Characterisation of secreted serine and metalloproteinases from the intestinal parasite *Giardia intestinalis*

Sascha Tobias Krakovka

Master Degree Project in Infection Biology, 45 credits. Autumn 2016 and Spring 2017

Department: ICM, Department of Cell and Molecular Biology

Examiner: Catharina Svensson

Supervisor: Staffan Svärd and Jingyi Liu
<table>
<thead>
<tr>
<th>Content</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>Popular science summary</td>
<td>4</td>
</tr>
<tr>
<td>1.0 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>1.1 Aim</td>
<td>8</td>
</tr>
<tr>
<td>2.0 Material and methods</td>
<td>9</td>
</tr>
<tr>
<td>2.1 Material</td>
<td>9</td>
</tr>
<tr>
<td>2.11 Media</td>
<td>9</td>
</tr>
<tr>
<td>2.12 Buffer</td>
<td>9</td>
</tr>
<tr>
<td>2.13 Staining solutions</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Methods</td>
<td>10</td>
</tr>
<tr>
<td>2.21 Bioinformatical comparison</td>
<td>10</td>
</tr>
<tr>
<td>2.22 Cloning</td>
<td>10</td>
</tr>
<tr>
<td>2.23 Expression and Purification</td>
<td>13</td>
</tr>
<tr>
<td>2.24 Protein detection and analysis</td>
<td>14</td>
</tr>
<tr>
<td>3.0 Results</td>
<td>16</td>
</tr>
<tr>
<td>3.1 Bioinformatical comparison</td>
<td>16</td>
</tr>
<tr>
<td>3.2 Characterisation of serine protease GL50803_6148 and GL50803_15574</td>
<td>24</td>
</tr>
<tr>
<td>3.3 Characterisation of metalloprotease GL50803_15832</td>
<td>24</td>
</tr>
<tr>
<td>3.4 Characterisation of metalloprotease GL50803_17327</td>
<td>25</td>
</tr>
<tr>
<td>4.0 Discussion</td>
<td>33</td>
</tr>
<tr>
<td>4.1 Cloning, expression and purification</td>
<td>33</td>
</tr>
<tr>
<td>4.2 Bioinformatical analysis and experimental characterisation</td>
<td>34</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>36</td>
</tr>
<tr>
<td>References</td>
<td>37</td>
</tr>
<tr>
<td>Appendix</td>
<td>39</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>41</td>
</tr>
</tbody>
</table>
Abstract

The intestinal pathogen *Giardia intestinalis* causes approximately 280 million symptomatic cases of giardiasis per year. As of now few virulence factors are characterised and exact disease patterns are unclear. Two serine and two metalloproteases were picked as potential virulence factors for characterisation after bioinformatical analysis of proteases secreted by *Giardia* during infection. These peptidases likely catalyse the last digestion steps of proline containing peptides into single amino acids. In the human intestine uptake of several amino acids including proline happens more efficiently on the di- and tripeptide level. Especially secretion of dipeptidases can convey an advantage for the pathogen in this context enabling *Giardia* to take up the needed amino acids before the host can access them. A function for the serine proteases can be immunomodulation as many chemokines have conserved proline motifs otherwise protecting them from digestion at their termini. All four proteases were cloned into the pPICZαA plasmid and expressed in *Pichia pastoris* X33. For prolidase-like GL50803_17327 protein could be purified by affinity chromatography and analysed using zymography, phage display and in-vitro activity assays. No clear activity could be shown in the assays used so far. Next a colorimetric assay for digestion of free proline will be set up to prove that it has prolidase activity.
**Giardia intestinalis- a tiny parasite with impressive weaponry**

If you want to go swimming or hiking this summer better be aware not to drink the water unfiltered. Diarrhoea lurks all around and one of its causative agents is evolved to not only infect you but trick your immune system at the same time as stopping you from taking up nutrients from your food.

*Giardia intestinalis* infection starts with the intake of infective cysts by drinking water or eating food contaminated by another human’s faeces beforehand. The cysts are too small to be seen, very few can infect you and they survive long time in fresh water. The parasite then lives in your gut. Every year many cases are reported when people using water for recreational purposes infect themselves. About two weeks later some experience heavy episodes of diarrhoea and abdominal cramping. On the other hand, it is estimated that many infected do not even express any symptoms, making it hard to track infections.

To date it is not fully understood how and when and which symptom is caused in every case. During my master thesis, I have worked on exactly this problem by characterising so called virulence factors. These virulence factors can be anything that helps a pathogen to infect its host including for example molecules that can enhance binding to host specific surfaces or enable the pathogen to open up host tissue and cells to move around more easily and access nutrients.

Proteins that can cut other proteins are called proteinases and they are omnipresent in nature. We humans have them for example in our gut for digestion of food with Pepsin being a well-known example. Proteinases have many more functions than simple digestion for nutrients in nature, they also degrade unneeded proteins as well as signal molecules to control signalling.

As you can see proteinases are a perfect example of virulence factors! They are needed by most pathogens and medication against infection often targets them. Therefore, we picked four promising proteinases secreted by *Giardia intestinalis* and had a look what they are doing in the intestine by comparing them to known human and bacterial proteins. Two of those proteinases are suspected to be involved in supressing our immune response, while the other two degrade nutrients before we can take them up from our guts and make them available to the parasite but not to us. We used a yeast to produce a pure version of these proteases and then tested them on the predicted targets. In this way, we could make sure it really is those specific proteins causing the exact effects we were observing. Further experiments are needed to get a full picture since time limitations did not allow to finish up this verification.

However, while we work on that you should keep yourself and others save by on one hand not drinking water you are not sure is safe to drink and on the other hand not polluting the water with faeces yourself, whether you experience diarrhoea or not.
1.0 Introduction

*Giardia intestinalis* (synonyms include *Giardia lamblia* and *Giardia duodenalis*) is an intestinal parasite from the order of Diplomonadida. *G. intestinalis* is distributed worldwide but the highest load of morbidity and mortality for giardiasis is reported in developing countries due to inability to purchase medication or predisposure by malnutrition. It is the most common parasitic gastrointestinal infection in children and one important causative agent of traveller’s diarrhoea.\(^1\) With an estimated 280 million symptomatic cases per year and being recognised as causative agent in 37% of outbreaks proven to be caused by waterborne parasites it remains a big burden. This burden is with high likelihood largely underestimated as mostly developed countries have the resources to put effective surveillance in place yet most infections are thought to take place in developing countries and huge proportion of infections all around the world is never reported due to the self-limiting nature of giardiasis. It is estimated that up to one third of the population in developing countries had had *G. intestinalis* at some point during their lives.\(^2,3\)

Infection is mostly caused by ingestion of water, either as drinking water or while using it for recreational purposes. The latter is the most important reason for infection in developed countries. Other sources of infection include food like green leafy plants or mussels.\(^1\) Giardiasis can be zoonotic; however, most outbreaks could be traced back to human faeces contaminating the water or food in question.

The *G. intestinalis* life cycle is direct and environmentally resistant cysts are shed by infected individuals a few weeks after initial infection. Cysts can survive direct sunlight, different osmotic pressures and temperatures but are metabolically downregulated and can e.g. not divide. They are the infectious stage and after cysts are ingested by a suitable host they excyst into trophozoits in the upper intestine (hence the name *G. duodenalis*) and attach to the gut walls without invading. Here they move around via flagella, grow and divide. It is also here that disease can be induced, yet due to it being a multifactorial disease it is hard to predict in which cases *G. intestinalis* causes disease.\(^4\)

Symptoms vary from asymptomatic to severe abdominal cramping, diarrhoea with malabsorption and weight loss and in children retarded growth and mental development. On the other side, *Giardia* infection can also protect from diarrhoea.\(^1,5\) This is most likely caused by *G. intestinalis* releasing anti-inflammatory factors in contrast to many other gut infectants.\(^4,6\) Infection is often cleared spontaneously but can be long-lasting or permanent as well. Post infection symptoms can include chronic fatigue, food intolerance or irritable bowel syndrome (IBS).\(^7–9\) The reasons why such a wide range of symptoms can be caused is not fully understood as many virulence factors remain uncharacterized and several factors like gut microflora, co-infections and nutritional status seem to play an important role.\(^1\) The diarrhoeal part can be explained by *G. intestinalis* opening up the tight junctions and inducing apoptosis in gut wall cells to be able to access the needed nutrients from those cells. Furthermore the gut microflora is changed during giardiasis and microvilli and crypt structure can be changed after attachment.\(^4,10\)

Treatment of giardiasis is usually done with nitroimidazoles. Other drugs in usage include benzimidazoles, acridines and quinolones. Failure of treatment with nitroimidazoles has been reported to be as high as 20% in some cases.\(^11\) Some of these drugs are seldom used because of strong side effects. This is caused by giardial cells being highly similar to human cells in many aspects as they are eukaryotic as well. Both issues can easily be addressed by investigating giardial proteins and pathways in more detail to understand both resistance mechanisms and potential targets for new classes of anti-giardial drugs in a better way.\(^12\)
Recently, a lot of research has been performed in order to understand the microbial and molecular aspects of giardiasis in more detail. *G. intestinalis* has several characteristic features starting with, as part of the Diplomonadida, having two nuclei with four copies of the genome in total. Due to their intestinal lifestyle in a microaerobic environment giardial cells do not have mitochondria but instead remnants of those in the form of mitosomes, small double enveloped compartments which have lost most features of mitochondria like independent genetic material and the ability for aerobic energy generation but still harbour iron sulphur cluster synthesis. Furthermore *G. intestinalis* lacks a typical Golgi-apparatus and peroxisomes. As trophozoite *G. intestinalis* has four pairs of flagella for movement in the intestine and a ventral disk that is used for attachment to the gut wall. Another interesting feature of this organism is that both genomes and proteomes are reduced to the absolute minimum in the sense that most pathways are simplified compared to other eukaryotic cells and introns are largely non-existent. Some pathways, namely *de novo* biosynthesis of pyrimidines and purines are lost completely; those molecules need to be acquired from the host. On the other hand, some pathways are imported via horizontal gene transfer from bacteria. All this makes *G. intestinalis* an interesting model organism and opens up targets for new drugs.\(^2\)

A classic target for drugs are virulence factors. Those include in *G. intestinalis* well-known proteins like variable surface proteins (VSPs) or the ventral disc and flagella system. VSPs are proteins used for immune system avoidance. They are exposed at the surface one at a time out of a pool of roughly 200 different genes. VSPs are changed via several mechanisms in an unknown pattern in some of the cells and guarantee that at least some are not picked up by a certain antibody targeting the VSP expressed in most cells previously thereby hindering complete clearance.\(^2,13\) Ventral disc and flagella work together to prevent *G. intestinalis* from getting carried away from the small intestine to parts of the gut were conditions are less favourable. However, since *G. intestinalis* is a non-invasive parasite there needs to be a plethora of other virulence factors to e.g. salvage nutrients from both food and host tissue, prevent immune responses, opening tight junctions and induce apoptosis. Indeed, there are 196 proteins known to be secreted into the gut lumen by *G. intestinalis*. (Ma’ayeh, S, unpublished)

Most of those secreted proteins have not been characterised yet but many were functionally predicted from their sequences similarities to known proteins. Analysing their function and differences will be a good start to improve understanding of the disease and enable different treatment options.

One big class of secreted proteins in *G. intestinalis* are proteases. (Ma’ayeh, S, unpublished) This is in good agreement with expectations since proteases can fulfil several of above mentioned roles for the parasite. Other parasites use secreted proteases to cut down potential nutrients in digestible pieces, cut immune relevant proteins, grant access to gut wall cells and deeper tissues by disrupting the extracellular matrix and cell-cell connections (for example tight junctions) and open up host cells.\(^14\) For a list of secreted proteases and their putative functions see table 1.
Proteases are traditionally divided into classes by their catalytic residue. *G. intestinalis* excretes proteases from three of those classes, namely cysteine, serine and metalloproteases using a cysteine residue, a serine residue or a metal cofactor to coordinate the proteolysis. Here only serine and metalloproteases will be covered; cysteine proteases in *G. intestinalis* are the topic of another thesis by Jingyi Liu.

Serine proteases can be found in great abundance in different organisms and have at least three different possible catalytic motifs and according functions. Those can be found in at least four different structural contexts indicating parallel evolution. In humans they are important in such diverse roles as immune response modulation, blood coagulation, reproduction and, with their most prominent member chymotrypsin, in digestion.

The two most abundantly secreted serine proteases in *G. intestinalis* were predicted to both be alanyl dipeptidyl peptidases by sequence analysis. (Ma’ayeh, S, unpublished) Alanyl dipeptidyl peptidases are mostly known for their unusual and rare substrate specificity cleaving N-terminal X-Pro or X-Ala motifs. In the context of the intestine this is on one hand interesting for close to final digestion steps of peptides in general and proline containing molecules like collagen which can only be done by few enzymes due to the ring structure of proline and the hereby implied restrictions on backbone flexibility of the peptide in question. However, even more interesting is that many chemokines, cytokines and other signal molecules have a conserved X-Pro motif at the N-terminal to resist degradation by other cellular proteases, making it a potential way of tightly controlling these signal molecules to possess such alanyl dipeptidyl peptidases. If those two protease are predicted correctly they would thereby give *Giardia intestinalis* a way to interfere with its host’s cell to cell communication and suppressing for example immune reactions in this way.

Metalloproteases are different to the other classes of proteases in that they need a metal cofactor, most often zinc, which is coordinated by the enzymes structure to then catalyse the reaction in question. Therefore, they can be readily inhibited by chelators. Metalloproteases can be divided again in more than 50 subclasses with extremely diverse functions including degradation of extracellular matrix proteins and signal molecules, modulation of transmembrane molecules or digestion of peptides for acquiring nutrients.

<table>
<thead>
<tr>
<th>Protein activity</th>
<th>Protein name</th>
<th>Open Reading Frame(ORF) in WB</th>
<th>Protein class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystein-type peptidase activity</td>
<td>Cathepsin B</td>
<td>GL50803_14019</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>GL50803_16779</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>GL50803_15564</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>GL50803_16468</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>GL50803_16160</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin L</td>
<td>GL50803_14983</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>GL50803_10217</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin L</td>
<td>GL50803_16380</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>GL50803_17516</td>
<td>Cysteine protease</td>
</tr>
<tr>
<td></td>
<td>Alanyl dipeptidyl peptidase</td>
<td>GL50803_15574</td>
<td>Serine protease</td>
</tr>
<tr>
<td></td>
<td>Alanyl dipeptidyl peptidase</td>
<td>GL50803_6148</td>
<td>Serine protease</td>
</tr>
<tr>
<td></td>
<td>Aminoacyl-histidine dipeptidase</td>
<td>GL50803_15832</td>
<td>Metalloprotease, deacetylase, hydrolase</td>
</tr>
<tr>
<td></td>
<td>Xaa-Pro dipeptidase</td>
<td>GL50803_17327</td>
<td>Transcription factor, metalloprotease</td>
</tr>
</tbody>
</table>

*Table 1:* Overview on secreted proteases, their ORF in the WB isolate of *Giardia intestinalis*, their classes and putative functions predicted by sequence analysis, modified after Ma’ayeh, S, unpublished
The two most abundantly secreted proteases fall into the last category according to sequence analysis. (Ma’ayeh, S, unpublished) One of them is predicted to be an aminoacyl histidine dipeptidase, cutting as said by its name dipeptides with an X-His motif, the other one is predicted to be closely related to human prolidase, the only known enzyme that catalyses degradation of X-Pro dipeptides.\textsuperscript{17} Notably both human equivalent enzymes are associated with severe symptoms if mutated, showing the high importance of correct dipeptide degradation.\textsuperscript{20,21}

For \textit{G. intestinalis} however, their products, single amino acids ready to be used again, are most likely more important. In the intestine, proteins are broken down by a cascade of proteases from big structured units via oligopeptides to di- and tripeptides and even single amino acids. Involved here are for example Pepsin, Trypsin and Chymotrypsin for the first steps. In humans, the uptake from the gut lumen into cells happens for several amino acids including proline with substantially greater efficiency at the Tri- and Dipeptide level. The final digestion steps by Tripeptidases, Dipeptidases and Prolidase take place intracellularly.\textsuperscript{22,23} The parasite needs to outcompete both host and other gut flora for nutrients, secretion of proteases to cleave those small peptides in its direct vicinity before they can be taken up by the host will assist in doing so.

Together those four enzymes enable \textit{Giardia intestinalis} to efficiently process the final steps of protein degradation of the stable proline motifs that can be found in many intestinal peptides including the omnipresent collagen. The malfunctions associated with mutation of the human equivalent proteins and the fact that large amounts of nutrients are needed to sustain \textit{G. intestinalis} at high levels render those proteins potential drug targets and classify them as virulence factors.

1.1 Aim

- Understand the pathogen-host interaction between \textit{Giardia intestinalis} and humans better by characterising important virulence factors
- Select potential virulence factors -> four secreted proteases
- Clone all four into pPICZαA vector and introduce into \textit{Pichia pastoris} X33 for expression
- Purify the protein in question in sufficient amounts for both biochemical and direct interaction studies
  - Test activity in vitro against suspected substrates
  - Phage display if activity is seen to find potential targets in the human proteome
  - Understand impact of those proteases in the intestine via testing with human cell lines
  - Knock out protease genes to see impact on \textit{Giardia intestinalis}
2.0 Material and methods

2.1 Material

2.11 Media

Low Salt LB Medium: 1% Tryptone, 0.5% Yeast Extract, 0.5% NaCl  
pH=7.5, in sterile H2O

Low Salt LB Agar: 1% Tryptone, 0.5% Yeast Extract, 0.5% NaCl, 15g/l agar  
pH=7.5, in sterile H2O

YPD: 1% Yeast extract, 2% peptone, 2% D-glucose in sterile H2O

YPDS Agar: 1% Yeast extract, 2% peptone, 2% D-glucose in sterile H2O  
2% Agar

BMGY: 1% Yeast extract, 2% peptone, 1.34% YNB, 4x10^{-5} biotin  
100mM K2HPO4/KH2PO4 pH=6.0, 1% glycerol

BMMY: 1% Yeast extract, 2% peptone, 1.34% YNB, 4x10^{-5} biotin  
100mM K2HPO4/KH2PO4 pH=6.0, 0.5% methanol

TBST: TBS with 0.1% Tween

2.12 Buffer

Wash buffer 20mM Imidazole, 500mM NaCl, 20mM NaH2PO4/ Na2HPO4, pH=7.4

Elution buffer 500mM Imidazole, 500mM NaCl, 20mM NaH2PO4/ Na2HPO4, pH=7.4

Exchange buffer 500mM NaCl, 20mM NaH2PO4/ Na2HPO4, pH= 7.4

Phage display washing buffer 1M NaCl, 0.1% Tween in PBS

2.13 Staining solutions

Coomassie staining solution 11.7ml H3PO4, 968.3ml H2O (100g (NH4)2SO4),  
Plus 20ml Serva Blue-G (5% stock)
2.2 Methods

2.21 Bioinformatical comparison of all target proteins to related and characterised proteinases

Alignment of protein sequences to related sequences and establishment of resulting distance trees of results was done by BLAST. Prediction of protein domain modules was done using the BLAST conserved domains tool. Alignment of several sequences was done using ClustalΩ. Structure modelling for Gl50803_17327 was performed using I-TASSER software. Signal peptides were predicted with the SignalP 4.1 Server.

2.22 Cloning

2.22.1 Preparation

All genes were cloned into the plasmid pPICZαA which has been constructed by Invitrogen. It contains an Escherichia coli origin of replication(ORI), several restriction sides that can be used for linearization and a Zeocin resistance gene for selection, which is controlled by three promoters so it can be expressed in both Escherichia coli (EM7 promoter) as well as Pichia pastoris (AOX1 and TEF1 promoters). It does not contain a yeast ORI in order to prevent plasmid amplification in yeast and make sure that it is integrated. pPICZαA also contains a cloning region consisting of a promoter which can be induced by methanol (AOX), an α-factor, which leads to secretion by Pichia pastoris, 10 different restriction enzyme target sequences that are not identical to the ones used for linearization, a c-myc epitope which can be used for Western blots and a polyhistidine tag that can be used for Western blots as well as protein purification using metal-chelating resins.

Primers for cloning were designed using DNA sequences collected on www.giardiadb.org. The sequences were screened for giardial secretion signals which were then removed to make sure that they do not interfere with the yeast secretion signals. In one case (GL50803_15832) no clear decision could be made if it has a signal peptide or not. Two constructs were therefore created. The remaining genetic information was extended by one restriction site at the very end and two bases as a small cap on the 5´-end. On the 3´-end the stop codon was removed, a HIS tag consisting of 6 histidines was added followed by a new stop codon. Another restriction site, which is downstream of the earlier chosen site and a small cap were added as well. The used primers were also designed to have similar annealing temperatures so PCR was possible.
All genes were taken from *Giardia intestinalis* isolate WB, bearing number GL50803_x.

PF = Primer forward, PR = Primer reverse, added restriction enzymes sites were marked in bold writing

<table>
<thead>
<tr>
<th>Gene/Primer ID</th>
<th>Primer</th>
<th>Restriction enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>6148/PF</td>
<td>5’-GGGGTACCGTCTCTAAACCGCCAGGAGGACAATG-3’</td>
<td>Knpl</td>
</tr>
<tr>
<td>6148/P599</td>
<td>5’-GGTCACTGAGTGATGACAGCA-3’</td>
<td>-</td>
</tr>
<tr>
<td>6148/P1250</td>
<td>5’-GCCACAACAAGATTITCTTGTGG-3’</td>
<td>-</td>
</tr>
<tr>
<td>6148/P1883</td>
<td>5’-ACGATGGATGTCCTGACCAC-3’</td>
<td>-</td>
</tr>
<tr>
<td>6148/PR</td>
<td>5’-GGTCTGAGCTAATGATGAGTATGAGTACGAGTG-3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>15574/PF</td>
<td>5’-CCGAATTCGACGTCTACGAGAACATTAAGC-3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>15574/P602</td>
<td>5’-ACGGAGATGCTCCGTTAAG-3’</td>
<td>-</td>
</tr>
<tr>
<td>15574/P1273</td>
<td>5’-GCAGATATAATGGGACCTG-3’</td>
<td>-</td>
</tr>
<tr>
<td>15574/PR</td>
<td>GGTCGTAGATCAAATGATGATGATGATGAGTACGAGTGCTCCTTGAGATATTATCTAAGG-3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>15832/PF1</td>
<td>5’-CCGAATTCATGTCCAATAAAATACAAT-3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>15832/PF2</td>
<td>5’-CCGAATTCGTCCTAAGGATGAGAAGCTCAG-3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>15832/P601</td>
<td>5’-GGTCTGAGAGGAGTTCAATG-3’</td>
<td>-</td>
</tr>
<tr>
<td>15832/P1126</td>
<td>5’-AGCTCGATGAATCTGGGAAG-3’</td>
<td>-</td>
</tr>
<tr>
<td>15832/PR</td>
<td>CCGCGGCGCGCTCAATGATGATGATGATGAGTACGAGTGCTCCTTGAGATATTATCTAAGG-3’</td>
<td>NotI</td>
</tr>
<tr>
<td>17327/PF</td>
<td>5’-GGGAATTCATGTCTTTTCTGAGGACACCGCA-3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>17327/P655</td>
<td>5’-GACAAGTCTACAAGCTAACC-3’</td>
<td>-</td>
</tr>
<tr>
<td>17327/PR</td>
<td>GGTCGTAGATCAAATGATGATGATGATGAGTACGAGTGCTCCTTGAG-3’</td>
<td>XbaI</td>
</tr>
</tbody>
</table>

### 2.222 PCR amplification of protease genes

The genes used for cloning were amplified using genomic DNA from *Giardia intestinalis* isolate WB as a template. The PCR was in a ThermoFisher Scientific PCR cycler, using Phusion Hot Start DNA Polymerase and connected products from ThermoFisher Scientific. The forward and reverse primers listed above were used in each reaction.

Initial and DNA denaturation were done at 98°C for 30s and 10s respectively, extension was done at 72°C for 80s. 40 cycles were done before final extension at 72°C for 5min.

All steps involving DNA were controlled using 0.8% agarose gels with SYBR Safe.

### 2.223 Digestion and Ligation

PCR amplified genes and empty pPICZαA were purified using kits from ThermoFisher Scientific for PCR products and the QIAGEN HiSpeedPlasmidMaxi for plasmids. Both were then digested using suitable restriction enzyme combinations. Enzymes used were FastDigest
Enzymes purchased from ThermoFisher Scientific. Digestion conditions were chosen according to restriction enzyme manufacturer’s instructions.

PCR products were then purified again using the same kit, plasmid digestion was controlled by running it on an agarose gel and purifying from there using the GeneJET Gel Extraction kit from ThermoFisher Scientific.

Ligation was done with a T4DNA Ligase from ThermoFisher Scientific, using about 80ng digested plasmid and 120ng PCR product and choosing conditions given by the company.

2.224 Transformation Escherichia coli

Competent E. coli TOP10 were aliquoted to 45µl each and mixed with 5µl ligation mixture. They were incubated on ice for 30min, heat shocked for 45s @ 42°C and then incubated on ice again for 2min. All 50µl were plated on low salt LB agar plates containing 50µg/ml Zeocin and incubated o/n @ 37°C in the dark.

2.225 Confirmation of correct gene insertion into the plasmid

To confirm single cell colonies on the plate contained plasmid with inserted gene a colony PCR was performed on colonies picked from those plates. The colonies were propagated by touching a new plate of on low salt LB agar plates containing 50µg/ml Zeocin and then mixed with ready-to-use PCR mix; all products were purchased from ThermoFisher Scientific. Primers used were forward and reverse primers for AOX, since it is flanking the cloning side of pPICZαA.

Initial and DNA denaturation were done at 95°C for 5min and 30s respectively, annealing at 56°C for 90s and extension was done at 72°C for 90s as well. 40 cycles were done before final extension at 72°C for 5min.

Single cell colonies containing the plasmid with insert were picked and used to inoculate low salt LB with 50µg/ml Zeocin to prepare bacterial stocks and purify plasmid with the GeneJET plasmid Miniprep Kit from ThermoFisher Scientific for sequencing. The purified plasmids were prepared for sequencing per instructions of the Mix2Seq Kit from Eurofins Genomics. For sequencing, additional primers had been designed (s.a.), so the whole gene could be sequenced. As most outward points AOX 5’ and AOX3’ primers were again included as well.

The sequencing data was analysed using ClustalΩ. Colonies with the correctly inserted gene inside pPICZαA were picked for another purification to create sufficient amounts of plasmid for linearization.

2.226 Linearization

Plasmids were purified using the HiSpeed Plasmid Midi Kit from Qiagen. They were then linearized using restriction enzyme BstX I (purchased from ThermoFisher Scientific) for GL50803_15574, GL50803_6148 and GL50803_15832 PF1 and PF2 and Pme I (purchased from NEB) for 50803_17327. Conditions were chosen per producer’s instructions. Linearization was confirmed by agarose gel electrophoresis on 0.8% agarose gels with SYBR Safe.
2.227 Purification for transfection

The linearized plasmids were purified using Phenol/Chloroform purification and resolved in H₂O. For this 500µl of linearization mixture were mixed with 500µl phenol/chloroform/isoamyl alcohol (25:24:1) (pH=7.38) then centrifuged at 13000g for 5min. The upper phase was transferred to a new tube and mixed with 500µl chiasm (24:1) and centrifuged again for 5min at 13000g. 1ml 100% ethanol, 50µl 3M NaAC (pH=5.2) and 5µl glycogen were added and all incubated at -80°C for 45min after mixing. Tubes were centrifuged at 21000g for 30min, the pellet was washed with 1ml 75% ethanol and centrifuged for 15min at 21000g again. It was then dissolved in 15µl H₂O

2.228 Preparation of competent yeast cells

_Pichia pastoris_ X-33 cells were made competent following the instructions in Invitrogen’s EasySelect _Pichia_ Expression Kit. (http://tools.thermofisher.com/content/sfs/manuals/easyselect_man.pdf)

2.229 Transformation by electroporation

Transformation was done via electroporation with minor changes as described in the EasySelect _Pichia_ Expression Kit purchased from Invitrogen. (http://tools.thermofisher.com/content/sfs/manuals/easyselect_man.pdf)

The amount of cells and linearised DNA used was different; in contrast to the original protocol 40µl cells and 3µg DNA were used. A control with PBS was transformed as well. Second instead of spreading different amounts of yeast cells on 100µg/ml Zeocin plates, identical amounts of yeast cells were spread on plates containing 100µg/ml, 200µg/ml and 500µg/ml Zeocin. Subsequent purification was done as described.

2.23 Expression and Purification

2.231 Test-expression

Test-expression was done using the protocol given by Invitrogen (http://tools.thermofisher.com/content/sfs/manuals/easyselect_man.pdf) changing the following details. As many clones had to be tested the volume was reduced to 3ml of BMGY for primary inoculation with one of them each, centrifuging the culture and replacing BMGY with 3ml of BMMY after one day of incubation. No different time points were checked, all expressions were stopped after roughly 72 hours of induction and analysed via SDS-PAGE and western blotting.

2.232 Test-purification

Colonies that showed good expression during the test-expression were picked for a test purification. This test purification was done using 30 µl Ni-NTA beads to bind the protein in 1ml supernatant from the test-expressions. This was incubated for 45min mixing at room temperature before washing off unbound protein with 4.5ml washing buffer and eluting with 30µl elution buffer. Analysis was done on Western blots.
2.233 SDS-PAGE/Western Blot

Protein samples were analysed using either SDS-PAGE gel electrophoresis or Western blotting, depending on the estimated concentration of protein in the sample in question.

SDS-PAGE was run using Mini-PROTEAN® TGX Stain-Free™ precast gels for all kD from BIORAD if not mentioned differently and running them at 100V for approximately 90min or until the loading dye migrate to the bottom of the gel. Gels were activated and documented using a BIORAD chamber.

For Western blotting similar gels were used. They were transferred on the Trans-Blot® Turbo™ Transfer System from BIORAD using MIDI PVDF transfer packs and the mixed MW(Turbo) setting of the device. Membranes were blocked in TBST with 5% BSA for >1h at RT. They were then incubated in primary antibody (1:6000 mouse anti-Histidine in TBST with 5% BSA o/n @4°C. Secondary antibody was HRP conjugated anti-mouse (1:5000) for 1h at RT. Detection was done using Clarity Western ECL substrate from BIORAD. Images were taken in a BIORAD chamber using the ChemiHiSensitivity setting for blots in the ImageLab software from BIORAD. Washing steps in between and after antibodies were done using TBST.

2.234 Expression and Purification

Expression was done using the same protocol used for the test expression scaling up both initial incubation volume to 200ml BMGY and incubation time to ca. 72h. The harvested cells were resuspended to a final OD$_{600}$= 1-1.5 in 1l BMMY and then kept and induced and harvested according to the protocol for a total induction time of 72h.

Purification begun by filtering the supernatant (pore size= 0.2µm) then binding all of it to 2ml/l Ni-NTA beads. Those beads were collected and transferred to a 10ml column made of filter paper in a 10ml syringe and washed with 4 CV (column volumes) washing buffer before eluting twice with 0.2 CV elution buffer. They were subsequently concentrated on a Pierce™ Protein Concentrator PES, 10K MWCO, 5-20 ml column to about 300µl and then the buffer was changed by loading one CV(=20ml) exchange buffer two times in succession and concentrating again to 200-300µl.

2.24 Protein detection and analysis

2.241 Protein detection

Protein was detected using the Thermo Scientific™ Pierce Coomassie Plus Protein Assay for general and quick assessment if protein had been purified. For this purified protein solution was mixed with CoomassiePlusProtein solution. If the solution turned blue, it was assumed protein had been purified. This was verified by both SDS-PAGE and a Qubit Fluorometric Quantitation System from ThermoFisher Scientific. The Qubit system was also used to quantify the protein amount by measuring three standards purchased from ThermoFisher Scientific and thereby establishing a standard curve before measuring the samples.
2.242 Zymography

It was decided to test protein activity using zymogram gels containing gelatine, since this substrate showed high likelihood of activity for GL50803_17327. The gels were cast using standard protocols. To the resolving gel gelatine or casein was added to a final concentration of 0.5%, collagen to a final concentration of 0.05%.

Both Trypsin as a positive control and the metalloprotease in question were mixed 1:5 with Pierce™ Lane Marker Non-Reducing Sample Buffer from ThermoFisher Scientific. Four times more protease than Trypsin was loaded. Before incubating the gel for digestion SDS was washed off with 2.5% Triton X-100. The gel was then incubated @37°C in developing buffer from ThermoFisher for 16-18h. The gel was stained with Coomassie Blue; pictures were taken in a BIORAD chamber using the Coomassie Fluor Orange setting.

2.243 Phage display

Phage display was done using a T7 phage library developed by Lars Hellman’s group at Uppsala University. This library contains approximately 5x10^7 variants and equals 10^9 plaque forming units (pfu). The phages were bound to Ni-NTA beads using histidine tags that had been connected to one surface protein via potential target sequences by incubating them together with the beads rotating at 4°C for one hour. Before adding the protease in question, the beads were washed extensively for 10 times with phage display washing buffer to remove unbound phages. They were then incubated with 4-16μg of the protease in PBS at 37°C for 2h rotating.

A dilution series of these phages was added to IPTG induced BLT5615 culture with an OD_600=0.5 and plated on LA plates with 50μg/ml ampicillin to see the number of phages that had been cleaved off the Ni-NTA beads. Some of the phages were propagated by adding them to IPTG induced BLT5615 culture and growing them until lysis, when phages were harvested, mixed with PBS and NaCl (C=0.5M) and stored at 4°C for up to two days. This was the sub-phage library for the next round of selection. BLT5615 is an *Escherichia coli* strain that has the gene for the unchanged phage surface protein controlled by a lacUV5 promoter and can be induced to express this protein. The used phage does not code for this native protein, only for an alternative coat protein that contains the target sequence; however, such phages are often growth inhibited hence a mixture of surface proteins is aimed for.

As a control, phages still bound to the Ni-NTA beads were eluted with imidazole and plated in the same way as the selected phages after diluting appropriately.

This was repeated up to four times and the differences between protease cleaved phage numbers and a PBS control were monitored. Differences higher than 10-fold on day 4 would have prompted to continues the experiment until up to seven days. Washing steps were increased after the first round of selection to give respect to the increased numbers of phages, dilution series were adjusted due to the same reason.

2.244 Testing possible substrates

Protein substrates were mixed in a ratio between 5:1 and 10:1 with protease; topped up to 12μl with PBS and then incubated for two hours at 37°C. Afterwards they were analysed using SDS-PAGE on Mini-PROTEAN® TGX Stain-Free™ precast gels. As a control substrate with PBS was incubated under identical conditions to make sure seen degradation is due to the protease used. Pictures were taking on the stain free setting for protein gels in a BIORAD chamber.
3.0 Results

3.1 Bioinformatical comparison of all target proteins to related and characterised proteases

The practical part of this project included a small bioinformatical analysis of the four proteases picked for characterisation. The first step here was to compare them to other described and characterised proteins. A protein BLAST (Basic Local Alignment Search Tool) showed for GL50803_6148, GL50803_15574 and GL50803_15832 that they are outside of the order of Diplomonadida closest related to bacterial proteins with only few closely related proteins known in eukaryotes (Data: Supplementary figure 1-3).

The two serine proteases that are part of this work are at 40% sequence identity closely related and both predicted to be alanyl-dipeptidyl peptidases. No structure is available for either of them. The human equivalent enzymes for both of those serine proteases are localised intracellularly in the intestine.

GL50803_6148 shows 40% sequence identity to the equivalent giardia assemblage GS protein GL50801_1499. In figure 1 it can be seen that residues 15-761 are predicted to be a dipeptidyl aminopeptidase, which is then specified to be involved in prolyl-degradation by subsets of residues. Residues 1-15 are predicted to be a signal peptide (Data not shown). In accordance with this it has been partly characterised before by Tous et al. in 2002, to be a surface associated Betstatin-blockable dipeptidyl-activity enzyme.\(^\text{30}\) It has also been shown that is important in encystation processes as it is two time upregulated in cysts compared to trophozoites and encystation gets blocked by Betstatin.\(^\text{30,31}\)

![Figure 1: Predicted functions for GL50803_6148 based on its sequence; DAP2= Dipeptidyl aminopeptidase, Peptidase_S9= Prolyl oligopeptidase family, AXE1 superfamily= Acetyl-xylan esterase, PRK05371= x-prolyl-dipeptidyl aminopeptidase](image)
GL50803_15574 has not been characterised before but is due to the high identity predicted to have a similar function. A putative transmembrane domain is predicted at the N-terminal at residues 13-53 (Data not shown). At the same time residues 1-30 are predicted to be a signal peptide (Data not shown). Prediction of function for this enzyme is very close to prediction of function for GL50803_6148, underlining their similarities. The major part of the enzyme after the TM region (94-728) is predicted to be a dipeptidyl aminopeptidase with a subset of residues again predicted to be more specifically a prolyl-peptidase (figure 2).

Figure 2: Predicted functions for GL50803_15574 based on its sequence; DAP2= Dipeptidyl aminopeptidase, Peptidase_S9= Prolyl oligopeptidase family, AXE1 superfamily= Acetyl-xylan esterase

For metalloprotease GL50803_15832 no structure is available either and no closely related protein has been characterised before. By sequence analysis it is predicted to be related to human carnosinase, a protein catalysing digestion of aminoacyl histidine dipeptides and involved in CREB signalling. During encystation expression of this protein is diminished to end at one fourth of the expression level in trophozoits. Exact function predictions can be seen in figure 3. Residues 26-520 are predicted on all levels to be a M20 peptidase D with specificity for beta-alanyl-L-histidine dipeptides. Five metal binding sites are predicted as well as several dimer interface locations yet without solved structure for a closely related protease those predictions are hard to judge for correctness.

Figure 3: Predicted functions for GL50803_15832 based on its sequence; M20_pepD= M20 Peptidase D, specificity for beta-alanyl-L-histidine dipeptide
For the last protease in this work, GL50803_17327, analogues in many different organisms could be found, including many other single and multicellular eukaryotes as well as bacteria. The protein is at 94% sequence identity highly conserved between giardia assemblages WB(GL50803) and GS(GL50801). Expression levels rise during early encystation for this protease but in cysts it is reduced to a level fourteen times lower than seen in trophozoits. The closest related non-diplomonad proteinases are from nematode intestinal parasites (figure 4).

Figure 4: Distance tree based on BLAST pairwise alignments for GL50803_17327 (marked in yellow) showing proteins with the biggest sequential identity and their evolutionary distance to this protease

By sequence function prediction subparts are identified as aminopeptidase N-terminal (residues 3-122) and prolidase (residues 145-425) and Xaa-Pro aminopeptidase (36-435) making this function very likely (figure 5). Six metal coordinating residues are predicted; however, for the solved structure of human prolidase (see below) only five are identified. In further analysis only the residues corresponding to these five residues (human prolidase residues Asp277, Asp288, His371, Glu413, and Glu453) are used. No Signal peptide or transmembrane regions were predicted (Data not shown).

Figure 5: Predicted functions for GL50803_17327 based on its sequence, AMP_N= aminopeptidase N-terminal domain, PepP= Xaa-Pro aminopeptidase
The Xaa-Pro dipeptidase from nematode parasites *Strongyloides ratti* was picked for sequence alignment using ClustalΩ (figure 6) showing good alignment around the five metal coordinating residues and the C-terminus in general. Most residues on this end are conserved or exchanged by strongly similar amino acids. On the N-terminal end the alignment is weaker and stretches of *S. ratti* protein have no match in GL50803_17327 due to the difference of 478 residues in the nematode protein to 444 residues in the giardial one.

*Figure 6*: ClustalΩ alignment of GL50803_17327 and *Strongyloides ratti* Xaa-Pro dipeptidase, colours on residues: Red= small and hydrophobic (incl. aromatic-Y) (AVPFMILW), Blue=acidic(DE), Magenta= basic-H(RK), Green=Hydroxyl, sulfhydryl, amine and G(STYHCNGQ), * (asterisk)= conserved residue, : (colon)=strongly similar residue, . (point)=weakly similar residue, red box= metal coordinating residues.
GL50803_17327 also has a closely related human equivalent enzyme, prolidase, at 39% sequence identity. Prolidase is localised intracellularly in the human intestine and has been structurally and functionally characterised. It is reported to be present in nature as a dimer of about 120kD and can be modelled from the human structure.

Amino acid sequences for GL50803_17327 and prolidase were aligned using ClustalΩ (figure 7). The human enzyme is 494 amino acid residues long compared to the 444 amino acids hence several parts on the N-terminal part of human prolidase do not have a match in GL50803_17327. In general differences are focused on the N-terminus while similarities outweigh them on the C-terminal end. All five metal coordinating residues marked with red boxes are conserved between the two enzymes. In GL50803_17327 those are residues number D233, D244, H327, E369 and E409. Residues to both sides of H327, E369 and E409 are as well conserved or at most substituted to highly similar amino acids. Amino acids around D233 and D244 are still similar between both proteins but differences are bigger here.

Figure 7: ClustalΩ alignment of GL50803_17327 and human Xaa-Pro dipeptidase, colours on residues: Red= small and hydrophobic (incl. aromatic-Y) (AVPFMILW), Blue=acidic(DE), Magenta= basic-(RK), Green=Hydroxyl, sulphydryl, amine and G(STYHCNGQ), * (asterisk)= conserved residue, : (colon)=strongly similar residue, . (point)=weakly similar residue, red box= metal coordinating residues
The distribution of conserved or highly similar residues compared to non-conserved residues becomes more apparent when colouring them differently on the structure of human prolidase. Human prolidase exists as homodimer in vivo, for better visualisation only one molecule (chain A) is pictured in figure 8. Given that the structure is similar not only in theory between the two enzymes red and orange residues representing the conserved or highly similar residues are concentrated around the active centre with the five metal coordinating residues in cyan. Green residues representing exchanged amino acids are more often seen at the outer surface of the molecule and distant from the catalytic centre.

Figure 8: Structure of human prolidase chain A, non-conserved residues in GL50803_17327= green, highly similar residues in GL50803_17327= orange, conserved residues in GL50803_17327= red, metal-coordinating residues= cyan, modified after: Crystal Structure of Human Prolidase: The Molecular Basis of Pd Disease FAU – Mueller et al. 32
Upon examination of the active site of human prolidase (figure 9) using the same colour code it can be seen that most residues in close sequential proximity are conserved or exchanged only to chemically highly similar amino acids. Furthermore, acid residues in close spatial proximity marked by black arrows tend to be conserved as well while those in spatial distance are more likely to be non-conserved in comparison.

*Figure 9*: Structure of human prolidase chain A close to the catalytic centre, non-conserved residues in GL50803_17327= green, highly similar residues in GL50803_17327= orange, conserved residues in GL50803_17327= red, metal-coordinating residues= cyan, 1=D233, 2=D244, 3=H327, 4=E369, 5=E409, black arrows= residues in spatial proximity to the catalytic centre, modified after: Crystal Structure of Human Prolidase: The Molecular Basis of Pd Disease FAU – Mueller et al. 32
Finally, an I-TASSER (Iterative Threading ASSEmbly Refinement) prediction was run with human prolidase structure as a template, showing that most parts can be well aligned with all restrictions from the amino acids that have been exchanged in GL50803_17327 considered. (figure 10).

*Figure 10:* Structure of human prolidase (2IW2) in red and I-TASSER prediction for GL50803_17327 in blue, using 2IW2 as template.
3.2 Characterisation of serine protease GL50803_6148 and GL50803_15574

3.21 Cloning of proteases GL50803_6148 and GL50803_15574

For both serine proteases secretion signals were predicted from the sequence. To prevent interaction with *Pichia pastoris* secretion system they were excluded when constructing primers for cloning of genes into the cloning site of pPICZαA, the plasmid chosen as vector. A HIS-tag was added right before the stop codon to enable purification and Western blotting. Successful cloning was ensured by sequencing the constructed plasmids before transfection into *Pichia pastoris*. Good quality was ensured via overlapping sequences so at least two independent sequencing results could be compared to each other and the original gene. All sequences were aligned with ClustalΩ.

3.22 Test-expressions and test-purifications for GL50803_6148 and GL50803_15574

50 clones growing on agar plates containing Zeocin two times in succession after transfection and thereby showing integration of the plasmid were tested in small scale test-expression for each of the two serine proteases. Clones showing good expression levels of protein in their supernatants on Western blots targeting the HIS-tags were picked for small scale test-purifications (Data not shown). None of these clones was confirmed to express either one of the serine proteases and the work on them was stopped due to time limitations (Compare figure 11 showing no HIS-tagged protein for clone 3 from the 500µg/ml YPDS agar expressing GL50803_6148 in the supernatant, purified supernatant or on the Nickel beads).

3.3 Characterisation of metalloprotease GL50803_15832

3.31 Cloning of protease GL50803_15832

For metalloproteinase GL50803_15832 no clear prediction could be done concerning the existence of a secretion signal sequence. Two primer sets were constructed to meet both possibilities, one cutting after the potential signal sequence the other covering the entire gene. Both constructs contain a HIS-tag and were cloned into the cloning site of pPICZαA in a similar way as the serine protease sequences. The correct insertion of the genes was confirmed by ClustalΩ alignment analysis and overlapping sequencing results ensured good quality over the entire gene length.

3.32 Test-expressions for GL50803_15832

Supernatants of 50 clones growing on agar plates containing Zeocin two times in succession after transfection were screened by Western blotting targeting the HIS-tags for both constructs. No clones with high expression and secretion of the target protein could be found (Data not shown). Work on them was stopped due to time limitations.
3.4 Characterisation of metalloprotease GL50803_17327

3.4.1 Cloning of protease GL50803_17327

The last protease that was included into this thesis project work was GL50803_17327, a prolidase-like metalloprotease. No secretion signal could be identified by sequence analysis hence the entire gene was cloned into the cloning site of pPICZαA with an HIS-tag added at the C-terminus. The correct insertion into the plasmid was confirmed by ClustalΩ alignment analysis with overlapping sequencing results to ensure good quality of sequencing at every point of the gene.

3.4.2 Test-expressions for GL50803_17327

50 clones were picked from an YPDS agar plate containing Zeocin as selection marker and purified again on a fresh YPDS agar plate. They were then tested on expression of the proteinase after induction with methanol by analysing the supernatant by both SDS-PAGE and Western blotting. Clones showing good expression were picked for a small-scale test-purification using Nickel beads based affinity chromatography.

![Western blot using antibodies against HIS-tag.](image)

*Figure 11:* Western blot using antibodies against HIS-tag. L= PageRuler Prestained Protein Ladder from ThermoFisher Scientific, E=supernatant from empty plasmid as negative control, P= HIS-tagged protein as positive control, SN= supernatant of the given clone, Pure= purified supernatant of given clone, NiCo= Nickel beads used for affinity chromatography after elution for given clone

Figure 11 is showing one of the western blots run during the screening for a clone with suitable expression and secretion levels. The correct function of the antibodies is shown by inclusion of
a positive control. Both supernatant and purified protease from clone 3 from the 200µg/ml YPDS agar expressing GL50803_17327 are visible at roughly 55kD, indicating that the protein is present as a monomer. No or very little protein remained bound to the Nickel bead column after elution. Some of the positive control spilled into the empty plasmids lane, where no protein of other size can be visualised by western blotting.

It was decided to continue with the clone 3 for GL50803_17327 and use it to scale up protein expression and purification.

3.43 Expression and purification of GL50803_17327

Supernatant from the chosen clone was harvested and purified via filtration and affinity chromatography. The concentrated protein solution was then run on a SDS-PAGE(Figure 12) together with 2µl each of increasing concentrations of BSA as a comparison to be able to estimate the protein amount in solution. It was estimated to be roughly 3µg that had been loaded in 5µl, giving a concentration of circa 0.6µg/µl.

![Figure 12](image)

A protein of the same size as has been previously seen in western blots could be purified in high amounts and at a good purity. The protein is visible in one strong band with only very weak accompanying bands indicating that it is stable and does not show degradation during the purification procedure which took place at room temperature.
3.44 Protein detection

With a Qubit Fluorometric Quantitation System, the protein amount purified was evaluated to be 0.4µg/µl giving a total protein amount of 120µg/l culture. This agrees with the earlier SDS-PAGE based estimation.

3.45 Zymography

The next experiment needed after successful purification of sufficient amounts of protein was to test for activity using a general substrate. Zymography using gels containing either gelatine, collagen or casein was chosen. The protein was incubated at physiological conditions in the gel, Trypsin was included as a positive control.

![Zymogram using casein as substrate](image)

*Figure 13: Zymogram using casein as substrate, L= PageRuler Prestained Protein Ladder from ThermoFisher Scientific, T= 1.5µg Trypsin, P= 2µg GL50803_17327*

In the zymogram using casein as substrate (Figure 13) it can be seen that the zymography worked. The ladder is well separated and Trypsin as positive control degraded casein completely leaving blank spaces in the otherwise completely stained gel. The giardial protease GL50803_17327 was separated by the gel, though it does not form as clear a band as can be seen when run on a SDS-PAGE. No degradation of casein by this protease can be observed. Due to long run time of nearly 2h at 100V time to be able to separate the higher molecular sizes the two smallest ladder sizes are not on the gel anymore. The long time was needed due to additional resistance to protein migration posed by the substrate.
Figure 14: Exemplary zymogram using gelatine as substrate, L= PageRuler Prestained Protein Ladder from ThermoFisher Scientific, T=0.5µg Trypsin, P=2µg GL50803_17327

In figure 14, a zymogram using gelatine as substrate, it can be seen that the zymography worked with gelatine as well. The ladder and protein are well separated, the giardial protease is a bit more concentrated around the size of 50kD compared to the zymograms run with casein as substrate. Again, the two smallest ladder sizes are not on the gel anymore since the zymogram had to be run for almost two hours to enable good separation of higher sized molecules. Trypsin shows again strong degradation of the substrate, proving that washing off SDS was done successfully leaving the protein in an active state. For gelatine as a substrate some activity for GL50803_17327 can be observed as well, though the activity seems to be shifted towards a higher size range ranging from 55kD to approximately 100kD. Several gelatine-based zymograms were run to verify this result, all showing the same pattern.
3.4.6 Phage display

After having shown that the protein is active it was decided to try and find out about the substrate specificity for this protease via phage display. In this variant of phage display the protease selectively cuts of those phages with matching sequence from a Nickel bead matrix they have been bound to beforehand. By this matching phage reproduction gets enhanced over several rounds of selection to ratios of more than 1000 times the control run with PBS. The selected phages are sequenced and the received sequence is compared to potential targets in the human genome or food.

In total two attempts of phage display were started, the first was run for three, the second for four days. Only data from the first round is shown, since the data for the second did not show a different picture. In the first attempt, constantly 4µg of protease were used for digestion of target sequences.

<table>
<thead>
<tr>
<th>Dilution factor</th>
<th>10⁻³</th>
<th>10⁻⁴</th>
<th>10⁻⁵</th>
<th>10⁻⁶</th>
<th>10⁻⁷ (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1 PBS</td>
<td>369</td>
<td>47</td>
<td>5</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>17327</td>
<td>605</td>
<td>69</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 2 PBS</td>
<td>49</td>
<td>9</td>
<td>1</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>17327</td>
<td>13</td>
<td>77</td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Day 3 PBS</td>
<td>729</td>
<td>98</td>
<td>7</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>17327</td>
<td>570</td>
<td>9</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Plaque numbers in first phage display after x days at different dilution levels, PBS=control, 17327=GL50803_17327, E= elution as control to see how many viruses have been in total binding to the Nickel beads, Dilution factors= phages that have been cut off by protease or were loosely bound to Nickel beads

In table 2 the counted number of phages in correlation with the dilution factors can be seen for the first round of phage display. An increase of phage number based on selection of sequences can be likely seen.
The total number of phages was estimated via multiplying the absolute number of phages on the plates with the corresponding dilution factors; results of this can be found in table 3 for the PBS control run and table 4 for the protease. The general number of phages increases over the course of the phage display for control and protease, however the increase for protease is higher. The number of elution (E) control stayed constant indicating the overall number of phages did not change between the rounds.

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Dilution factor</th>
<th>Colony number on plate</th>
<th>Total phage number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^3$</td>
<td>369</td>
<td>369.000</td>
</tr>
<tr>
<td></td>
<td>$10^4$</td>
<td>47</td>
<td>470.000</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>5</td>
<td>500.000</td>
</tr>
<tr>
<td>Day 2</td>
<td>$10^4$</td>
<td>49</td>
<td>490.000</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>9</td>
<td>900.000</td>
</tr>
<tr>
<td></td>
<td>$10^6$</td>
<td>1</td>
<td>1.000.000</td>
</tr>
<tr>
<td>Day 3</td>
<td>$10^4$</td>
<td>729</td>
<td>729.000</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>98</td>
<td>980.000</td>
</tr>
<tr>
<td></td>
<td>$10^6$</td>
<td>7</td>
<td>700.000</td>
</tr>
</tbody>
</table>

Table 3: Numbers of phages in the PBS control over all selection rounds for the first phage display round, red numbers not included in calculation colony numbers between ca. 40 and 1000 are most reliable

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Dilution factor</th>
<th>Colony number on plate</th>
<th>Total phage number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^3$</td>
<td>605</td>
<td>605.000</td>
</tr>
<tr>
<td></td>
<td>$10^4$</td>
<td>69</td>
<td>690.000</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>11</td>
<td>110.000</td>
</tr>
<tr>
<td>Day 2</td>
<td>$10^4$</td>
<td>13</td>
<td>130.000</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>77</td>
<td>7.700.000</td>
</tr>
<tr>
<td></td>
<td>$10^6$</td>
<td>8</td>
<td>8.000.000</td>
</tr>
<tr>
<td>Day 3</td>
<td>$10^4$</td>
<td>570</td>
<td>5.700.000</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>9</td>
<td>9.000.000</td>
</tr>
<tr>
<td></td>
<td>$10^6$</td>
<td>4</td>
<td>4.000.000</td>
</tr>
</tbody>
</table>

Table 4: Numbers of phages for GL50803_17327 over all selection rounds for the first phage display round, red numbers not included in calculation as colony numbers between ca. 40 and 1000 are most reliable

To statistically evaluate the increase which can be observed when comparing phage numbers for control and GL50803_17327 arithmetic averages for both have been calculated for every selection round(=day). Red marked numbers have excluded from the calculation either because they were derived from plates with very low colony numbers and therefore more prone to statistical variation or because they were heave outliers compared to the other phage numbers. The arithmetic values can be found in table 5 together with the ratio of sample over control which has been calculated to give a comparable measure since phages numbers even in the control selection vary substantially over several rounds of selection. The ratio increases by a factor 10 from day 1 to day 2 to a ratio of 16 from 1.5. On the third day of phage display this trend did not continue, manifesting itself in a drop of ratio mean phage number protease over mean phage number PBS to 6,67.

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Sample</th>
<th>Mean phage number</th>
<th>Ratio Sample/PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PBS</td>
<td>419.500</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GL50803_17327</td>
<td>647.500</td>
<td>1,543504172</td>
</tr>
<tr>
<td>Day 2</td>
<td>PBS</td>
<td>490.000</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GL50803_17327</td>
<td>7850000</td>
<td>16,02040816</td>
</tr>
<tr>
<td>Day 3</td>
<td>PBS</td>
<td>854500</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GL50803_17327</td>
<td>5.700.000</td>
<td>6,670567583</td>
</tr>
</tbody>
</table>

Table 5: Mean phage numbers and ratios for PBS and GL50803_17327 first round of phage display
Graph 1: Graphical evaluation of the development of protease driven selection over several selection rounds of phage display for first run of GL50803_17327. X-axis=round of selection, Y-axis= Ratio sample/control

The ratio of sample over PBS control is graphically presented in graph 1 showing the strong increase from day 1 to day 2 as well as the following decrease. It was decided to stop this round of phage display after day 3 and start new since most likely the selected phages had been at least partly lost on day 3. The next round of phage display was done with increased amounts of protease to increase the changes of selecting phages.

A second round of phage display was done in the exact same way. (Data not shown)

Due to time and enzyme limitations it was decided to stop phage display for this project and instead try testing possible substrates directly. Those substrates were chosen by their sequence containing proline or them having shown to be cleaved by giardial proteases earlier.
Testing of possible substrates

Several different likely substrates were in vitro tested on degradation, incubating them alone and together with the protease at 37°C for 2h. They were then run on SDS-PAGE for analysis of possible degradation.

No degradation for BSA or GL50803_17327 was seen (part A of figure 15). However, it was not possible to show degradation of BSA by the protease either. The same is true for all three different target sequences which were taken from studies on another giardial protease (part B of figure 11). Those target sequences are placed inside an artificial protein consisting of two thioredoxin residues to simulate a structured environment in which those sequences are likely to be found in vivo. The two residues are slightly different in size to make detection on the gel easier and are expected around 12-14kD. None of the sequences was cleaved by the protease, no degradation due to conditions was visible either. This indicates as well that for GL50803_17327 autoproteolytic activity is low even at 37°C.
4.0 Discussion

How exactly disease is caused by *Giardia intestinalis* is still unclear. Very few virulence factors are characterised as of now, yet it is known that the pathogen manipulates its host environment in manifold ways. With this thesis work on secreted serine and metalloproteases a small step towards understanding the exact molecular basis of the disease was done.

4.1 Cloning, expression and purification

Cloning could be done after protocols established for introduction of cysteine proteases into *Pichia pastoris*. It was confirmed via sequencing showing the correct insertion of all genetic elements into reading frame on the plasmid. Transfection into *Pichia pastoris* was done after established protocols as well and verified by restreaking yeast clones on selective marker Zeocin containing agar plates before beginning with test expressions. *Pichia pastoris* was chosen as expression vector since it has been shown to be able to express and secrete eukaryotic proteases fast, correctly translated, folded and modified.

Test expressions for selection of yeast clones that show high expression levels had to be slightly modified since neither serine protease nor metalloprotease inhibitors could be added. Usually those inhibitors are added to prevent yeast proteases from degrading the product and might here be needed as well to prevent auto digestion. This degradation is one possible reason why expression only for one of the four selected proteases has been successful so far. Other than that, protocols established for test expressions of cysteine proteases have been used.

Another reason why no high expression clones could be found present itself in the fact that especially the serine proteases are quite big compared to the cysteine proteases that have been expressed with the herein used protocols before. Modifications of those expression protocols might be needed to be able to find clones expressing those proteases at sufficient levels for purification and characterisation. These modifications can be for example concerning the time needed to express protein in high amounts after induction using methanol which might well be different from 72h. Another factor that can be changed in future expression attempts is temperature at which expression takes place with a slightly lower temperature resulting in slower expression and thereby possibly more exact folding of the protein making it more stable.

A different problem that might occur is that the other proteases might be toxic to yeast cells, resulting in death of high expressing clones rather than selection for them; indeed, in some cases slow growth of clones has been observed during test expression (data not shown).

Furthermore, it was not tested if the proteases were secreted or not. The used secretion signal usually leads to secretion in *Pichia pastoris* but it might be covered by some part of the protease or changed structurally in such a way that low or no secretion can occur. In such a case, the protease might be expressed at a high rate but will not be found in the supernatant and hence not be picked up by this method of screening for clones.

Finally, the rate of clones that express protein at high rate is quiet low in *P. pastoris* shown in the fact that by screening 50 clones for each construct only one clone in total was found expressing one of those proteases at a sufficiently high level. There is a good chance that additional testing of clones will also yield high level expression clones for the other proteases.

The purification of GL50803_17327 was successful using the established protocols. Work was carried out at room temperature. Loss of major amounts of protein was not observed at any single point during the process. The protease was of high purity after one step purification and showed little degradation indicating that it is very stable towards all present enzymes in the yeast supernatant and the temperature of 30°C used during expression. Later incubation at 37°C
confirmed this. These results are matching expectation for an enzyme that will be excreted into the intestine, a harsh environment with many proteases and degrading molecules as well as a constant high temperature. To fulfil its role, it needs to able to withstand those hazards. The additions of HIS-tag and yeast secretion signal do not seem to change those properties.

The amount of purified protein was very low at only 120µg/l culture. Optimisation of protocols for both expression and purification is therefore likely possible for Gl50803_17327 as well.

4.2 Bioinformatical analysis and experimental characterisation

The bioinformatical comparison in this work can give some inside on potential targets for all four proteases. This is even more so the case for GL50803_17327 were the human equivalent enzyme is well characterised and a structure is available. Human prolidase has 39% overall identity to GL50803_17327 and 98% of the giardial protein is covered. Human prolidase is a 494 residues long enzyme which is found to be active as a homodimer. It mostly cleaves Xaa-Pro dipeptides with a preference for Gly-Pro as substrate and is with this activity needed in final degradation of proline rich polypeptides like gelatine. Activity has also been reported against organophosphorus compounds detoxifying those. 33

Similar enzymes are well conserved over the entire palette of life including pro-and eukaryotes highlighting the importance of this protease and protein identity is at 94% high between assemblages WB and GS.

Sequence comparison via ClustalΩ alignment shows that the conserved residues are clustered in the C-terminal half of this protein between GL50803_17327 and both human prolidase and the identical gene in Strongyloides ratti. Here cluster with residues D277, D288, H371, E413 and E453 the metal-coordinating and thereby catalytic most important parts of this enzyme as well. All five residues are conserved in Gl50803_17327 here represented by respective residues D233, D244, H327, E369 and E409. Remarkably not only those residues but many around them are identical between all three sequences; those residues that have been exchanged are often highly similar indicating that their properties are important in creating a functional protease as well. Marking those conserved and highly similar residues on the structure of human prolidase shows a more differentiated picture of why certain amino acid residues had to be conserved and others could be replaced. Exchange of amino acids seems to have taken place mostly at the active site distant N-terminus as well as outer surfaces of the enzyme. Residues that were conserved are mostly found near the C-terminus where the active site is found and on the inner surfaces of the structure of human prolidase. This holds true for both sequential and structural proximity. Therefore, the metal coordinating pocket and the substrate binding pocket should be highly similar between those two enzymes as well.

This information together hints at GL50803_17327 being a protease with function very close to human prolidase likely to share substrate specificity as well. This would explain why no or little activity can be seen for the substrates tested as well as zymography and phage display. Next step of in-vitro analysis is an assay specifically to test for existence of free proline after cleavage of Gly-Pro dipeptides. A fitting ninhydrin-colorimetric assay resulting in increasing absorption at 515nm with increasing free proline in solution has been used to test activity of human prolidase before. 34-36 If this assay can verify activity of GL50803_17327 against Gly-Pro dipeptides it is highly likely that the final degradation step of proline containing polypeptides is the function of this protease in vivo as well since the digestion of proline requires highly specialised enzymes.
All in all, yeast is a suitable host for expression for these four proteases which are likely to be important virulence factors in giardiasis as well. For GL50803_6148, GL50803_15574 and GL50803_15832 additional work optimizing expression conditions and finding clones with high expression and secretion must be done as a mandatory step to be able to characterise them.

GL50803_17327 activity on the other hand will be tested by zymography again using commercial gels as well as predigesting for example collagen with the cysteine proteases found in *Giardia* before incubating it with this enzyme.

The ninhydrin based colorimetric assay will be established to test if the bioinformatical predictions hold true. As next steps after that activity of this proteinase can be characterised concerning different parameters like temperature, pH value, preferred metal ion and metal ion optimum by usage of that assay.

Once all four proteases are characterised a complete image of the final steps of digestion of peptides containing proline carried out by *Giardia* during infection will be visible. These peptides include not only digestion products of for example collagen but also chemokines which are often protected from degradation by proline motifs at their termini. This can explain the immunosuppression reported in some cases of giardiasis. On the other hand, the complete digestion of peptides to single amino acids is helping *Giardia intestinalis* to get an edge over host uptake. The malabsorption seen in infected can be partly explained by this digestive activity since human intestinal cells preferably take up amino acids at the dipeptide or tripeptide level and have their equivalent proteinases located intracellularly. Together those proteases can thereby help explain one small part of the broad range of symptoms seen in giardiasis.
**Abbreviations**

Amino acids | One and three letter code follow standard nomenclature
---|---
Bis | Bis(hydroxymethyl)aminomethane
BLAST | Basic Local Alignment Search Tool
BMGY | Buffered Glycerol-complex Medium
BMMY | Buffered Methanol-complex Medium
BSA | Bovine serum albumin
CREB | cAMP response element-binding protein
DNA | Deoxyribonucleic acid
HF | High-fidelity
IBS | Irritable bowel syndrome
IPTG | Isopropyl β-D-1-thiogalactopyranoside
I-TASSER | Iterative Threading ASSEMBly Refinement
LB | Lysogeny broth
MW | Molecular weight
Ni-NTA | Nickel-Nitritoltriacetic acid
ORI | Origin of replication
PCR | Polymerase chain reaction
PF | Primer forward
PR | Primer reverse
PVDF | Polyvinylidene fluoride
RT | Room temperature
SDS-PAGE | Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
TBST | Tris-buffered saline Tween 20 (=polysorbate 20)
Tris | Tris(hydroxymethyl)aminomethane
VSP | Variable surface protein
YPD | Yeast extract peptone dextrose medium
YPDS | Yeast extract peptone dextrose sorbitol medium
References


34. Chinard, F. P. PHOTOMETRIC ESTIMATION OF PROLINE AND ORNITHINE. Journal of Biological Chemistry 199, 91–95 (1952) http://www.jbc.org/content/199/1/91.short.
Appendix

Figure S1: Distance tree based on BLAST pairwise alignments for GL50803_6148 (marked in yellow) showing proteins with the biggest sequential identity and their evolutionary distance to this protease.

Figure S2: Distance tree based on BLAST pairwise alignments for GL50803_15574 (marked in yellow) showing proteins with the biggest sequential identity and their evolutionary distance to this protease.
Figure S 3: Distance tree based on BLAST pairwise alignments for GL50803_15832 (marked in yellow) showing proteins with the biggest sequential identity and their evolutionary distance to this protease
Acknowledgements

I want to thank first and foremost Staffan Svärd for accepting me as student for this master thesis and providing theoretical background whenever needed as well as an open ear for whatever problem might come up during the months of this project.

A similar big thank you goes out to Jingyi Liu for teaching me everything needed to practically execute this project without complaining if I ask the same questions three or more times. Also thank you for stopping me from exhausting myself unneeded by working while sick.

Every other member of the Giardia group: Thank you! I felt welcome from the first week and every single one was always willing to help if I needed some. Be it like Elin with general knowledge about anything that can be needed in this laboratory, like Showgy who always had a friendly word and protein background knowledge, or Asgeir with sequencing related help, like Dimitra with musical entertainment at the bench or Laura for being the social Mexican you are keeping care of us “children”. Special thanks to my fellow child Jana for countless coffee gossip breaks and constant bench neighbourship as well as permanent discussion about both our projects.

The social part is of great importance for a workplace to be somewhere you can work well- the Micro corridor is providing plenty of this with regular common coffee breaks, pub nights and overall social activities. Thanks to everyone who is helping to make it such a friendly and entertaining place you are way too many to be listed here.

Furthermore, I can hardly express the thankfulness I feel towards my family, who have supported me on my way until here and never asked why I need to study so far away from home but instead pushed me to go wherever I wanted to go.

Finally, all my friends, those from earlier education and those I made here and that provide an excellent counterweight to the closed off science world I am moving in for the majority of my time, be it by dragging me out into nature, visiting festivals and concerts or playing music and theatre with me.

THANK YOU!