum*: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp. *Parasite. Vector.* 11, 1–17. doi: 10.1186/s13071-018-2666-2

Jílková, A., Horn, M., Řezáčová, P., Marešová, L., Fajtová, P., Brynda, J., et al. (2014). Activation route of the *Schistosoma mansoni* cathepsin B1 drug target: structural map with a glycosaminoglycan switch. *Structure* 22, 1786–1798. doi: 10.1016/j.str.2014.09.015

Kašný, M., Haas, W., Jamieson, B. G. M., and Horák, P. (2016). "Cercaria of *Schistosoma*," in *Schistosoma*: *Biology, pathology and control*, ed B. G. M. Jamieson (Boca Raton, FL: CRC Press; Taylor and Francis Group), 149–183. doi: 10.1201/9781315368900-9

Krupa, J. C., Hasnain, S., Nägler, D. K., Ménard, R., and Mort, J. S. (2002). S2' substrate specificity and the role of His110 and His111 in

the exopeptidase activity of human cathepsin B. *Biochem. J.* 361, 613–619. doi: 10.1042/bj3610613

Law, R. H. P., Smooker, P. M., Irving, J. A., Piedrafita, D., Ponting, R., Kennedy, N. J., et al. (2003). Cloning and expression of the major secreted cathepsin B-like protein from juvenile *Fasciola hepatica* and analysis of immunogenicity following liver fluke infection. *Infect. Immun.* 71, 6921–6932. doi: 10.1128/IAI.71.12.6921-6932.2003

Leontovyč, R., Young, N. D., Korhonen, P. K., Hall, R. S., Bulantová, J., Jeřábková, V., et al. (2019). Molecular evidence for distinct modes of nutrient acquisition between visceral and neurotropic schistosomes of birds. *Sci. Rep.* 9:1347. doi: 10.1038/s41598-018-37669-2

Leontovyč, R., Young, N. D., Korhonen, P. K., Hall, R. S., Tan, P., Mikeš, L., et al. (2016). Comparative transcriptomic exploration reveals unique molecular adaptations of neuropathogenic *Trichobilharzia* to invade and parasitize its avian definitive host. *PLoS Neglect. Trop. Dis.* 10, 1–24. doi: 10.1371/journal.pntd.0004406

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Lichtenbergová, L., Lassmann, H., Jones, M. K., Kolářová, L., and Horák, P. (2011). *Trichobilharzia regenti*: host immune response in the pathogenesis of neuroinfection in mice. *Exp. Parasitol.* 128, 328–335. doi: 10.1016/j.exppara.2011.04.006

Lipps, G., Fü, R., and Beck, E. (1996). Cathepsin B of *Schistosoma mansoni* - Purification and activation of the recombinant proenzyme secreted by *Saccharomyces cerevisiae*. *J. Biol. Chem.* 271, 1717–1725. doi: 10.1074/jbc.271.3.1717

Macháček, T., Panská, L., Dvořáková, H., and Horák, P. (2016). Nitric oxide and cytokine production by glial cells exposed *in vitro* to neuropathogenic schistosome *Trichobilharzia regenti*. *Parasite. Vector.* 9:579. doi: 10.1186/s13071-016-1869-7

Macháček, T., Turjanicová, L., Bulantová, J., Hrdý, J., Horák, P., and Mikeš, L. (2018). Cercarial dermatitis: a systematic follow-up study of human cases with implications for diagnostics. *Parasitol. Res.* 117, 3881–3895. doi: 10.1007/s00436-018-6095-0

McCoubrie, J. E., Miller, S. K., Sargeant, T., Good, R. T., Hodder, A. N., Speed, T. P., et al. (2007). Evidence for a common role for the serine-type *Plasmodium falciparum* serine repeat antigen proteases: implications for vaccine and drug design. *Infect. Immun.* 75, 5565–5574. doi: 10.1128/IAI.00405-07

Mendoza-Palomares, C., Biteau, N., Giroud, C., Coustou, V., Coetzer, T., Authié, E., et al. (2008). Molecular and biochemical characterization of a cathepsin B-like protease family unique to *Trypanosoma congolense*. *Eukaryot. Cell* 7, 684–697. doi: 10.1128/EC.00405-07

Merckelbach, A., Hasse, S., Dell, R., Eschlbeck, A., and Ruppel, A. (1994). cDNA sequences of *Schistosoma japonicum* coding for 2 cathepsin B-like proteins and Sj32. *Trop. Med. Parasitol.* 45, 193–198.

Mort, J. S., and Buttle, D. J. (1997). Cathepsin B. *Int. J. Biochem. Cell Biol.* 29, 715–720. doi: 10.1016/S1357-2725(96)00152-5

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., et al. (1991). The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.* 10, 2321–2330. doi: 10.1002/j.1460-2075.1991.tb07771.x

Pavlova, A., Krupa, J. C., Mort, J. S., Abrahamson, M., and Björk, I. (2000). Cystatin inhibition of cathepsin B requires dislocation of the proteinase occluding loop. Demonstration by release of loop anchoring through mutation of His110. *FEBS Lett.* 487, 156–160. doi: 10.1016/S0014-5793(00)02337-1

Pillay, D., Boulangé, A. F., and Coetzer, T. H. T. (2010). Expression, purification and characterisation of two variant cysteine peptidases from *Trypanosoma congolense* with active site substitutions. *Protein Expres. Purif.* 74, 264–271. doi: 10.1016/j.pep.2010.06.021

R Development Core Team (2015). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Reynolds, S. L., Pike, R. N., Mika, A., Blom, A. M., Hofmann, A., Wijeyewickrema, L. C., et al. (2014). Scabies mite inactive serine proteases are potent inhibitors of the human complement lectin pathway. *PLoS Neglect. Trop. Dis.* 8:e2872. doi: 10.1371/journal.pntd.0002872

Řimnáčová, J., Mikeš, L., Turjanicová, L., Bulantová, J, and Horák, P. (2017): Changes in surface glycosylation and glycocalyx shedding in *Trichobilharzia regenti* (Schistosomatidae) during the transformation of cercaria to schistosomulum. *PLoS ONE* 12:e0173217. doi: 10.1371/journal.pone.0173217

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I., et al. (2003). Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol.* 131, 65–75. doi: 10.1016/S0166-6851(03)00194-4

Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K. C., Craik, C. S., et al. (2002). Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. *J. Biol. Chem.* 277, 24618–24624. doi: 10.1074/jbc.M202364200

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157–62. doi: 10.1016/S0006-291X(67)80055-X

Sojka, D., Hajdušek, O., Dvořák, J., Sajid, M., Franta, Z., Schneider, E. L., et al. (2007). IrAE - An asparaginyl endopeptidase (legumain) in the gut of the hard tick *Ixodes ricinus*. *Int. J. Parasitol.* 37, 713–724. doi: 10.1016/j.ijpara.2006.12.020

Soloviova, K., Fox, E. C., Dalton, J. P., Caffrey, C. R., and Davies, S. J. (2019). A secreted schistosome cathepsin B1 cysteine protease and acute schistosome infection induce a transient T helper 17 response. *PLoS Neglect. Trop. Dis.* 13:e0007070. doi: 10.1371/journal.pntd.0007070

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. doi: 10.1016/S1471-4922(03)00189-2

Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., et al. (2004). A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *J. Biol. Chem.* 279, 35950–35957. doi: 10.1074/jbc.M405842200

Yamamoto, A., Hara, T., Tomoo, K., Ishida, T., Fujii, T., Hata, Y., et al. (1997). Binding mode of CA074, a specific irreversible inhibitor, to bovine cathepsin B as determined by X-Ray crystal analysis of the complex. *J. Biochem.* 121, 974–977. doi: 10.1093/oxfordjournals.jbchem.a021682

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Dvořáková, Leontovyč, Macháček, O'Donoghue, Šedo, Zdráhal, Craik, Caffrey, Horák and Mikeš. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Concluding remarks

In contrast to the comprehensive knowledge concerning peptidases from medically or economically important parasites, only very limited information is available about those enzymes derived from monogeneans. Our work thus represents the first complex bioinformatic and biochemical characterization of acid endopeptidases of monogenean origin. Unfortunately, the lack of peptidase research in case of other monogeneans does not allow any comparisons to be made. According to our work, we believe that cathepsin L-like cysteine peptidases, in cooperation with the aspartic peptidase cathepsin D, play a leading role in the digestion of haemoglobin and other proteins in the sanguinivorous monogenean *E. nipponicum*. The dominance of cathepsins L in *E. nipponicum* recalls the situation in, e.g., fasciolid flukes. We have also shown that *E. nipponicum* adults express many different cathepsin L transcripts; some of them display structural features rather unusual in cathepsins L., especially in the composition of S2 subsites. This finding expands general understanding of structural features of cathepsins L. Since the S2 subsites of cysteine peptidases are responsible for interaction with substrates, it seems that this diversity of *E. nipponicum* cathepsins L gives the parasite a wide array of proteolytic tools to degrade host-derived macromolecules more efficiently. However, although we know that ten cathepsin L types/isoforms are expressed in the adult *E. nipponicum*, the localisation in the parasite and the biochemical characterisation have been performed only with 2 out of 10 enzymes. These two cathepsins L termed EnCL1 and EnCL3are probably involved in digestion. The other cathepsin L types/isoforms of *E. nipponicum* may have different localisation and thus may be involved in other biological processes in the parasite body. According to pioneering ultrastructural and histochemical studies, blood processing by monogeneans has been believed to take place in the lysosomes, i.e., intracellularly, within the specialized types of gut cells. However, the visualisation of two cathepsins L in the lumen of the gut *E. nipponicum*, together with the presence of a signal peptide in the EnCL3 molecule, create support for the possibility of extracellular part of the digestive process in diplozoid monogeneans. Moreover, the enzymes may be eventually released from the gut lumen into host blood circulation where they could potentially participate in parasite-host interactions by affecting, e.g., haemocoagulation, fibrinolysis or immune response: recombinant cathepsins L of *E. nipponicum* were able to cleave fibrinogen and immunoglobulins in vitro. In contrast to a wide repertoire of cathepsins L, a single cathepsin B was found in *E. nipponicum* transcriptome. Despite its ability to hydrolyse haemoglobin in vitro, its localisation in vesicular structures in vitelline cells and not in digestive tract, suggests that it is not involved in digestion, but probably in vitellogenesis.

The possible lack of exopeptidase activity during digestion due to the absence of cathepsin B in the digestive tract may be compensated by the activity of cathepsins C (dipeptidyl peptidases) and other exopeptidases such as leucine aminopeptidase (leucyl aminopeptidase), aminopeptidase P3 (Xaa-Pro aminopeptidase), amino-peptidase A (glutamyl aminopeptidase) and serine carboxypeptidase S28, which were also found in *E. nipponicum* transcriptome (Vorel et al., 2021) and in tissue-specific proteomes of the worm's gut (Roudnický et al., 2020). Nevertheless, this statement is only speculative, and it would be good to verify it experimentally.

Compared to the number of molecules investigated in medically or veterinary important parasites, very little research has been focused on functional proteins derived from monogeneans, although they include highly important pathogens such as *Gyrodactylus salaris*. This situation seems to be improved a little by several currently published papers focused on interesting bioactive molecules of *E. nipponicum*, namely cysteine peptidases cathepsins L and B, aspartic peptidases cathepsins D, and also peptidase inhibitors such as a Kunitz protein, serpin and cystatin (Ilgová et al., 2020, 2017; Jedličková et al., 2019, 2018, 2016; Roudnicky et al., 2020; Roudnický et al., 2018).This was easier in particular due to the fact that we have participated in the transcriptome project of the adult *E. nipponicum* (Vorel et al., 2021).Therefore it was possible to mine the data for the presence of various functional protein which could be further studied. The detailed characterization of the structures and functions of these molecules in monogeneans may lead to understanding molecular interactions between the fish host and this group of parasites, which would be of interesting especially in the context of pathogenesis and fish immunity (see in Ilgová et al., 2021). Additionally, such research is likely to provide data interesting from an evolutionary point of view and can open a possibility for future comparative biochemical research of the two monogenean trophic strategies, mucophagy and haematophagy, characteristic of the two groups of monogeneans, Polyonchoinea and Heteronchoinea, respectively.

Unlike the studied cathepsin B of *E. nipponicum* (EnCB), cathepsin B1of the avian schistosome *Trichobilhalzia regenti* (TrCB1) is responsible for the digestion of host protein in parasite's gut. Six forms (paralogs) of TrCB1with high sequence similarity were previously identified (Dvořák et al., 2005). However, even small changes in the amino acid sequence of an enzyme can affect the protein's overall structure and it is likely to alter substrate specificity. Therefore. I focused on in depth comparison of substrate preferences of the two most divergent paralogs, TrCB1.1 and TrCB1.4. The comparison confirmed that the enzymes have different affinities for peptide substrates thereby facilitating diversity in

protein-derived nutrition for the parasite. Detailed study of *T. regenti* transcriptomes (*s*chistosomula and cercariae) revealed that TrCB1.1 was the top expressed active isoform of TrCB1. The high expression of TrCB1.1, its poor ability to degrade haemoglobin and a high capacity to cleave myelin basic protein of the nervous tissue indicate that schistosomula of *T. regenti* employs TrCB1.1 as the main enzyme facilitating migration and nutrition up take within the CNS. On the other hand, our analysis revealed that transcripts of TrCB1.4 were not present at detectable levels in the transcriptome of cercariae or schistosomula. We suggested that TrCB1.4 is not stably expressed in schistosomula, and it is likely not essential for migrating worms in the definitive host. This enzyme form may be expressed even later in adult worms; unfortunately, data on the level of TrCB1.4 expression in other life stages is not available, therefore this supposition needs to be further verified.

An interesting phenomenon is that two of these six paralogs (namely TrCB1.5 and TrCB1.6) possess a substitution of the catalytic Cys for Gly. As a result, these forms are enzymatically inactive and their function in the biology of the parasite is still debatable. Surprising and unexpected was the relatively high expression of those inactive paralogs, especially TrCB1.6, which was comparable to TrCB1.1 in the stage of the schistosomulum. This finding indicates that the catalytically inactive forms play a significant role in the biology of *T. regenti* schistosomula. The ability of the TrCB1.6wt inactive paralog to bind cystatin supports the hypothesis that peptidases with mutation-abolished active sites could potentially regulate the activity of the enzymatically active forms by competition for substrates. They could also protect them, and potentially even other cysteine peptidases from being inhibited by host inhibitors such as cystatins. The site-directed mutation of TrCB1.6 active site resulted in restoration of endopeptidase but not exopeptidase (dipeptidase) activity of the paralog. We assume that the other mutations in the occluding loop (namely the His^{111} to Asn^{111}) and the substitution Asp^{224} to Gly^{224} (human cathepsin B numbering) in the main body of the processed TrCB1.6wt are probably responsible for the disruption of exopeptidase activity and also the lack of susceptibility to inhibitor CA-074 in its site-directed mutant TrCB1.6G/C. Nevertheless, the knowledge of 3-D structure of TrCB1.6 cystatin complex could help to elucidate the nature of the binding interactions. Despite the results of a previous study, where TrCB1.1 and TrCB2 (an enzyme from post-acetabular penetration glands of *T. regenti* cercariae) were shown to trigger nitric oxide production in murine astrocyte cultures (Macháček et al., 2016), we found out that TrCB1.6 paralog has no effect on cell viability and nitric oxide production of selected

immune cells (astrocytes, microglia, and macrophages). Perhaps, the complete active site and therefore the hydrolytic activity are somehow responsible for the above stated effect.

Even though the occurrence of enzymes with the active site mutations is more a rule than an exception in organisms (including parasites), their function remains unclear. This study represents to the best of my knowledge; the first deeper characterisation of an inactive paralog in the context of its biological functions and supports the hypothesis of inhibitor sequestration. However, it does not rule out other possible roles of this interesting molecule in parasite-host interactions and should be a subject to further research in the unique neurotropic *T. regenti.*

References

Abdulla, M.-H., Lim, K.-C., Sajid, M., McKerrow, J.H., Caffrey, C.R., 2007. Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. PLoS Med. 4, e14. https://doi.org/10.1371/journal.pmed.0040014

Almeida, P.C., Nantes, I.L., Chagas, J.R., Udia, C., Rizzi, C.A., Faljoni-Alario, A., Carmona, E., Julianoţ, L., Nader, H.B., Tersariol, I.L.S., 2001. Cathepsin B activity regulation. Heparin-like glycosaminoglycans protect human cathepsin b from alkaline ph-induced inactivation. J. Biol. Chem. 276, 944–951. https://doi.org/10.1074/jbc.M003820200

Atkinson, H.J., Babbitt, P.C., Sajid, M., 2009. The global cysteine peptidase landscape in parasites. Trends Mol. Med. 25, 573–581. https://doi.org/10.1016/j.pt.2009.09.006.The

Baici, A., Mü, K., Willimann, A., Zwicky, R., 2006. Regulation of human cathepsin B by alternative mRNA splicing: homeostasis, fatal errors and cell death. Biol. Chem 387, 1017–1021. https://doi.org/10.1515/BC.2006.125

Barr, S.C., Warner, K.L., Kornreic, B.G., Piscitelli, J., Wolfe, A., Benet, L., McKerrow, J.H., 2005. A cysteine protease inhibitor protects dogs from cardiac damage during infection by Trypanosoma cruzi. Antimicrob. Agents Chemother. 49, 5160–1. https://doi.org/10.1128/AAC.49.12.5160-5161.2005

Beckham, S.A., Law, R.H.-P., Smooker, P.M., Quinsey, N.S., Caffrey, C.R., McKerrow, J.H., Pike, R.N., Spithill, T.W., 2006. Production and processing of a recombinant Fasciola hepatica cathepsin B-like enzyme (FhcatB1) reveals potential processing mechanisms in the parasite. Biol. Chem. 387, 1053–1061. https://doi.org/10.1515/BC.2006.130

Benes, P., Vetvicka, V., Fusek, M., 2008. Cathepsin D-many functions of one aspartic protease. Crit. Rev. Oncol. Hematol. 68, 12–28. https://doi.org/10.1016/j.critrevonc.2008.02.008

Bergström, F.C., Reynolds, S., Johnstone, M., Pike, R.N., Buckle, A.M., Kemp, D.J., Fischer, K., Blom, A.M., 2009. Scabies Mite Inactivated Serine Protease Paralogs Inhibit the Human Complement System. J. Immunol. 182, 7809–7817. https://doi.org/10.4049/jimmunol.0804205

Bogitsh, B.J., Dalton, J.P., Brady, C.P., Brindley, P.J., 2001. Gut-associated immunolocalization of the Schistosoma mansoni cysteine proteases, smcl1 and Smcl2. J. Parasitol. 87, 237–241. https://doi.org/10.1645/0022-3395(2001)087[0237:GAIOTS]2.0.CO;2

Bovet, J., 1967. Contribution a la morphologie et a la biologie de Diplozoon paradoxum v. Nordmann, 1832. Bull. la Société Neuchatel. des Sci. Nat. 90, 63–159.

Brindley, P.J., Kalinna, B.H., Dalton, J.P., Day, S.R., Wong, J.Y., Smythe, M.L., McManus, D.P., 1997. Proteolytic degradation of host hemoglobin by schistosomes. Mol. Biochem. Parasitol. 89, 1–9.

Brindley, P.J., Smyth, Danielle Julianna, Fairlie, D., Milburn, P., Kalinna, B.H., Wong, J.Y.M., Bogitsh, B.J., King, L.T., Smyth, Danielle J, Verity, C.K., Abbenante, G., Brinkworth, R.I., Fairlie, D.P., Smythe, M.L., Milburn, P.J., Bielefeldt-Ohmann, H., Zheng, Y., Mcmanus, D.P., 2001. Proteolysis of human hemoglobin by schistosome cathepsin D. Artic. Mol. Biochem. Parasitol. 112, 103–112. https://doi.org/10.1016/S0166-6851(00)00351-0

Brinkworth, R.I., Prociv, P., Loukas, A., Brindley, P.J., 2001. Hemoglobin-degrading, aspartic proteases of blood-feeding parasites. Substrate specificity revealed by homology models. J. Biol. Chem. 276, 38844–38851. https://doi.org/10.1074/jbc.M101934200

Buchmann, K., Bresciani, J., 2006. Monogenea (phylum Platyhelminthes)., in: Woo,

P.T.K. (Ed.), Fish Diseases and Disorders: Protozoan and Metazoan Infections. CABI Publishing, Wallingford, pp. 297–344.
Butler, R., Michel, A., Kunz, W., Klinkert, M., 1995. Sequence of Schistosoma mansoni cathepsin C and its structural comparison with papain and cathepsins B and L of the parasite. Protein Pept. Lett. 2.
Buttle, D., Mort, J., 2004. Cysteine proteases, in: Lennarz, W., Lane, M. (Eds.), Encyclopedia of Biological Chemistry - Vol_1. Academic Press, pp. 516–520.
Caffrey, C.R., Britton, C., McKerrow, J.H., 2013. Helminth cysteine proteases, in: Rawlings, N., Salvesen, G. (Eds.), Handbook of Proteolytic Enzymes. Elsevier Science & Technology, pp. 1949–1957. https://doi.org/10.1016/B978-0-12-382219-2.00444-0
Caffrey, Conor R., Mathieu, M.A., Gaffney, A.M., Salter, J.P., Sajid, M., Lucas, K.D., Franklin, C., Bogyo, M., McKerrow, J.H., 2000. Identification of a cDNA encoding an active asparaginyl endopeptidase of Schistosoma mansoni and its expression in Pichia pastoris. FEBS Lett. 466, 244–248. https://doi.org/10.1016/S0014-5793(99)01798-6
Caffrey, C.R., McKerrow, J.H., Salter, J.P., Sajid, M., 2004. Blood "n" guts: An update on schistosome digestive peptidases. Trends Parasitol. 20, 241–248. https://doi.org/10.1016/j.pt.2004.03.004
Caffrey, C R, Scory, S., Steverding, D., 2000. Cysteine proteinases of trypanosome parasites: novel targets for chemotherapy. Curr. Drug Targets 1, 155–62.
Caglič, D., Pungerčar, J.R., Pejler, G., Turk, V., Turk, B., 2007. Glycosaminoglycans facilitate procathepsin B activation through disruption of propeptide-mature enzyme interactions. J. Biol. Chem. 282, 33076–33085. https://doi.org/10.1074/jbc.M705761200
Cancela, M., Acosta, D., Rinaldi, G., Silva, E., Durán, R., Roche, L., Zaha, A., Carmona, C., Tort, J.F., 2008. A distinctive repertoire of cathepsins is expressed by juvenile invasive Fasciola hepatica. Biochimie 90, 1461–1475. https://doi.org/10.1016/j.biochi.2008.04.020
Cappetta, M., Díaz, A., Roche, L., 2002. Role of the prosegment of fasciola hepatica cathepsin L1 in folding of the catalytic domain. Biol. Chem. 383, 1215–21. https://doi.org/10.1515/BC.2002.134
Carmona, E., Dufour, É., Plouffe, C., Takebe, S., Mason, P., Mort, J.S., Ménard, R., 1996. Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. Biochemistry 35, 8149–8157. https://doi.org/10.1021/bi952736s
Chan, V.J., Selzer, P.M., McKerrow, J.H., Sakanari, J.A., 1999. Expression and alteration of the S2 subsite of the Leishmania major cathepsin B-like cysteine protease. Biochem. J. 340 (Pt 1), 113–7.
Chaparro, J.D., Cheng, T., Tran, U.P., Andrade, R.M., Brenner, S.B.T., Hwang, G., Cohn, S., Hirata, K., McKerrow, J.H., Reed, S.L., 2018. Two key cathepsins, TgCPB and TgCPL, are targeted by the vinyl sulfone inhibitor K11777 in in vitro and in vivo models of toxoplasmosis. PLoS One 13, e0193982. https://doi.org/10.1371/journal.pone.0193982
Choe, Y., Leonetti, F., Greenbaum, D.C., Lecaille, F., Bogyo, M., Brömme, D., Ellman, J. a., Craik, C.S., 2006. Substrate Profiling of Cysteine Proteases Using a Combinatorial Peptide Library Identifies Functionally Unique Specificities. J. Biol. Chem. 281, 12824–12832. https://doi.org/10.1074/jbc.M513331200
Chowdhury, S.F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Ménard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T., Purisima, E.O., 2002. Design of noncovalent inhibitors of human cathepsin L. from the 96-residue proregion to optimized tripeptides. J. Med. Chem. 45, 5321–5329. https://doi.org/10.1021/jm020238t

Chung, Y.B., Kim, T.S., Yang, H.J., 2005. Early Cysteine Protease Activity in Excretory Bladder Triggers Metacercaria Excystment of Paragonimus westermani. J. Parasitol. 91, 953–954. https://doi.org/10.1645/GE-475R.1

Cornelissen, J.B.W.J., Gaasenbeek, C.P.H., Boersma, W., Borgsteede, F.H.M., Van Milligen, F.J., 1999. Use of a pre-selected epitope of cathepsin-L 1 in a highly specific peptide-based immunoassay for the diagnosis of Fasciola hepatica infections in cattle. Int. J. Parasitol. 29, 685–696.

Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J.S., Cygler, M., 1996. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. EMBO J. 15, 5492–5503. https://doi.org/066/14

Dalton, J.P., Brindley, P.J., 1996. Schistosome asparaginyl endopeptidase SM32 in hemoglobin digestion. Parasitol. Today 12, 125. https://doi.org/10.1016/0169-4758(96)80676-4

Dalton, J.P., Caffrey, C.R., Sajid, M., Stack, C., Donnelly, S., Loukas, A., Don, T., McKerrow, J., Halton, D.W., Brindley, P.J., 2006. Proteases in trematode biology., in: Maule, A., Marks, N. (Eds.), Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology. CAB International, Wallingford, pp. 348–368. https://doi.org/10.1079/9780851990279.0348

Delcroix, M., Sajid, M., Caffrey, C.R., Lim, K.C., Dvořák, J., Hsieh, I., Bahgat, M., Dissous, C., McKerrow, J.H., 2006. A Multienzyme Network Functions in Intestinal Protein Digestion by a Platyhelminth Parasite. J. Biol. Chem. 281, 39316–39329. https://doi.org/10.1074/jbc.M607128200

Denis, A., Gabrion, C., Lambert, A., 1983. Présence en france de deux parasites d'origine est-asiatique: Diplozoon nipponicum Goto, 1891 (Monogenea) et Bothriocephalus acheilognathi Yamaguti, 1934 (Cestoda) chez Cyprinus carpio (Teleostei, Cyprinidae). Bull. français la pêche la Piscic. 289, 128–134.

Dickinson, D.P., 2002. Cysteine peptidases of mammals: their biological roles and potential effects in the oral cavity and other tissues in health and disease. Crit. Rev. Oral Biol. Med. 13, 238–275. https://doi.org/10.1177/154411130201300304

Dolečková, K., Kašný, M., Mikeš, L., Cartwright, J., Jedelský, P., Schneider, E.L., Dvořák, J., Mountford, A.P., Craik, C.S., Horák, P., 2009. The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome Trichobilharzia regenti. Int. J. Parasitol. 39, 201–211. https://doi.org/10.1016/j.ijpara.2008.06.010

Doyle, P.S., Zhou, Y.M., Engel, J.C., McKerrow, J.H., 2007. A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. Antimicrob. Agents Chemother. 51, 3932–9. https://doi.org/10.1128/AAC.00436-07

Duncan, E.M., Muratore-Schroeder, T.L., Cook, R.G., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., 2008. Cathepsin L Proteolytically Processes Histone H3 During Mouse Embryonic Stem Cell Differentiation. Cell 135, 284–294. https://doi.org/10.1016/j.cell.2008.09.055

Dunn, B.M., Scarborough, P.E., Lowther, W.T., Rao-Naik, C., 1995. Comparison of the active site specificity of the aspartic proteinases based on a systematic series of peptide substrates. Adv. Exp. Med. Biol. 362, 1–9.

Dvořák, J., Delcroix, M., Rossi, A., Vopálenský, V., Pospíšek, M., Šedinová, M., Mikeš, L., Sajid, M., Sali, A., McKerrow, J.H., Horák, P., Caffrey, C.R., 2005. Multiple cathepsin B isoforms in schistosomula of Trichobilharzia regenti: Identification, characterisation and putative role in migration and nutrition. Int. J. Parasitol. 35, 895–910. https://doi.org/10.1016/j.ijpara.2005.02.018

Dvořák, J., Mashiyama, S.T., Sajid, M., Braschi, S., Delcroix, M., Schneider, E.L., McKerrow, W.H., Bahgat, M., Hansell, E., Babbitt, P.C., Craik, C.S., McKerrow, J.H., Caffrey, C.R., 2009. SmCL3, a gastrodermal cysteine protease of the human

blood fluke Schistosoma mansoni. PLoS Negl. Trop. Dis. 3. https://doi.org/10.1371/journal.pntd.0000449
Dvořáková, H., Leontovyč, R., Macháček, T., O'Donoghue, A.J., Šedo, O., Zdráhal, Z., Craik, C.S., Caffrey, C.R., Horák, P., Mikeš, L., 2020. Isoforms of Cathepsin B1 in Neurotropic Schistosomula of Trichobilharzia regenti Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin. Front. Cell. Infect. Microbiol. 10, 1–17. https://doi.org/10.3389/fcimb.2020.00066
Engel, J.C., Doyle, P.S., Hsieh, I., McKerrow, J.H., 1998. Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection. J. Exp. Med. 188, 725–34.
Ermolieffs, J., Boudiers, C., Laineg, A., Meyers, B., Bieths, J.G., 1994. Heparin Protects Cathepsin G against Inhibition by Protein Proteinase Inhibitors. J. Biol. Chem. 269, 29502–29508.
Fairhead, M., Kelly, S.M., Van Der Walle, C.F., 2007. A heparin binding motif on the pro-domain of human procathepsin L mediates zymogen destabilization and activation. https://doi.org/10.1016/j.bbrc.2007.12.062
Farady, C.J., Craik, C.S., 2010. Mechanisms Of Macromolecular Protease Inhibitors. Chembiochem 11, 2341. https://doi.org/10.1002/CBIC.201000442
Fox, T., De Miguel, E., Mort, J.S., Storer, a. C., 1992. Potent slow-binding inhibition of cathepsin B by its propeptide. Biochemistry 31, 12571–12576. https://doi.org/10.1021/bi00165a005
Gallego, G.S., Slade, R.W., Brindley, P.J., 1998. A cDNA encoding a pepsinogen-like, aspartic protease from the human roundworm parasite Strongyloides stercoralis. Acta Trop. 71, 17–26. https://doi.org/10.1016/S0001-706X(98)00050-3
Gillmor,', S.A., Craik, C.S., Fletterick3, R.J., 1997. Structural determinants of specificity in the cysteine protease cruzain, Protein Science. Cambridge University Press.
Goto, 1891. On Diplozoon nipponicum,n. sp. J. Coll. Sci. Imp. Univ. Tokyo 4, 151–192.
Goulet, B., Baruch, A., Moon, N.S., Poirier, M., Sansregret, L.L., Erickson, A., Bogyo, M., Nepveu, A., 2004. A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. Mol. Cell 14, 207–219. https://doi.org/10.1016/S1097-2765(04)00209-6
Guay, J., Falgueyret, J.P., Ducret, A., Percival, M.D., Mancini, J.A., 2000. Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. Eur. J. Biochem. 267, 6311–6318. https://doi.org/10.1046/j.1432-1327.2000.01730.x
Guiliano, D.B., Hong, X., Mckerrow, J.H., Blaxter, M.L., Oksov, Y., Liu, J., Ghedin, E., Lustigman, S., 2004. A gene family of cathepsin L-like proteases of filarial nematodes are associated with larval molting and cuticle and eggshell remodeling. Mol. Biochem. Parasitol. 136, 227–242. https://doi.org/10.1016/j.molbiopara.2004.03.015
Guncar, G., Klemencic, I., Turk, B., Turk, V., Karaoglanovic-Carmona, A., Juliano, L., Turk, D., 2000. Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease. Structure 8, 305–13.
Guncar, G., Podobnik, M., Pungercar, J., Strukelj, B., Turk, V., Turk, D., 1998. Crystal structure of porcine cathepsin H determined at 2.1 A resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. Structure 6, 51–61. https://doi.org/10.1016/S0969-2126(98)00007-0
Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J., Tanaka, I., 1978. Isolation and Characterization of E-64, a New Thiol Protease Inhibitor. Agric. Biol. Chem 42, 523–528.
Hánová, I., Brynda, J., Houštecká, R., Alam, N., Sojka, D., Kopáček, P., Marešová, L., Vondrášek, J., Horn, M., Schueler-Furman, O., Mareš, M., 2018. Novel Structural Mechanism of Allosteric Regulation of Aspartic Peptidases via an Evolutionarily Conserved Exosite. Cell Chem. Biol. 25, 318-329.e4.

https://doi.org/10.1016/J.CHEMBIOL.2018.01.001
Hanpude, P., Bhattacharya, S., Dey, A.K., Maiti, T.K., 2015. Deubiquitinating enzymes in cellular signaling and disease regulation. IUBMB Life 67, 544–555. https://doi.org/10.1002/iub.1402
Hart, T.C., Hart, P.S., Bowden, D.W., Michalec, M.D., Callison, S.A., Walker, S.J., Zhang, Y., Firatli, E., 1999. Mutations of the cathepsin C gene are responsible for Papillon-Lefèvre syndrome. J. Med. Genet. 36, 881–7.
Hartmann, S., Schönemeyer, A., Sonnenburg, B., Vray, B., Lucius, R., 2002. Cystatins of filarial nematodes up-regulate the nitric oxide production of interferon-gamma-activated murine macrophages. Parasite Immunol. 24, 253–62.
Hasnains, S., Hirama, T., Huber, C.P., Mason, P., Mort, J.S., 1993. Characterization of Cathepsin B Specificity by Site-directed Mutagenesis. J. Biol. Chem. 268, 235–240.
He, W., Ohashi, K., Sugimoto, C., Onuma, M., 2005. Theileria orientalis : Cloning a cDNA encoding a protein similar to thiol protease with haemoglobin-binding activity. Exp. Parasitol. 111, 143–153. https://doi.org/10.1016/j.exppara.2005.06.003
Hewitson, J.P., Maizels, R.M., 2014. Vaccination against helminth parasite infections. Expert Rev. vaccines 473–487. https://doi.org/10.1586/14760584.2014.893195
Hola-Jamriska, L., Dalton, J., Aaskov, J., Brindley, P.J., 1999. Dipeptidyl peptidase I and III activities of adult schistosomes. Parasitology 118, 275–282. https://doi.org/10.1017/S0031182098003746
Holt, D.C., Fischer, K., Allen, G.E., Wilson, D., Wilson, P., Slade, R., Currie, B.J., Walton, S.F., Kemp, D.J., 2003. Mechanisms for a Novel Immune Evasion Strategy in the Scabies Mite Sarcoptes Scabiei: A Multigene Family of Inactivated Serine Proteases. Soc. Investig. Dermatology 121, 1419–1424. https://doi.org/10.1046/j.1523-1747.2003.12621.x
Holt, D.C., Fischer, K., Pizzutto, S.J., Currie, B.J., Walton, S.F., Kemp, D.J., 2004. A Multigene Family of Inactivated Cysteine Proteases in Sarcoptes scabiei. Soc. Investig. Dermatology 123, 240–241. https://doi.org/10.1111/j.0022-202X.2004.22716.x
Horák, P., Bulantová, J., Mikeš, L., 2019. Schistosomatoidea and Diplostomoidea. Adv. Exp. Med. Biol. 1154, 217–254. https://doi.org/10.1007/978-3-030-18616-6_8
Horák, P., Dvorák, J., Kolárová, L., Trefil, L., 1999. Trichobilharzia regenti, a pathogen of the avian and mammalian central nervous systems. Parasitology 119, 577–581.
Horák, P., Kolářová, L., Adema, C.M., 2002. Biology of the schistosome genus Trichobilharzia. Adv. Parasitol. https://doi.org/10.1016/S0065-308X(02)52012-1
Horák, P., Kolárová, L., Dvorák, J., 1998. Trichobilharzia regenti n. sp. (Schistosomatidae, Bilharziellinae), a new nasal schistosome from Europe. Parasite 5, 349–357.
Horák, P., Mikeš, L., Lichtenbergová, L., Skála, V., Soldánová, M., Brant, S.V., 2015. Avian Schistosomes and Outbreaks of Cercarial Dermatitis. Clin. Microbiol. Rev. 28, 165. https://doi.org/10.1128/CMR.00043-14
Horn, M., Jílková, A., Vondrášek, J., Marešová, L., Caffrey, C.R., Mareš, M., 2011. Mapping the Pro-Peptide of the Schistosoma mansoni Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors. ACS Chem. Biol. 6, 609–617. https://doi.org/10.1021/cb100411v
Hrádková, K., Horák, P., 2002. Neurotropic behaviour of Trichobilharzia regenti in ducks and mice. J. Helminthol. 76, 137–141. https://doi.org/10.1079/JOH2002113
Huete-Pérez, J.A., Engel, J.C., Brinen, L.S., Mottram, J.C., McKerrow, J.H., 1999. Protease Trafficking in Two Primitive Eukaryotes Is Mediated by a Prodomain Protein Motif. J. Biol. Chem. 274, 16249–16256. https://doi.org/10.1074/JBC.274.23.16249
Ilgová, J., Jedličková, L., Dvořáková, H., Benovics, M., Mikeš, L., Janda, L., Vorel, J., Roudnický, P., Potěšil, D., Zdráhal, Z., Gelnar, M., Kašný, M., 2017. A novel type I

cystatin of parasite origin with atypical legumain- binding domain. Sci. Rep. 7, 1–12. https://doi.org/10.1038/s41598-017-17598-2
Ilgová, J., Kavanová, L., Matiašková, K., Salát, J., Kašný, M., 2020. Effect of cysteine peptidase inhibitor of Eudiplozoon nipponicum (Monogenea) on cytokine expression of macrophages in vitro. Mol. Biochem. Parasitol. 235, 111248. https://doi.org/10.1016/J.MOLBIOPARA.2019.111248
Ilgová, J., Salát, J., Kašný, M., 2021. Molecular communication between the monogenea and fish immune system. Fish Shellfish Immunol. 112, 179–190. https://doi.org/10.1016/J.FSI.2020.08.023
Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., Mort, J.S., 1997. Role of the Occluding Loop in Cathepsin B Activity. J. Biol. Chem. 272, 1197–1202. https://doi.org/10.1074/jbc.272.2.1197
Inagami, T., Murakami, K., 1977. Pure renin isolation from hog kidney and characterization. J. Biol. Chem. 252, 2978–2983.
Jasmer, D.P., Mitreva, M.D., McCarter, J.P., 2004. mRNA sequences for Haemonchus contortus intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared with other adult mammalian parasitic nematodes. Mol. Biochem. Parasitol. 137, 297–305. https://doi.org/10.1016/j.molbiopara.2004.06.010
Jedličková, L., Dvořák, J., Hrachovinová, I., Ulrychová, L., Kašný, M., Mikeš, L., 2019. A novel Kunitz protein with proposed dual function from Eudiplozoon nipponicum (Monogenea) impairs haemostasis and action of complement in vitro. Int. J. Parasitol. 49, 337–346. https://doi.org/10.1016/J.IJPARA.2018.11.010
Jedličková, L., Dvořáková, H., Dvořák, J., Kašný, M., Ulrychová, L., Vorel, J., Žárský, V., Mikeš, L., 2018. Cysteine peptidases of Eudiplozoon nipponicum: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp. Parasit. Vectors 11, 1–17. https://doi.org/10.1186/s13071-018-2666-2
Jedličková, L., Dvořáková, H., Kašný, M., Ilgová, J., Potěšil, D., Zdráhal, Z., Mikeš, L., 2016. Major acid endopeptidases of the blood-feeding monogenean Eudiplozoon nipponicum (Heteronchoinea: Diplozoidae). Parasitology 143, 494–506. https://doi.org/10.1017/S0031182015001808
Jílková, A., Horn, M., Řezáčová, P., Marešová, L., Fajtová, P., Brynda, J., Vondrášek, J., McKerrow, J.H., Caffrey, C.R., Mareš, M., 2014. Activation route of the Schistosoma mansoni cathepsin B1 drug target: Structural map with a glycosaminoglycan switch. Structure 22, 1786–1798. https://doi.org/10.1016/j.str.2014.09.015
Jolodar, A., Fischer, P., Bü Ttner, D.W., Miller, D.J., Schmetz, C., Brattig, N.W., 2004. Onchocerca volvulus: expression and immunolocalization of a nematode cathepsin D-like lysosomal aspartic protease. https://doi.org/10.1016/j.exppara.2004.06.006
Kang, J.-M., Ju, H.-L., Lee, K.H., Kim, T.-S., Pak, J.H., Sohn, W.-M., Na, B.-K., 2014. Identification and characterization of the second cysteine protease inhibitor of Clonorchis sinensis (CsStefin-2). Parasitol. Res. 113, 47–58. https://doi.org/10.1007/s00436-013-3624-8
Kang, J.-M., Yoo, W.-G., Le, H.G., Thai, T.L., Hong, S.-J., Sohn, W.-M., Na, B.-K., 2019. Partial Characterization of Two Cathepsin D Family Aspartic Peptidases of Clonorchis sinensis. Korean J Parasitol 57, 671–680. https://doi.org/10.3347/kjp.2019.57.6.671
Kang, S.Y., Cho, M.S., Chung, Y.B., Kong, Y., Cho, S.Y., 1995. A cysteine protease of Paragonimus westermani eggs. Korean J. Parasitol. 33, 323. https://doi.org/10.3347/kjp.1995.33.4.323
Kawatsu, 1978. Studies on the anemia of fish-ix hypochromic microcytic anemia of crucian carp caused by infestation with a trematode, Diplozoon nipponicum. Bull.

Japanese Soc. Sci. Fish. 44, 1315–1319. https://doi.org/10.1001/archpedi.1954.02050100569001
Khotenovsky, I.A., 1977. Life cycle of several monogeneans of the genus Diplozoon. Parazitol. Sb. 27, 35–43.
Kirschke, H., Wiederanders, B., Bromme, D., Rinnet, A., 1989. Cathepsin S from bovine spleen Purification, distribution, intracellular localization and action on proteins. Biochem. J. 264, 467–473.
Kolářová, L., Horák, P., Čada, F., 2001. Histopathology of CNS and nasal infections caused by Trichobilharzia regenti in vertebrates. Parasitol. Res. 87, 644–650. https://doi.org/10.1007/s004360100431
Konstanzová, V., Koubková, B., Kašný, M., Ilgová, J., Dzika, E., Gelnar, M., 2015. Ultrastructure of the digestive tract of Paradiplozoon homoion (Monogenea). Parasitol. Res. 114, 1485–1494. https://doi.org/10.1007/s00436-015-4331-4
Kouřilová, P., Hogg, K., Kolářová, L., Mountford, A., 2004a. Cercarial dermatitis caused by bird schistosomes comprises both immediate and late phase cutaneous hypersensitivity reactions. J. Immunol. 172, 3766–3774. https://doi.org/10.4049/jimmunol.172.6.3766
Kouřilová, P., Syrůček, M., Kolářová, L., 2004b. The severity of mouse pathologies caused by the bird schistosome Trichobilharzia regenti in relation to host immune status. Parasitol. Res. 93, 8–16. https://doi.org/10.1007/s00436-004-1079-7
Kramer, L., Turk, D., Turk, B., 2017. The Future of Cysteine Cathepsins in Disease Management. https://doi.org/10.1016/j.tips.2017.06.003
Lecaille, F., Choe, Y.§, Brandt, W., Li, Z., Craik, C.S. and, Brömme*, D., 2002. Selective Inhibition of the Collagenolytic Activity of Human Cathepsin K by Altering Its S2 Subsite Specificity†. Biochemistry 8447–8454. https://doi.org/10.1021/BI025638X
Lehane, M.J., Müller, H.M., Crisanti, A., 1996. Mechanisms controlling the synthesis and secretion of digestive enzymes in insects, in: Biology of the Insect Midgut. Springer Netherlands, pp. 195–205. https://doi.org/10.1007/978-94-009-1519-0_7
Li, S., Chung, Y.-B., Chung, B.-S., Choi, M.-H., Yu, J.-R., Hong, S.-T., 2004. The involvement of the cysteine proteases of Clonorchis sinensis metacercariae in excystment. Parasitol. Res. 93, 36–40. https://doi.org/10.1007/s00436-004-1097-5
Li, Y.L., Idris, M.A., Corachan, M., Han, J.J., Kirschfink, M., Ruppel, A., 1996. Circulating antigens in schistosomiasis: detection of 31/32-kDa proteins in sera from patients infected with Schistosoma japonicum, S. mansoni, S. haematobium, or S. intercalatum. Parasitol. Res. 82, 14–18. https://doi.org/10.1007/s004360050060
Li, Z., Hou, W.-S., Escalante-Torres, C.R., Gelb, B.D., Brö Mme, D., 2002. Collagenase Activity of Cathepsin K Depends on Complex Formation with Chondroitin Sulfate*. https://doi.org/10.1074/jbc.M204004200
Li, Z., Yasuda, Y., Li, W., Bogyo, M., Katz, N., Gordon, R.E., Fields, G.B., Brömme, D., 2004. Regulation of Collagenase Activities of Human Cathepsins by Glycosaminoglycans. J. Biol. Chem. 279, 5470–5479. https://doi.org/10.1074/jbc.M310349200
Lichtenbergová, L., Kolbeková, P., Kouřilová, P., Kašný, M., Mikeš, L., Haas, H., Schramm, G., Horák, P., Kolářová, L., Mountford, A.P., 2008. Antibody responses induced by *Trichobilharzia regenti* antigens in murine and human hosts exhibiting cercarial dermatitis. Parasite Immunol. 30, 585–595. https://doi.org/10.1111/j.1365-3024.2008.01059.x
Liu, J., Istvan, E.S., Gluzman, I.Y., Gross, J., Goldberg, D.E., 2006. Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. Proc. Natl. Acad. Sci. U. S. A. 103, 8840–8845. https://doi.org/10.1073/pnas.0601876103
Longbottom, D., Redmond, D.L., Russell, M., Liddell, S., Smith, W.D., Knox, D.P., 1997.

Molecular cloning and characterisation of a putative aspartate proteinase associated with a gut membrane protein complex from adult Haemonchus contortus. Mol. Biochem. Parasitol. 88, 63–72. https://doi.org/10.1016/S0166-6851(97)00074-1
Loukas, A., Gaze, S., Mulvenna, J.P., Gasser, R.B., Brindley, P.J., Doolan, D.L., Bethony, J.M., Jones, M.K., Gobert, G.N., Driguez, P., Mcmanus, D.P., Hotez, P.J., 2011. Vaccinomics for the Major Blood Feeding Helminths of Humans. A J. Integr. Biol. 15, 567–577. https://doi.org/10.1089/omi.2010.0150
Lustigman, S., Zhang, J., Liu, J., Oksov, Y., Hashmi, S., 2004. RNA interference targeting cathepsin L and Z-like cysteine proteases of Onchocerca volvulus confirmed their essential function during L3 molting. Mol. Biochem. Parasitol. 138, 165–170. https://doi.org/10.1016/J.MOLBIOPARA.2004.08.003
Macháček, T., Panská, L., Dvořáková, H., Horák, P., 2016. Nitric oxide and cytokine production by glial cells exposed in vitro to neuropathogenic schistosome Trichobilharzia regenti. Parasit. Vectors 9, 1–10. https://doi.org/10.1186/s13071-016-1869-7
Macháček, T., Turjanicová, L., Bulantová, J., Hrdý, J., Horák, P., Mikeš, L., 2018. Cercarial dermatitis: a systematic follow-up study of human cases with implications for diagnostics. Parasitol. Res. 117, 3881–3895. https://doi.org/10.1007/s00436-018-6095-0
Martin, S., 1884. Papain-digestion. J. Physiol. 5, 213.
Masa, M., Maresova, L., Vondrasek, J., Horn, M., Jezek, J., Mares, M., 2006. Propeptide interaction with cathepsin D: Mechanism and regulation. Biochemistry 12, 15474–15482.
Masson, O., Bach, A.S., Derocq, D., Prébois, C., Laurent-Matha, V., Pattingre, S., Liaudet-Coopman, E., 2010. Pathophysiological functions of cathepsin D: Targeting its catalytic activity versus its protein binding activity? Biochimie 92, 1635–1643. https://doi.org/10.1016/j.biochi.2010.05.009
Matějusová, I., Koubkova, B., D 'amelio, S., Cunningham, C.O., 2001. Genetic characterization of six species of diplozoids (Monogenea ; Diplozoidae). Parasitology 123, 463–474. https://doi.org/10.1017/S0031182001008617
McCarthy, E., Stack, C., Donnelly, S.M., Doyle, S., Mann, V.H., Brindley, P.J., Stewart, M., Day, T.A., Maule, A.G., Dalton, J.P., 2004. Leucine aminopeptidase of the human blood flukes, Schistosoma mansoni and Schistosoma japonicum. Int. J. Parasitol. 34, 703–714. https://doi.org/10.1016/j.ijpara.2004.01.008
McCoubrie, J.E., Miller, S.K., Sargeant, T., Good, R.T., Hodder, A.N., Speed, T.P., De Koning-Ward, T.F., Crabb, B.S., 2007. Evidence for a Common Role for the Serine-Type Plasmodium falciparum Serine Repeat Antigen Proteases: Implications for Vaccine and Drug Design. Infect. Immun. 75, 5565–5574. https://doi.org/10.1128/IAI.00405-07
McGonigle, L., Mousley, A., Marks, N.J., Brennan, G.P., Dalton, J.P., Spithill, T.W., Day, T.A., Maule, A.G., 2008. The silencing of cysteine proteases in Fasciola hepatica newly excysted juveniles using RNA interference reduces gut penetration. Int. J. Parasitol. 38, 149–155. https://doi.org/10.1016/J.IJPARA.2007.10.007
McGrath, M.E., 1999. The lysosomal cysteine proteases. Annu. Rev. Biophys. Biomol. Struct 28, 181–204.
Mcintyre, G.F., Godbold, G.D., Ericksons, A.H., 1994. The pH-dependent Membrane Association of Procathepsin L Is Mediated by a 9-Residue Sequence within the Propeptide* 269, 567–572. https://doi.org/10.1016/S0021-9258(17)42386-6
McKerrow, J.H., 2018a. The diverse roles of cysteine proteases in parasites and their suitability as drug targets. PLoS Negl. Trop. Dis. 12, 3–5.
McKerrow, J.H., 2018b. Update on drug development targeting parasite cysteine proteases. https://doi.org/10.1371/journal.pntd.0005850

McKerrow, J.H., Caffrey, C., Kelly, B., Loke, P., Sajid, M., 2006. Proteases in parasitic diseases. Annu. Rev. Pathol. Mech. Dis. 1, 497–536. https://doi.org/10.1146/ANNUREV.PATHOL.1.110304.100151

McVeigh, P., Maule, A.G., Dalton, J.P., Robinson, M.W., 2012. Fasciola hepatica virulence-associated cysteine peptidases: A systems biology perspective. Microbes Infect. 14, 301–310. https://doi.org/10.1016/j.micinf.2011.11.012

Meléndez-López, S.G., Herdman, S., Hirata, K., Choi, M.H., Choe, Y., Craik, C., Caffrey, C.R., Hansell, E., Chávez-Munguía, B., Yen, T.C., Roush, W.R., McKerrow, J., Eckmann, L., Guo, J., Stanley, S.L., Reed, S.L., 2007. Use of recombinant Entamoeba histolytica cysteine proteinase 1 to identify a potent inhibitor of amebic invasion in a human colonic model. Eukaryot. Cell 6, 1130–1136. https://doi.org/10.1128/EC.00094-07

Menard, R., Carmona, E., Takebe, S., Dufour, R., Plouffe, C., Mason, P., Mort, J.S., 1998. Autocatalytic Processing of Recombinant Human Procathepsin L: Contribution of both intermolecular and unimolecular events in the processing of procathepsin L in vitro*. J. Biol. Chem. 273, 4478–4484.

Mendiola, J., Alonso, M., Marquetti, M.C., Finlay, C., 1996. Boophilus microplus: multiple proteolytic activities in the midgut. Exp. Parasitol. 82, 27–33. https://doi.org/S0014-4894(96)90004-1 [pii]\r10.1006/expr.1996.0004

Mendoza-Palomares, C., Biteau, N., Giroud, C., Coustou, V., Coetzer, T., Authié, E., Boulangé, A., Baltz, T., 2008. Molecular and biochemical characterization of a cathepsin B-like protease family unique to Trypanosoma congolense. Eukaryot Cell 7, 684–697. https://doi.org/10.1128/EC.00405-07

Merckelbach, A., Hasse, S., Dell, R., Eschlbeck, A., Ruppel, A., 1994. cDNA sequences of Schistosoma japonicum coding for two cathepsin B-like proteins and Sj32. Trop. Med. Parasitol 45, 193–198.

Michel, A., Ghoneim, H., Resto, M., Klinkert, M.-Q., Kunz, W., 1995. Sequence, characterization and localization of a cysteine proteinase cathepsin L in Schistosoma mansoni. Mol. Biochem. Parasitol. 73, 7–18.

Morales, F.C., Furtado, D.R., Rumjanek, F.D., 2004. The N-terminus moiety of the cystatin SmCys from Schistosoma mansoni regulates its inhibitory activity in vitro and in vivo. Mol. Biochem. Parasitol. 134, 65–73. https://doi.org/10.1016/j.molbiopara.2003.10.016

Müntener, K., Zwicky, R., Csucs, G., Rohrer, J., Baici, A., 2004. Exon skipping of cathepsin B: mitochondrial targeting of a lysosomal peptidase provokes cell death. J. Biol. Chem. 279, 41012–41017.

Murata, M., Miyashita, S., Yokoo, C., Tamai, M., Hanada, K., Hatayama, K., Towatari, T., Nikawa, T., Katunuma, N., 1991. Novel epoxysuccinyl peptides. Selective inhibitors of cathepsin B, in vitro. FEBS Lett. 280, 307–10.

Musil, D., Zucic, D., Turk, D., Engh, R. a, Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., 1991. The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. EMBO J. 10, 2321–2330.

Na, B.-K., Kim, S.-H., Lee, E.-G., Kim, T.-S., Bae, Y.-A., Kang, I., Yu, J.-R., Sohn, W.-M., Cho, S.-Y., Kong, Y., 2006. Critical roles for excretory-secretory cysteine proteases during tissue invasion of Paragonimus westermani newly excysted metacercariae. Cell. Microbiol. 8, 1034–1046. https://doi.org/10.1111/j.1462-5822.2006.00685.x

Nägler, D.K., Zhang, R., Tam, W., Sulea, T., Purisima, E.O., Ménard, R., 1999. Human Cathepsin X: A Cysteine Protease with Unique Carboxypeptidase Activity. Biochemistry 38, 12648–12654. https://doi.org/10.1021/bi991371z

Naudin, C., Lecaille, F., Chowdhury, S., Krupa, J.C., Purisima, E., Mort, J.S., Lalmanach,

G., 2010. The Occluding Loop of Cathepsin B Prevents Its Effective Inhibition by Human Kininogens. J. Mol. Biol. 400, 1022–1035. https://doi.org/10.1016/j.jmb.2010.06.006
Ndao, M., Nath-Chowdhury, M., Sajid, M., Marcus, V., Mashiyama, S.T., Sakanari, J., Chow, E., Mackey, Z., Land, K.M., Jacobson, M.P., Kalyanaraman, C., McKerrow, J.H., Arrowood, M.J., Caffrey, C.R., 2013. A cysteine protease inhibitor rescues mice from a lethal Cryptosporidium parvum infection. Antimicrob. Agents Chemother. 57, 6063–6073. https://doi.org/10.1128/AAC.00734-13
Ni, F., Wang, Y., Zhang, J., Yu, L., Fang, W., Luo, D., 2012. Cathepsin B-like and hemoglobin-type cysteine proteases: Stage-specific gene expression in Angiostrongy cantonensis. Exp. Parasitol. 131, 433–441. https://doi.org/10.1016/j.exppara.2012.05.014
Nishihira, T., Urabe, M., 2020. Morphological and molecular studies of Eudiplozoon nipponicum (Goto, 1891) and Eudiplozoon kamegaii sp. n. (Monogenea; Diplozoidae). Folia Parasitol. . 67, 1–16. https://doi.org/10.14411/FP.2020.018
Nishimura, Y., Kawabata, T., Furuno, K., Kato, K., 1989. Evidence that aspartic proteinase is involved in the proteolytic processing event of procathepsin L in lysosomes. Arch. Biochem. Biophys. 271, 400–406. https://doi.org/10.1016/0003-9861(89)90289-0
Nycander, M., Estrada, S., Mort, J.S., Abrahamson, M., Björk, I., 1998. Two-step mechanism of inhibition of cathepsin B by cystatin C due to displacement of the proteinase occluding loop. FEBS Lett. 422, 61–64. https://doi.org/10.1016/S0014-5793(97)01604-9
O'Neill, S.M., Mills, K.H.G., Dalton, J.P., 2001. Fasciola hepatica cathepsin L cysteine proteinase suppresses Bordetella pertussis-specific interferon-gamma production in vivo. Parasite Immunol. 23, 541–547. https://doi.org/10.1046/j.1365-3024.2001.00411.x
Palmer, J.T., Rasnick, D., Klaus, J.L., Bromme, D., 1995. Vinyl Sulfones as Mechanism-Based Cysteine Protease Inhibitors. J . Med. Chem 38, 3193–3196.
Paul, T.J., Barman, A., Ozbil, M., Bora, R.P., Zhang, T., Sharma, G., Hoffmann, Z., Prabhakar, R., 2016. Mechanisms of peptide hydrolysis by aspartyl and metalloproteases. Phys. Chem. Chem. Phys. 18, 24790–24801. https://doi.org/10.1039/c6cp02097f
Pavlova, A., Krupa, J.C., Mort, J.S., Abrahamson, M., Björk, I., 2000. Cystatin inhibition of cathepsin B requires dislocation of the proteinase occluding loop. Demonstration by release of loop anchoring through mutation of His110. FEBS Lett. 487, 156–160. https://doi.org/10.1016/S0014-5793(00)02337-1
Pečínková, M., Matějusová, I., Koubková, B., Gelnar, M., 2007. Investigation of Paradiplozoon homoion (Monogenea, Diplozoidae) life cycle under experimental conditions. https://doi.org/10.1016/j.parint.2007.01.010
Pillay, D., Boulangé, A.F., Coetzer, T.H.T., 2010. Expression, purification and characterisation of two variant cysteine peptidases from Trypanosoma congolense with active site substitutions. Protein Expr. Purif. 74, 264–271. https://doi.org/10.1016/j.pep.2010.06.021
Podobnik, M., Kuhelj, R., Turk, V., Turk, D., 1997. Crystal structure of the wild-type human procathepsin B at 2.5 A resolution reveals the native active site of a papain-like cysteine protease zymogen. J. Mol. Biol. 271, 774–788. https://doi.org/10.1006/jmbi.1997.1218
Ranjit, N., Zhan, B., Stenzel, D.J., Mulvenna, J., Fujiwara, R., Hotez, P.J., Loukas, A., 2008. A family of cathepsin B cysteine proteases expressed in the gut of the human hookworm, Necator americanus. Mol. Biochem. Parasitol. 160, 90–99. https://doi.org/10.1016/j.molbiopara.2008.04.008
Rawlings, N., Barrett, A.J., 2013. Introduction: Aspartic and Glutamic Peptidases and

Their Clans, in: Rawlings, N., Salvesen, G. (Eds.), Handbook of Proteolytic Enzymes. Elsevier Science & Technology, pp. 3–19.
Rawlings, N., Salvesen, G. (Eds.), 2013. Handbook of Proteolytic enzymes, 3rd ed. Elsavier, Amsterdam.
Rawlings, N.D., Barrett, A.J., 1993. Evolutionary families of peptidases. Biochem. J. 290, 205–218. https://doi.org/10.1042/BJ2900205
Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46, D624–D632. https://doi.org/10.1093/nar/gkx1134
Rawlings, N.D., Bateman, A., 2009. Pepsin homologues in bacteria. BMC Genomics 10, 437. https://doi.org/10.1186/1471-2164-10-437
Rawlings, N.D., Tolle, D.P., Barrett, A.J., 2004. Evolutionary families of peptidase inhibitors. Biochem. J 378, 705–716.
Reiser, J., Adair, B., Reinheckel, T., 2010. Specialized roles for cysteine cathepsins in health and disease. J. Clin. Invest. https://doi.org/10.1172/JCI42918
Reynolds, S.L., Pike, R.N., Mika, A., Blom, A.M., Hofmann, A., Wijeyewickrema, L.C., Kemp 1{, D., Fischer, K., 2014. Scabies Mite Inactive Serine Proteases Are Potent Inhibitors of the Human Complement Lectin Pathway. PLoS Negl. Trop. Dis. 8. https://doi.org/10.1371/journal.pntd.0002872
Robinson, M.W., Dalton, J.P., Donnelly, S., 2008. Helminth pathogen cathepsin proteases: it's a family affair. Trends Biochem. Sci. https://doi.org/10.1016/j.tibs.2008.09.001
Rosenthal, P.J., 2004. Cysteine proteases of malaria parasites 34, 1489–1499. https://doi.org/10.1016/j.ijpara.2004.10.003
Rossi, A., Deveraux, Q., Turk, B., Sali, A., 2004. Comprehensive search for cysteine cathepsins in the human genome. Biol. Chem. 385, 363–72. https://doi.org/10.1515/BC.2004.040
Roudnicky, P., Potěšil, D., Zdráhal, Z., Gelnar, M., Kašný, M., 2020. Laser capture microdissection in combination with mass spectrometry: Approach to characterization of tissue-specific proteomes of Eudiplozoon nipponicum (Monogenea, Polyopisthocotylea). PLoS One 15, 1–15. https://doi.org/10.1371/journal.pone.0231681
Roudnický, P., Vorel, J., Ilgová, J., Benovics, M., Norek, A., Jedličková, L., Mikeš, L., Potěšil, D., Zdráhal, Z., Dvořák, J., Gelnar, M., Kašný, M., 2018. Identification and partial characterization of a novel serpin from Eudiplozoon nipponicum (Monogenea, Polyopisthocotylea). Parasite 25. https://doi.org/10.1051/PARASITE/2018062
Rowan, A.D., Mason, P., Mach, L., Mort, J.S., 1992. Rat procathepsin B: Proteolytic processing to the mature form in vitro. J. Biol. Chem. 267, 15993–15999.
Sajid, M., McKerrow, J.H., 2002. Cysteine proteases of parasitic organisms. Mol. Biochem. Parasitol. 120, 1–21. https://doi.org/10.1016/S0166-6851(02)00043-9
Sajid, M., McKerrow, J.H., Hansell, E., Mathieu, M. a., Lucas, K.D., Hsieh, I., Greenbaum, D., Bogyo, M., Salter, J.P., Lim, K.C., Franklin, C., Kim, J.H., Caffrey, C.R., 2003. Functional expression and characterization of Schistosoma mansoni cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. Mol. Biochem. Parasitol. 131, 65–75. https://doi.org/10.1016/S0166-6851(03)00194-4
Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27, 157–62.
Schönemeyer, A., Lucius, R., Sonnenburg, B., Brattig, N., Sabat, R., Schilling, K., Bradley, J., Hartmann, S., 2001. Modulation of human T cell responses and macrophage functions by onchocystatin, a secreted protein of the filarial nematode Onchocerca volvulus. J. Immunol. 167, 3207–15. https://doi.org/10.4049/JIMMUNOL.167.6.3207

Shin, M.H., Kita, H., Park, H.Y., Seoh, J.Y., 2001. Cysteine Protease Secreted by Paragonimus westermani Attenuates Effector Functions of Human Eosinophils Stimulated with Immunoglobulin G. Infect. Immun. 69, 1599–1604. https://doi.org/10.1128/IAI.69.3.1599-1604.2001

Siklos, M., BenAissa, M., Thatcher, G.R.J., 2015. Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors. Acta Pharm. Sin. B 5, 506–19. https://doi.org/10.1016/j.apsb.2015.08.001

Simões, I., Faro, R., Bur, D., Kay, J., Faro, C., 2011. Shewasin A, an active pepsin homolog from the bacterium Shewanella amazonensis. FEBS J. 278, 3177–3186. https://doi.org/10.1111/j.1742-4658.2011.08243.x

Smyth, D.J., Halton, D.W., 1983. The Physiology of Trematodes, 2nd. editi. ed. Cambridge University Press;, Cambridge.

Sojka, D., Franta, Z., Horn, M., Caffrey, C.R., Mareš, M., Kopáček, P., 2013. New insights into the machinery of blood digestion by ticks. Trends Parasitol. 29, 276–285. https://doi.org/10.1016/j.pt.2013.04.002

Sojka, D., Franta, Z., Horn, M., Hajdusek, O., Caffrey, C.R., Mares, M., 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. https://doi.org/10.1186/1756-3305-1-7

Sojka, D., Hartmann, D., Bartoš Ová-Sojková, P., Dvor ̌ák, J., 2016. Parasite Cathepsin D-Like Peptidases and Their Relevance as Therapeutic Targets. https://doi.org/10.1016/j.pt.2016.05.015

Suttiprapa, S., Mulvenna, J., Huong, N.T., Pearson, M.S., Brindley, P.J., Laha, T., Wongkham, S., Kaewkes, S., Sripa, B., Loukas, A., 2009. Ov-APR-1, an aspartic protease from the carcinogenic liver fluke, Opisthorchis viverrini: functional expression, immunolocalization and subsite specificity. Int J Biochem Cell Biol 41, 1148–1156. https://doi.org/10.1016/j.biocel.2008.10.013

Szabelski, M., Rogiewicz, M., Wiczk, W., 2005. Fluorogenic peptide substrates containing benzoxazol-5-yl-alanine derivatives for kinetic assay of cysteine proteases. https://doi.org/10.1016/j.ab.2005.04.014

Tao, K., Stearns, N.A., Dong, J.M., Wu, Q.I., Sahagian, G.G., 1994. The Proregion of Cathepsin L Is Required for Proper Folding, Stability, and ER Exit. Arch. Biochem. Biophys. 311, 19–27. https://doi.org/10.1006/ABBI.1994.1203

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalization and function. Biochem. Physiol 109. https://doi.org/10.1016/0305-0491(94)90141-4

Towatari, T., Nikawa, T., Murata, M., Yokoo, C., Tamai, M., Hanada, K., Katunuma, N., 1991. Novel epoxysuccinyl peptides: A selective inhibitor of cathepsin B, in vivo. FEBS Lett. 280, 311–315. https://doi.org/10.1016/0014-5793(91)80319-X

Turk, B., Bieth, J.G., Björk, I., Dolenc, I., Turk, D., Cimerman, N., Kos, J., Colic, A., Stoka, V., Turk, V., 1995. Regulation of the activity of lysosomal cysteine proteinases by pH-induced inactivation and/or endogenous protein inhibitors, cystatins. Biol. Chem. Hoppe. Seyler. 376, 225–30.

Turk, D., Guncar, G., Podobnik, M., Turk, B., 1998. Revised definition of substrate binding sites of papain-like cysteine proteases. Biol. Chem. 379, 137–47.

Turk, D., Janjić, V., Stern, I., Podobnik, M., Lamba, D., Dahl, S.W., Lauritzen, C., Pedersen, J., Turk, V., Turk, B., 2001. Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. EMBO J. 20, 6570–82. https://doi.org/10.1093/emboj/20.23.6570

Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R., Turk, V., 1995. Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. Biochemistry 34, 4791–4797.

Turk, D., Turk, B., Turk, V., 2003. Papain-like lysosomal cysteine proteases and their inhibitors: Drug discovery targets?, in: Biochemical Society Symposium. Portland Press Ltd, pp. 15–30. https://doi.org/10.1042/bss0700015
Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M., Hamada, M., 1970. Pepstatin, a new pepsin inhibitor produced by Actinomycetes. J. Antibiot. (Tokyo). 23, 259–62.
Valigurová, A., Hodová, I., Sonnek, R., Koubková, B., Gelnar, M., 2011. Eudiplozoon nipponicum in focus: Monogenean exhibiting a highly specialized adaptation for ectoparasitic lifestyle. Parasitol. Res. 108, 383–394. https://doi.org/10.1007/s00436-010-2077-6
Vasiljeva, O., Dolinar, M., Pungerčar, J.R., Turk, V., Turk, B., 2005. Recombinant human procathepsin S is capable of autocatalytic processing at neutral pH in the presence of glycosaminoglycans. https://doi.org/10.1016/j.febslet.2004.12.093
Verity, C.K., Loukas, A., Mcmanus, D.P., Brindley, P.J., 2001. Schistosoma japonicum cathepsin D aspartic protease cleaves human IgG and other serum components. Parasitology 122, 415–421. https://doi.org/10.1017/S0031182001007521
Verity, C.K., McManus, D.P., Brindley, P.J., 1999. Developmental expression of cathepsin D aspartic protease in Schistosoma japonicum. Int. J. Parasitol. 29, 1819–1824. https://doi.org/10.1016/S0020-7519(99)00126-5
Verma, S., Dixit, R., Pandey, K.C., 2016. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. Front. Pharmacol. 7, 107. https://doi.org/10.3389/fphar.2016.00107
Vermeire, J.J., Lantz, L.D., Caffrey, C.R., 2012. Cure of Hookworm infection with a cysteine protease inhibitor. PLoS Negl. Trop. Dis. 6, 1–8. https://doi.org/10.1371/journal.pntd.0001680
Vorel, J., Cwiklinski, K., Roudnický, P., Ilgová, J., Jedličková, L., Dalton, J.P., Mikeš, L., Gelnar, M., Kašný, M., 2021. Eudiplozoon nipponicum (Monogenea, Diplozoidae) and its adaptation to haematophagy as revealed by transcriptome and secretome profiling. BMC Genomics 22. https://doi.org/10.1186/s12864-021-07589-z
Wiederanders, B., Kaulmann, G., Schilling, K., 2003. Functions of Propeptide Parts in Cysteine Proteases. Curr. Protein Pept. Sci. 4, 309–326. https://doi.org/10.2174/1389203033487081
Williamson, Angela L, Brindley, P.J., Abbenante, G., Datu, B.J.D., Prociv, P., Berry, C., Girdwood, K., Pritchard, D.I., Fairlie, D.P., Hotez, P.J., Zhan, B., Loukas, A., 2003. Hookworm Aspartic Protease, Na-APR-2, Cleaves Human Hemoglobin and Serum Proteins in a Host-Specific Fashion.
Williamson, A.L., Brindley, P.J., Abbenante, G., Prociv, P., Berry, C., Girdwood, K., Pritchard, D.I., Fairlie, D.P., Hotez, P.J., Dalton, J.P., Loukas, A., 2002. Cleavage of hemoglobin by hookworm cathepsin D aspartic proteases and its potential contribution to host specificity. FASEB J. 16 (11), 1458–1460. https://doi.org/10.1096/fj.02-0181fje
Williamson, Angela L., Brindley, P.J., Knox, D.P., Hotez, P.J., Loukas, A., 2003. Digestive proteases of blood-feeding nematodes. Trends Parasitol. 19, 417–423. https://doi.org/10.1016/S1471-4922(03)00189-2
Williamson, A.L., Lecchi, P., Turk, B.E., Choe, Y., Hotez, P.J., McKerrow, J.H., Cantley, L.C., Sajid, M., Craik, C.S., Loukas, A., 2004. A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. J. Biol. Chem. 279, 35950–35957. https://doi.org/10.1074/jbc.M405842200
Willstätter, R., Bamann, E., 1929. Über die Proteasen der Magenschleimhaut. Erste Abhandlung über die Enzyme der Leukocyten. Hoppe-Seyler´s Zeitschrift für Physiol. Chemie 180, 127–143.
Wong, J.Y.M., Harrop, S.A., Day, S.R., Brindley, P.J., 1997. Schistosomes express two forms of cathepsin D. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1338,

156–160. https://doi.org/10.1016/S0167-4838(97)00019-8
Yamamoto, A., Hara, T., Tomoo, K., Ishida, T., Fujii, T., Hata, Y., Murata, M., Kitamurat, K., 1997. Binding mode of CA074, a specific irreversible inhibitor, to bovine cathepsin B as determined by X-Ray crystal analysis of the complex. J. Biochem. 121, 974–977.
Yan, C., Boyd, D.D., 2007. Regulation of matrix metalloproteinase gene expression. J. Cell. Physiol. 211, 19–26. https://doi.org/10.1002/jcp.20948
Yang, Y., Qin, W., Wei, H., Ying, J., Zhen, J., 2011. Characterization of cathepsin B proteinase (AcCP-2) in eggs and larvae stages of hookworm Ancylostoma caninum. Exp. Parasitol. 129, 215–220. https://doi.org/10.1016/j.exppara.2011.08.018
Žebrakovská, I., Máša, M., Srp, J., Horn, M., Vávrová, K., Mareš, M., 2011. Complex modulation of peptidolytic activity of cathepsin D by sphingolipids. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1811, 1097–1104. https://doi.org/10.1016/j.bbalip.2011.09.005