CHAPTER VI
DISCUSSION

1. Cloning and sequence analysis of the cDNAs encoding legumain-like proteases isolated from *F. gigantica*

Two novel asparginyl endopeptidases or legumain-like proteases and were cloned from *F. gigantica* and named FgLGMN-1 and FgLGMN-2. An adult stage *F. gigantica* cDNA library (Meemon *et al*., 2004) was screened by using a PCR-generated probe with primers against highly conserved regions (NYRHQAD and IYFTDHG) of the legumain family. Full-length cDNAs of 1,438 bp for FgLGMN-1 and 1,365 for FgLGMN-2 were obtained. Although full sequences are different in length, the open reading frames are identical in size (1,275 bp). However, the deduced amino acid sequences showed only 47% identity to each other which implies that they encode distinct members of the legumain family. The deduced amino acid sequences were compared to sequences in the GenBank database. The results revealed that the two legumains of *F. gigantica* belong to the C13 family with similarity to legumains of other species: helminthes (*F. hepatica*, *O. viverrini*, *S. mansoni* and *H. contortus*), mammalian (mouse and human), and plant (*V. mungo* and *A. thaliana*). The FgLGMN-1 and 2 are closely related to FhLGMN with 50% and 42% identity (Table 2), respectively. The legumains in the genus *Fasciola* show less similarity than those of *Schistosoma* which implies that the two legumains of *F. gigantica* should be classified as novel isoenzymes which means that at least three isoenzymes of legumain exist in the genus *Fasciola*. The similarity is high in the related parasites *Opisthrochis* and *Schistosoma*, but lower in mammalians and plants.

The evolutionary relationship was shown by multiple sequence alignment (Figure 15) and the homology was calculated to reconstruct the phylogenetic tree by using the bootstrap-neighbor joining tree method (Figure 18). The phylogenetic tree shows that the FgLGMNs have the closest relationship to the legumains from helminthes (*F. hepatica*, *O. viverrini* and *S. mansoni*) which are categorized in the same clade and are distinguished from the mammalian legumain clade and the plant legumain clade. In contrast, the legumains of the parasitic protozoan, *T. vaginalis*, were classified in

outgroups which were previously proposed to be a new subgroup of legumains (León-Félix *et al*. 2004). The close relationship between parasite and mammalian legumains may be explained in terms of co-evolution caused by host-parasite interaction. The multiple alignment shows several conserved regions including six cysteine residues that are required to maintain the basic structure and essential properties. In addition, the conserved regions in the catalytic site of the legumain family show a catalytic dyad of histidine and cysteine that are characteristics of proteases of the C13 family (Figures 16 and 17). In contrast, the major group of cysteine proteases in parasites, cathepsin B and cathepsin L, which belong to the C1 family, contain the catalytic residues in the order cysteine and histidine (Rawlings and Barrett, 1993). The conserved motif in the catalytic site of the legumain family is composed of two blocks containing predominant hydrophobic amino acid residues, one N- terminus of the catalytic residue is histidine and the other N- terminus of the catalytic residue is cysteine (Chen *et al*., 1998; León-Félix *et al*., 2004; Oliver *et al*., 2006) (Figure 16). The conserved amino acid composition and hydrophobicity in the active site of the enzyme are required for the conformation of the active site including enzymatic function and substrate specificity. However, the crystallography of the enzyme structure should be studied to elucidate the basic properties and active domain conformation further. Fascinatingly, the cysteine residue in the legumain of *F. hepatica* is replaced by serine which is different from all selected orthologs. Consistently, the *S. mansoni* legumain inactive form containing an asparagines residue in place of cysteine shows no enzymatic activity (Caffrey *et al*., 2002). Probably, the serine-form of *F. hepatica* is the inactive form because no enzymatic activity was shown in mutagenesis experiments with mouse legumain when cysteine at the catalytic residue was replaced by serine (Chen *et al*., 1998). However, the biochemical activity should be analyzed in the future.

2. Characterization of FgLGMN genes and their transcripts

The study presented here shows that *F. gigantica* legumains are encoded by two genes and represent isoenzymes. Two isoenzymes are also identified and characterized in *T. vaginalis* (León-Félix *et al*., 2004). Southern analysis was used to identify the number of legumain genes that encode the isoenzymes. No cross-

hybridization between FgLGMN-1 and -2 cDNAs was found and the different Southern hybridization patterns suggest that FgLGMN-1 and -2 genes belong to different gene families. Several bands detected with FgLGMN-1 and -2 DNA probes could indicate that the isoenzyme encoding genes belong to a polygene family. Unfortunately, no information about genomic DNA of legumain genes is available from other organisms especially from related parasites. According to the number of bands in Southern analysis it can be conclude that FgLGMN-1 and -2 belong to a gene family at least five related genes (Figure 20). However, the specific probe detected only one band that showed a strong signal over the others. The variation in strength of the hybridization signal may depend on the similarity between DNA sequences of the gene member and the probe.

Characterization of mRNA transcription products was performed by Northern analysis. The mRNA of FgLGMN-1 is approximately 2,400 nucleotides in length (Figure 21), of which 1,275 nucleotides represent the coding region. The 3'-untranslated region, not including the poly(A)$^+$ tail is 108 nucleotides long. This indicates a quite long 5'-untranslated region of FgLGMN-1 mRNA. FgLGMN-2 mRNA with a size of 2,100 nucleotides contains also a long 5'-untranslated region (Figure 22). The coding region is 1,275 nucleotides and containes a short 3'-untranslated region of only 47 nucleotides. Northern hybridization was done with RNA from several developmental stages: two-, four- and six-week old juvenile and adult parasites. Both FgLGMN specific probes hybridized to RNA of all developmental stages. This indicates that FgLGMN-1 and -2 are expressed in juvenile and adult parasites and may play an essential role in parasite survival.

RT-PCR was done to detect the mRNA in all developmental stages to prove the results of the Northern analysis. Total RNA from all developmental stages, including metacercariae, 1-, 2-, 4- and 6-week old juvenile and adult parasites, was used to amplify FgLGMN-1 and -2 cDNAs. FgLGMN-1 mRNA could be detected in juvenile and adult parasites but not in metacercariae, whereas FgLGMN-2 was detected in all stages. Detection of FgLGMN-2 in metacercariae suggests that this isoenzyme may be the first isoenzyme in the legumain family that plays some essential function in the parasite such as activation of other proteases in excystation and metabolism, *etc*. The abundant transcription of FgLGMN-1 only in juvenile and adult parasites implies that

this isoenzyme is transcribed after excystation and maintains necessary functions until full development to the adult parasite. Not only legumains, but also other proteases were found to be expressed in several isoenzymes as well, such as cathepsin B 1, 2 and 3 in *F. gigantica* (Meemon *et al*., 2004), and a widely variety of cathepsin L isoenzymes in the genus *Fasciola* (Grams *et al*., 2001). The expression of multi-isoenzymes in organisms may be due to the variety of activity in multipurpose for survival.

3. Detection of FgLGMN-1 and FgLGMN-2 expression in tissues of *F. gigantica*

Tissue-specific distribution of *F. gigantica* legumain transcripts exhibited that the two legumains are specifically transcribed in the gastrodermal epithelial cells of adults (Figures 24 and 27) and 4-week old juveniles (Figures 26, 29 and 30). In consistence with this, the immunohistochemistry with anti-rFgLGMN-1 and -2 sera shows localization of the proteins in the gut in dense granules accumulated on the microvilli surface (Figures 46, 47, 48 and 49). The localization of mRNA transcripts and proteins indicates that legumains are produced for specific purposes in the digestive tract of the parasites, for example digestion or activation of other digestive zymogens to active enzymes (Sajid *et al*., 2003). The gut-specific expression was found in other helminthes (*S. mansoni* and *H. contortus*) as well (Skelly and Shoemaker, 2001; Oliver *et al*., 2006). The expression of parasite legumains is co-localized with those of other essential proteases belonging to the cysteine protease family, cathepsin B and cathepsin L, which are abundantly expressed in the gut. In *S. mansoni*, Sajid *et al*. (2003) demonstrated that the processing of the premature *S. mansoni* cathepsin B1 was rapidly accelerated to its mature catalytic form by *S. mansoni* legumain *in vitro*. The trans-activation of cathepsin B by legumain was demonstrated as well in *F. hepatica* by incubation of rFhcatB1 with rSmLGMN. The rFhcatB1 was processed to the correct size and digested a specific peptide substrate by activation of rSmLGMN (Beckham *et al*., 2006). In *F. gigantica*, cathepsin B1 is transcribed in the gastrodermal epithelial cells (Meemon *et al*., 2004) where it is co-localized with legumains. Probably, the *F. gigantica* cathepsin B1 is activated by legumain-like proteases in the same process of trans-activation in *S. mansoni* and *F. hepatica*. An *Ex vivo* study, where interference RNA (iRNA) knocked out the legumain, suggested an

essential function in the *S. mansoni* gut enzyme network. The reduction of SmCatB1 activity and more than 50% inhibition of albumin degradation suggests that the *in vivo* function of legumains is associated with the activation of cathepsin B, probably together with other proteases, to activate mature enzymes for serum protein digestion and nutrition uptake (Delecroix *et al*., 2006).

4. Expression and purification of recombinant FgLGMN proteins

The isolated *F. gigantica* cDNA fragments were subcloned into the prokaryotic expression vector pET20b(+) and expressed in *E. coli* BL21(DE3)pLysS to obtain recombinant proteins. The recombinant proteins were expressed at a high level in *E. coli* and were stored in the insoluble material (Figure 33). The recombinant FgLGMN-1 and -2 proteins are at the size of approximately 49 and 47 kDa, respectively, which are the were expected sizes. The His-tagged proteins were purified by using Ni-NTA chromatography under denaturing condition. The purified proteins, analyzed by SDS-PAGE showed trace amount of contaminating bacterial proteins (Figures 36 and 37). To improve the quality of recombinant proteins other additional purification steps would be necessary such as second affinity purification by a different tag.

5. Characterization of FgLGMN-1 and FgLGMN-2 proteins

To characterize the native legumains, recombinant FgLGMN proteins were used to immunize experimental animals and specific anti-FgLGMNs sera were obtained. Although FgLGMN-1 and -2 showed 47% identity, no cross-reaction between anti-FgLGMN-1 and -2 was detected (Figure 41). However, the capacity of the detection method used in this study may limit the sensitivity. By immunoblot analysis a native protein was detected with anti-rFgLGMN-1 antiserum of 40 kDa in size and anti-rFgLGMN-2 antiserum of 30 kDa (Figure 42). The recombinant proteins were detected with the expected sizes: 49 and 47 kDa, respectively. The detected products in parasite extracts may correspond to the mature enzymes that are modified by post-translational processing or activation. In *S. mansoni*, the mature protease is present in a molecular size of 32 kDa while the recombinant pro-preprotein is 50 kDa in size (Götz and Klinkert, 1993; Caffrey *et al*., 2000). In mammals, the *in vitro* autocatalytic form of legumain processes the 56 kDa of the pro-preform to 46 kDa while the *in vivo*

mature protein was detected at a molecular size of 36 kDa (Li *et al*., 2003). Not only enzyme activation, but also N-glycosylation was proposed to modify legumains. The N-glycosylation was found in mammalian legumains (Chen *et al*., 1997) but not in *F. gigantica* (Figure 45), though the N-glycosylation sites were predicted in the deduced amino acid sequences (Figure 13). However, the post-translational processing would be elucidated in further studies. The reason, why the two legumain isoforms detected in crude parasite extracts are different in size may depend on different sites for autocatalytic processing or unknown intracellular processing. In addition, each isoform can have different activity for different purposes. The correct functions should be studied in the future.

6. Immune response against recombinant FgLGMN proteins

The specific antibody of infected animals detects the recombinant legumains four weeks after infection (Figures 39 and 40). This shows that legumains of the parasites can stimulate a host immune response but the mechanisms are still not clear because in excretory-secretory products no legumains were detected. The quality of the excretory-secretory products was tested by probing with a monoclonal antibody against FgCatL1 which shows a strong positive band at the expected size of cathepsin L1 (Figure 43). Usually, the presence of a signal peptide indicates that protein is excreted or secreted, but a negative result may be due to a low amount of secreted proteins. Probably, the presence of traces could not be detected because of a limitation of detection sensitivity. Recently, excretory-secretory products from *F. hepatica* were analyzed by proteomics *in vivo* and *in vitro*. Only cathepsin proteases from *F. hepatica* were identified in sheep bile preparations. Other proteins were identified as host proteins including albumin and enolase. Analysis of excretory-secretory products prepared by classical *in vitro* culture indicated that cathepsin L proteases are the major constituents of the *in vitro* ES proteome. In addition, detoxification proteins (glutathione transferase and fatty acid binding protein), actin, the glycolytic enzymes enolase and glyceraldehyde-3P-dehydrogenase were all identified *in vitro* (Morphew *et al*., 2007). In *in vivo* and *in vitro* analyze of *F. hepatica* excretory-secretory products no legumains were found which is consistent to our study.

Variations in the signal intensity of rFgLGMN-1 and -2 were observed, although the same amount of proteins was loaded and hybridized with the same animal serum at the same dilution. These may be caused by different immune response and/or different amounts of the endogenous native proteins in *Fasciola*. However, the immunogenicity variation may explain the evolution any mechanism for survival of the parasite in the host. While high immunogenic isoenzymes are blocked by the antibody, low immunogenic isoenzymes are still active. Although legumains were not detected in excretion-secretion products, the molecular and immunological data of *F. gigantica* legumains suggest that legumains may be the candidates for vaccine or the target of therapeutic drugs. In comparison to leucine aminopeptidases, the peptidases do not have a signal peptide and there is no evidence for secretion but they elicit high immune response and high potential protection (Lucía *et al*., 1999; McCarthy *et al*., 2004). The possibility for vaccine development was demonstrated by vaccination of mice with *S. mansoni* legumain DNA vaccine, which showed no significant decreas in worm burden but decrease in egg production (Chliclia *et al*., 2002). Although, there is no strong evidence that legumains could be ideal candidates for vaccine development, vaccine experiments with subunit peptides or a combination of other candidates (cocktail vaccine) may exhibit the potential protection. Moreover, development of vaccines against fasciolosis has to be studied.

7. Autocatalytic processing and functional assay of recombinant FgLGMN and native proteins

The autocatalytic processing was studied in recombinant FgLGMN-1 and -2 proteins. Recombinant proteins were incubated at different pH in sodium citrate buffer containing 1 mM DTT. At neutral conditions (pH 6 and 7), the autocatalytic processing was inhibited. At acidic conditions (pH 4 and 5), the recombinant proteins processed themselves to smaller products. The activity assay was performed in all conditions to verify that the autoactivation of recombinant FgLGMN proteins is a pH dependent mechanism. Unfortunately, the released pNA value of processed recombinant proteins showed no significant difference to the control (boiled recombinant proteins at different pH value). Previously, the success of *in vitro* autocatalytic processing and the functional assay of recombinant legumain proteins

was studied for mammalian legumains by using a mammalian expression system (Li *et al*., 2003) and *S. mansoni* legumain was expressed in a yeast expression system (Caffrey *et al*., 2000). These experiments indicated that the correct function may require post-translation modification including correct folding and these modification are present in eukaryotic cells but not in prokaryotic cells.

Legumain activity was detected in crude parasite extracts as well. Antigens were incubated with the synthetic peptide substrate (Bz-Asn-pNA) that is specific to legumain and then the released pNA was measured by spectrophotometer at 405 nm. Although, the specific activity value was quite low (Absorbance: mOD), it showed a significant higher value than the negative control (boiled crude worm extract). The low sensitivity of this substrate was reported in other publications as well (Outchkourov *et al*., 2003). To improve the sensitivity of detection, a fluorogenic substrate or other substrates that can elicit higher sensitivity should be used. However, these measurements require special equipment.

Excretory-secretory products were tested for legumain activity by incubating with the legumain specific substrate Bz-Asn-pNA. Legumain activity could not be detected in freshly prepared ESAg. This is consistent with immunoblot analysis which confirmed that legumains in adult *F. gigantica* may not be secreted as ESAg or are present in rare amounts.

The activity of legumain in crude worm extracts showed inhibitor sensitivity with iodoacetamide, a cysteine protease inhibitor. Almost all activity was inhibited which suggested that *F. gigantica* legumains are cysteine proteases. A partial inhibition was detected when PMSF (serine proteases inhibitor) is present and partially inhibits the cysteine protease by reversible reducing the disulphide bond. The chelating agent, EDTA, did not affect the legumain activity which implies that the activity of legumain may not require ion cofactors. Although, *in vitro* studies of legumain activity give clear results, the understanding of function and activity *in vivo* is poor. Consequently, the *in vivo* studies of legumain properties, activation and function have to be proceeded further on.