Helminth pathogen cathepsin proteases: it's a family affair

Mark W. Robinson, John P. Dalton and Sheila Donnelly

Institute for the Biotechnology of Infectious Diseases, University of Technology Sydney, Level 6, Building 4, Corner of Thomas and Harris Street, Ultimo, Sydney, NSW 2007, Australia

Corresponding author: Robinson, M.W. (mark.robinson@uts.edu.au).

Helminth pathogens express papain-like cysteine peptidases, termed cathepsins, which have important roles in virulence, including host entry, tissue migration and the suppression of host immune responses. The liver fluke *Fasciola hepatica*, an emerging human pathogen, expresses the largest cathepsin L cysteine protease family yet described. Recent phylogenetic, biochemical and structural studies indicate that this family contains five separate clades, which exhibit overlapping but distinct substrate specificities created by a process of gene duplication followed by subtle residue divergence within the protease active site. The developmentally regulated expression of these proteases correlates with the passage of the parasite through host tissues and its encounters with different host macromolecules.

Papain-like cysteine proteases

Papain-like lysosomal cysteine proteases of the clan CA are expressed by all organisms from bacteria to mammals [1]. Humans express 11 different functional papain-like cysteine proteases (cathepsins L, S, K, F, W, V, O, B, C, H and X) that share a high degree of homology and substrate specificity because they arose from a common ancestor [2]. Some, including cathepsins B, L and S, are ubiquitously expressed in tissues and perform general household functions in protein turnover, although in immune cells they participate in antigen or major histocompatibility complex (MHC) II invariant chain processing [3]. Others display more restricted expression patterns and perform more specialized functions; for example, in osteoblasts, cathepsin K participates in bone re-modelling [4]. The papain-like cysteine proteases are of particular importance in medicine because their uncontrolled activities have been associated with diverse pathological disease states in humans including cancer metastasis, Alzheimer's disease, muscular dystrophy, osteoporosis, rheumatoid arthritis and osteoarthritis [5]. Therefore, several research programs aim to develop selective inhibitors of these proteases; some, such as cathepsin K inhibitors for osteoporosis treatment, have reached clinical testing [6].

Papain-like cysteine proteases are also used by helminth, or worm, parasites in many aspects of their relationship with their animal and/or human hosts; for example, cathepsins facilitate skin and intestine infection, tissue migration, feeding and suppression of host immune effector cell functions [7]. Helminths are complex multi-cellular organisms that include parasitic roundworms (nematodes) and flatworms (trematodes and cestodes). Human diseases caused by helminth parasites are among the most prevalent on earth, with >1 billion people infected worldwide, predominantly in poverty-stricken regions in developing countries. Despite the enormous amount of disease they cause and the burden they place on the global economy, the armoury of chemotherapeutic drugs available to combat these parasites is limited, consisting predominantly of praziquantel, ivermectin, albendazole and oxamniquine (for a review, see Ref. [8]). As with many other forms of pathogens, drug-resistant helminths have emerged and, therefore, there is an urgent need to identify additional strategies and target molecules for the development of new anti-parasitic drugs.

Owing to the roles of papain-like cysteine proteases in fundamental aspects of the parasitic way of life, cathepsins are lead targets for the development of new chemotherapeutic interventions. Recent studies justify this approach; cysteine protease inhibitors impair the growth and fecundity of the liver fluke *Fasciola hepatica* in infected sheep [9] and the vinyl sulfone cysteine protease inhibitor K11777 causes considerable reductions of worm burden and parasite egg output in mice infected with the blood fluke *Schistosoma mansoni* [10]. K11777 also exhibits potent activity against

some non-helminth parasites including *Trypanosoma cruzi* (the causative agent of Chagas' disease) [11] and *Entamoeba histolytica* [12], thereby indicating that it might be possible to develop broad-spectrum anti-parasitics that could be employed in the treatment of individuals co-infected with pathogenic helminths and protozoans. However, whereas the development of anti-parasitic drugs that target cysteine proteases might seem tangible, the fact that a large number of homologous proteases with overlapping specificity, as mentioned earlier, also have crucial functions in human tissues and are associated with various disease conditions presents a major challenge in identifying and developing parasite-specific inhibitors. Accordingly, it is imperative that we learn more about the biochemistry, structure and function of the parasitic proteases to enable comparisons to the human enzymes; these comparisons should facilitate the discovery of compounds that selectively inhibit the parasite enzymes.

Much of our biochemical knowledge of the papain-like cathepsin cysteine proteases of helminth parasites comes from studies of trematodes, or flukes, of medical and veterinary significance. These include the blood flukes of the genus *Schistosoma* that cause bilharzia, or schistosomiasis, a disease afflicting >200 million people in >70 tropical countries and resulting in >280 000 deaths annually in sub-Saharan Africa alone [13]. Also included are several fish-borne infections caused by liver flukes of the genuses *Clonorchis*, *Paragonimus* and *Opisthorchis*, which put >600 million people at risk of infection across Asia [14,15], and liver flukes of the genus *Fasciola* that cause fasciolosis in animals such as sheep, cattle and water buffalo. Although fasciolosis was traditionally regarded as a world-wide disease of livestock, the disease has recently emerged as an important human zoonosis in South America, Egypt, Iran and Vietnam. Estimates indicate that 2.4–17.0 million people are infected worldwide, with a further 91.1 million people currently living at risk of infection [14,16,17].

Here, we discuss how the biochemical and structural properties of papain-like cathepsin cysteine proteases of helminth parasites have evolved and adapted to perform their particular function in host–parasite interaction. We focus primarily on the cathepsin L proteases of *F. hepatica* because we have a better understanding of their phylogeny, biochemistry and function. We describe how subtle changes that occurred in the residues constituting the active site of the enzymes resulted in dramatic functional adaptations that enable the proteases to deal with the host macromolecules they encounter during migration and development of the parasite.

Trematode (fluke) cysteine proteases

Phylogenetic studies indicate that, at the time of helminth worm emergence, 480–540 million years ago (MYA) [18] several cathepsin classes had evolved: (i) two with endopeptidase activity (cathepsin L and F); (ii) one with both exo- and endopeptidase activity (cathepsin B); and (iii) one with dipeptidylpeptidase activity (cathepsin C) [19–21] (Figure 1). This comparatively small array of papain-like cysteine proteases had a major role in the evolution and success of parasitism. Indeed, further expansion of these protease classes into families took place within individual parasitic lineages, thus, reflecting an adaptation to a new way of life in which nutrients were readily available but, at the same time, in which more energy was required to protect against host responses and to produce large numbers of offspring [20]. Interestingly, however, a clear divergence in the type of cathepsins expressed by different trematode species is observed: *Fasciola* spp. and *Schistosoma* spp. predominantly express cathepsin L-like proteases [22,23], whereas cathepsin F-like proteases are the major proteases expressed by *Clonorchis* spp., *Paragonimus* spp. and *Opisthorchis* spp. [24,25]. The reason(s) for this difference is unclear at present but, unlike *Fasciola* spp. and *Schistosoma* spp., the species of the other three parasitic genuses use fish as a second intermediate host, which could expose these pathogens to unique selective pressures that resulted in the evolution of cathepsin F-like protease families rather than cathepsin L-like enzymes.

Cysteine proteases make up a large proportion of the total transcriptome of each of these trematode parasites, further emphasizing their importance in parasitism. For example, nearly 15% of the transcripts derived from adult *F. hepatica* [7] (http://www.sanger.ac.uk/Projects/Helminths/), 10% from adult *C. sinensis* [26,27] and 18% and 27% from adult diploid and triploid *Paragonimus westermani*, respectively [28], encode cysteine proteases. Immunocytochemical studies show that helminth cysteine proteases are predominantly associated with the cells lining the gut lumen (gastrodermis), the reproductive structures (vitellaria) and the

surface layers (tegument) of parasites, sites at which rapid protein degradation and synthesis takes place [29]. Cathepsin L, F and B proteases are also commonly found among the profile of molecules secreted by helminths, termed their secretome, which enables them to perform their functions in parasite–host interactions, for example, facilitating tissue penetration and disarming the soluble and cellular arms of the host immune system [23,30]. Cathepsin C, by contrast, functions in the hydrolysis of peptide products of host tissues and macromolecules that are degraded in the gut lumen and absorbed into the gastrodermis [7,31]. Further hydrolysis by aminopeptidases within the same cells leads to the release of free amino acids that are then distributed throughout the tissues of the parasite and used for growth and development [7].

Fasciola hepatica *papain-like cysteine proteases*

F. hepatica expresses and secretes cathepsin L and B proteases, but not cathepsin F. Cysteine protease production is highly regulated during the development of the parasite in its hosts and correlates with its migration and maturation. Dormant larvae emerge from cysts in the duodenum and infect their mammalian hosts by penetrating the intestinal wall (Figure 2). Recently, McGonigle *et al.* [32] showed that RNA interference-mediated knock-down of either cathepsin B or cathepsin L transcripts in these infective juveniles blocked their ability to penetrate the host intestinal wall, thus, demonstrating that both enzyme types are required for host entry. Studies on the infective larvae of other trematodes, including those of *Schistosoma* spp., *Trichobilharzia* spp., *Paragonimus* spp. and *Clonorchis* spp., point to a universal use of cysteine proteases in host infection [19,33,34].

As the *F. hepatica* parasites enter the liver parenchyma and begin feeding on the tissue, cathepsin B expression diminishes and expression of cathepsin L proteases is upregulated. Biochemical and immunological studies indicate that, by the time the parasites have moved into the immunologically safe environment of the bile ducts, they become reliant solely on cathepsin L proteases [23]. Within the bile ducts, the parasite is an obligate blood-feeder and, hence, its nutrients are acquired predominantly from host haemoglobin and serum proteins.

F. hepatica cathepsins are synthesized within the gastrodermal epithelial cells of the parasite and are stored in specialized secretory vesicles as inactive zymogens [35]. Recently, a 1.4 Å atomic structure of the major *F. hepatica* cathepsin L1 (FhCL1) zymogen was determined [36]; it is almost identical in tertiary structure to human cathepsin L1, although their primary structures exhibit only 35% identity [37]. The mature domains of the enzymes are arranged in two lobes: the left-hand lobe is predominantly composed of α-helices, whereas the right-hand lobe contains several elements of β-sheet. The substrate-binding cleft, containing the catalytic machinery of the enzyme, runs between the two lobes (Figure 3). Similarly to their mammalian counterparts, the *Fasciola* spp. zymogens contain an N-terminal extension, or prosegment, that binds the substrate cleft and prevents uncontrolled proteolysis during trafficking and storage [35–37] (Figures 3,4). Removal of the prosegment occurs by specific autocatalytic processing events and is facilitated by a low pH environment. In mammalian cells, this environment is provided by the acidic lysosome but, in helminth parasites, activation occurs within the slightly acidic milieu of the parasite gut lumen [20,35]. In the case of *F. hepatica* cathepsin L, we have shown that auto-activation from a pro-enzyme to a mature active enzyme occurs over a broad pH range (4.0–7.3), but is ~40-fold faster at pH 4.5 than at pH 7.0 [38]. Thus, the acidic gut of the parasite provides a perfect environment to facilitate denaturing of ingested host macromolecules and activation of the digestive enzymes (not dissimilar to the mammalian gut). A remarkable feature of FhCL1 is its broad pH range for activity (pH 3.0–9.0) coupled with high stability at neutral pH (up to 10 days at 37°C); by contrast, the human cathepsin L homologue is most active at pH 4.5 and inactivates within minutes at pH 7.0 [39]. These properties enable the parasite enzyme to function for long periods within the parasite gut (pH 5.5), in which 1.5×10^8 red blood cells are degraded $hour^{-1}$ $worm^{-1}$, and in the host tissue (pH 7.3), which is degraded to facilitate their migration.

Phylogenetic studies reveal a large family of F. hepatica *cathepsin L proteases*

The cathepsin L proteases have undergone a great expansion within the *Fasciola* spp. A phylogenetic analysis of *F. hepatica* genes deposited in public databases revealed that this parasite expresses the largest monophyletic cysteine protease family described to date [23,40]. The family can be

grouped into five distinct phylogenetic clades; complementary DNAs (cDNAs) for two of these are found only in the early infective larvae (clades FhCL3 and FhCL4), whereas the remaining three clades (clades FhCL1, FhCL2 and FhCL5) were derived from adult parasites taken from bile ducts [23,40] (Figure 5). This cathepsin L family is estimated to have expanded within the genus *Fasciola* by a series of gene duplications followed by divergence beginning 135 MYA. Irving *et al.* [40] estimated that most of the expansion took place ~25 MYA, which was co-incident with climatic conditions that favoured the development of grasslands and the expansion of common *F. hepatica* mammalian hosts (e.g. cattle and sheep). Accordingly, the divergence of the cathepsin L protease family seems to have been important in the evolution and adaptation of the parasite to a wider host range [23,40].

Proteomics studies of the cysteine proteases secreted by mature liver flukes in culture and *in situ* within the bile duct support the earlier phylogenetic analysis and highlight several important observations [23,41] (Figure 5). First, FhCL3 and FhCL4 cathepsin L proteases were not detected in the adult stage secretory proteins, supporting a specific role for these proteases in the early-stage juvenile flukes that initiate infection through the intestinal wall [32]. Second, adult parasites secreted protease members of all remaining clades (i.e. FhCL1, FhCL2 and FhCL5); however, the clade FhCL1 and FhCL2 proteases were, by far, the most predominantly expressed proteins (accounting for 67.39% and 27.63% of total secreted cathepsin Ls, respectively). This finding indicates that clades FhCL1 and FhCL2 are most crucial for parasite survival and adaptation. Third, proteases of a FhCL1 sub-clade, FgCL1C, were not present in the *F. hepatica* secretome, consistent with our findings [23] that genes in this sub-clade are present in the related tropical liver fluke, *Fasciola gigantica*, but not in the *F. hepatica* nucleotide databases. It seems that the FgCL1C sub-clade expanded most recently and was important in *F. hepatica* and *F. gigantica* speciation ~17 MYA and their subsequent adaptation to their specific hosts [40].

Mature *F. hepatica* did not secrete any proteases other than cathepsin L cysteine proteases, confirming that adult parasites living in the bile duct rely only on these enzymes for feeding on host blood [23,41,42]. The parasite draws blood through punctures it makes in the duct wall and the nutrients obtained from this process are used to produce an enormous number of progeny (30–50 000 eggs $day^{-1}worm^{-1}$) [7]. Accordingly, FhCL1, FhCL2 and FhCL5 cathepsin L proteases must collectively possess hydrolytic machinery capable of digesting the macromolecules in the bloodmeal to peptides that are sufficiently small to be absorbed into the parasite gastrodermis. Recently, we showed that one of these enzymes, FhCL1, can degrade haemoglobin to peptides of between 3–14 amino acids; these presumably are absorbed into the gastrodermis for further catabolism to amino acids by intracellular cathepsin C and aminopeptidases [7,21,38].

All cathepsin L proteases detected in the secretions of adult *F. hepatica* were fully processed to mature enzymes, indicating that the events that lead to prosegment removal is completed within the parasite gut lumen [23]. A comparison of the primary structures of the prosegments of the various phylogenetic clades shows that they are highly conserved, except within a 21 amino acid stretch before the junction between the prosegment and mature enzyme [23,37,43] (Figure 4). The FhCL1 atomic structure shows that the final 12 residues in this prosegment C-terminal region do not make contacts with the main body of the enzyme but, rather, move freely in space, creating a 'protease-accessible' region that facilitates prosegment removal [43]. We have proposed that the cleavage events leading to prosegment removal take place in two steps [23]. Initially, in a bimolecular process, a small number of inactive cathepsin L zymogens are *trans*-activated by another enzyme located in the gastrodermal cells, an asparaginyl endopeptidase [44], that cleaves peptide bonds involving Asn residues close to the junction of the prosegment and mature domain. Cathepsin Ls activated by this process are then free to remove prosegments from other cathepsin Ls through specific cleavage at the Leu-Ser↓His motif and, thus, the process increases exponentially [36]. Mutations that alter this motif to a Pro-Ser↓His motif prevent cathepsin L activation [43]. This proposal for enzyme activation is supported by sequence alignments of the five *F. hepatica* cathepsin L clades (Figure 5); despite the lack of conservation at the prosegment C terminus, the cleavage site for asparaginyl endopeptidase and the autocatalytic cleavage Leu-Ser↓His motif are preserved in all phylogenetic clades.

Fasciola cathepsins have varying but overlapping substrate specificities

Gene duplication followed by divergence is one of the primary means by which organisms generate proteins with new functions [45]. The tight regulation of the expression of the various *F. hepatica* cathepsin L clades as the parasite develops implies that the specific biochemical activity of certain clades are required at different stages of development as the parasite moves to new environments and encounters different macromolecular substrates (Figure 2). Understanding the functional diversity of the various members of the gene family, therefore, is important for understanding pathogen virulence and adaptation.

The active site of papain-like cysteine proteases is short; although it comprises four subsites (S2–S1–S1′–S2′) that interact with the corresponding P2–P1–P1′–P2′ residues of the protein substrate, the scissile bond being between P1–P1′ [1,46]. Substrate binding specificity is predominantly governed by the residues that form a deep S2 subsite pocket capable of holding the substrate P2 amino acid and positioning the scissile bond P1–P1′ into the S1 subsite for cleavage [1,46]. Within the 3D structures of FhCL1, human cathepsin L and cathepsin K, the S2 subsite is composed of residues occupying positions 67, 68, 133, 157, 160 and 205, which exhibit variation among members of the papain superfamily. An analysis of these residues in the various *Fasciola* spp. cathepsin L clades reveals divergence within the S2 subsite, in particular at the three positions that have the greatest influence on P2 binding (i.e. at residues 67, 157 and 205) [36] (Figure 3). Our biochemical data using recombinant versions of FhCL1 and FhCL2 proteases show how important changes in these positions can be; for example, FhCL1 (Leu67, Val157 and Leu205) cleaves substrates with hydrophobic residues (Phe and Leu) in the P2 position with catalytic rates (k_{cat}/K_m) that are 25- and eightfold greater, respectively, than FhCL2 (Tyr67, Leu157 and Leu205). Furthermore, in comparison to human cathepsin L (Leu67, Met157, Ala205), which can accommodate a wide range of amino acids in the S2 subsite, the S2 subsite of FhCL1 is restricted. Using mass spectrometry, we found that those residues in host haemoglobin that were most susceptible to cleavage by FhCL1 were invariably hydrophobic residues and in the order Leu>Val>Ala>Phe. Together, these four residues make up ~42% of the haemoglobin molecule and, therefore, it seems that FhCL1 has been specifically adapted to degrade the host substrate, which it exploits as nutrient [38].

A comparison with the parasite and human cysteine protease homologues shows that the FhCL2 S2 subsite is identical to that of human cathepsin K (Figure 3). Biochemical studies indicate that, like cathepsin K, FhCL2 readily accommodates substrates with a bulky proline residue in the P2 position; by contrast, these substrates are poorly cleaved by FhCL1 and not at all by human cathepsin L [43]. Both cathepsin K and FhCL2 also possess the unique ability to cleave native collagen, which contains a repeat motif of Gly-Pro-Xaa (in which Xaa is any amino acid). Comparison of the atomic structures of the two enzymes show that this unique collagenase-like activity is, in part, due to the presence of Tyr67 (not found at this position in any other cathepsin proteases) at the entrance of the S2 pocket; this site can form stabilizing ring–ring interactions between the five-membered proline ring of the substrate [36]. This is an example of convergent evolution, whereby positive selection acting on the residues within the active site of two vastly distant enzymes has endowed them with similar biochemical and functional properties. In humans, cathepsin-K-mediated collagen cleavage is essential to its function in bone remodelling [47]; by contrast, in *Fasciola* spp., which must penetrate and migrate through large host organs including the liver, this FhCL2 activity is crucial for degrading collagen-rich interstitial matrices.

Interestingly, adult clade FhCL5 that diverged from FhCL2 before its divergence from FhCL1 (Figure 5) exhibits an intermediate S2 subsite (Leu67, Leu157 and Leu205). Like clade FhCL1, it does not readily accommodate substrates with a P2 proline residue [48]. Thus, separation of the FhCL1 and FhCL5 clades from the FhCL2 clade was accompanied by the loss of the ability to cleave substrates, like collagen, with a P2 proline. At the same time, this alteration enabled the S2 pocket of FhCL1 to open up so that it could accommodate hydrophobic residues that were most common in host haemoglobin. No biochemical data are available for the infective larval proteases of clades FhCL3 and FhCL4 but, because these are considered the most ancient of the clades from which the FhCL2 and FhCL5–FhCL1 lineages diverged, it is likely that they possess hybrid-like specificities [40].

Concluding remarks

Helminth parasites have exploited papain-like cysteine proteases to perform a wide

range of functions that enable them to reside in their mammalian hosts. Despite the evolutionary distance between helminth parasites and humans, their papain-like cysteine proteases share many biochemical and structural similarities. Given that substrate specificity in these proteases is largely governed by a few interactions between the substrate P2 amino acid and residues within the S2 subsite of the enzyme, it is not surprising that the differences observed between the parasite and human enzymes are minor. Turk and Guncar [1] suggest that it might be difficult to design inhibitors that can distinguish between the human enzymes because they differ only subtly in their substrate specificity. Finding inhibitors that potently inactivate parasite papain-like cysteine proteases without affecting one or more host enzymes will pose an even greater challenge. However, drug design efforts could exploit bioavailability and pharmo-kinetic and -dynamic properties to enhance the probability of inhibitors preferentially reaching their parasite target. Availability of recombinant functional parasite proteases will be invaluable in the search for selective lead-like compounds with these characteristics within synthetic or natural chemical libraries. Additionally, or alternatively, we could take advantage of the exquisite selectivity by which the host immune system can distinguish foreign molecules and epitopes to target the papain-like cathepsin cysteine proteases of helminth parasites as vaccine molecules. Such an approach has been successfully taken in the path towards the development of a first-generation recombinant FhCL1 vaccine that protects domestic animals against infection with the liver fluke *F. hepatica* [49].

References

1 Turk, D. and Guncar, G. (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr. D Biol. Crystallogr.* 59, 203–213

2 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

3 Podolin, P.L. *et al.* (2008) Inhibition of invariant chain processing, antigen-induced proliferative responses, and the development of collagen-induced arthritis and experimental autoimmune encephalomyelitis by a small molecule cysteine protease inhibitor. *J. Immunol.* 180, 7989–8003

4 Avnet, S. *et al.* (2006) Effects of antisense mediated inhibition of cathepsin K on human osteoclasts obtained from peripheral blood. *J. Orthop. Res.* 24, 1699–1708

5 Vasiljeva, O. *et al.* (2007) Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr. Pharm. Des.* 13, 387–403

6 Stoch, S.A. and Wagner, J.A. (2008) Cathepsin K inhibitors: a novel target for osteoporosis therapy. *Clin. Pharmacol. Ther.* 83, 172–176

7 Dalton, J.P. *et al.* (2006) Proteases in trematode biology. In *Parasitic Flatworms: Molecular biology, Biochemistry, Immunology and Physiology* (Maule, A.G. and Marks, N.J., eds), pp. 348–368, CAB International

8 Hotez, P.J. *et al.* (2008) Helminth infections: the great neglected tropical diseases. *J. Clin. Invest.* 118, 1311–1321

9 Alcala-Canto, Y. *et al.* (2007) Effect of a cysteine protease inhibitor on *Fasciola hepatica* (liver fluke) fecundity, egg viability, parasite burden, and size in experimentally infected sheep. *Parasitol. Res.* 100, 461–465

10 Abdulla, M.H. *et al.* (2007) *Schistosomiasis mansoni*: novel chemotherapy using a cysteine protease inhibitor. *PLoS Med.* 4, e14

11 Barr, S.C. *et al.* (2005) A cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi*. *Antimicrob. Agents Chemother.* 49, 5160–5161

12 Melendez-Lopez, S.G. *et al.* (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

13 Gryseels, B. *et al.* (2006) Human schistosomiasis. *Lancet* 368, 1106–1118

14 Keiser, J. and Utzinger, J. (2005) Emerging foodborne trematodiasis. *Emerg. Infect. Dis.* 11, 1507–1514

15 Lun, Z.R. *et al.* (2005) Clonorchiasis: a key foodborne zoonosis in China. *Lancet Infect. Dis.* 5, 31–41

16 Mas-Coma, S. *et al.* (2005) Fascioliasis and other plant-borne trematode zoonoses. *Int. J. Parasitol.* 35, 1255–1278

17 Parkinson, M. *et al.* (2007) Endemic human fasciolosis in the Bolivian Altiplano. *Epidemiol. Infect.* 135, 669–674

18 Littlewood, D.T.J. (2006) The evolution of parasitism in flatworms. In *Parasitic Flatworms: Molecular biology, Biochemistry, Immunology and Physiology* (Maule, A.G. and Marks, N.J. eds), pp. 1–36, CAB International

19 Tort, J. *et al.* (1999) Proteinases and associated genes of parasitic helminths. *Adv. Parasitol.* 43, 161–266

20 Dalton, J.P. *et al.* (2004) Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. *Can. J. Zool.* 82, 211–232

21 Caffrey, C.R. *et al.* (2004) Blood 'n' guts: an update on schistosome digestive peptidases. *Trends Parasitol.* 20, 241–248

22 Dvorak, J. *et al.* (2008) Differential use of protease families for invasion by schistosome cercariae. *Biochemie* 90, 345–358

23 Robinson, M.W. *et al.* (2008) Proteomic and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen, *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol. Cell. Proteomics* 7, 1111–1123

24 Park, H. *et al.* (2001) *Paragonimus westermani*: cloning of a cathepsin F-like cysteine proteinase from the adult worm. *Exp. Parasitol.* 98, 223–227

25 Na, B.K. *et al.* (2007) CsCF-6, a novel cathepsin F-like cysteine protease for nutrient uptake of *Clonorchis sinensis*. *Int. J. Parasitol.* 38, 493–502

26 Cho, P.Y. *et al.* (2006) Expressed sequence tag analysis of adult *Clonorchis sinensis*, the Chinese liver fluke. *Parasitol. Res.* 99, 602–608

27 Cho, P.Y. *et al.* (2008) Gene expression profile of *Clonorchis sinensis* metacercariae. *Parasitol. Res.* 102, 277–282

28 Kim, T.S. *et al.* (2006) Comparison of gene representation between diploid and triploid *Paragonimus westermani* by expressed sequence tag analyses. *J. Parasitol.* 92, 803–816

29 Fairweather, I. *et al.* (1999) Development of *Fasciola hepatica* in the mammalian host. In *Fasciolosis* (Dalton, J.P., ed.), pp. 47–112, CABI

30 Curwen, R.S. *et al.* (2006) Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Mol. Cell. Proteomics* 5, 835–844

31 Brindley, P.J. *et al.* (1997) Proteolytic degradation of host hemoglobin by schistosomes. *Mol. Biochem. Parasitol.* 89, 1–9

32 McGonigle, L. *et al.* (2008) The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int. J. Parasitol.* 38, 149–155

33 Dalton, J.P. *et al.* (1997) The cysteine proteinases of *Schistosoma mansoni* cercariae. *Parasitology* 114, 105–112

34 Kašný, M. *et al.* (2007) Comparison of cysteine and serine peptidase activities in *Trichobilharzia regenti* and *Schistosoma mansoni* cercariae. *Parasitology* 134, 1599–1609

35 Collins, P.R. *et al.* (2004) Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J. Biol. Chem.* 279, 17038–17046

36 Stack, C.M. *et al.* (2008) Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. *J. Biol. Chem.* 283, 9896–9908

37 Coulombe, R. *et al.* (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* 15, 5492–5503

38 Lowther, J. *et al.* The importance of pH in regulating the function of *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Negl. Trop. Dis.* (in press)

39 Dowd, A.J. *et al.* (2000) Stability studies on the cathepsin L proteinase of the helminth parasite, *Fasciola hepatica*. *Enzyme Microb. Technol.* 27, 599–604

40 Irving, J.A. *et al.* (2003) The evolution of enzyme specificity in *Fasciola* spp. *J. Mol. Evol.* 57, 1–15

41 Morphew, R.M. *et al.* (2007) Comparative proteomics of excretory-secretory proteins released by the liver fluke *Fasciola hepatica* in sheep host bile and during *in vitro* culture ex host. *Mol. Cell. Proteomics* 6, 963–972

42 Lee, E.G. *et al.* (2006) Identification of immunodominant excretory-secretory cysteine proteases of adult *Paragonimus westermani* by proteome analysis. *Proteomics* 6, 1290–1300

43 Stack, C.M. *et al.* (2007) The major secreted cathepsin L1 protease of the liver fluke, *Fasciola hepatica*: a Leu-12 to Pro-12 replacement in the nonconserved C-terminal region of the prosegment prevents complete enzyme autoactivation and allows definition of the molecular events in prosegment removal. *J. Biol. Chem.* 282, 16532–16543

44 Adisakwattana, P. *et al.* (2007) Comparative molecular analysis of two asparaginyl endopeptidases and encoding genes from *Fasciola gigantica*. *Mol. Biochem. Parasitol.* 156, 102–116

45 Lynch, M. (2002) Gene duplication and evolution. *Science* 297, 945–947

46 Turk, D. *et al.* (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* 379, 137–147

47 Atley, L.M. *et al.* (2000) Proteolysis of human bone collagen by cathepsin K: characterization of the cleavage sites generating the cross-linked N-telopeptide neoepitope. *Bone* 26, 241–247

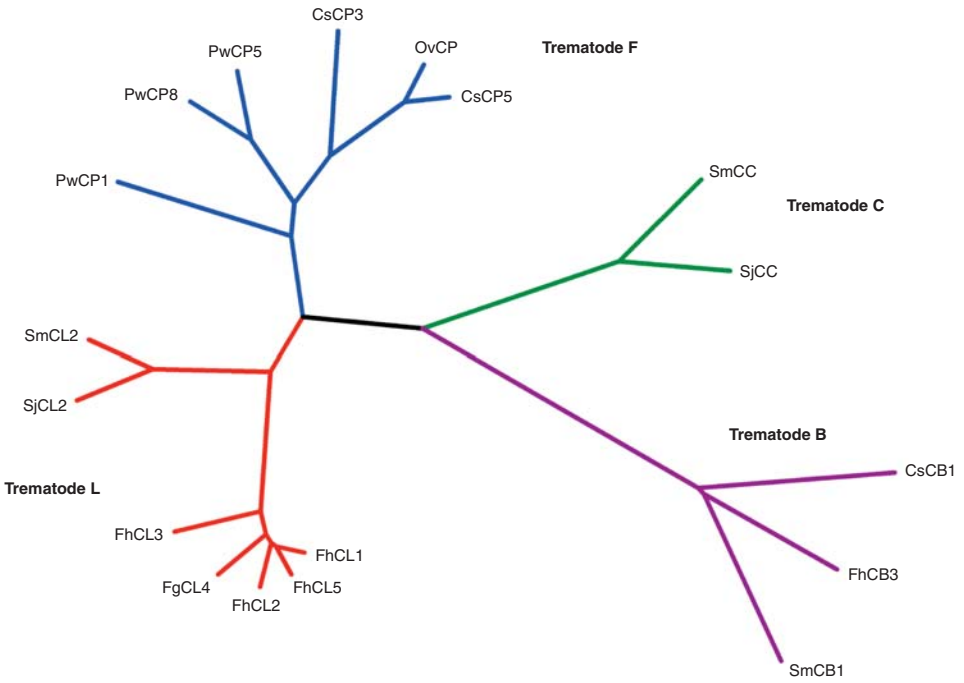
48 Smooker, P.M. *et al.* (2000) A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Sci.* 9, 2567–2572

49 Dalton, J.P. *et al.* (2003) *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *Int. J. Parasitol.* 33, 1173–1181

(a)

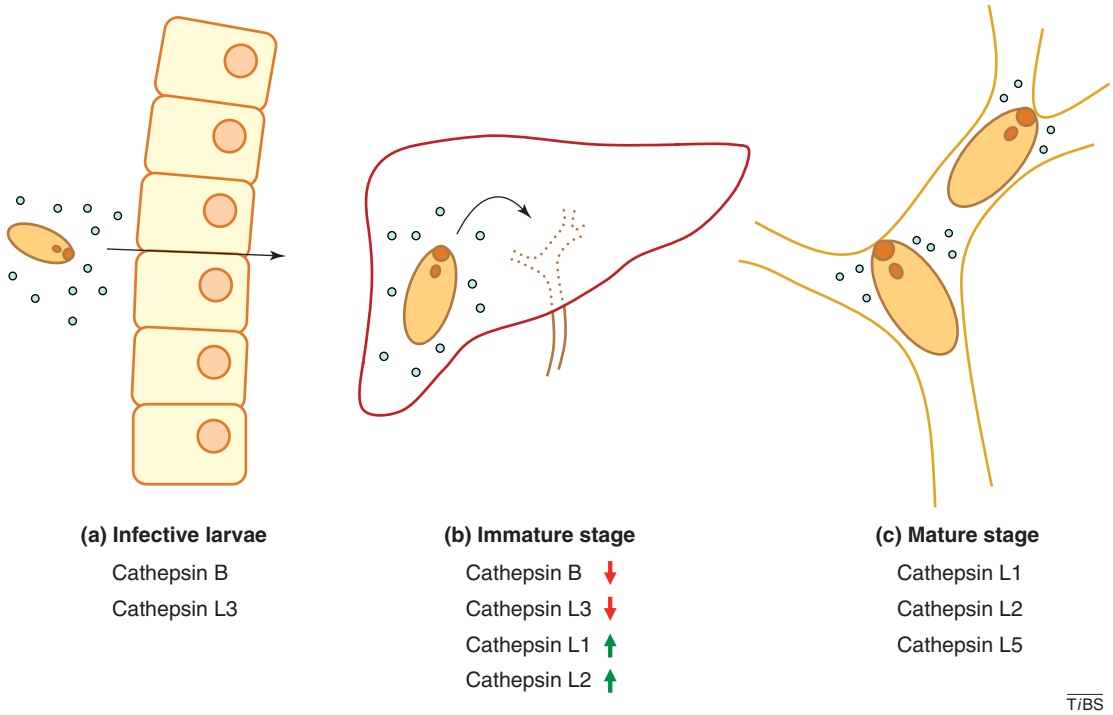
Infection	Trematode	Abbrev.	Definitive hosts	Intermediate hosts	Distribution
Fasciolosis	*Fasciola hepatica*	Fh	Herbivores: sheep, cattle, humans	Freshwater snails	World-wide
	Fasciola gigantica	Fg			Africa, Asia
Schistosomiasis	*Schistosoma japonicum*	Sj	Sheep, cattle, horses, humans	Freshwater snails	Asia-Pacific
	Schistosoma mansoni	Sm			South America, Africa
Clonorchiasis	*Clonorchis sinensis*	Cs	Dogs, pigs, cats, camels, humans	Freshwater snails, carp	Asia-Pacific
Opisthorchiasis	*Opisthorchis viverrini*	Ov	Dogs, cats, pigs, humans	Freshwater snails, fish	Eastern Europe, Asia-Pacific
Paragonimiasis	*Paragonimus westermani*	Pw	Wild and domestic cats, humans	Freshwater snails, crabs	Asia-Pacific

(b)



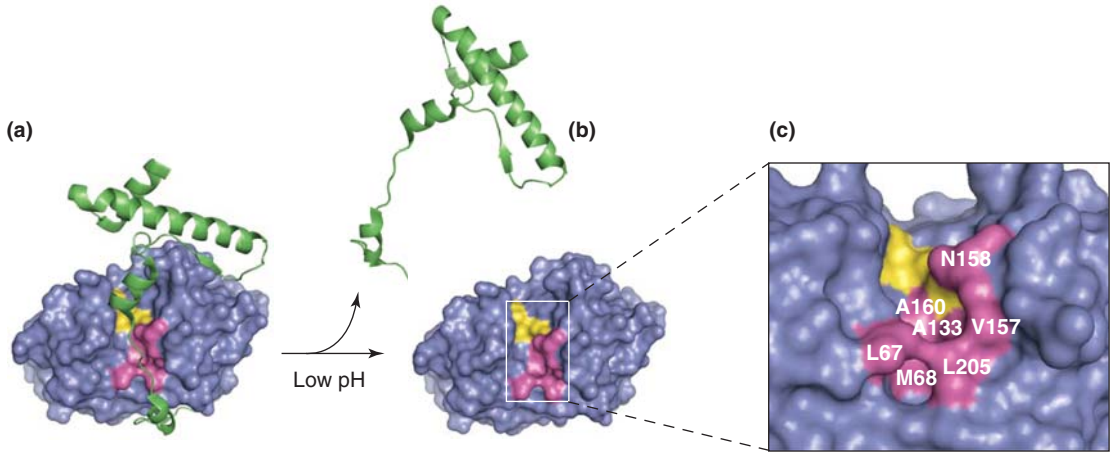
T*i*BS

Figure 1. Papain-like cysteine proteases of trematode pathogens. Human diseases caused by trematode (flatworm) parasites are among the most prevalent on earth and predominantly occur in developing countries. Trematode pathogens express four major classes of cathepsin proteases (L, B, F and C) that have functions in parasite virulence including tissue invasion and feeding. **(a)** The major trematode pathogens of humans and animals. **(b)** Bootstrapped (1000 trials) neighbour-joining phylogenetic tree showing the evolutionary relationship of the trematode cathepsin superfamily. The tree was constructed using trematode cDNA sequences (including the prosegment region). The branches are coloured according to the four major enzyme classes. Abbreviations: CB, cathepsin B; CC, cathepsin C; CL, cathepsin L; CP, cysteine protease; Cs, *Clonorchis sinensis*; Fg, *Fasciola gigantica*; Fh, *Fasciola hepatica*; Ov, *Opisthorchis viverrini*; Pw, *Paragonimus westermani*; Sj, *Schistosoma japonicum*; Sm, *Schistosoma mansoni*. Part (b) adapted from Refs [7,19].



TiBS

Figure 2. Developmental regulation of *F. hepatica* cathepsin proteases. Trematode pathogens such as *F. hepatica* undergo complex life-cycles involving transitions between host organ systems and tissues. Movement from one site to another, as the life-cycle progresses, is associated with the differential expression of cathepsin enzymes. **(a)** Penetration of the host intestinal wall by newly excysted juvenile (NEJ) *F. hepatica* is facilitated by the secretion of proteolytic enzymes. This process involves cathepsin B and clade 3 cathepsin Ls. **(b)** The migration of immature flukes through host liver tissue corresponds to a period of rapid growth and development of the parasite. At this stage, host protein degradation is achieved primarily by *Fasciola* clade 1 and 2 cathepsin Ls, although some cathepsin B and clade 3 and 4 cathepsin L activity could remain. **(c)** Adult *F. hepatica* resides within the bile ducts where it feeds on host red blood cells. Proteomics analyses have shown that clade 1, 2 and 5 cathepsin Ls are the only proteases secreted by adult flukes, thus, indicating crucial roles for these enzymes in penetrating the bile duct wall and digesting host haemoglobin.



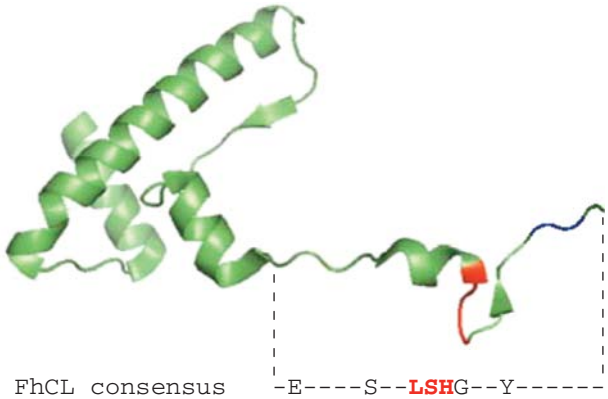
(d)

		Residues						
		67	68	133	157	158	160	205
Human cathepsin L		Leu	Met	Ala	Met	Asp	Gly	Ala
Human cathepsin K		Tyr	Met	Ala	Leu	Asn	Ala	Leu
Adult:	FhCL 1A	Leu	Met	Ala	Val	Asn	Ala	Leu
	FhCL 1B	Leu	Met	Ala	Leu	Asn	Ala	Leu
	FhCL 2	Tyr	Met	Ala	Leu	Thr	Ala	Leu
	FhCL 5	Leu	Met	Ala	Leu	Asn	Gly	Leu
Juvenile:	FhCL 3	Trp	Met	Ala	Val	Thr	Ala	Val
	FhCL 4	Phe	Met	Ala	Leu	Asn	Ala	Phe

TiBS

Figure 3. A structural analysis of *F. hepatica* cathepsin L proteases. *Fasciola* spp. cathepsin Ls are stored in specialized secretory vesicles within the gut epithelial cells of the parasite as inactive zymogens, which consist of a prosegment and mature enzyme domain. The prosegment is removed by catalytic cleavage after secretion into the parasite intestine to reveal an active mature protease. **(a)** Surface representation of the major *F. hepatica* secreted cathepsin L protease (FhCL1) (Protein Data Bank [PDB] ID: 2O6X). The prosegment is shown as a green ribbon and the catalytic machinery is shown in yellow (P1 residues) and magenta (P2 residues). **(b)** At low pH (4.0–4.5), the prosegment is auto-catalytically removed to produce the mature active enzyme. **(c)** Detail from the active site of the enzyme. S2 residues that determine substrate-binding specificity are labelled. Figures were created using Pymol (http://www.pymol.org). **(d)** Comparison of the residues from the S2 active site that contribute to differential substrate-binding in *F. hepatica* cathepsin Ls (clades FhCL1–FhCL5) and human cathepsins L and K. The S2 residues that confer substrate specificity are in red. Residues were identified using primary sequence alignments and analyses of the FhCL1 atomic structure (PDB ID: 2O6X) [36].

(a)



(b)

```
FhCL consensus    -E----S--LSHG--Y------
FhCL 1A_ie1       T.MSRA.DI..H.VP.EANNRa
FhCL 1A_pe        T.MSRA.DI..H.VP.EANNRa
FhCL 1A_au3       T.MSRA.DI..H.VP.EANNRa
FhCL 1A_pl        T.MSRA.DI..H.VP.ETNNRa
FhCL 1A_pt1       T.MSRA.DI..H.VP.ETNNRa
FhCL 1A_tr        T.MPRA.DI..H.IP.EANNRa
CL1A consensus    T.MSRA.DI..H.VP.EANNRa

FhCL 1B_n11       T.MPRA.DI..H.IP.EANNRa
FhCL 1B_ar        T.MPRA.DI..H.IP.EANNRa
CL1B consensus    T.MPRA.DI..H.IP.EANNRa

FhCL 2_jn3        I.MPRS.EL..H.IP.KAKNRa
FhCL 2_ie2        I.IPRS.EL..R.IP.KANKLa
FhCL 2_chC        I.IPRS.EL..R.IP.KANKLa
FgCL 2_thD        I.IPRS.EL..R.IP.KANKPa
CL2 consensus     I.IPRS.EL..R.IP.KANKLa

FhCL 5_au4        T.MPRA.EL..H.IP.KANKRa
FgCL 5_thC        T.MPHR.DI..H.IP.EANKRa
FhCL 5_au5        T.MPRA.EL..H.IP.KANKRa
CL5 consensus     T.MPRA.EL..H.IP.KANKRa

FhCL 3_nl64       M.MSPV.ES..D.IS.EAEGKd
FgCL 3_thG        I.MSPE.ES..D.IS.EAEGNd
FhCL 3_nl22       M.MSPV.ES..D.VS.EAEGNd
FhCL 3_uy3        M.MSPE.ES..D.IS.EAEGNd
CL3 consensus     M.MSP-.ES..D.IS.EAEGNd

FhCL 4_uy4        R.IPRA.DM..H.IP.EANDRa
FgCL 4_thH        R.IPRA.DIH.H.IP.EANDRa
CL4 consensus     R.IPRA.D--.H.IP.EANDRa
                                  TiBS
```

Figure 4. A structural and sequence analysis of *Fasciola* cathepsin L prosegments. Cathepsin L prosegments regulate enzyme activity by binding to the substrate cleft and act as molecular chaperones to ensure correct folding of the enzyme. Prosegment removal is mediated by cleavage at two conserved motifs towards the C-terminal end of the molecule. **(a)** The atomic structure of the *F. hepatica* major secreted cathepsin L protease prosegment region (PDB ID: 2O6X). The conserved Leu-Ser↓His motif required for zymogen auto-catalytic processing is shown in red and the asparagine residues required for *trans*-activation by apsaraginyl endopeptidases are shown in blue. The region containing the Leu-Ser↓His motif and asparagine residues in the atomic structure of the molecule is disordered; however, they have been drawn for illustrative purposes only. The figure was created using Pymol (http://www.pymol.org). **(b)** Primary sequence alignment of the non-conserved *Fasciola* cathepsin L protease prosegment C-terminal regions. A consensus sequence for all *Fasciola* cathepsin Ls (FhCL consensus) is shown at the top of the alignment as are consensus sequences for individual clades (e.g. CL1A consensus, CL1B consensus, etc.). Gaps in the alignment are represented by a dash (–) and amino acids that are conserved in all sequences are indicated by a dot (·). Prosegment residues

are in capital letters, whereas the first amino acids of the mature enzymes are in lower case letters. The colouring of residues follows that of part (a).

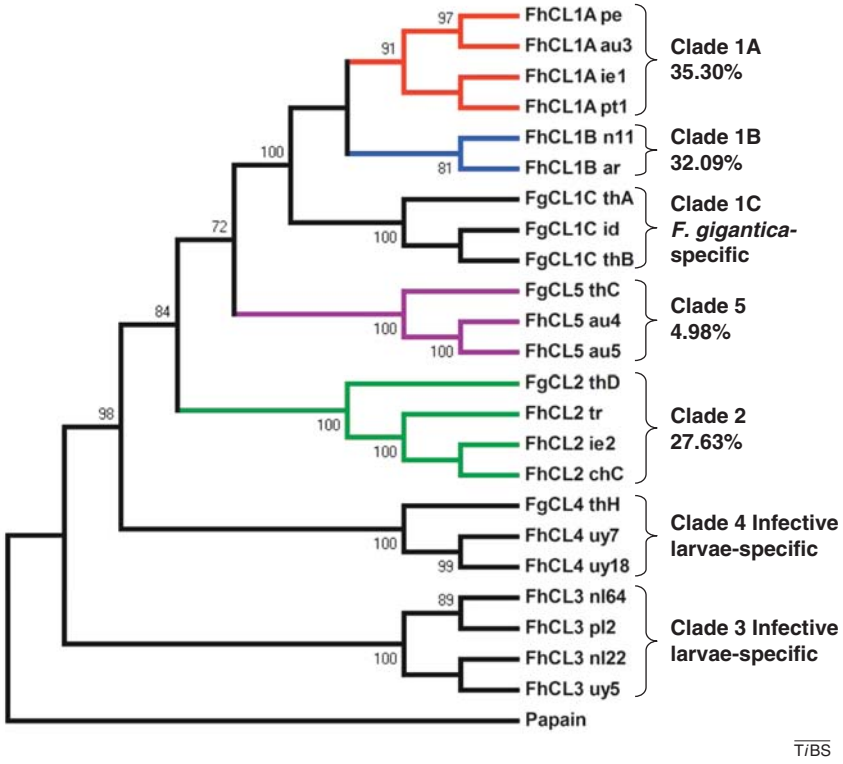


Figure 5. A phylogenetic analysis of the *Fasciola* cathepsin L gene family. The *Fasciola* cathepsin L gene family expanded by a series of gene duplications, followed by divergence of residues in the active site, which gave rise to three clades associated with mature adult worms and two clades specific to infective juvenile stages. These changes resulted in a repertoire of cathepsin L proteases with overlapping and complementary substrate specificities. Bootstrapped (1000 trials) neighbour-joining tree constructed using *F. hepatica* and *F. gigantica* cathepsin L cDNA sequences. The tree is rooted to *Carica papaya* papain. The cathepsin L proteases expressed and secreted by adult flukes are coloured and the brackets and figures represent their expression levels. Figure adapted from Ref. [23].