Comparative proteomics of *Giardia duodenalis* from humans and cattle

Robert Francis Lascelles Steuart
BSc Hons

This thesis is presented for the degree of Doctor of Philosophy
of Murdoch University
2012
I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

Robert Francis Lascelles Steuart
Abstract

*Giardia duodenalis* is a gastrointestinal parasite capable of infecting humans as well domesticated animals, for example cattle. There are seven distinct genetic groups, termed assemblages A-G, with assemblages A and B able to infect humans and assemblage E specific to livestock, including cattle. The level of genetic variation between the assemblages has been studied over multiple loci and the phylogenetic relationship between these assemblages is well known. There is however, little information available on the protein differences between the assemblages. Proteins of the human infective assemblages A and B were compared using SDS-PAGE and 2D-PAGE to determine proteins of difference. Proteins determined to be assemblage-specific were then identified using mass spectrometry. In total, eleven proteins of difference were identified between assemblages A and B. Four proteins; alpha 2 giardin, GASP-180, UPL-1 and GLORF-C4, were chosen for further characterisation. Genetic analysis confirmed that alpha 2 giardin is absent from assemblage B and that the size variation seen in the GASP-180 protein is mirrored by a series of indels in a portion of the gene sequence. The UPL-1 gene did not show any variation indicating the protein variation seen is likely due to post translational modification. The GLORF-C4 protein, which is involved in the formation of cysts, was only identified in assemblage B. Therefore, the ability of assemblages A and B to undergo the encystment process was also studied. The assemblage B isolates produced fully formed cysts 24 hrs faster than assemblage A isolates. Analysis of levels of GLORF-C4 mRNA indicated that the gene is constitutively expressed in assemblage B and induced in assemblage A. The proteins of the human infective assemblages were then compared to those of the livestock infective assemblage E, with thirteen protein variants identified, the majority of which are the same as those
identified between assemblages A and B. The proteins identified in this study are the first protein variants documented between assemblages of *G. duodenalis*. 
# Table of Contents

Abstract ........................................................................................................................................ iii

Table of Contents .......................................................................................................................... v

List of Figures .................................................................................................................................... ix

List of Tables ...................................................................................................................................... xi

Acknowledgements .......................................................................................................................... xii

1 Introduction .................................................................................................................................... 1
   1.1 General Introduction .................................................................................................................. 1
   1.2 Life Cycle ................................................................................................................................. 4
   1.3 Transmission ............................................................................................................................. 6
   1.4 Pathogenicity ............................................................................................................................ 7
   1.5 Variation in Giardia duodenalis ............................................................................................... 10
   1.6 Proteomics ............................................................................................................................... 12
   1.7 Proteomics of Giardia ................................................................................................................ 14
      1.7.1 Cell Structure and Biology ............................................................................................... 15
      1.7.2 Host-Parasite Interactions ............................................................................................... 19
      1.7.3 Protein Variation ............................................................................................................... 21
   1.8 Objective and aims ................................................................................................................... 24

Chapter 2 General Materials and Methods ..................................................................................... 26
   2.1 Isolates .................................................................................................................................... 26
   2.2 Culture Medium ....................................................................................................................... 27
   2.3 Culture Maintenance ................................................................................................................ 28
   2.4 Cryopreservation ................................................................................................................... 29
   2.5 Mass Culture .......................................................................................................................... 29
   2.6 Sample Preparation ................................................................................................................ 30

Chapter 3 One Dimensional Polyacrylamide Gel Electrophoresis .................................................... 32
   3.1 Introduction ............................................................................................................................. 32
   3.2 Materials and methods ............................................................................................................ 33
      3.2.1 Sample preparation ........................................................................................................... 33
      3.2.2 One dimensional electrophoresis .................................................................................... 34
      3.2.3 Gel Staining .................................................................................................................... 37
      3.2.4 Gel Imaging ..................................................................................................................... 37
      3.2.5 Comparison of Gel Images ............................................................................................. 38

Results 3.3 ........................................................................................................................................ 38
   3.3.1 Preliminary SDS-PAGE Analysis ....................................................................................... 38
3.3.2 Comparative SDS-PAGE Analysis .................................................. 40
3.4 Discussion ..................................................................................... 43
  3.4.1 Intra-assemblage Variation ............................................... 43
  3.4.2 Inter-assemblage variation ................................................. 44
  3.4.3 Comparison with published data ........................................ 44
Chapter 4 Two Dimensional Polyacrylamide Gel Electrophoresis ............. 47
  4.1 Introduction ............................................................................. 47
  4.2 Material and Methods ............................................................. 49
    4.2.1 Sample preparation ......................................................... 49
    4.2.2 Isoelectric Focusing ........................................................ 49
    4.2.3 SDS Polyacrylamide Gel Electrophoresis ....................... 51
    4.2.4 Protein Visualisation ....................................................... 53
  4.3 Results ..................................................................................... 53
    4.3.1 Protein Solubilisation Optimisation .................................. 53
    4.3.2 Protein Concentration Optimisation ................................. 54
    4.3.3 Gel Reproducibility ......................................................... 56
    4.3.3 Inter Assemblage Variation ............................................. 59
  4.4 Discussion .............................................................................. 60
    4.4.1 Comparison of protein resolution between 2D-PAGE and SDS-PAGE ... 60
    4.4.2 Optimisation of protein solubilisation ............................... 62
    4.4.3 Optimisation of the protein concentration required for visualisation ..... 63
    4.4.4 Reproducibility of the 2D-PAGE system ......................... 64
    4.4.5 Inter-assemblage protein variation ................................... 65
Chapter 5 Identification of Proteins ....................................................... 67
  5.1 Introduction ............................................................................. 67
  5.2 Material and Methods ............................................................. 68
    5.2.1 Gel spot excision .............................................................. 68
    5.2.2 Gel Sample Processing .................................................... 69
    5.2.3 Mass Spectrometry ........................................................... 70
    5.2.4 Bioinformatics .................................................................. 70
  5.3 Results ..................................................................................... 71
    5.3.1 Proteins Identified from SDS-PAGE .................................. 71
    5.3.2 Proteins Identified from 2D-PAGE .................................... 75
  5.4 Discussion .............................................................................. 79
    5.4.1 Proteins Identified with Mass spectrometry ....................... 79
    5.4.2 Success of Proteins Identified from 2D-PAGE Gels ............. 84
    5.4.3 Protein Variation .............................................................. 85
List of Figures

Figure 1.1 Cysts (▲) and trophozoites (●) of *G. duodenalis* derived from culture. ...5
Figure 1.2 Transmission Dynamics of *G. duodenalis*.......................................................8
Figure 3.1 Schematic of Gel Assembly..................................................................................36
Figure 3.2 Trial SDS-PAGE for Protein Separation.................................................................39
Figure 3.3 12.5% PAGE of TTP from *Giardia duodenalis* Assemblages B and A...41
Figure 3.4 10% PAGE of TTP from *Giardia duodenalis* Assemblages A and B......42
Figure 4.1 2D-PAGE of *Giardia duodenalis* isolate BAH 34c8 TTP extracted with MSS, pH 3-10 IPG strip and 12.5% polyacrylamide gel..........................55
Figure 4.2 2D-PAGE of *Giardia duodenalis* isolate BAH 34c8 TTP Extracted with PS2 Protocol pH 3-10 IPG strip and 12.5% polyacrylamide gel...............56
Figure 4.3 Protein Concentration Optimisation.................................................................57
Figure 4.4 Reproducibility of 2D Gel Images.......................................................................58
Figure 4.5 2D-PAGE of isolate BAH 2c2 (A).................................................................59
Figure 4.6 2D-PAGE of isolate BAH 34c8 (B).................................................................60
Figure 5.1 Protein bands of assemblage A and B isolates BAH 2c2 and BAH 34c8 excised and identified from SDS-PAGE..........................................................72
Figure 5.2 Proteins excised and identified from assemblage A isolate BAH 2c2 2D-PAGE ..............................................................................................................75
Figure 5.3 Protein spots of assemblage B isolate 34c8 excised and identified via mass spectrometry..................................................................................77
Figure 6.1 A2GEX PCR products run in a 1% agarose TAE gel stained with Sybr Safe..........................................................100
Figure 6.2 Alph2ex PCR products run in a 1% agarose TAE gel stained with Sybr Safe..........................................................100
Figure 6.3 Alph2int PCR products run in a 1% agarose TAE gel stained with Sybr Safe..........................................................101
Figure 6.4 Alignment of assemblage B sequences obtained from the 1000bp amplicon of the Alpha2int PCR with the genomic Assemblage A Wee Kinase sequence..........................................................103
Figure 6.5 Axo3 PCR products run in a 1% agarose TAE gel stained with Sybr Safe .................................................................................................104
Figure 6.6 Alignment of Axo3 PCR amplicon sequence from assemblages A and B with the assemblage A genomic sequence..........................................................105
Figure 6.7 UDP PCR products run in a 1% agarose TAE gel stained with Sybr Safe..106
Figure 6.8 Alignment of UDP PCR amplicon sequences from Assemblages A and B with the assemblage A genomic sequence..........................................................107
Figure 7.1 Percentage Encystation of Trophozoites incubated with 1% w/v Porcine and Bovine bile .................................................................124
Figure 7.3 Effect of Variation in pH on Encystation of *Giardia duodenalis* from Assemblages A and B .................................................................125
Figure 7.2 Effect of New Born Serum and Foetal Bovine Serum on the Percentage of Trophozoites Undergoing Encystation .....................................126
Figure 7.4 Encystation Rate for Assemblage A Isolates BAH 2c16 and BAH 3c3. 127
Figure 7.5 Encystation Rate for Assemblage B Isolates BAH 34c8 and BAH 12c14 .........................................................................................128
Figure 7.6 Comparison of *glorf-c4* expression between Assemblages A and B trophozoites ...........................................................................129
Figure 7.7 Expression of *glorf-c4* before and 8 and 18 hours post induction for encystment .............................................................................130
Figure 8.1 SDS-PAGE of Assemblages A and B and Comparing with Assemblage E .................................................................................140
Figure 8.2 Protein Bands of Assemblage E Isolate P15c1 that were submitted for mass spectrometry ......................................................................141
List of Tables

Table 1.1 Morphological Characterisation of *Giardia* sp. ....................................................... 1
Table 2.1 *Giardia duodenalis* isolates used in this study ......................................................... 27
Table 2.2 BS-I-33 Giardia culture medium ................................................................................. 28
Table 3.1 Polyacrylamide Gel Formulations .............................................................................. 35
Table 5.1 Proteins of assemblage A isolate BAH 2c2 visualised on SDS PAGE and identified with mass spectrometry ......................................................................................... 73
Table 5.2 Proteins of assemblage B isolate BAH 34c8 visualised on SDS-PAGE and identified with mass spectrometry ......................................................................................... 74
Table 5.3 Proteins of Assemblage A isolate BAH 2c2 visualised on 2D-PAGE identified by Mass Spectrometry ................................................................................................. 76
Table 5.4 Proteins of Assemblage B isolate BAH 34c8 visualised on 2D-PAGE identified by Mass Spectrometry ................................................................................................. 78
Table 5.5 Protein Variants of Assemblages A and B Identified Using SDS-PAGE and 2D-PAGE Coupled With Mass Spectrometry ................................................................. 86
Table 6.1 Primers used for PCR analysis of variant protein genes .............................................. 94
Table 7.1 Primers Used for Quantitative Real Time PCR Assay .................................................. 123
Table 7.2 Summary of data for the variation in encystment and expression of *glorf c4* ................................................................................................................................................. 131
Table 8.1 Proteins of Assemblage E Isolate P15c1 Identified Using Mass Spectrometry from SDS-PAGE Gels ............................................................................................................. 142
Table 8.2 Variant Proteins Identified by Mass Spectrometry and Their Position on Gels ................................................................................................................................. 144
Acknowledgements

There are many people I would like to thank for helping me to produce this thesis and the body of work that sits behind it. Firstly to my supervisors Prof Andrew Thompson, Dr Ryan O’Handley and Dr Richard Lipscombe whose support and encouragement through this process has been greatly appreciated, especially our meetings which would often leave me with many more ideas than I had time for. I would particularly like to thank Andy (Prof. Thompson) whose mentorship, guidance and ultimately friendship has been of tremendous benefit to me. I have had the great benefit to work with several visiting international scientists as part of this work and on other experiments surrounding it. The visits of Prof Timothy Paget were tremendously insightful and the work I conducted with Tim on GLORF-C4 forms a strong part of this thesis. My thanks also to Prof. Andre Buret, Dr. Thomas Geurden and Prof. John Barta who also visited during my studies and broadened my interests in parasitology. I have been very lucky to share an office space with some other fantastic students and their support and sometimes distraction was much needed, particular thanks to Drs. Carly Palmer, Nevi Parameswaran and Amanda Ash. Finally to my family, my grandparents who both died during my studies and who I know would be proud of me for completing; they were and still are an inspiration. To my mum, brother and sister for always being interested, asking and supporting me and to my adopted family for their support as well. Ultimately, my deepest thanks go to my own special family, to my wife who has put up with Giardia for way too long (and will probably have to do so for a lot longer) and the two little boys we have brought in to this world during my studies who have been corrupted from the start.

Thank you all!
1 Introduction

1.1 General Introduction

*Giardia* is a genus of flagellated protozoan organisms belonging to the Order Diplomonadida. They colonize the intestinal tract of a wide range of organisms, in some cases causing intestinal disease. There are six distinct species within the genus *G. duodenalis, G. muris, G. agilis, G. ardea, G. microti* and *G. pistacci* which are delineated on the basis of morphological characteristics (Table 1.1) (Thompson and Monis, 2011).

Table 1.1 Morphological Characterisation of *Giardia* sp.
Adapted from (Adam, 2001)

<table>
<thead>
<tr>
<th>Species Name</th>
<th>Hosts</th>
<th>Light Microscopy</th>
<th>Electron Microscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>G. duodenalis</em></td>
<td>Mammals</td>
<td>Pear shaped; claw hammer shaped median body</td>
<td></td>
</tr>
<tr>
<td><em>G. muris</em></td>
<td>Rodents</td>
<td>Short and rounded; small rounded median body</td>
<td></td>
</tr>
<tr>
<td><em>G. agilis</em></td>
<td>amphibians</td>
<td>Long and slender; teardrop shaped median body</td>
<td>Ventral disk and caudal flagella similar to <em>G. muris</em></td>
</tr>
<tr>
<td><em>G. ardea</em></td>
<td>Herons</td>
<td>Same as <em>G. duodenalis</em></td>
<td>Incomplete ventrolateral falange, absent marginal groove</td>
</tr>
<tr>
<td><em>G. psittaci</em></td>
<td>Psittacine birds</td>
<td>Same as <em>G. duodenalis</em></td>
<td></td>
</tr>
<tr>
<td><em>G. microti</em></td>
<td>Voles and muskrats</td>
<td>Same as <em>G. duodenalis</em></td>
<td>Cysts contain two trophozoites with mature adhesive disks</td>
</tr>
</tbody>
</table>
Giardia duodenalis (synonymous with G. intestinalis and G. lamblia) is the species of Giardia associated with human infection and is a major cause of diarrhea (Thompson, 2004). There are approximately 700 cases of G. duodenalis infection reported annually in Western Australia, with over half of these cases occurring in the Greater Perth area (Anon, 2006). In the United States of America it is estimated that there are 2,000,000 infections every year with 5,000 of these requiring hospitalization (Mead et al., 1999). Infection with G. duodenalis follows the faecal oral route. As such, a majority of cases are seen in the developing world where sanitation and hygiene are lacking. In the developed world the majority of cases are seen in infants and toddlers, and their care givers, be they parents or child care professionals (Read et al., 2002).

Due to its importance to human health, most of the information we have on the genus Giardia is based on studies conducted on G. duodenalis (Adam, 2001). As well as being able to infect humans, G. duodenalis has been isolated from a wide variety of other mammalian hosts and there is a large amount of variation within the species (Thompson and Monis, 2004). Genetically, the Giardia duodenalis morphological group is differentiated into eight distinct genetic assemblages. These assemblages designated A-H, show varying degrees of host specificity. Assemblages A and B are the only assemblages able to infect humans, however they have been identified in a wide range of different mammalian hosts (Thompson and Monis, 2012). Assemblages C and D are specific to canids, and assemblage E to hoofed animals (Monis et al., 1999). Assemblage F is the felid genotype and G is found in rats (Thompson and Monis, 2004). Assemblage H has only recently been described and
was identified in samples from marine vertebrates (Lasek-Nesselquist et al., 2010). Currently only assemblages A and B are able to be routinely cultured and as such, the majority of in depth biologically research is conducted on isolates from these assemblages. There is one isolate of assemblage E in culture which was collected from the intestine of a pig in the former Czechoslovakia (Koudela et al., 1991).

The two human infective assemblages have been further split, each into two sub-assemblages, AI and AII and BIII and BIV (Mayrhofer et al., 1995b). The phylogenetic investigations of assemblage A indicate that it has a much more clonal ancestry with limited variation between individual isolates (Logsdon., 2008). Assemblage B on the other hand, displays a large degree of genetic variation between isolates. The reason behind this is unclear; however the increased divergence between isolates in assemblage B is used as an example of a possible sexual life cycle in *G. duodenalis*.

There is increasing evidence supporting revision of the taxonomy of the *G. duodenalis* assemblages as separate species (Monis et al., 2009). The level of genetic diversity seen between the assemblages of *G. duodenalis* is as great as between assemblage A and *G. muris*. It has been proposed that assemblage A *G. duodenalis* should retain the original species name proposed by Filice (1952) and that assemblage B take the name *G. enterica* (Monis et al., 2009). Assemblages C and D are proposed to be combined as one species resurrecting the name *G. canis* with assemblage E named *G. bovis*. Finally assemblage F is proposed as *G. felis* and assemblage G *G. simondi* (Monis et al., 2009). This revision is based on the phylogenetic distance between the assemblages and the host specificity of some of
the assemblages. The new species names for the assemblages of *G. duodenalis* are yet to be accepted and further information as to the level of variation between these assemblages would be of great benefit.

### 1.2 Life Cycle

All *Giardia* species follow a similar two stage life cycle involving an infective cyst and a trophozoite found in the small intestine of the host. As stated previously, infection follows a faecal oral route with infective cysts shed in the faeces of a host (Figure 1.1). Cyst are environmentally resistant and consist of a trophozoite that has undergone partial replication (4N instead of 2N) wrapped in a polymer of N-acetyl galactosamine along with two proteins, termed the cyst wall proteins (CWP-1, CWP-2) (Lujan *et al.*, 1998). The cysts are ingested and travel through the stomach into the small intestine. The low pH environment in the stomach triggers excystation with the release of several factors including an acid phosphotase that breaks down the cell wall (Slavin *et al.*, 2002). The trophozoites are then released in the small intestine where they attach to epithelial cells.

As the trophozoites move down the small intestine the environment changes and triggers encystation. Originally this was thought to be due to an increase in the concentration of bile; however it has been shown that a decrease in the level of free cholesterol in the environment is the key trigger for encystation (Lujan *et al.*, 1998). Encystation begins with the production of the components for the cyst wall, such as CWP-1 and CWP-2, and their compartmentalisation in encystation specific vesicles (ESVs) within the cytoplasm (Hehl *et al.*, 2000). The ESVs are transported to the trophozoite membrane for incorporation into the cyst wall along with other material.
described above. As the cyst wall is formed around the trophozoite it begins a final round of binary fission which is left uncompleted and is restarted when excystation is triggered.

Figure 1.1 Cysts (▲) and trophozoites (●) of *G. duodenalis* derived from culture.

All species of *Giardia* have trophozoites with similar morphological characteristics; they all possess two nuclei, 4 pairs of flagella (caudal, ventral, anterolateral and posterolateral), an adhesive disc (syn. sucking or ventral disc) and median bodies (Benchimol and De Souza, 2011). *Giardia duodenalis* trophozoites are characterised as having a pear-shaped body with claw hammer shaped median bodies. In the small intestine, the trophozoites attach to the epithelial cell surface using the adhesive disc. The exact mode of attachment is unclear, however several theories exist. The longest held position is that the movement of the flagella pushes the disk on to the epithelial
cell surface causing a vacuum like attachment of the trophozoite to the cell surface (Holberton, 1974). This theory was developed given the imprint of the disc that is sometimes left on the cell surface following detachment of trophozoites (Holberton, 1974). A recent study has shown that trophozoites with mutations to the flagella proteins that have impaired or even lack motility are still able to attach to a surface indicating there may be a conformational change to the disk which enables attachment (House et al., 2011).

In the small intestine the trophozoites divide via binary fission with a single trophozoite producing two daughter cells. Encystation is triggered by a decrease in the level of free cholesterol in the small intestine (Lujan et al., 1996). This triggers the production of cellular components required for encystation including the encystation specific vesicles (ESV) (Hehl et al., 2000). The ESVs contain CWP-1 and CWP-2 and are trafficked to the trophozoite surface so that the CWPs can be used to form the cyst wall. The environmentally resistant cysts, which are infective immediately, are then shed in the faeces of the host.

1.3 Transmission

Transmission *G. duodenalis* follows the faecal oral route with the process of infection described above. The source of the cysts that infect an individual can vary. As two of the assemblages are able to infect a variety of mammalian species there has been much debate as to the zoonotic significance of *G. duodenalis*. There is evidence of humans and dogs within the same household carrying genetically similar strains of *G. duodenalis* (Traub et al., 2003), however there is no direct evidence of infection of humans from close association with infected animals and vice-versa.
(Geurden and Olson, 2011). The potential interplay between humans, animals and the environment gives rise to a very complex transmission cycle for *G. duodenalis* infection that may contain both zoonotic, zooanthroponotic and anthropoontic disease routes (Figure 1.2).

### 1.4 Pathogenicity

The exact disease mechanisms involved in clinical giardiasis, the disease caused by infection with *G. duodenalis*, are not clearly understood. Clinical symptoms are not always present; however, those who do display clinical symptoms can experience diarrhoea, nausea, bloating, excessive flatulence, vomiting and steatorrhea (Buret and Cotton, 2011). In extreme cases malnutrition and failure to thrive can develop, particularly in children. Those infected with assemblage B *G. duodenalis* are more likely to display chronic infections, whereas those with assemblage A infections are more likely to be acutely affected (Read *et al.*, 2002, Haque *et al.*, 2005, Sahaqun *et al.*, 2008). Chronically infected individuals were characterised as having a prolonged history of symptoms associated with giardiasis, while acute infections had no previous history of abnormal gut function (Read *et al.*, 2002, Haque *et al.*, 2005, Sahaqun *et al.*, 2008).
Several potential disease processes have been identified. Enterocytes in the small intestine act as a barrier to prevent molecules and organisms entering the interstitial space and to prevent loss of water and solutes from interstitial fluid into the gut lumen (Buret and Cotton, 2011). Tight junctions are formed between the enterocytes to prevent this from occurring. The integrity of the tight junctions has been shown to be compromised during infection with *G. duodenalis* (Buret et al., 2002, Chin et al., 2002). The protein zona-occludens 1, which is involved in tight junction formation,
displays reduced expression at the cell margins during infection with \textit{G. duodenalis}.
This allows molecules to move freely between the gut lumen and intra-cellular matrix, which could lead to diarrhoea (Buret \textit{et al.}, 2002).

There is also a diffuse shortening of the microvilli, this reduces the surface area the epithelial cells can use to absorb nutrients as such there is a build-up of nutrients in the gut lumen. Associated with the decreased surface area of the gut lumen there is a reduction in the active transport of water, sodium and glucose into the cells (Holberton, 1974, Nores \textit{et al.}, 2009). The decrease in nutrient absorption leads to a concentration gradient within the gut and water moves from the tissue into the gut leading to the watery diarrhoea often associated with giardiasis (Nores \textit{et al.}, 2009).

Clinical symptoms are not seen in every case, even when different patients are infected with the same isolate (Nash \textit{et al.}, 1987). This indicates that the parasite does not solely dictate pathogenesis. It is likely that the immune response mounted against trophozoites has some detrimental effect on the small intestine itself. Difference in an individual’s immune response to infection would therefore affect the clinical symptoms that are manifested.
1.5 Variation in Giardia duodenalis

There is a large amount of variation within the species *G. duodenalis*, as such it is often described as a “species complex” (Andrews et al., 1989). This variation has been demonstrated by many different authors using a vast array of techniques including morphometrics, DNA sequencing and isoenzyme analysis. Binz (1996) conducted a thorough study of several *G. duodenalis* isolates, collating data from caudal flagella length, median body shape, trophozoite size, mean generation time and DNA content. All of the isolates used in this study were also characterized genetically. Although there was a large degree of variation between isolates there was no correlation between genetic group and any of the characteristics tested (Binz, 1996).

As previously stated the species *G. duodenalis* comprises eight distinct genetic groupings, termed assemblages A to H. Four genetic loci are commonly used to distinguish between the assemblages; 18S ribosomal RNA (18S rRNA), β giardin, glutamate dehydrogenase (*gdh*) and triosephosphate isomerase (*tpi*). The level of genetic variation between the assemblages at each of these loci varies (Wielinga and Thompson, 2007). The 18S rRNA locus, which is commonly used in phylogenetic analysis due to its conserved nature, shows several single base substitutions between the assemblages. The *tpi* locus has a much greater level of diversity and is used to sub-genotype assemblages A and B (Wielinga and Thompson, 2007). The variation at the genetic level for *tpi* and *gdh* also translates to variation in the amino acid sequence whereas the differences in the β giardin and 18s rRNA sequences are silent mutations causing no change in the amino acid sequence (Wielinga and Thompson, 2007).
Isoenzyme analysis was commonly used for analysis of *G. duodenalis* isolates before PCR analysis became popular (Meloni *et al.*, 1988, Meloni *et al.*, 1991, Moss *et al.*, 1992, Mayrhofer *et al.*, 1995a, Monis *et al.*, 2003b). The technique uses isoelectric focusing to separate proteins according to their pI, the proteins are electrophoresed on cellulose-acetate membranes which is then stained for the presence of a particular enzyme. Staining is performed by incubating the membrane with the substrate of a particular enzyme. Variation in the migration of the enzyme of interest is used to group the isolates for phylogenetic analysis. All studies found variation with the majority of enzymes tested (Meloni *et al.*, 1988, Meloni *et al.*, 1991, Moss *et al.*, 1992, Mayrhofer *et al.*, 1995a, Monis *et al.*, 2003b) and where phylogenetic relationships between isolates were tested these correlate with PCR based analyses (Monis *et al.*, 1999, Monis *et al.*, 2003b).

With the advent of PCR, the focus of research on variation within *G. duodenalis* has concentrated further and further on genetic sequence analysis. The outcome of the isoenzyme studies shows that proteins can also be used to generate meaningful data on variation within *G. duodenalis*. Isoenzyme analysis is however a highly directed methodology which requires a protocol for staining the protein of interest whereas global proteome studies a much broader range of proteins to be studied not just enzymes. To truly understand the level of protein variation within a species it is best to take a global view of the proteins produced.
1.6 Proteomics

Proteomics is the study of the proteins within an organism or sub cellular fraction under a given set of conditions at a specific point in time. This protein complement is termed the proteome (Barrett et al., 2000). The tools used in proteomics have been available for the study of organisms since the 1970’s, however it is only in the last ten years that its true potential has been demonstrated. Advances in genomics and computational biology have opened the proteome up for in-depth study. With the publication of an organism’s genome it is now possible to identify proteins back to a gene in a very short period of time, using the latest mass spectrometry (MS) techniques and search algorithms (Burlingame et al., 1998).

Proteomics provides biological data as to the proteins being expressed and the level of expression within the cell. These levels are often found to differ from those generated on the basis of messenger RNA (mRNA), as other regulatory factors contribute to the translation of mRNA into protein (Cordwell et al., 2001). Levels of mRNA and protein can differ up to as much as 30 fold in yeast models (Gygi et al., 1999). Proteomics also allows for the identification of post translation modification which can not be measured using technologies based on DNA and mRNA analysis. Measuring the degree of post translation modifications such as phosphorylation is important in understanding the flux of enzyme activation of a cell (Barrett et al., 2000).

The techniques used in proteomics revolve around the separation of proteins and their subsequent identification. Separation can be performed in gel based systems, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) or two
dimensional PAGE (2D-PAGE), or using high performance liquid chromatography (HPLC). SDS-PAGE separates proteins on the basis of size and 2D-PAGE separates proteins firstly by their isoelectric point (pI) and secondly by their size. HPLC utilises columns that retain proteins according to a physical or chemical property. By changing the conditions of the solution running through the column the proteins can be sequentially eluted (Wang and Hanash, 2003). Multi-dimensional HPLC analysis is now being used extensively for large scale proteomic analysis. This technique uses HPLC columns that separate the proteins according to different properties; fractions from the first column are placed onto a second column and then put through the MS workflow to identify proteins (Wang and Hanash, 2003). Multi-dimensional HPLC produces a large amount of bioinformatics data which needs to be processed. The HPLC system does identify more proteins than SDS-PAGE or 2D-PAGE; however no technique is able to identify all proteins. Identification was originally performed using Edman degradation; however MS is now the primary tool for protein identification.

Mass spectrometry is a technique used to determine the molecular mass of an ion, by determining it’s mass to charge ratio, m/z. Mass spectrometers consist of three parts, an ionisation source, a mass analyser and a detector. The ionisation source places a charge on the sample to be studied. There are two major methods used to ionise the sample, electrospray ionisation and matrix assisted laser desorption/ionisation (MALDI). The mass analyser separates the sample into the individual ions with the two most common analysers being time of flight and quadrapole. The ions reach a detector and a spectra is generated with the ions displayed based on their mass to charge ratio.
For the identification of proteins, protein samples are first digested with trypsin which cleaves the protein at the carboxyl side of arginine and lysine residues and then put through a mass spectrometer. The spectra generated is searched against an in silico database of gene products theoretically digested with trypsin. Proteins are identified based on the spectra matching. Many proteins are now identified using de novo peptide sequencing using tandem mass spectrometry. This utilises two rounds of mass analysis, after the first round of mass analysis a peptide ion of interest is selected for in an ion trap. The peptide is then broken up with an inert gas and a second round of mass analysis is performed. The inert gas has the potential to break all of the bonds in the peptide but it will not break every bond in every peptide. When the spectra is analysed peaks which differ by the mass of one amino acid are selected for. These peaks represent the peptide with an extra amino acid removed between each peak. By determining the mass difference between two adjacent peaks the amino acid lost is inferred. This is repeated for the entire spectra and the peptide is sequenced. This peptide sequence can then be used in sequence homology searches.

1.7 Proteomics of Giardia

Proteomics has been used for the study of G. duodenalis since the 1980’s. Initial investigations revolved around identifying the nature of the proteins prevalent in the cytoskeleton and those proteins released by trophozoites. Later the focus shifted to identification of the dominant proteins recognised by the host immune system. In recent years the development of the G. duodenalis genome project and advances in
mass spectrometry and bioinformatics have allowed for the identification of *G. duodenalis* proteins back to a gene.

### 1.7.1 Cell Structure and Biology

Several studies have aimed to identify the proteins of the cytoskeleton of *G. duodenalis* trophozoites (Holberton and Ward, 1981, Crossley and Holberton, 1983, Crossley *et al.*, 1986, Peattie *et al.*, 1989, Palm *et al.*, 2005, Hernandez-Sanchez *et al.*, 2008, Hagen *et al.*, 2011). The first, by Holberton and Ward (1981) examined proteins extracted from the axonemes and adhesive disks of trophozoites, with two major proteins identified; one that co-migrated with tubulin and a second protein at around 30 kDa. A second study visualised 20 individual protein bands on SDS-PAGE ranging from >120 kDa to 17 kDa from similar cytoskeletal preps (Crossley and Holberton, 1983). Three protein bands were identified on basis of co-migration with known proteins, these being alpha and beta tubulin as well as actin. A new class of structural proteins that migrated at 30 kDa, the giardins, were also described due to the level of difference in amino acid composition when compared to both rat brain and *G. duodenalis* tubulin (Crossley and Holberton, 1983). This was the same protein band previously identified by Holberton and Ward (1981). The name giardins was later used to include those cytoskeletal proteins that migrated between 29 and 38 kDa (Peattie *et al.*, 1989), with the 30 kDa giardin designated as beta giardin.

Peattie *et al.* (1989) continued work on the cytoskeleton of *G. duodenalis* with the first 2D-PAGE analysis. Nine individual proteins were separated with several proteins displaying multiple isoforms and a new sub class of the giardins was described, alpha giardin, which migrated at 33 kDa. The overall number of proteins and the molecular mass range was decreased when compared to the SDS-PAGE
analysis of Crossley and Holberton (1983), however this is not unexpected due to the difference in solubilisation techniques used in 2D-PAGE and SDS-PAGE (Rabilloud, 1996). Sodium dodecyl sulphate is one of the most effective protein solubilisation agents, however it coats the protein molecules in a negative charge (Rabilloud, 1996). This charge would interfere with the isoelectric focusing used in the first dimension of 2D-PAGE, which relies on proteins having no net charge at their pI, therefore ruling out SDS for protein solubilisation for 2D-PAGE. The buffers for 2D-PAGE must use zwitterionic detergents, which are not as effective at solubilising proteins, resulting in the decreased number of proteins visible in 2D-PAGE (Rabilloud, 1996).

The work of Peattie et al. (1989) was later confirmed in a second investigation of the adhesive disk using 2D-PAGE (Palm et al., 2005). Seven separate proteins were identified including alpha and beta tubulin, alpha-1 giardin, beta giardin, delta giardin, gamma giardin and a new cytoskeletal protein, SF-assemblin like protein-1 (SALP-1) using 2D-PAGE coupled with MS (Palm et al., 2005). The expression of the cytoskeletal proteins during encystation of some of the cytoskeletal proteins were also tested by Palm et al. (2005). Of the three proteins investigated (beta giardin, delta giardin and SALP-1) there was no marked change in protein abundance; however there was a significant decrease in the levels of mRNA detected for these proteins (Palm et al., 2005). This highlights a key point in proteomic studies that the level of mRNA is not always indicative of the amount of protein within the system. Other regulatory factors can affect the translation of mRNA into its corresponding protein. Therefore a change in the gene expression profile does not necessarily
correlate with an increase in cellular protein levels (Gygi et al., 1999). Proteomics measures the actual level of protein flux within a cell.

Recently a study was conducted using non-gel based proteomics to identify the proteins in the adhesive disc (Hagen et al., 2011). The investigators utilised liquid chromatography coupled with mass spectrometry (LC-MS) to identify proteins back to the *G. duodenalis* genome. In total 57 putative disc-associated proteins were identified (Hagen et al., 2011). This LC-MS based system was able to identify significantly more proteins than the gel base studies described above. The LC-MS workflow is very sensitive as it does not rely on pre-visualisation of proteins on gels so more proteins are traditionally identified using these systems.

Subcellular fractionation to identify the proteome of specific organelles has also been used for *G. duodenalis*. Studies have involved the encystation specific vesicles (ESVs) (Luján et al., 1995, Stefanic et al., 2006), the ribosomes (Shirakura et al., 2001) the 20S proteasome (Emmerlich et al., 1999) and mitosomes (Jedelský et al., 2011). The ESVs of *G. duodenalis* trophozoites are produced after induction of encystation and contain the cyst wall proteins (CWP1, CWP2) as well as other proteins (Luján et al., 1995). Using 2D-PAGE, 16 separate proteins were identified from the ESV/Golgi fraction derived from encysting trophozoites. Proteins found included several alpha giardins, heat shock proteins, metabolic enzymes, components of the 20S proteasome and the trophozoite antigens, GTA-1 and GTA-2 (Stefanic et al., 2006). Neither CWP1 nor CWP2 were identified on the 2D-PAGE gel however their presence was confirmed using Western blots analysis of ESV protein separated using SDS-PAGE. Again the solubilisation techniques needed for 2D-PAGE
analysis were not enough to effectively solubilise the CWPs (Rabilloud, 1996). A recent study found 139 proteins from the mitosomes of *G. duodenalis* using a modified HPLC MS based system (Jedelský *et al.*, 2011).

The susceptibility of isolates to metronidazole has also been linked to the protein complement of the trophozoite. A study of two *G. duodenalis* isolates axenised from a patient prior to and following treatment with metronidazole showed variation in surface proteins and genetic composition (Butcher *et al.*, 1994). Radioiodination of the surface proteins of the trophozoites displayed a major protein band at 70 kDa and several minor bands of 82, 58, 31 and 22 kDa for the metronidazole susceptible isolate. The isolate axenised following treatment with metronidazole had a major protein of 24 kDa with minor bands of 28 and the 70 kDa proteins (Butcher *et al.*, 1994). None of these protein differences were characterised further, although DNA fingerprinting with an M-13 bacteriophage probe did highlight some genetic differences between the isolates (Butcher *et al.*, 1994). Unfortunately genotyping was not performed so it is unclear if the variation is due to a mixed assemblage infection or if there is enough natural variation in a single isolate population to allow for selection of resistant trophozoites to occur. There is *in vitro* based evidence suggestive of there being enough variation within an axenised clonal population of trophozoites to give rise to a drug resistant sub-populations (Argüello-García *et al.*, 2004).

The proteome of the two lifecycle stages has also been compared using 2D-PAGE and the variably expressed proteins identified using mass spectrometry. A total of twenty up regulated proteins were identifies in cysts when compared to trophozoites
(Kim et al., 2009). Not surprisingly, many structural proteins appeared to be up-regulated, as the cyst contains a trophozoite that has partially divided an increase in the amount of structural proteins is expected. Several metabolic enzymes and proteins involved in translation and protein protection were also up-regulated (Kim et al., 2009). The mRNA levels of several of these differences were also examined and found to correlate with the protein data.

1.7.2 Host-Parasite Interactions

Understanding how a parasite interacts with host cells and how the host tries to combat the infection are keys to studying pathogenesis. Proteomics allows us to identify the proteins produced by trophozoites during interaction with host cells as well as those proteins that are targeted by antibodies. One potential toxic protein has been identified in the excretory/secretory products (ESP) of *G. duodenalis* trophozoites (Shant et al., 2002). A 58 kDa protein was identified in the ESP which caused morphological changes to intestinal cell monolayers and caused fluid accumulation in the small intestine of mice (Shant et al., 2002). Edman degradation was performed to identify the protein. However the amino acid sequence derived, ADFVPQVST, did not match any sequence in the *G. duodenalis* genome or the NCBI database (Shant et al., 2002).

The ESP was further characterised by Ringqvist et al. (2008), who elucidated some of the proteins released by trophozoites upon incubation with host cells. Trophozoites were incubated with CaCo-2 cells in serum free medium and proteins extracted from the growth medium and analysed using 2D-PAGE. Five protein spots were identified by mass spectrometry, 3 of which are *G. duodenalis* proteins, with the other two being human derived (Ringqvist et al., 2008). The three *G. duodenalis*
proteins, ornithine carbamoyl transferase (OCT), arginine deiminase (ADI) and enolase, are all metabolic enzymes and were also found in the intestines of mice infected with *G. duodenalis* (Ringqvist *et al.*, 2008). The ADI and OCT may have a role in reducing arginine levels within the small intestine, promoting apoptosis of host cells and reducing the amount of nitric oxide produced in the small intestine (Ringqvist *et al.*, 2008). Overall, the level of ESP increased dramatically when the trophozoites were exposed to the monolayers, indicating that the trophozoites are actively releasing these proteins into the medium in a response to the cells.

The mechanism behind attachment of the trophozoite to a surface is also poorly defined. A recent study using SDS-PAGE found that trophozoite clones of the WB isolate that showed decreased efficiency in adhesion to cell monolayers were missing a 200 kDa protein (Hernandez-Sanchez *et al.*, 2008). When Western blots were performed on SDS-PAGE separated proteins with an antibody produced against the 200 kDa protein, no reaction was seen in the adhesion deficient clones (Hernandez-Sanchez *et al.*, 2008). The clones that lacked the 200 kDa protein also had a reduced ability to establish infection in gerbils, indicating that this protein plays a key role in the attachment of trophozoites to the intestine and possibly in preventing clearing of the parasite from the gut.

Utilising proteomics it is also possible to identify the proteins that the host mounts an immune response against during an infection. These proteins may be used as targets for vaccine or drug development. Separated proteins from both SDS-PAGE and 2D-PAGE can be transferred to membranes and reacted with patient sera to identify those proteins the host antibodies are able to bind with. The major *G.*
*duodenalis* trophozoite proteins recognised by the human immune system have been investigated in two studies using IgG (Palm *et al.*, 2003) and IgA (Tellez *et al.*, 2005) from patients and 2D-PAGE Western blots. Proteins identified included several structural proteins (tubulin, alpha giardins, beta giardin and SALP-1), metabolic enzymes (ADI, OCT, enolase, uridine phosphorylase like protein and fructose-1,6 bisphosphate aldolase), the 2 *Giardia* trophozoite antigens (GTA-1, GTA-2) and a variant surface protein (Palm *et al.*, 2003, Tellez *et al.*, 2005). Three of the metabolic enzymes (ADI, OCT and enolase) were also identified by Ringqvist *et al.* (2008) as being released by trophozoites in the presence of human intestinal cells. Other immunodominant proteins of *G. duodenalis* identified from SDS-PAGE and Western blotting include a 170 and a 155 kDa protein (Torian *et al.*, 1984), a 57 kDa protein (Char *et al.*, 1991), a range of proteins between 30 and 34 kDa (Janoff *et al.*, 1989, Wenman *et al.*, 1993) and an 8 kDa fatty acid binding protein (Hasan *et al.*, 2002).

### 1.7.3 Protein Variation

Protein variation, as detected through SDS-PAGE, was commonly used in the pre-PCR era to study diversity amongst isolates of *G. duodenalis* (Moore *et al.*, 1982, Smith *et al.*, 1982, Nash and Keister, 1985, Wenman *et al.*, 1986, Adam *et al.*, 1988, Aggarwal and Nash, 1988, Capon *et al.*, 1989, Bruderer *et al.*, 1993). Protein variation is separate from genetic variation as many genetic differences have no effect on the protein encoded (silent mutations) and some protein variants can be due to post translational modifications which cannot be determined from the genetic code. There has also been a concerted effort to characterise a group of proteins known as the variant-specific surface proteins or VSPs (Nash *et al.*, 1988). The VSPs are of interest as isolates are able to change the major VSP proteins exposed
on their surface, in an attempt to evade the host immune system. This group of proteins and the variation seen within and between isolates has been reviewed previously (Nash, 2002).

In trying to determine protein variation between isolates using SDS-PAGE to separate proteins, several criteria have been used to group isolates together in an attempt to find variant proteins, including geographical location, host species, clinical signs and more recently assemblage. No significant correlation has been seen according to the geographical location or the symptomatology of the isolates under investigation (Smith, 1985, Capon et al., 1989). There is contradictory evidence as to the correlation between host species and SDS-PAGE protein profiles. A study of isolates from beaver, muskrat, dog, sheep and human sources in Alberta, Canada found no significant difference in SDS-PAGE protein profiles (Wenman et al., 1986). This is contrasted by a study of human, cat, beaver, sheep, dog and muskrat isolates that identified a 35 kDa protein missing from the three cat isolates tested, which was observed in the other 48 isolates tested again utilising SDS-PAGE protein profiles (Capon et al., 1989). Interestingly, these three isolates all belong to assemblage A; it is not known whether the loss of the 35 kDa protein is required for maintaining infection in cats, due to host pressure, or if it is just coincidence. In general, proteins within the area of 30 and 35 kDa have been seen to show more variation between isolates, however individual proteins were not identified (Smith et al., 1982, Wenman et al., 1986, Capon et al., 1989).

A larger degree of variation between isolates of *Giardia duodenalis* has been observed using Western blot analysis. One study was unable to find variation at the
protein level using SDS-PAGE analysis. The same protein preparations were then electrophoresed with antibodies raised in rabbits against the WB isolate with several protein differences seen between isolates (Smith et al., 1982). Further to this, very little variation was detected among 37 human derived isolates by Capon et al. (1989) using SDS-PAGE, whereas Western blotting with rabbit antibodies raised against a clone of the BRIS/83/HEPU/99 isolate revealed significant variation. One isolate, BAH 12, lacked seven antigens that were present in all other isolates. Variation in antigenic response has also been demonstrated for the excretory secretory products, where the cross reactivity of antibodies raised against specific isolates was tested (Nash and Keister, 1985). The groupings based on cross reactivity correlated with those derived from whole genome restriction fragment length polymorphism analysis (Nash and Keister, 1985). Several factors could account for the difference in results based on standard SDS-PAGE or Western blot analysis. Commonly with SDS-PAGE more than one protein will migrate at a given point, this may mask protein variation if the missing protein has the same molecular mass as other proteins. The antibody response could be directed against a post translational modification which may not show variation using SDS-PAGE, again masking the variation. The variation in protein profile could be due to variation in the variable surface proteins (VSPs) being expressed by the different isolates.

With the increasing power of mass spectrometry and bioinformatics, the use of gels to separate proteins for comparative studies is diminishing. Mass spectrometry can be used to produce a protein fingerprint for a given organism which can be compared between isolates or species. This is being used as a method for identifying variation and for the typing of some organisms (Villegas et al., 2006). The protein fingerprints
of *G. duodenalis* and the murine specific species, *Giardia muris*, have been compared to illicit species specific peaks. *Giardia duodenalis* cysts elicited 20 major peaks ranging from 3.5 to 18 kDa, eight of which were species-specific and 16 major peaks for the *G. muris* cysts 4 of which were species specific (Villegas et al., 2006). This technique has the ability to be used for high throughput identification and typing of cysts from semi purified samples, as it has been used for identification of intact micro-organisms in previous studies (Fenselau and Demirev, 2001, Shaw et al., 2004).

There has not been a definitive study to investigate the level of protein variation between the assemblages of *G. duodenalis*. Understanding the level of protein difference between the assemblages may aid in the production of novel typing methods, novel diagnostic markers, identify vaccine candidates and ultimately provide further evidence to support the revision of the taxonomy of the *G. duodenalis* species complex as several distinct species.

### 1.8 Objective and aims

The aim of this study is to identify the level of protein variation between isolates of *Giardia duodenalis* based on their genetic assemblage, symptomatology and geographical region. Gel electrophoresis will be utilised to separate proteins produced by trophozoites of three assemblage A and three assemblage B isolates and mass spectrometry performed to identify these proteins back to the *G. duodenalis* genome database. A selection of the proteins of difference will be further characterised at the genetic level to determine if the variation in protein carries over to the genetic level. Finally, the human infective assemblages will be compared at
the protein level to the only culturable non-human infective isolate, P15c1, from assemblage E. The hypotheses for this study are as follow.

1. Different genetic assemblages of *Giardia duodenalis* will have differences in protein complement.
2. Isolates of *Giardia duodenalis* from symptomatic and asymptomatic infections will show the same protein profiles.
3. Where protein variation exists between the assemblages of *Giardia duodenalis* there will be a corresponding variation in the genetic sequence.
Chapter 2 General Materials and Methods

2.1 Isolates

Isolates for this study were sourced from a reference collection at Murdoch University with the exception of isolate P15c1, which was obtained from Graham Mayrhofer at the University of Adelaide, and was first described by Koudela et al. (1991). Isolates from the Murdoch reference collection were chosen on the basis of assemblage and use in previous studies (Table 2.1). The Murdoch isolates were originally established from positive faecal samples of individuals around Western Australia. The isolates were designated as being symptomatic or asymptomatic based on survey data collected from each of the patients the isolates were collected from. Symptomatic patients displayed any of the symptoms outlined in section 1.4 and included; diarrhoea, nausea, bloating, excessive flatulence, vomiting and steatorrhea.
Table 2.1 *Giardia duodenalis* isolates used in this study

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Assemblage</th>
<th>Location</th>
<th>Clinical Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAH 2c2</td>
<td>A I</td>
<td>Woodanilling</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>BAH 26c11</td>
<td>A II</td>
<td>Rockingham</td>
<td>Symptomatic</td>
</tr>
<tr>
<td>BAH 40c10</td>
<td>A II</td>
<td>Perth</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>BAH 3c3</td>
<td>A II</td>
<td>Perth</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>BAH 34c8</td>
<td>B IV</td>
<td>Perth</td>
<td>Symptomatic</td>
</tr>
<tr>
<td>BAH 12c14</td>
<td>B III</td>
<td>Wyndham</td>
<td>Symptomatic</td>
</tr>
<tr>
<td>BAH 15c1</td>
<td>B IV</td>
<td>Kununnura</td>
<td>Symptomatic</td>
</tr>
<tr>
<td>P15c1</td>
<td>E</td>
<td>Czechoslovakia</td>
<td>NA</td>
</tr>
</tbody>
</table>

2.2 Culture Medium

*Giardia duodenalis* trophozoites were routinely cultured in BS-I-33 medium (Table 2.2) supplemented with 10% v/v new born calf serum (GibCo), which had been heat inactivated for 30 mins at 56°C. The pH of the medium was adjusted to 7.2 using 1 M sodium hydroxide before filter sterilising through a 0.2µm polyethersulfone membrane (Nalgene). The medium was stored in 50mL aliquots at -20°C until needed.
<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration (g/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biosate Peptone</td>
<td>20</td>
</tr>
<tr>
<td>Glucose</td>
<td>10</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>2</td>
</tr>
<tr>
<td>Cysteine hydrochloride</td>
<td>2</td>
</tr>
<tr>
<td>diPotassium hydrogen phosphate</td>
<td>1</td>
</tr>
<tr>
<td>Potassium dihydrogen phosphate</td>
<td>0.6</td>
</tr>
<tr>
<td>Ferric ammonium citrate</td>
<td>0.01</td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>0.2</td>
</tr>
<tr>
<td>Bovine bile</td>
<td>0.5</td>
</tr>
</tbody>
</table>

### 2.3 Culture Maintenance

Stock cultures of trophozoites were maintained in duplicate in 10mL Nunclon® flat bottom screw cap culture tubes. Tubes were filled to the top to minimise oxygen stress. Isolates were subcultured once they had reached confluency. If a culture had not reached confluency within 5 days, the culture medium was removed and fresh warm medium added to replenish spent nutrients. Once confluent the culture was chilled on ice for 10 mins, rolled between the hands and inverted to disperse the trophozoites and 0.5-1mL of cell suspension was used to inoculate a fresh tube of pre-warmed culture medium. Isolates were subcultured every 3-4 days in this manner.
2.4 Cryopreservation

To reduce the risk of losing a cell line, cultures were routinely cryopreserved for storage in liquid nitrogen. Trophozoites were grown to confluency in 10mL flat bottomed tubes then chilled on ice for 30 minutes. The tubes were inverted to dislodge the trophozoites and the culture fluid transferred to a 10mL centrifuge tube. The culture fluid was centrifuged for 5 minutes at 2,000 x g to pellet the trophozoites in order to increase the concentration of cells in each aliquot for cryopreservation, therefore increasing the likelihood of strong growth upon starting a new culture from liquid nitrogen stocks. The supernatant was removed and the cell pellet resuspended in 10mL of fresh medium containing 7.5% v/v dimethyl sulfoxide (Merck). Aliquots of 1.8mL were placed into 2mL screw capped cryopreservation tubes (Nalgene) and transferred to a Nalgene Cryo 1°C Freezing Container™ filled with isopropanol and left overnight at -80°C. The isopropanol allows for the gradual cooling of the tubes at a rate of 1°C per minute. The vials were then transferred to liquid nitrogen canisters for long term storage.

2.5 Mass Culture

To obtain the large number of cells necessary for proteomic analysis, *G. duodenalis* trophozoites were mass cultured in large culture vessels. Glass Schott bottles were filled with borosilicate glass culture tubes and autoclaved to sterilise. Medium was prepared as before except that the medium was filter sterilised directly into the culture vessels.

Isolates were grown as normal in flat bottom 10mL tubes. Once cells were confluent the tubes were placed on ice for 30 minutes to detach the cells. The entire contents of
the tube were then used to inoculate a culture vessel filled with new pre-warmed medium. The culture vessel was incubated at 37°C. The large inoculum ensured confluency was reached as quickly as possible, usually taking around 7 days.

To harvest the trophozoites the culture vessel was immersed in ice for 30 minutes to detach the cells. The culture vessel was then inverted to disperse the cells and the culture liquid decanted into 50mL centrifuge tubes. The tubes were centrifuged at 2,000 x g at 4°C for 5 minutes and the supernatant removed. The cell pellet was washed with cold phosphate buffered saline (PBS) and centrifuged at 2,000 x g at 4°C for 5 minutes. The trophozoites were washed a further two times to remove as much culture medium as possible as the calf serum contains many proteins that could contaminate the proteomic analysis. Pellets were resuspended in approximately 3mL of PBS supplemented with mini-Complete ™ protease inhibitor cocktail (Roche) and stored at -20°C.

2.6 Sample Preparation

Cells were lysed by sonication to release cellular contents including DNA and proteins. Cell suspensions were sonicated on a probe tip Misonix XL2015 set at an amplitude of 5 and a 50% s⁻¹ pulsar duty cycle for three 30 second bursts, cells were kept on ice during the sonication to prevent them overheating and denaturing the proteins. Cell suspensions were examined microscopically for the presence of intact trophozoites. If whole trophozoites were visible the sonication process was repeated. The sonicated trophozoite extracts were transferred to microcentrifuge tubes for storage at -80°C.
Bradford assays were performed to determine the relative protein concentration of the sample to the bovine serum albumin standard used. The Quick Start™ Bradford reagent (BioRad) was used and the assay performed following the manufacturer’s instructions. The trophozoite extracts used were serially diluted with doubling dilutions to ensure that at least one data point that fell within the linear range of the test. All tests were performed in triplicate to ensure reliability of results.
Chapter 3 One Dimensional Polyacrylamide Gel Electrophoresis

3.1 Introduction

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) has been used in several studies to determine the level of protein complexity in *Giardia duodenalis* (Smith et al., 1982, Nash et al., 1983, Wenman et al., 1986, Nash et al., 1988, Capon et al., 1989, Butcher et al., 1994, Luján et al., 1995, Guimaraes et al., 1999). These studies used SDS-PAGE to determine a correlation between the protein banding and factors including geographical region and host of an isolate. Only three studies identified any major differences between isolates that could be correlated to another factor.

One study has identified a difference between isolates based on the species of the host. In a study of 51 isolates sourced from a variety of hosts, three cat isolates all lacked a protein of 35 kDa in size that was present in all other samples tested (Capon et al., 1989). All three isolates were later genotyped to the human infective assemblage A (Hopkins et al., 1999) It is however highly possible that the protein is not being expressed as it is needed by the parasite when infecting feline hosts. This could be related to direct interaction with host cells or could be due to different metabolic conditions experienced in different hosts. No further characterisation of the protein was undertaken as to its role or position in the trophozoite.

The immune response to proteins has also been used to separate groups of *G. duodenalis*. The excretory secretory products of *G. duodenalis* trophozoites were
examined by Nash and Keister (1985) with no major differences seen between isolates. When the same protein gels were blotted with a polyclonal antibody produced against the type strain WB, which belongs to assemblage A, two clear groups were delineated (Nash and Keister, 1985). Genetic characterisation using whole genome restriction fragment length polymorphism demonstrated significant genetic variation between these isolates as well.

The aim of this study was to determine if SDS-PAGE could be used to identify protein differences between various genetic assemblages of *Giardia duodenalis*. To achieve this, total trophozoite proteins of isolates from specific assemblages (namely A, B and E) were electrophoresed in varying concentrations of polyacrylamide gel. The protein profiles were then visualised and compared to identify specific proteins. Isolates also varied according to clinical symptoms and geographical location within Western Australia. The protein profiles were examined to determine if there are proteins unique to the symptomatic/asymptomatic isolates and general geographical regions (Perth or Far North W.A.).

### 3.2 Materials and methods

#### 3.2.1 Sample preparation

For standard SDS-PAGE, a total of 50 µg of total trophozoite protein (TTP) was adjusted to 40 µL with 1D sample buffer 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 100 mM tris, 69 mM SDS and bromophenol blue to colour; (pH 6.8) and incubated at 100°C for 5 minutes to solubilize the proteins. The β-mercaptoethanol serves to reduce disulphide binds within the protein and the heat...
denatures the proteins, the proteins are unlikely to refold after the heat denature step. Samples were allowed to cool then loaded into gels. Initial experiments were performed using 50 µg of protein in 20 µL of buffer in a smaller gel system. When the experiment was carried out using the larger gel equipment the amount of protein was kept at 50 µg and the amount of buffer was increased. This was done to overcome distortion of the gels seen in the preliminary studies.

3.2.2 One dimensional electrophoresis

For the initial proof of concept experiments, 12% polyacrylamide Criterion XT™ gels were run using the MOPS buffer system (BioRad, Hercules, U.S.A.). The pre-poured gel cassette was placed into the tank and the tank plus top reservoir were filled with 1X MOPS running buffer (BioRad). All 20 µL of sample was loaded into the well with broad range molecular mass markers run on either end of the gel (BioRad). The gels were run at 200 V for 1 hour or until the bromophenol blue dye front had run off the bottom of the gel.

After the initial gel runs using the criterion system, the method was scaled up to use 16 cm long polyacrylamide gels. Gel cassettes were assembled as per the manufacturers’ instructions with 1D gel spacers and clamps (Figure 3.1) (BioRad). The gel cassette assembly was locked into a pouring stand and the seals checked by filling the cassette with methanol, if any leaking occurred the cassette was taken apart and reassembled and tested again. The methanol was then decanted and the gel cassette dried, inverted, at 37°C. The resolving gel solution was prepared as per Table 3.1. The different percentage polyacrylamide used allows for better resolution in different molecular mass regions. The acrylamide, buffer and water were added to a Buckner flask mixed and degassed by vacuum for 30 minutes. Once degassed the
sodium thiosulphate TEMED and ammonium persulphate were added to the gel solution and mixed. The solution was immediately transferred to the dried gel cassette using a disposable 25 mL serological pipette. The cassette was filled to within 5 cm of the top of the short plate and the gel surface overlayed with a 0.1% w/v SDS solution that was sprayed onto the long plate of the gel cassette and allowed to run down onto the gel surface, this ensured a flat gel surface and effective polymerisation. Gels were left to polymerise for 2 hours and the excess buffer and SDS was poured off. The top of the gel was sealed with plastic wrap and the gels left overnight at 4°C to completely polymerise.

<table>
<thead>
<tr>
<th>Table 3.1 Polyacrylamide Gel Formulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagents</td>
</tr>
<tr>
<td>Acrylamide stock</td>
</tr>
<tr>
<td>1.9 M Tris pH 8.8</td>
</tr>
<tr>
<td>hp H₂O</td>
</tr>
</tbody>
</table>

The following day the stacking gel was made by mixing 5.5 mL of high pure water (hp H₂O), 1 mL of acrylamide solution (40% acrylamide, 1% w/v piperazine diacrylamide) and 1.5 mL of 5X stacking buffer (1 M tris, 3% w/v SDS; pH 6.9) in a Buckner flask and degassing by vacuum for 15 minutes. Once degassed 75 µL of 10% ammonium persulphate and 7.5 µL of TEMED were added to the gel mix and it was transferred to the gel cassette using a glass pasteur pipette. The gel comb was placed into the stacking gel leaving approximately one centimetre between the bottom of the wells and the start of the resolving gel. The stacking gel was the left to set for 30 minutes and the comb was removed.
The gel cassette was removed from the pouring stand and clipped into a Protean II xi™ central cooling core (BioRad). A blank cassette assembly (prepared as for the gel except the spacers are removed) was clipped into the other side of the cooling core and the cooling core filled with water. The cooling pack was placed into the Protean II xi tank and connected to a circulating refrigeration unit, which was switched on and set to 14°C to maintain the water circulating through the cooling core at 14°C. The gel tank was half filled and the top reservoir filled with running buffer (192 mM Glycine, 25 mM Tris, 1% w/v SDS). The samples were loaded into the wells, with broad range molecular mass markers loaded at either end. Gels were run at 100 volts until the bromophenol blue dye front had run off the gel, approximately 18-20 hours.
3.2.3 Gel Staining

After the gels had finished running they were removed from the gel cassette and in the case of the large format gels, the stacking gel was cut away from the resolving gel using a scalpel. The gels were placed into a pyrex container and a modified Coomassie G-250 stain was added to cover the gels and placed on an orbital shacking platform overnight to stain (Candiano et al., 2004). The stain was poured off the gel and replaced with 10% v/v acetic acid solution, to decrease the background, and returned to the orbital shaker. The gels were destained until the background was significantly reduced, if the background was very high the acetic acid solution was replaced with a fresh solution after an hour of destaining.

3.2.4 Gel Imaging

Gels were imaged with a ProXpress™ CCD camera system (Perkin Elmer, Waltham, U.S.A.). The glass cassette used with the camera was washed with methanol and dried with lint free tissues to remove any dust that may affect the quality of the gel image. The camera was calibrated before every use by taking a flat field image of a clean glass cassette, using bottom illumination, with the UV diffuser plate in position, excitation of 0/0 and emission of 680 nm. The flat field image was exposed for 1300 milliseconds per frame. The gel was loaded into the glass cassette and all the air bubbles removed from under the gel surface. The gel image was taken as per the flat field except that the exposure time was increased to 1500 milliseconds.
3.2.5 Comparison of Gel Images

Due to a lack of software available for the comparison of SDS-PAGE gel images, all comparisons were performed by eye. Every image was viewed as both a hard and soft copy to determine proteins of difference between the isolates. A band was considered unique for an assemblage if it appeared in 2 out of three isolates on a gel. All unique bands were cross checked on replicate gels. Areas of difference were noted on hardcopies and used as a template for excision from gels (See Chapter 5).

Results 3.3

3.3.1 Preliminary SDS-PAGE Analysis

This experiment was designed as a proof of concept to show that SDS-PAGE was an effective tool for identifying proteins of difference between the assemblages, as previous studies found no repeatable differences between isolates. Figure 3.2 demonstrates the reproducibility within assemblages and some of the variation between the assemblages. Although the resolution is poor, the technique was considered useful for the discrimination of assemblage specific proteins.
Figure 3.2 Trial SDS-PAGE for Protein Separation
Lane 1 BAH 2c2 (A), Lane 2 BAH 26c11 (A), Lane 3 BAH 34c8 (B), Lane 4 BAH 12c14 (B), Lane 5 molecular weight marker in kDa.
3.3.2 Comparative SDS-PAGE Analysis

The comparative analysis was firstly performed to compare assemblages A and B of *G. duodenalis*. One dimensional gel analysis using 12.5% polyacrylamide gels showed highly conserved protein banding within the assemblages and several differences between assemblages (Figure 3.3). The differences highlighted in Figure 3.3 are investigated further in Chapters 5, 6 and 7. The resolution compared to the preparative gel is greatly increased. Proteins are visible over an extended size range, form over 200 kDa to 7.5 kDa, with the highest degree of resolution in the region from 20 to 60 kDa. The lower molecular weight bands, those below 25 kDa, are at much lower intensity than those in the higher mass range.

Different percentages of acrylamide, 10% and 15%, were trialled to increase the resolution. The 10% acrylamide gels allowed for a clearer delineation of proteins between 30 and 100 kDa (see Figure 3.5 for an example). A 15% acrylamide gel was also trialled to allow for visualisation of the lower molecular weight proteins, however they were not high enough in abundance to be visualised effectively (data not shown). The concentration of protein loaded was doubled to try and overcome this. Unfortunately, this failed to increase the intensity of these bands as the gel was overloaded with higher molecular weight proteins that prevented efficient running of the gel.
Figure 3.3 12.5% PAGE of TTP from *Giardia duodenalis* Assemblages B and A
Lane 1 BAH 15c1, Lane 2 BAH 12c14, Lane 3 BAH 34c8, Lane 4 BAH 2c2, Lane 5 BAH 26c11, Lane 6 BAH 40c10. i) Whole 12.5% gel highlighting major protein difference (A). ii) Close up of rectangle in part (i), B and C mark an area of difference.

Due to the homogeneity within assemblages, subsequent experiments were conducted using only one isolate per assemblage. An acrylamide percentage of 10% was selected to separate proteins as the majority of high abundance proteins are above 20kDa and a 10% PAGE gel gives good resolution between 20kDa and 100kDa. All gels were repeated at least three times, with the figures shown representing the most commonly seen protein profile.
Figure 3.4 10% PAGE of TTP from *Giardia duodenalis* Assemblages A and B
Lane 1 Molecular mass marker (size given in kDa), Lane 2 BAH 2c2 (A), Lane 3 BAH 34c8 (B). The numbers indicate areas of difference between the gels.
The 10% acrylamide gel gave good separation between 31 and 100 kDa (Fig. 3.4). There is a large degree of protein variation seen throughout the gel. Marked on Figure 3.4 are the positions of six possible protein variants. Band 1 appears to be a size variation between the assemblages A and B, with the assemblage A isolate having a band below 200kDa and the assemblage B isolate a band above 200kDa. Regions 2, 3, 5 and 6 all have single bands in the assemblage A isolate and apparent doublets in the assemblage B isolate. Band 4 is a potential size variant with the assemblage B isolate having a band with a higher molecular mass when compared to assemblage A.

### 3.4 Discussion

#### 3.4.1 Intra-assemblage Variation

There was no variation in protein banding between isolates of the same assemblage for the SDS-PAGE analysis as seen in Figure 3.3. This lack of intra-assemblage variation is interesting due to the amount of heterogeneity seen at the genomic level especially for assemblage B isolates (Caccio et al., 2008). From genetic analysis of loci such as beta giardin the assemblage B isolates are delineated into multiple subtypes (Lalle et al., 2005), however examination of the derived amino acid sequence greatly reduces the number of subtypes present, indicating that many of the nucleotide substitutions are not having an impact on the amino acid sequence (Wielinga and Thompson, 2007).
3.4.2 Inter-assemblage variation

In contrast to the homogeneity seen within the assemblages the level of protein variation between the assemblages was quite large. There were few presence/absence variants; however there were what appear to be size variants of several proteins (Figure 3.4). A variation in size implies a difference in the amino acid sequence, and therefore the genetic sequence between the assemblages. According to the genetic data, assemblages A and B are quite distinct with the level of genetic distance between them greater than that seen between other genera of protozoa (Mayrhofer et al., 1995a). The number of visible protein differences from SDS-PAGE correlates with this.

3.4.3 Comparison with published data

The level of homogeneity within the assemblages and heterogeneity between them also gives some insight into previous studies. With the exception of Capon et al. (1989), studies of protein variation in G. duodenalis using SDS-PAGE have not discovered major protein differences (Moore et al., 1982, Smith et al., 1982, Nash and Keister, 1985, Wenman et al., 1986). As these studies were not using genetically characterised isolates it is likely they were comparing isolates from the same assemblage, based on the large amount of protein variation we observed between assemblages A and B. The axenisation process itself has been shown to be biased for the isolation of assemblage A strains (Andrews et al., 1992). The one human isolate, BAH 12c14, that yielded a unique banding pattern in the study by Capon et al. (1989) was also used in our study and gave the same banding pattern as all other assemblage B isolates. It is possible that BAH 12c14 was the only assemblage B isolate used by Capon et al. (1989), explaining why it gave a distinct banding pattern in their study.
Similar to previous studies no difference was found between isolates that caused symptomatic or asymptomatic infections (Smith et al., 1982). Two studies have used human infection models to try and unravel the variation in clinical symptoms (Rendtorff, 1954, Nash et al., 1987). A study by Rendtorff found that out of 15 individuals infected with cysts derived from the same source only 9 displayed clinical symptoms, defined as in increase in frequency of bowel movements and looser stools. With those displaying clinical symptoms there were individuals who were more severely affected than others. A later study confirmed these results with 3 out of 5 individuals infected with trophozoites of a selected strain of *G. duodenalis* displaying clinical symptoms (Nash et al., 1987). This indicates that the parasite on its own is not the cause of disease and that the host’s response to the infection must play some role. A study by Scott et al. in 2004 found in mouse models that the hosts own CD8+ T lymphocytes migrating to the small intestine during infection with *G. duodenalis* reduced the surface area of the small intestine and the level of disaccharidase activity (Scott et al., 2004). This indicates that differences in an individual’s T cell response to infection with *G. duodenalis* could be the cause of variation in clinical presentation when individuals are infected with known pathogenic *G. duodenalis* strains.

Overall SDS-PAGE is an effective method for comparing proteins produced by the different assemblages of *Giardia duodenalis*, with several protein variants highlighted in all three assemblages tested. There was no protein variation seen within assemblages or between isolates with different clinical presentations, confirming previous reports. This is the first time that a difference at the protein
level has been linked directly to the assemblage of the *G. duodenalis* isolates under investigation
Chapter 4 Two Dimensional Polyacrylamide Gel Electrophoresis

4.1 Introduction

Although SDS-PAGE allows for the separation of proteins on the basis of size many bands contain several proteins all with the same molecular mass. The use of two dimensional polyacrylamide gel electrophoresis (2D-PAGE) allows for the resolution of single protein spots, increasing the level of discrimination when comparing protein profiles. This allows for a more complex view of the proteome. The process is based around the separation of proteins according to two physical properties, their isoelectric point (pI) and their size. The pI of a protein is defined as the pH at which the protein has no net charge, and is determined by the amino acid sequence of the protein and its structure. The technique of isoelectric focusing (IEF) utilises the pI to separate proteins. Samples are incubated with an acrylamide strip with an immobilized pH gradient running along it. The samples are then placed under the influence of an electric field; proteins will therefore migrate along the acrylamide strip until they reach the position in the acrylamide strip where the pH corresponds to their pI. At this point they have no net charge and are therefore no longer under the influence of an electric field so they stop migrating.

The acrylamide strip from isoelectric focusing is then loaded onto a standard polyacrylamide gel and the proteins can be separated on the basis of size, with smaller proteins migrating through the gel at a faster rate than larger proteins. This allows for a much finer level of resolution when compared to SDS-PAGE as proteins with similar molecular mass will be separated according to pI. It also allows for the
identification of isoelectric variants of the same protein, a common phenomenon caused by variation in the post translation modification of proteins.

Two dimensional polyacrylamide gel electrophoresis has been used in several studies of *G. duodenalis*. The first were conducted to investigate the proteins of the cytoskeleton and adhesive disc of trophozoites with nine protein groups described (Peattie *et al.*, 1989). This work was repeated in 2005 by Palm *et al.* who identified seven separate proteins back to *G. duodenalis* genes using mass spectrometry. The proteins within the encystation specific vesicles were also investigated using 2D-PAGE (Stefanic *et al.*, 2006), as were the trophozoite proteins recognised by patient IgG (Palm *et al.*, 2003) and IgA (Tellez *et al.*, 2005). A recent study used 2D-PAGE to identify the proteins up regulated during encystation (Kim *et al.*, 2009). A total of 14 proteins were found to be up-regulated with the majority being structural proteins. There is yet to be a detailed description of the proteome of *G. duodenalis* using 2D-PAGE.

The major aim of this series of experiments was to determine if 2D-PAGE is an effective tool for the identification of assemblage specific proteins. Several sub aims included the optimisation of protein solubilisation, the optimisation of the protein concentration required for visualisation, determination of the reproducibility of the 2D-PAGE system and comparison of the resolution of 2D-PAGE with SDS-PAGE.
4.2 Material and Methods

4.2.1 Sample preparation

Two protein solubilisation methods were trialled, the first is a standard protein solubilisation technique for proteomic analysis, termed multiple surfactant solution (MSS: 40 mM tris, 65 mM SB 3-10, 32.5 mM CHAPS, 5 M urea, 2 M thiourea 0.5% (v/v) ampholytes, 0.05% (v/v) tributyl phosphine) and the second termed PS2 buffer, developed by Palm et al. (2003) for use with *G. duodenalis* (PS2: 9.9 M Urea, 4% v/v Igepal CA630 and 2% v/v Pharmalytes pH 3-10), both techniques were tested with multiple isolates. To determine the best solubilisation method to use, 500 µg of total trophozoite protein was made up to 400 µL with solubilisation buffer and run through the 2D-PAGE workflow described below. The optimal loading concentration was determined by preparing samples with 300 µg, 500 µg, 800 µg and 1 mg of total trophozoite protein with the MSS sample preparation method and running them through the 2D-PAGE workflow described below. For the 2D-PAGE analysis sample BAH 2c2 was used for assemblage A and sample BAH 34c8 was used for assemblage B.

4.2.2 Isoelectric Focusing

The first stage of 2D-PAGE requires the isoelectric focusing of proteins. The isoelectric focusing uses immobilised pH gradient (IPG) strips which are dehydrated strips of acrylamide that contain a pH gradient running along them, this pH gradient does not change when an electric current is applied to the gel. The acrylamide is rehydrated with the sample in specific buffers (see above). An 18cm pH 3-10 IPG strip (GE Life Science) was placed into a 2 mL serological pipette, cut down to 1 cm longer than the strip, with the gel side down. One end of the tube was sealed with
parafilm, the prepared protein sample was pipetted into the tube between the gel surface and the edge of the tube and the open end of the tube was sealed with parafilm. The sample was checked for the presence of bubbles on the gel surface if present the tube was placed on its end and tapped to remove the bubbles from the sample then placed down again so that the gel surface was facing down with the sample underneath. The presence of air bubbles on the gel surface can cause uneven distribution of the sample in the gel which may lead to increased resistance in the strip and the strip to burn. The IPG strips were incubated with the sample overnight at room temperature for the gel to rehydrate, in some cases it took up to 24 hours for the entire sample to be adsorbed into the gel.

Once the entire sample had been incorporated into the IPG strip the Multiphor II system (GE Life Sciences) was set up for the isoelectric focusing. The Multiphor tank was placed on a flat surface and the cooling plate, filled with water and connected to a circulating refrigeration unit, was placed on top. A spirit level was then used to determine whether the unit was completely flat. A small amount of paraffin oil was placed onto the cooling plate at one end and the gel tray was placed onto the cooling plate with a rocking motion starting at the end with the oil, so as to ensure even coverage of the oil between the gel tray and the cooling plate. Any air bubbles were pushed out from the between the cooling plate and the gel tray, the oil ensures good thermal conduction between the cooling plate and the gel tray, air bubbles can cause localised areas of heating and even burning on the gel surface. The electrodes on the gel tray were plugged into the sockets attached to the Multiphor tank. More oil was placed into the gel tray at one end and the well sheet was inserted into the gel tray, again starting at the end with the oil to ensure even
distribution of the oil under the well sheet. As before, any air bubbles are removed to ensure effective heat transfer.

The rehydrated IPG strip containing the sample was placed into one of the wells with the gel surface facing upwards, making sure to observe polarity. Two filter paper strips approximately 2 cm by 5 cm were moistened with hp H₂O and placed, one at either end of the strip so that they covered 0.25-0.5 cm of the strip. The filter paper acts as a wick to absorb any salts that migrate out of the sample as these salts can cause distortions at the pH extremes of the gel. Electrodes were placed over the wicks making sure that they were also over the IPG strip. The tray was then ¾ filled with paraffin oil and the lid positioned on top making sure that all the electrodes were connected. The lid electrodes were connected to a PowerPac 3000 power supply unit (BioRad) and the gels were run at 300 V for 30 min, 1000 V for 30 min and 3,500 V for 18 hr. Once the run had finished IPG strips were removed from the Multiphor II, drained of oil and placed in a plastic Petri dish. If not used straight away the Petri dish containing the IPG strip was sealed with parafilm and stored at -20°C.

4.2.3 SDS Polyacrylamide Gel Electrophoresis

Polyacrylamide gels were set up as for SDS-PAGE (Chapter 3, section 3.2.2) for the 16 cm gels except that 2D gel spacers and gel clamps were used, both 10% and 12.5% acrylamide gels were used. Instead of filling the gel to within 5cm of the top of the short plate, the gel solution was added to within 0.5 cm of the top of the short glass plate. As before gels were allowed to set overnight at +4°C.
The first dimension IEF strips were incubated in re-equilibration buffer (20% v/v 1.9 M tris pH 8.8, 2.6% v/v acrylamide, 20% v/v glycerol, 6 M urea, 2% w/v SDS, 0.25% v/v tributylphosphine) on a rocking platform for 30 minutes. This step reduces and alkylates the protein for the 2nd dimension and the inclusion of SDS gives a net negative charge to all proteins so they can be resolved according to size. After re-equilibration the strips were rinsed in cathode buffer (192 mM glycine, 2.5 mM tris, 1% w/v SDS in hp H₂O) and the plastic ends of the strip were removed. A thin piece of filter paper, approximately 2 mm wide, was cut and 10 µL of All Blue Precision Plus Protein™ standard (BioRad) was allowed to absorb into the filter paper which was inserted on the far left side of a pre-prepared gel. The rinsed IPG strip was then loaded in to the gel with the positive end of the strip next to the piece of filter paper containing the mass standard. The IPG strip was then overlayed with molten 0.5% w/v agarose (Invitrogen, U.S.A.) made up in cathode buffer, coloured with bromophenol blue, so as to measure the progression of electrophoresis.

Once the agarose had set the gel cassette was clipped into the cooling core as for the SDS-PAGE, using 2D gel gaskets in the cooling core to ensure an effective seal and prevent cathode buffer leakage. The cooling core was filled with water and the seals of the top reservoir were checked by filling the reservoir with hp H₂O, which was then discarded. The cooling core with the gels attached was placed into the Protean Xi running tank (BioRad), which was one third filled with anode buffer (740 mM tris pH 8.8). The cooling core was connected to a refrigerated water cycler set to 14°C. The top reservoir was filled with anode buffer and the gels were electrophoresed at 25 mA per gel for 5 hours or until the bromophenol blue dye front had run off the gel.
4.2.4 Protein Visualisation

Once finished, gel cassettes were removed from the cooling core and the gels removed from the cassettes. The IPG strip was cut from the acrylamide gel using a scalpel blade and placed into a Pyrex® dish containing approximately 250 mL of blue silver modified Coomassie stain, to check if there had been efficient transfer from the first dimension gel into the second dimension (Candiano et al., 2004). The gel was then transferred to the dish containing the protein stain and left to stain overnight on a rocking platform.

The stain was poured off the gel and replaced with approximately 250 mL of 10% v/v acetic acid, in order to reduce the level of background stain. The gel was left to destain for several hours until the level of background stain was reduced significantly. If the background of the gel was heavily stained then the acetic acid solution was changed to avoid it becoming saturated and preventing further destaining. Once sufficiently destained the gel was imaged using the ProXpress system (Perkin Elmer) and gel images compared according to the method for SDS-PAGE.

4.3 Results

4.3.1 Protein Solubilisation Optimisation

Two protein solubilisation methods were trialled to determine the method that gave the best resolution and spread of proteins. Both methods gave reproducible 2D gel
images for *G. duodenalis* trophozoites. The two methods showed the same dominant spots on the gel, however a larger spread of proteins was seen with the MSS buffer with less streaking in the gel (Figure 4.1 and 4.2). The MSS sample preparation also gave better resolution of proteins within the higher molecular weight region. The MSS sample preparation procedure was therefore used for all subsequent experiments.

### 4.3.2 Protein Concentration Optimisation

In order to determine the best protein concentration for the 2D-PAGE four separate protein concentrations were trialled ranging from 300 µg to 1 mg. Figure 4.3 depicts representative gels for the four concentrations used. The 300 µg of TTP showed good resolution however the protein concentration was not sufficient, by increasing the concentration to 500 µg a larger number of spots became visible with the same level of resolution for the major spots. At 800 µg of protein the concentration starts to affect the proteins migration in the IEF. There is a clear reduction in the resolution around the pI of the dominant proteins on the gel, (Fig 4.3 panels (iii) and (iv)), due to the precipitation of these proteins in the first dimension and insufficient transferral into the second dimension gel. In particular the series of dominant spots in the bottom right quadrant of the other three protein concentrations are absent in both the 800 µg and 1 mg sample.
Figure 4.1 2D-PAGE of *Giardia duodenalis* isolate BAH 34c8 TTP extracted with MSS, pH 3-10 IPG strip and 12.5% polyacrylamide gel
4.3.3 Gel Reproducibility

In order to test reproducibility all gelws were run in at least triplicate, replicate gels were examined to detect any variation. Gels that showed high degree of variation from the norm were discarded from further analysis. These aberrant protein maps were mainly due to failures in the IEF procedure as opposed to problems with the sample. Figure 4.4 shows three representative gels of an assemblage A isolate run over a two week period. There is very little variation between the three gels. The IEF and 2D gels were prepared and run separately.
Figure 4.3 Protein Concentration Optimisation.
i) BAH 34c8 300 µg, ii) BAH 34c8 500 µg, iii) BAH 34c8 800 µg, iv) BAH 34c8 1mg. TTP was extracted using MSS first dimension is a pH 3-10 IPG strips with the second dimension run on a 10% polyacrylamide gel.
Figure 4.4 Reproducibility of 2D Gel Images
BAH 40c10 (A) gel images produced over several weeks. Gels loaded with 500 µg of BAH 40c10 TTP. First dimension run on pH 3-10 IPG strips with second dimension on 12.5% polyacrylamide gels.
4.3.3 Inter Assemblage Variation

Overall the 2D-PAGE showed a high degree of similarity between the assemblages however some distinct protein variants were evident. Both gels had dominant spots in the higher pI range around 40 kDa labelled area A on the gels. In the assemblage B gel the protein spots have clearly shifted to the more basic end of the pH range (to the right). The second area of difference, marked at B, is for protein around 31 kDa in size. In the assemblage A isolate there is a single spot at this position, whereas in the assemblage B isolate there are three clearly defined spots. Finally in assemblage A there are four spots at point C which do not appear in assemblage B.

Figure 4.5 2D-PAGE of isolate BAH 2c2 (A)

500µg of TTP run in the first dimension on 18 cm pH 3-10 linear IPG strips and in the second dimension on a 12.5% polyacrylamide gel. Labelled circles indicate areas of difference between assemblages. Mass on side of gel is in kDa.
**Figure 4.6 2D-PAGE of isolate BAH 34c8 (B)**

500µg of TTP run in the first dimension on 18 cm pH 3-10 linear IPG strips and in the second dimension on a 12.5% polyacrylamide gel. Labelled circles indicate areas of difference between assemblages. Mass on side of gel is in kDa.

### 4.4 Discussion

#### 4.4.1 Comparison of protein resolution between 2D-PAGE and SDS-PAGE

Overall 2D-PAGE proved effective at resolving proteins of *G. duodenalis* trophozoites. The largest amount of resolution was seen within the pI range of 4 and 8 and mass range between 10 and 60 kDa. The overall resolving power was reduced
compared to the SDS-PAGE which showed a spread of proteins from over 200 kDa to 10 kDa. The reduction in resolution is likely due to the difference in the protein solubilisation methods used for the two techniques. For SDS-PAGE, proteins are solubilised by boiling the samples in 2% SDS, which collapses the proteins’ structure and coats the protein in a negative charge. This negative charge would interfere with the isoelectric focusing which relies on the proteins’ change in charge distribution with differing pH. For 2D-PAGE protein solubilisation it is therefore necessary to use zwitterionic detergents that do not interfere with the native charge of the proteins. These detergents are not as effective at solubilising many of the larger structural proteins of *G. duodenalis* and hence the reduction in the number and spread of proteins. Boiling samples, which helps to disrupt intramolecular bonds is also seldom used in 2D-PAGE as removing structure will alter the proteins pI. This has been seen in other investigations of *G. duodenalis*, when comparing the SDS-PAGE gels of Crossley and Holberton (1983) with 2D-PAGE gels of Peattie *et al.* (1989) and Palm *et al.* (2005) there is a significant reduction in the molecular mass range of the proteins, consistent with that seen in this study (Crossley and Holberton, 1983, Peattie *et al*., 1989, Palm *et al*., 2005).

For the proteins that are soluble there is a significant increase in resolution when compared with SDS-PAGE. In the region around 30 kDa which on the SDS-PAGE was very convoluted and hard to differentiate clear bands, there are a series of different protein spots across the whole pI range. Some of these proteins are probably the giardins, based on migration compared to other studies (Peattie *et al*., 1989, Palm *et al*., 2005). Although the overall resolving power of 2D-PAGE is
reduced, the increased resolution within molecular mass range makes it a useful tool for visualising proteins.

**4.4.2 Optimisation of protein solubilisation**

As there are inherent issues with the solubilisation of proteins for 2D-PAGE, as described above, it is important to optimise the solubilisation process to ensure that the maximum numbers of protein spots are resolved. Two different protein solubilisation techniques were used, one taken from previously published work on 2D-PAGE of *G. duodenalis* termed PS-2 (Palm *et al.*, 2003) and the MSS method developed in house. The PS-2 method utilises a very high urea concentration (9.9 M) in order to solubilise the proteins with the addition of a single zwitterionic detergent, Igepal CA630 (at 4% v/v) which is chemically indistinguishable from Nonidet P-40. The MSS method was developed at the Australian Proteomic Analysis Facility, it utilises both urea (at 5 M) and thiourea (at 2 M), two zwitterionic detergents derived from the sulfobetaines (SB 3-10 and CHAPS) and tributyl phosphine as a reducing agent.

The MSS method gave better resolution of proteins in the higher molecular mass region of the gel as well as at lower molecular mass range. Both methods showed similar protein patterns within the 30 kDa range, where the majority of resolved spots occur. The difference in resolution can be attributed to different extraction protocols. The high urea concentration in the PS-2 protocol is effective at disrupting covalent bonds and weakening hydrophobic bonding, however the addition of thiourea in the MSS method has been show to improve the solubilisation of
membrane proteins allowing for an increase in total protein solubilisation (Rabilloud, 1996, Gorg et al., 2004).

The detergent used between the two methods also differs. The PS-2 method utilises Igepal CA630 which is chemically indistinct from Nonidet P-40. The MSS buffer contains both SB 3-10 and CHAPS which are sulfobetaine derivative detergents. In a review of protein solubilisation techniques by Rabilloud (1996) these detergents are classed as the most effective detergents for protein solubilisation. SB 3-10 is not soluble in the high concentration of urea normally used for sample preparation; however the ability of SB 3-10 to solubilize outweighs the reduction in urea concentration (Herbert, 1999). The MSS protocol also includes a reducing agent in the form of tributyl phosphine, which further helps to denature the protein by reducing the disulphide bridges. The PS-2 protocol does not include any reducing agent, however the standard chemical used is dithiothreitol (DTT). The tributyl phosphine used in the MSS protocol has the advantage in that it is an uncharged molecule so that it stays in the IPG strip during IEF whereas DTT is a charged molecule and will migrate out during IEF resulting in a loss of protein solubility.

4.4.3 Optimisation of the protein concentration required for visualisation

The concentration of the protein loaded into the IPG strip for IEF has an impact on the resolution of the proteins within the sample. As stated previously the pI of a protein is the pH at which the protein has no net charge. At this point a protein will be least soluble as it can no longer interact with polar solvents. If the concentration of the protein in the solution is relatively low then it will still be soluble at its pI. However at higher concentrations the protein molecules will aggregate together and
precipitate out of solution preventing transfer into the second dimension. These precipitated proteins will also interfere with the transfer of proteins with a similar pI from the first dimension into the second dimension. It is therefore necessary to optimise the protein concentration added to the IPG strip to prevent precipitation of the more abundant proteins.

Figure 4.3 demonstrates the precipitation of proteins at their pI. There are blank areas in panels (iii) and (iv), especially in the bottom right quadrant where there are three major proteins seen in panels (i) and (ii). These proteins are seen in high abundance in panels (i) and (ii) of Figure 4.3, making it highly likely that they have precipitated in the first dimension of the higher protein concentration gels. The only way to overcome this is to use lower protein concentrations. From Figure 4.3 a concentration of 500 mg of TTP (panel (ii)) gives the highest amount of resolution without detrimental effects, although some protein spots evident in the 800 µg and 1 mg gels are lost. This could be compensated by using lower protein concentrations and staining with a silver based protein stain, however the likelihood of identifying spots using silver stained gels is decreased due to lower amounts of protein in the sample.

4.4.4 Reproducibility of the 2D-PAGE system

The reproducibility of the system needed to be tested in order to determine if samples prepared at different times and gels run on different days could still be compared. Figure 4.4 depicts three gels run separately over a three week period. The sample for the gel displayed in panel (i) was prepared separately to those prepared for panels (ii) and (iii). There is minor variation between the gels, but this appears to be due to variation in protein loading causing some of the lower abundance proteins
in panels (i) and (iii) to be absent in panel (ii), all of the dominant proteins are present in all three gels. Overall the system appears to be robust and reproducible allowing for comparison between gels produced independently of each other.

4.4.5 Inter-assemblage protein variation

The major aim of this series of experiments was to determine protein spots of difference between assemblages A and B of *G. duodenalis* and this was achieved. Potentially assemblage-specific spots have been identified for each of the assemblages; however their status as assemblage-specific proteins cannot yet be confirmed. Assemblage A showed a series of four spots at around 31 kDa in size (Figure 4.5 C) that were absent in assemblage B. These are dominant spots in the assemblage A gels so their absence in assemblage B may be significant.

The assemblage B isolate had two protein spots that were unique and an area that showed variation when compared to assemblage A. The unique proteins can be seen at point B in Figure 4.6. They are approximately 31 kDa in size with a pI of 5.5. In this area (Figure 4.5 B) the assemblage A isolate has only one protein spot whereas the assemblage B isolate has three proteins. The possible variant proteins are seen in the area designate as A on Figure 4.6. These protein spots are at a higher pI when compared to the same region of assemblage A (Figure 4.5 A).

Overall the level of variation seen with 2D-PAGE was less than that observed for the SDS-PAGE analysis, although this is likely due to the difference in solubilisation techniques used affecting the overall number of protein spots visible. The advantage of 2D-PAGE is the visualisation of pI variants of proteins. Many of the proteins on the protein gels exist as chains of spots which are likely to be pI variants of the one
protein caused by variation in the post translation modification of the proteins. The difference highlighted at point A in Figures 4.5 and 4.6 is most likely due to a difference in the post translational modification status of the protein. As with the SDS-PAGE these protein differences can only be seen as provisional and require further characterisation to identify them and to ascertain whether these differences are due to post translational modifications or variation at the genetic level.
Chapter 5 Identification of Proteins

5.1 Introduction

After visualising a protein band or spot on a gel it is necessary to identify that protein back to the genome of the organism of interest so that the biological importance of the protein can be determined. Prior to genomic information this was performed using co-migration studies with known proteins and by using anti-sera raised against known proteins, for example beta tubulin of Giardia was identified through co-migration with beta-tubulin standards (Holberton and Ward, 1981). With modern genomic information and mass spectrometry based techniques it is possible to generate de novo peptide sequence from proteins extracted from gels then search this against genomic databases to identify a protein of interest. (Coon et al., 2005). From this point the underlying gene of the protein can be investigated, the level of expression tested using quantitative PCR or the gene can be cloned and the function or structure of the protein examined.

The mass spectrometry based identification of proteins in a sample does require knowledge of the genomic sequence of the organism under investigation. In the case of Giardia duodenalis the genome has been published for the isolate WB from assemblage A (Morrison et al., 2007), from the assemblage B GS isolate (Franzen et al., 2009) and from the assemblage E isolate P15 (Jerlstrom-Hultqvist et al., 2010). The WB genome has been extensively annotated whereas the assemblage B and E genomes are still in a draft form. Several studies have used mass spectrometry to identify important proteins of G. duodenalis, the majority of these were all conducted on isolates from assemblage A (Emmerlich et al., 1999, Shirakura et al.,
with one study including the assemblage B GS isolate (Palm et al., 2003). The aim of this study was to identify the proteins of difference between assemblages A and B of *G. duodenalis* highlighted in chapters 3 and 4. A sub aim was to identify some of the major proteins from the 2D analysis of *G. duodenalis* assemblage A, as this is the fully annotated genome we have available to work with. This will help to confirm the genomic data currently available.

### 5.2 Material and Methods

#### 5.2.1 Gel spot excision

Gel bands and spots identified from Chapters 3 and 4 as being unique to a given assemblage were excised from gels using a scalpel blade and transferred to a clean 1.7 mL microcentrifuge tube or if processing a large number of samples into the well of a 96 well ‘v’ bottomed plate (See figures 5.1-5.6 and Table 5.1). A selection of proteins common to all assemblages was also excised as controls to test if the assemblage A genomic database could be used to identify proteins from other assemblages. As all the gel slices were not excised at the same time several protein bands and spots were sampled repeatedly as internal controls (identified in figures) as well as bands from the molecular mass markers, as the identity of these proteins is known. All of the major spots from the assemblage A 2D gels were excised to identify the major trophozoite proteins.

Due to problems associated with contamination with keratin, a protein component of human skin, every care was taken to reduce the possibility of contamination. Nitrile
gloves were used whenever handling gels, as the standard latex gloves can bind protein therefore contaminating gels. Gel bands were excised on clean glass plates and the scalpel blades were changed frequently to avoid cross contamination of samples. Face masks were also worn to avoid keratin contamination due to breathing over the gels.

5.2.2 Gel Sample Processing

Gel plugs were first incubated with 50 µL of a 25 mM ammonium bicarbonate solution prepared in a 50% v/v acetonitrile solution at 37°C on a rocking platform for 45 min, to remove the Coomassie stain. The supernatant was removed from the gel plug and replaced with a second 50 µL of the ammonium bicarbonate and incubated for 45 min under the same conditions. This was repeated until all of the stain had been removed from the gel plugs. The gel plugs were then dried using a centrifugal vacuum dryer. A total of 10 µL of trypsin solution (12.5 µg/mL trypsin in 25 mM ammonium bicarbonate with 0.01% v/v trifluoracetic acid) was added to each gel plug, which were left at 37°C overnight to digest the proteins.

The next morning the peptides were extracted by adding 25 µL of 50% v/v acetonitrile to each plug and incubating at room temperature for 15 minutes. The supernatant was removed and placed into a separate clean 1.8 mL microcentrifuge tube and a second 25 µL volume of 50% v/v acetonitrile was added to the gel plugs and incubated as before. The supernatant was again removed and pooled with the previous sample. The extracted peptides were dried using a centrifugal vacuum dryer.
5.2.3 Mass Spectrometry

Dried peptide samples were resuspended in 10 µL of 50% v/v acetonitrile with 1% v/v formic acid. A microlitre of this was spotted onto a piece of parafilm and was mixed with 1µL of 10 mg/mL α-cyano-4-hydroxycinnamic acid in 50% v/v acetonitrile with 1% v/v formic acid. An aliquot of 0.6 µL of this mixture was spotted onto the well of a MALDI-ToF sample plate and allowed to air dry.

The mass spectrometric analysis was performed at the Western Australian Proteomics Facility at the Western Australian Institute of Medical Research using a 4800 MALDI-ToF-ToF mass spectrometer (Applied Biosystems, Foster City, U.S.A.) in MS/MS mode by staff from Proteomics International Pty. Ltd..

5.2.4 Bioinformatics

The mass spectra from the mass spectrometry were searched using the Mascot algorithm (Perkins et al., 1999) against the National Centre for Biotechnology Information non-redundant (NCBI_nr) database. Results were considered significant if they matched a *Giardia duodenalis* protein with an e-value close to zero and at least two peptide matches. If fewer than two peptide matches were achieved the identity was listed as provisional. For each protein the *G. duodenalis* locus tag, score and number of matches were recorded.
5.3 Results

5.3.1 Proteins Identified from SDS-PAGE

From the SDS-PAGE gels a total of 16 bands were excised and submitted for identification by mass spectrometry from both assemblage A isolate BAH 2c2 and assemblage B isolate BAH 34c8 (Figure 5.1). These were a mixture of potentially assemblage specific protein bands and protein bands common between the two assemblages. For each of the assemblages proteins were identified back to the genome for every band. For the assemblage A isolate a total of 43 protein identities were achieved which are outlined in Table 5.1. Of the 43 proteins identified there were 41 individual gene products present (39 individual proteins with peroxiredoxin 1 identified from three separate loci). The 16 samples analysed from assemblage B isolate BAH 34c8 resulted in 27 protein identities covering 21 individual proteins (Table 5.2). The XP prefix to the accession numbers is used to denote proteins that have come from the WB isolate genome.
Figure 5.1 Protein bands of assemblage A and B isolates BAH 2c2 and BAH 34c8 excised and identified from SDS-PAGE
The proteins identified from assemblages A and B can be broadly placed into 5 categories; metabolic proteins, structural proteins, proteins involved in cell cycling (division and en/ex-cystation), proteins involved in protein synthesis and protein structure and hypothetical proteins which show no homology to any known proteins.

Table 5.1 Proteins of assemblage A isolate BAH 2c2 visualised on SDS PAGE and identified with mass spectrometry

<table>
<thead>
<tr>
<th>Number</th>
<th>Protein Identity</th>
<th>Score</th>
<th>Peptide Matches a</th>
<th>Accession Number b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>axoneme associated protein</td>
<td>352</td>
<td>10</td>
<td>XP_001709214</td>
</tr>
<tr>
<td>A2</td>
<td>pyruvate (flavodoxin) dehydrogenase</td>
<td>444</td>
<td>11</td>
<td>XP_001708704</td>
</tr>
<tr>
<td></td>
<td>pyruvate flavodoxin oxidoreductase</td>
<td>145</td>
<td>4</td>
<td>XP_001708948</td>
</tr>
<tr>
<td>A3</td>
<td>dynamin</td>
<td>579</td>
<td>12</td>
<td>XP_001704962</td>
</tr>
<tr>
<td></td>
<td>phosphoenolpyruvate carboxykinase c</td>
<td>58</td>
<td>1</td>
<td>XP_001709869</td>
</tr>
<tr>
<td>A4</td>
<td>arginine deiminase</td>
<td>571</td>
<td>12</td>
<td>XP_001705755</td>
</tr>
<tr>
<td></td>
<td>2,3-bisphosphoglycerate-independent phosphoglycerate mutase</td>
<td>85</td>
<td>2</td>
<td>XP_001706738</td>
</tr>
<tr>
<td>A5</td>
<td>NEK Kinase</td>
<td>300</td>
<td>6</td>
<td>XP_001707030</td>
</tr>
<tr>
<td></td>
<td>Beta tubulin</td>
<td>265</td>
<td>6</td>
<td>XP_001707372</td>
</tr>
<tr>
<td>A6</td>
<td>Beta tubulin</td>
<td>265</td>
<td>6</td>
<td>XP_001707372</td>
</tr>
<tr>
<td>A7</td>
<td>NADP-specific glutamate dehydrogenase</td>
<td>985</td>
<td>16</td>
<td>XP_001707257</td>
</tr>
<tr>
<td>A8</td>
<td>Hypothetical protein</td>
<td>404</td>
<td>8</td>
<td>XP_001706436</td>
</tr>
<tr>
<td></td>
<td>Hypothetical protein</td>
<td>353</td>
<td>7</td>
<td>XP_001706898</td>
</tr>
<tr>
<td>A9</td>
<td>Uridine phosphorylase like protein 1</td>
<td>262</td>
<td>5</td>
<td>XP_001707342</td>
</tr>
<tr>
<td></td>
<td>carbamate kinase</td>
<td>239</td>
<td>4</td>
<td>XP_001708232</td>
</tr>
<tr>
<td></td>
<td>glucokinase</td>
<td>91</td>
<td>2</td>
<td>XP_001706734</td>
</tr>
<tr>
<td>A10</td>
<td>14-3-3 protein</td>
<td>335</td>
<td>6</td>
<td>XP_001706755</td>
</tr>
<tr>
<td></td>
<td>Alpha 7.3 Giardin</td>
<td>145</td>
<td>3</td>
<td>XP_001708455</td>
</tr>
<tr>
<td></td>
<td>Hypothetical protein</td>
<td>52</td>
<td>2</td>
<td>XP_001707037</td>
</tr>
<tr>
<td>A11</td>
<td>Beta Giardin</td>
<td>623</td>
<td>11</td>
<td>XP_001705425</td>
</tr>
<tr>
<td></td>
<td>delta giardin c</td>
<td>61</td>
<td>1</td>
<td>XP_001707449</td>
</tr>
<tr>
<td>A12</td>
<td>ribosomal protein S5</td>
<td>333</td>
<td>5</td>
<td>XP_001705992</td>
</tr>
<tr>
<td></td>
<td>14-3-3 protein</td>
<td>62</td>
<td>2</td>
<td>XP_001706755</td>
</tr>
<tr>
<td></td>
<td>Onithine Carbomoyl Transferase</td>
<td>50</td>
<td>2</td>
<td>XP_001709979</td>
</tr>
<tr>
<td>A13</td>
<td>SALP-1</td>
<td>179</td>
<td>3</td>
<td>XP_001708306</td>
</tr>
<tr>
<td></td>
<td>Ribosomal protein L7</td>
<td>116</td>
<td>3</td>
<td>XP_001705231</td>
</tr>
<tr>
<td></td>
<td>Alpha 2 Giardin</td>
<td>108</td>
<td>2</td>
<td>XP_001706958</td>
</tr>
<tr>
<td></td>
<td>GTP-binding nuclear protein</td>
<td>98</td>
<td>3</td>
<td>XP_001705159</td>
</tr>
<tr>
<td></td>
<td>Mitotic spindle checkpoint protein MAD2</td>
<td>85</td>
<td>3</td>
<td>XP_001705618</td>
</tr>
<tr>
<td></td>
<td>Peroxiredoxin 1 c</td>
<td>67</td>
<td>1</td>
<td>XP_001704832</td>
</tr>
<tr>
<td></td>
<td>Malic enzyme c</td>
<td>54</td>
<td>1</td>
<td>XP_001707581</td>
</tr>
<tr>
<td>A14</td>
<td>Peroxiredoxin 1</td>
<td>336</td>
<td>5</td>
<td>XP_001704199</td>
</tr>
</tbody>
</table>
Table 5.2 Proteins of assemblage B isolate BAH 34c8 visualised on SDS-PAGE and identified with mass spectrometry

<table>
<thead>
<tr>
<th>Number</th>
<th>Protein Identity</th>
<th>Score</th>
<th>Peptide Matches (^a)</th>
<th>Accession Number (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Axoneme associated protein</td>
<td>112</td>
<td>3</td>
<td>XP_001709214</td>
</tr>
<tr>
<td>B2</td>
<td>Dynamin</td>
<td>188</td>
<td>4</td>
<td>XP_001704962</td>
</tr>
<tr>
<td>B3</td>
<td>Transketolase</td>
<td>99</td>
<td>2</td>
<td>XP_001704614</td>
</tr>
<tr>
<td>B4</td>
<td>Protein 21.1</td>
<td>71</td>
<td>2</td>
<td>XP_001708927</td>
</tr>
<tr>
<td>B5</td>
<td>2,3-bisphosphoglycerate-independent phosphoglycerate mutase</td>
<td>92</td>
<td>2</td>
<td>XP_001706738</td>
</tr>
<tr>
<td>B6</td>
<td>Alpha tubulin</td>
<td>89</td>
<td>2</td>
<td>XP_001705720</td>
</tr>
<tr>
<td>B7</td>
<td>Beta tubulin</td>
<td>868</td>
<td>14</td>
<td>XP_001707372</td>
</tr>
<tr>
<td>B8</td>
<td>NADP-specific glutamate dehydrogenase</td>
<td>328</td>
<td>7</td>
<td>XP_001707257</td>
</tr>
<tr>
<td>B9</td>
<td>NADP-specific glutamate dehydrogenase</td>
<td>131</td>
<td>3</td>
<td>XP_001707257</td>
</tr>
<tr>
<td>B10</td>
<td>Phosphoglycerate kinase</td>
<td>105</td>
<td>3</td>
<td>XP_001706998</td>
</tr>
<tr>
<td>B11</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
<td>127</td>
<td>3</td>
<td>XP_001705042</td>
</tr>
<tr>
<td>B12</td>
<td>Hypothetical protein</td>
<td>106</td>
<td>3</td>
<td>XP_001706436</td>
</tr>
<tr>
<td>B13</td>
<td>Hypothetical protein</td>
<td>61</td>
<td>2</td>
<td>XP_001706436</td>
</tr>
<tr>
<td>B14</td>
<td>Uridine phosphorylase like protein 1</td>
<td>186</td>
<td>4</td>
<td>XP_001707342</td>
</tr>
<tr>
<td>B15</td>
<td>Carbamate kinase (^c)</td>
<td>68</td>
<td>1</td>
<td>XP_001708232</td>
</tr>
<tr>
<td>B16</td>
<td>GLORF-C4</td>
<td>115</td>
<td>2</td>
<td>XP_001704917</td>
</tr>
<tr>
<td>B17</td>
<td>Ornithine carbamoyltransferase (^c)</td>
<td>102</td>
<td>1</td>
<td>XP_001709979</td>
</tr>
</tbody>
</table>

\(^a\) Number of peptides matched to the protein  
\(^b\) Accession number for protein in NCBI database  
\(^c\) Provisional identification as only one peptide match
<table>
<thead>
<tr>
<th>Protein Description</th>
<th>Peptide Matches</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-3-3 protein</td>
<td>83</td>
<td>XP_001706755</td>
</tr>
<tr>
<td>B16 SALP-1</td>
<td>334</td>
<td>XP_001708306</td>
</tr>
<tr>
<td>GLORF-C4</td>
<td>115</td>
<td>XP_001704917</td>
</tr>
<tr>
<td>Ribosomal protein L7</td>
<td>81</td>
<td>XP_001705231</td>
</tr>
<tr>
<td>Eukaryotic translation elongation factor</td>
<td>60</td>
<td>XP_001704069</td>
</tr>
</tbody>
</table>

* a Number of peptides matched to the protein  
* b Accession number for protein in NCBI database  
* c Provisional identification as only one peptide match

### 5.3.2 Proteins Identified from 2D-PAGE

![Proteins excised and identified from assemblage A isolate BAH 2c2 2D-PAGE](image)

In total 86 discrete spots were cut from 2D-PAGE gels of assemblage A isolate BAH 2c2. These samples were submitted for mass spectrometry analysis and of the 86 samples, meaningful data was generated for 27 of the protein spots. The protein spots that were identified are shown in Figure 5.2 with the corresponding protein
information summarised in Table 5.2. Of the 27 spots, 4 were found to contain more than one protein and multiple spots contained the same protein. The proteins identified are varied including; metabolic enzymes, structural proteins and proteins involved in protein production and cycling. The XP prefix to the accession number denotes that these have matched back to the assemblage A WB isolate genome sequence.

### Table 5.3 Proteins of Assemblage A isolate BAH 2c2 visualised on 2D-PAGE identified by Mass Spectrometry

<table>
<thead>
<tr>
<th>Number</th>
<th>Protein Identity</th>
<th>Score</th>
<th>Peptide Matches</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RNA Binding Protein £</td>
<td>39</td>
<td>1</td>
<td>XP_001708843</td>
</tr>
<tr>
<td>2</td>
<td>Fructose-1,6-bisphosphate aldolase</td>
<td>194</td>
<td>3</td>
<td>XP_001710050</td>
</tr>
<tr>
<td>3</td>
<td>Uridine phosphorylase like protein 1</td>
<td>168</td>
<td>3</td>
<td>XP_001707342</td>
</tr>
<tr>
<td></td>
<td>Fructose-1,6-bisphosphate aldolase</td>
<td>91</td>
<td>2</td>
<td>XP_001710050</td>
</tr>
<tr>
<td>4</td>
<td>Carbamate kinase</td>
<td>337</td>
<td>6</td>
<td>XP_001708232</td>
</tr>
<tr>
<td></td>
<td>Glyceraldehyde-3-phosphate dehydrogenase £</td>
<td>72</td>
<td>1</td>
<td>XP_001705042</td>
</tr>
<tr>
<td>5</td>
<td>Gamma giardin</td>
<td>564</td>
<td>10</td>
<td>XP_001706274</td>
</tr>
<tr>
<td>6</td>
<td>Alpha-11 giardin</td>
<td>63</td>
<td>2</td>
<td>XP_001705101</td>
</tr>
<tr>
<td>7</td>
<td>Alpha-11 giardin</td>
<td>144</td>
<td>3</td>
<td>XP_001705101</td>
</tr>
<tr>
<td>8</td>
<td>Beta-giardin</td>
<td>191</td>
<td>3</td>
<td>XP_001705425</td>
</tr>
<tr>
<td>9</td>
<td>Delta giardin</td>
<td>159</td>
<td>4</td>
<td>XP_001707449</td>
</tr>
<tr>
<td>10</td>
<td>Alpha-2 giardin</td>
<td>66</td>
<td>3</td>
<td>XP_001706958</td>
</tr>
<tr>
<td>11</td>
<td>Alpha-7.3 giardin</td>
<td>295</td>
<td>7</td>
<td>XP_001708455</td>
</tr>
<tr>
<td></td>
<td>Alpha 7.1 giardin</td>
<td>175</td>
<td>4</td>
<td>XP_001704112</td>
</tr>
<tr>
<td></td>
<td>Alpha-2 giardin £</td>
<td>72</td>
<td>1</td>
<td>XP_001706958</td>
</tr>
<tr>
<td>12</td>
<td>Alpha-2 giardin £</td>
<td>76</td>
<td>1</td>
<td>XP_001706958</td>
</tr>
<tr>
<td>13</td>
<td>Arginine deiminase</td>
<td>89</td>
<td>2</td>
<td>XP_001705755</td>
</tr>
<tr>
<td>14</td>
<td>Caltractin</td>
<td>151</td>
<td>3</td>
<td>XP_001705487</td>
</tr>
<tr>
<td>15</td>
<td>Centrin</td>
<td>159</td>
<td>2</td>
<td>XP_001707577</td>
</tr>
<tr>
<td>16</td>
<td>Peptidyl-prolyl cis-trans isomerase B precursor</td>
<td>164</td>
<td>4</td>
<td>XP_001707838</td>
</tr>
<tr>
<td>17</td>
<td>Peptidyl-prolyl cis-trans isomerase B precursor</td>
<td>113</td>
<td>2</td>
<td>XP_001707838</td>
</tr>
<tr>
<td>18</td>
<td>Eukaryotic translation initiation factor 5A</td>
<td>311</td>
<td>3</td>
<td>XP_001704055</td>
</tr>
<tr>
<td>19</td>
<td>Nascent polypeptide-associated complex protein transciption factor</td>
<td>153</td>
<td>3</td>
<td>XP_001704531</td>
</tr>
<tr>
<td>20</td>
<td>Peptidyl-prolyl cis-trans isomerase B precursor</td>
<td>257</td>
<td>4</td>
<td>XP_001707838</td>
</tr>
<tr>
<td>21</td>
<td>Histone H2B</td>
<td>117</td>
<td>2</td>
<td>XP_001706788</td>
</tr>
<tr>
<td>22</td>
<td>Thioredoxin</td>
<td>398</td>
<td>5</td>
<td>XP_001709555</td>
</tr>
<tr>
<td></td>
<td>Protein Name</td>
<td>Peptides Matched</td>
<td>Matched Peptides</td>
<td>Accession Number</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>23</td>
<td>Pyridoxamine 5'-phosphate oxidase</td>
<td>111</td>
<td>3</td>
<td>XP_001705592</td>
</tr>
<tr>
<td>24</td>
<td>Pyridoxamine 5'-phosphate oxidase</td>
<td>137</td>
<td>4</td>
<td>XP_001705592</td>
</tr>
<tr>
<td>25</td>
<td>Rhodanese</td>
<td>262</td>
<td>5</td>
<td>XP_001710142</td>
</tr>
<tr>
<td></td>
<td>Pyridoxamine 5'-phosphate oxidase (c)</td>
<td>65</td>
<td>1</td>
<td>XP_001705592</td>
</tr>
<tr>
<td>26</td>
<td>Hypothetical protein</td>
<td>115</td>
<td>3</td>
<td>XP_001704574</td>
</tr>
<tr>
<td>27</td>
<td>vacuolar ATP synthase, subunit G (c)</td>
<td>42</td>
<td>1</td>
<td>XP_001706291</td>
</tr>
</tbody>
</table>

*a* Number of peptides matched to the protein  
*b* Accession number for protein in NCBI database  
*c* Provisional identification as only one peptide match

Figure 5.3 Protein spots of assemblage B isolate 34c8 excised and identified via mass spectrometry
Table 5.4 Proteins of Assemblage B isolate BAH 34c8 visualised on 2D-PAGE identified by Mass Spectrometry

<table>
<thead>
<tr>
<th>Number</th>
<th>Protein Identity</th>
<th>Score</th>
<th>Peptide Matches a</th>
<th>Accession Number b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fructose-1,6-bisphosphate aldolase c</td>
<td>57</td>
<td>1</td>
<td>XP_001710050</td>
</tr>
<tr>
<td>2</td>
<td>Uridine phosphorylase like protein 1</td>
<td>164</td>
<td>4</td>
<td>EET01236</td>
</tr>
<tr>
<td></td>
<td>Beta tubulin</td>
<td>119</td>
<td>2</td>
<td>CAA29923</td>
</tr>
<tr>
<td>3</td>
<td>Carbamate kinase</td>
<td>104</td>
<td>2</td>
<td>EET00411</td>
</tr>
<tr>
<td>4</td>
<td>Gamma giardin</td>
<td>76</td>
<td>1</td>
<td>EES98269</td>
</tr>
<tr>
<td>5</td>
<td>Carbamate kinase</td>
<td>188</td>
<td>3</td>
<td>EET00411</td>
</tr>
<tr>
<td>6</td>
<td>Uridine phosphorylase like protein 1</td>
<td>162</td>
<td>5</td>
<td>EET01236</td>
</tr>
<tr>
<td>7</td>
<td>Uridine phosphorylase like protein 1</td>
<td>155</td>
<td>4</td>
<td>EET01236</td>
</tr>
<tr>
<td>8</td>
<td>Beta giardin</td>
<td>99</td>
<td>3</td>
<td>XP_001705425</td>
</tr>
<tr>
<td>9</td>
<td>delta giardin</td>
<td>135</td>
<td>3</td>
<td>EES99594</td>
</tr>
<tr>
<td>10</td>
<td>ornithine carbamoyl transferase</td>
<td>85</td>
<td>2</td>
<td>EET02056</td>
</tr>
<tr>
<td>11</td>
<td>ornithine carbamoyl transferase</td>
<td>137</td>
<td>3</td>
<td>EET02056</td>
</tr>
<tr>
<td></td>
<td>Uridine phosphorylase like protein 1</td>
<td>107</td>
<td>2</td>
<td>EET01236</td>
</tr>
<tr>
<td>12</td>
<td>SALP-1</td>
<td>200</td>
<td>4</td>
<td>XP_001708306</td>
</tr>
<tr>
<td>13</td>
<td>Beta Giardin</td>
<td>131</td>
<td>3</td>
<td>XP_001705425</td>
</tr>
<tr>
<td>14</td>
<td>Ornithine carbamoyl transferase</td>
<td>137</td>
<td>3</td>
<td>EET02056</td>
</tr>
<tr>
<td>15</td>
<td>Giardia trophozoite antigen 1 c</td>
<td>127</td>
<td>1</td>
<td>XP_001705774</td>
</tr>
<tr>
<td>16</td>
<td>Thioredoxin</td>
<td>80</td>
<td>2</td>
<td>EET02063</td>
</tr>
</tbody>
</table>

a Number of peptides matched to the protein  

b Accession number for protein in NCBI database  

c Provisional identification as only one peptide match

For the assemblage B isolate BAH 34c8, 76 individual protein spots were excised and submitted for analysis by mass spectrometry. Of the samples submitted protein identities were obtained for 16 of the protein spots, a 21% success rate, these spots are highlighted in Figure 5.3. The protein identities for each of the protein spots are listed in Table 5.3, as for the assemblage A isolate there is a mix of metabolic proteins, structural proteins and one hypothetical protein. Several gel spots were found to contain the same protein leading to 11 different proteins identified for isolate BAH 34c8. The EET prefix to the accession numbers denotes that these proteins have matched to the assemblage B GS isolate genome and the XP prefix is for proteins matched to the WB genome.
5.4 Discussion

5.4.1 Proteins Identified with Mass spectrometry

In total, one hundred and ninety four protein spots or bands were excised from either 2D or SDS-PAGE gels and submitted for identification via mass spectrometry. Of these samples seventy five returned meaningful protein identification data with one hundred and twenty protein identities, 21 of which are provisional based on the number of peptides present. Combining all of the data, sixty three separate gene products were determined. The proteins identified can be placed into five major groups; metabolic proteins, structural proteins, proteins involved in protein synthesis and structure, cell cycling proteins and hypothetical proteins with no known function. The SDS-PAGE analysis was performed prior to the publication of the assemblage B genomic information hence why proteins from assemblage B have all been identified back to the assemblage A genome. When performing the 2D-PAGE analysis the assemblage B genomic information was available so the majority of protein spots mapped back to the B genome.

5.4.1.1 Metabolic Enzymes

Fourteen of the proteins can be classified as metabolic proteins. Five of these proteins are involved in the glycolysis pathway. The first step in glycolysis is the phosphorylation of glucose by glucokinase, identified from SDS-PAGE gels of isolate BAH 2c2 (Henze et al., 2001). From this phosphorylated glucose molecule there are several steps which result in the production of D-fructose 1,6 bisphosphate which is cleaves by fructose 1, 6 bisphosphate aldolase in a reversible reaction to
form dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate (Galkin et al., 2007). The fructose 1, 6 bisphosphate adolase enzyme was identified on 2D-PAGE gels of assemblage A and B isolates. Glyceraldehyde 3 phosphate dehydrogenase cleaves the glyceraldehyde 3 phosphate produced by fructose 1, 6 bisphosphate aldolase, generating 1, 3 diphosphoglycerate (Yang et al., 2002). Phosphoglycerate kinase, seen on SDS-PAGE gels of assemblage B, then dephosphorylates the resulting glycerate 1, 3 phosphate to glycerate 3 phosphate (Morrison et al., 2007). The enzyme 2, 3 bisphosphoglycerate independent phosphoglycerate mutase, detected in assemblage A and B isolated in SDS-PAGE, utilises the glycerate 3 phosphate to produce glycerate 2 phosphate, which is eventually converted into pyruvate (Morrison et al., 2007). When in axenic culture glucose is the only metabolite used for energy production, hence the proteins used in this pathway are likely to be found in high abundance as they are required by the trophozoites (Jarroll et al., 2011).

Another set of three interrelated proteins involved in the arginine dihydrolase pathway were also identified from both the SDS and 2D-PAGE gels. This is thought to be a significant source of energy for the cell, relying on the conversion of arginine into carbon dioxide, ammonium and ornithine, with citrulline and carbomoyl-P as intermediates (Jarroll et al., 2011). Arginine is initially converted to citrulline by arginine deiminase which was identified from SDS-PAGE and 2D gels. Ornithine carbomoyl transferase, also identified from the 2D gels, then converts the citrulline to carbamoyl P and finally the carbamoyl P is used to produce adenosine triphosphate through the generation of ammonium ions by the enzyme carbamate kinase, again identified on 2D gels. Ornithine carbomoyl transferase and arginine
deiminase are both excreted by *G. duodenalis* trophozoites upon exposure to host cells in cell culture experiments (Ringqvist *et al.*, 2008).

As well as using the arginine as an energy source *G. duodenalis* may be actively reducing the amount of arginine available to the host. Arginine is used by host cells to produce nitric oxide which is used by the cells as an antimicrobial agent, and has been shown to limit growth of *G. duodenalis* trophozoites (Eckmann *et al.*, 2000). By releasing metabolic enzymes into their external environment, trophozoites of *G. duodenalis* would actively decrease the amount of arginine and therefore the potential for nitric oxide to be produced, protecting themselves from its antimicrobial activity (Ringqvist *et al.*, 2008).

### 5.4.1.2 Structural Proteins

A total of 19 structural proteins were identified from the 2D gels, just under a third of all of the proteins identified. Of these, eight belong to the giardins, a group of structural proteins unique to *Giardia* spp. (alpha 2, alpha 7.1, alpha 7.3, alpha 10 alpha 11, beta, delta and gamma giardin). The alpha giardins associate with the adhesive disk, flagella and the plasma membrane of trophozoites (Weiland *et al.*, 2005). They are related to the annexins which are found in many eukaryotic organisms including humans. The annexins bind phospholipids in a calcium dependant manner and are thought to play a role in anchoring the cytoskeleton to the plasma membrane with a similar role proposed in *G. duodenalis* (Weiland *et al.*, 2005). Alpha 2 giardin was routinely identified from both SDS-PAGE and 2D-PAGE gels of assemblage A isolate BAH 2c2; however the protein was never
identified in assemblage B gels. Alpha 2 giardin may therefore represent an assemblage A specific protein.

Beta giardin is the major protein in the adhesive disk and is related to the striated fibre assemblins (SF-assemblins) (Elmendorf et al., 2003). Beta giardin was seen as a dominant spot in both assemblages A and B SDS and 2D-PAGE gels (Fig. 5.1 A11, B14; Fig. 5.2 spot 8; Fig. 5.3 spot 8). In assemblage B a second spot was identified as beta giardin (Fig. 5.3, spot 13). This appears to be assemblage B specific and may represent an assemblage based post-translational modification variance. Beta giardin is proposed to be one of three proteins that make up the structural core of the microribbons which form a base for other structural proteins (Dawson, 2011). The host mounts a strong antibody mediated immune response against beta giardin which may indicate that it has the potential to be used as a vaccine candidate to block attachment (Palm et al., 2003).

5.4.1.3 Protein Synthesis and structure

Proteins involved in protein synthesis and structure accounted for 18 of the proteins identified via mass spectrometry. Half of these proteins are ribosomal proteins and are therefore involved in the translation of proteins. The ribosomal proteins of *G. duodenalis* are of interest in understanding the divergence of the eukaryotic lineage from the prokaryotic lineage (Hartman and Fedorov, 2002). Of the 72 ribosomal proteins in *G. duodenalis*, 61 have a prokaryotic or archaeal lineage, with the remaining 11 showing homology to other eukaryotic proteins (Hartman and Fedorov, 2002).
5.4.1.4 Cell Cycling

A total of 6 proteins involved in the cell cycle process, being either mitotic division or the encystation and excystation processes, were determined from the protein gels of assemblages A and B. One such protein was a 14-3-3 protein homologue, a eukaryotic protein involved in cell signalling, differentiation and apoptosis (Lalle et al., 2006). Another cell cycling protein identified from the gels was GLORF-C4 which has an early role in the encystation process (Nores et al., 2009).

5.4.1.4 Hypothetical Proteins

There were four proteins from the assemblage A and B protein gels that did not show any homology to any known eukaryotic or prokaryotic proteins (XP_001706898, XP_001706436, XP_001707037 and XP_001704574). These proteins potentially represent novel classes of proteins. Detecting these proteins from protein gels helps to validate the gene prediction undertaken as part of the Giardia genome project.
5.4.2 Success of Proteins Identified from 2D-PAGE Gels

The 32 protein bands excised from SDS-PAGE gels of assemblage and B could all be identified back to the *G. duodenalis* genome, although the analysis of proteins from 2D-PAGE was not as productive. Between the two isolates analysed using 2D-PAGE, 162 separate spots were excised from the gels and submitted for identification by mass spectrometry with 43 being identified back to a protein in the *G. duodenalis* genome. This represents a successful identification rate of 26.5% which is lower than expected; the reason for this is unclear. For some samples, duplicate spots were sent to Proteomics International for processing without an increase in the number of spots identified. Similarly in some instances peptide mixtures were loaded onto MALDI plates by staff from Proteomics International without an increase in the number of spots identified. This would suggest that there is an inherent issue with the sample or the preparation process.

There are several potential issues with the analysis that would affect the identification of proteins. Firstly, the matrix used may not be appropriate for the peptides being analysed. Certain types of peptides, e.g. hydrophobic, ionize better with matrices other than the α-cyano-4-hydroxycinnamic acid used (Lubec and Afjehi-Sadat, 2007). Alternatively, the mass spectra generated may not be the problem and the issue is matching the peptide sequence back to the *G. duodenalis* genome. Both of these explanations are, however, highly unlikely given the success of identifying proteins from the SDS-PAGE gels. In fact with the analysis of the 2D-PAGE data a draft of an assemblage B genome was included in the search database (Franzen *et al.*, 2009). Any incompatibilities with the matrix or with the
bioinformatics should affect the 2D-PAGE and SDS-PAGE proteins equally. The reason behind this lower level of protein identification remains unclear. The third possibility is that the 2D-PAGE protocol has limited the ability of the proteins to be identified by mass spectrometry.

5.4.3 Protein Variation

There were eleven proteins identified as being assemblage variants between assemblages A and B, and the proteins of difference are outlined in Table 5.5. The majority of these differences are size variants between the two assemblages with a few proteins that appear to be assemblage specific. *Giardia* axoneme associated protein 180 has size variants between assemblages A and B, with the assemblage A protein below 200 kDa and the assemblage B variant above 200 kDa. *Giardia* axoneme associated protein 180 is a head stalk protein consisting of an N-terminus with 5 repeating ankyrin elements and a coiled coil carboxyl tail (Elmendorf *et al.*, 2005). The ankyrin repeats are thought to bind to beta tubulin in the microtubules of the cytoskeleton and the coiled coil domain arrangement seen in GASP 180 indicates that it may associate with other coiled coil domains, similar to that seen in human muscle proteins. This implies that GASP 180 is involved in driving the flagella (Elmendorf *et al.*, 2005). Variation in this protein between isolates could impact the strength of flagella movement, which would in turn affect how an isolate is able to move within the gut. An isolate with a stronger flagella movement may be able to resist the peristaltic movement within the gut and therefore persist for a longer period in the host.
Table 5.5 Protein Variants of Assemblages A and B Identified Using SDS-PAGE and 2D-PAGE Coupled With Mass Spectrometry

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Source</th>
<th>Assemblage A</th>
<th>Assemblage B</th>
</tr>
</thead>
<tbody>
<tr>
<td>GASP 180</td>
<td>SDS-PAGE</td>
<td>A1</td>
<td>B1</td>
</tr>
<tr>
<td>2,3-bisphosphoglycerate independent phosphoglycerate mutase</td>
<td>SDS-PAGE</td>
<td>A4</td>
<td>B5</td>
</tr>
<tr>
<td>NADP-specific glutamate dehydrogenase</td>
<td>SDS-PAGE</td>
<td>A7</td>
<td>B8, B9</td>
</tr>
<tr>
<td>Hypothetical Protein XP_001706436</td>
<td>SDS-PAGE</td>
<td>A8</td>
<td>B10, B11</td>
</tr>
<tr>
<td>Hypothetical Protein XP_001706898</td>
<td>SDS-PAGE</td>
<td>A8</td>
<td>NI</td>
</tr>
<tr>
<td>Alpha 7.3 giardin</td>
<td>SDS-PAGE</td>
<td>A10</td>
<td>NI</td>
</tr>
<tr>
<td>Alpha 10 giardin</td>
<td>SDS-PAGE</td>
<td>NI</td>
<td>B13</td>
</tr>
<tr>
<td>Alpha 11 giardin</td>
<td>SDS-PAGE 2D-PAGE</td>
<td>2DA 6, 7</td>
<td>B13</td>
</tr>
<tr>
<td>GLORF-C4</td>
<td>SDS-PAGE</td>
<td>NI</td>
<td>B15 B16</td>
</tr>
<tr>
<td>Alpha 2 giardin</td>
<td>SDS-PAGE 2D-PAGE</td>
<td>A13  2DA 11, 12</td>
<td>NI</td>
</tr>
<tr>
<td>Uridine phosphorylase like proteins 1</td>
<td>2D-PAGE</td>
<td>2DA 3</td>
<td>2DB 2, 6, 7</td>
</tr>
</tbody>
</table>

NI not identified
SDS-PAGE numbers refer to Figure 5.1
2DA 2D-PAGE assemblage A, numbers refer to spots on Figure 5.2
2DB 2D-PAGE assemblage B, numbers refer to spots on Figure 5.3

The glycolytic enzyme 2,3-bisphosphoglycerate independent phosphoglycerate mutase had two size variants in assemblage B and one in assemblage A. Neither of the assemblage B bands was the same size as the assemblage A band. Band A8 in assemblage A SDS-PAGE had two protein identities, hypothetical proteins XP_001706436 and XP_001706898. Only the XP_001706436 protein was identified in assemblage B. However rather than one band it was found in two bands, one of which was the same size as the A8 band. This indicates that either assemblage B has two copies of the gene one of which is truncated or that assemblage B has another variant of this hypothetical protein that is related to hypothetical protein.
XP_001706898 and has sufficient amino acid sequence similarity to be incorrectly identified as XP_001706898.

The NADP specific glutamate dehydrogenase (GDH) also showed size variation for assemblages A and B, as well as a potential copy number variation. One band was identified as GDH for assemblage A whereas two bands were identified for assemblage B, one with a higher molecular mass than the assemblage A band and the other a lower. The GDH gene is commonly used for phylogenetic studies and there has been no reported evidence of multiple genes (Monis et al., 1996, Read et al., 2004, Wielinga and Thompson, 2007). In assemblage B one of these protein bands may be the native enzyme and the other a post-translationally modified isoform.

Several of the alpha giardins appear to be assemblage specific. Alpha 7.3 giardin was only identified in assemblage A (band A10) and alpha 10 and alpha 11 giardin were both identified from assemblage B in a band (band B13) that migrated alongside band A10. Alpha 11 giardin was identified from 2D-PAGE gels of assemblage A isolate BAH 2c2 (Spots 6, 7). This would indicate that alpha 7.3 giardin is only being expressed in assemblage A and alpha 10 giardin is being expressed in assemblage B. In phylogenetic studies of the alpha giardins, alpha 7.3 and alpha 10 giardin belong to the same lineage so there could be cross over in action to the point where alpha 10 could replace alpha 7.3 (Weiland et al., 2005). A study of expression of some of the alpha giardins in assemblage A isolates found that alpha 10 giardin had negligible expression when compared to beta giardin (Weiland et al., 2005).
*Giardia lamblia* open reading frame C4 (GLORF-C4) appears to be specific to assemblage B (Fig. 5.3.1 B15 and B16). The protein has been routinely identified from assemblage B SDS-PAGE gels but not from assemblage A. This protein has been implicated in the encystation process with a delay in encystation resulting from a reduction in the GLORF-C4 mRNA levels (Nores *et al.*, 2009). The gene for GLORF-C4 is present in both human infective assemblages (Yong *et al.*, 2002), however the mRNA transcripts were originally identified as only being expressed at the trophozoite stage in assemblage B (Nash and Mowatt, 1992).

This indicates that there is variation in the way in which the gene is expressed between assemblages A and B. As the protein and the mRNA of GLORF-C4 have been identified in assemblage B trophozoites and not in assemblage A, this implies that the assemblage B isolates constitutively express the GLORF-C4 protein. As expression of GLORF-C4 has been directly related to encystation, assemblage B isolates may be able to encyst at a faster rate to assemblage A isolates, which would first need to begin expressing GLORF-C4.

*Giardia duodenalis* lacks the ability to *de novo* synthesise its own pyrimidine and purine bases and must therefore scavenge them from the environment. One of the nucleotide scavenging enzymes is uridine phosphorylase like protein 1 (UPL-1) which was identified in assemblages A and B in 2D and SDS-PAGE gels. In the assemblage B isolate 2D gels there were three different spots identified as UPL-1 (Fig. 5.3.3; spots 6, 7 and 11) whereas in assemblage A there was only one identified (Fig. 5.3.2; spot 3). The three spots in assemblage B are likely isoforms of the same protein. Based on the comparative 2D work outlined in Chapter 4, spot 7 in Figure
5.3 is assemblage B specific and represents an isoform of UPL-1 only found in assemblage B. Spot 3 in Figure 5.2 is the assemblage A variant of UPL-1 and is also assemblage specific in respect to its migration on a gel. Little is known about variation in UPL-1 of *G. duodenalis* so it is uncertain if the difference seen on the gel between assemblages A and B is due to variation at the genetic level or post translational modification.

### 5.4.4 Conclusions

Protein bands and spots of isolates from assemblages A and B of *G. duodenalis* were successfully identified back to the genome using mass spectrometry. One of the major aims of this study was to identify protein variants between the human infective assemblages of *G. duodenalis* and this has been achieved. Four major protein variants between assemblages A and B have been characterised back to the genome. Based on gel and mass spectrometry data: alpha 2 giardin is an assemblage A specific protein; there is major size variation in GASP-180 between the assemblages; UPL-1 has isoelectric point variances between the assemblages; and GLORF-C4 is constitutively expressed in assemblage B. In the case of alpha 2 giardin, GASP 180 and UPL-1, molecular characterisation of the corresponding genes is needed to confirm the basis of the protein variation seen. For GLORF-C4, an investigation into the mRNA levels present in the two assemblages is required. Studies into the timing of the encystation process would also be useful to ascertain if the variation in GLORF-C4 expression has a biological impact. The protein variants identified above represent the first reported protein differences between the human infective assemblages of *G. duodenalis*. 
Chapter 6 Genetic Characterisation of Protein Variation

6.1 Introduction

Variation at the genetic level is not always indicative of a change to the protein produced, as more than one DNA codon can encode the same amino acid. Characterising protein differences at the genetic level is important to confirm the results from proteomic analysis and to understand the phylogenetic relationships of isolates based on the nucleotide sequence of the variant proteins.

For effective phylogenetic analysis it is important to look at conserved genes. If there is too much variation at the genetic level then there may be insufficient similarity to construct a relationship. However, if genes are closely related phylogenetically there may be few differences at the protein level. Therefore the migration of a protein in either SDS-PAGE or 2D-PAGE will remain constant. Conversely if there is a large amount of difference in a particular protein between two isolates, there should be an even greater degree of genetic diversity at the same locus.

The genes of proteins that show high levels of diversity could be used as novel targets for assemblage specific PCR assays making genotyping of samples more streamlined. The aim of this study was to amplify the genes of three proteins highlighted in the comparative protein analysis as having variation between the assemblages, in order to determine if the level of protein variation correlated with a genetic difference. The three proteins examined were alpha 2 giardin, the GASP-180 axoneme associated protein and uridine phosphorylase like protein-1. These three proteins all showed significant variation between assemblages at the protein level.
Alpha 2 giardin was chosen as it has been previously shown to be an immunodominant protein during infection (Palm et al., 2003) and the potential assemblage specific nature of alpha 2 giardin is significant given the role it is thought to play in anchoring the plasma membrane to the cytoskeleton (Weiland et al., 2005). The GASP-180 gene sequence was chosen due to the large size variation between assemblages A and B and the impact this may have on motility of the trophozoite (Elmendorf et al., 2005). Since G. duodenalis lacks the ability to de novo synthesise its’ own nucleotides the variation in the uridine phosphorylase like protein 1 gene may affect how pyrimidine bases are scavenged by the two assemblages, hence why it was chosen for further characterisation (Jarroll et al., 2011).

6.2 Materials and Methods

6.2.1 Gene sequence collection

Three proteins of difference between the assemblages were chosen based on the variation seen at the protein level. The first alpha 2 giardin was found in assemblage A, but was absent from assemblage B. The gene sequence was taken from the NCBI gene database from accession number GL50803_7796. The nucleotide sequence M34550 was also used as it includes approximately 400bp up and down stream of the alpha 2 giardin gene. The second protein to be examined, Giardia axoneme associated protein (GASP-180) showed size variation between the assemblages examined. The gene sequence was obtained from the NCBI database from accession number GL50803_137716. The third gene examined was the uridine phosphorylase like-protein which showed pI variation between the assemblages.
The gene sequence, accession number GL50803_9779, was also obtained from the NCBI database.

6.2.2 Primer Design

Gene sequences were submitted to the Primer3 software package accessed through Biology Workbench using the default settings for primer construction, except that the product size range was varied (http://workbench.sdsc.edu/). For the alpha 2 giardin PCR primers already exist to amplify the whole gene, A2GEX5’ and A2GEX3’ (Table 6.1), as such these were used to amplify the whole gene (Palm et al., 2003). A second set of primers were designed internal of the Palm et al. primers (2003) to amplify approximately 500 bp of the alpha 2 giardin gene, A2int5’ and A2int3’ (Table 6.1). A third set of primers were designed using the M34550 sequence of alpha 2 giardin sequence that amplified the whole gene plus a 200 bp flanking region up and down stream of the gene. For the GASP-180, 5 sets of primers were designed. Each set of primers were designed to amplify approximately 1000 bp of the 4000 bp gene with approximately 100 bp overlap between primer sets, Axo1-5 forward and reverse (Table 6.1). Two sets of primers were used to amplify the uridine phosphorylase like protein-1 gene sequence. The first set were from Palm et al. (2003), UPL1-5’ and UPL1-3’ (Table 6.1) and the second set were developed using the Primer3 software to amplify approximately 1000 bp of the uridine phosphorylase gene, UDP-5’ and UDP-3’ (Table 6.1). Primers were searched against the NCBI nucleotide database to ensure that they were specific for the G. duodenalis target gene. All primers were manufactured by Geneworks Australia.
6.2.3 DNA Extraction

Trophozoites were prepared by sonication as per the protein preparation method in Chapter 2, except protein inhibitor cocktail (Roche) was not added to the PBS and the resuspended trophozoites were made up to 2 mL. To the sonicated trophozoites, 11 µL of proteinase K (27mg/mL) was added and the samples incubated at 37°C for 4 hr after which 2 µL of RNase was added and incubated at 37°C for a further 30 min. The samples were then cooled on ice for 10 minutes followed by the addition of 1 mL of 7.5 M ammonium acetate to precipitate any protein. The protein contaminants were pelleted by centrifugation and the supernatant retained and kept on ice. One volume of isopropanol was added to the supernatant to precipitate the DNA, which was collected by centrifugation. The resulting pellet was washed with 2 mL of ethanol and left to air dry for a maximum of 24 hr then resuspended in 50 µL of TE buffer.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Target Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2GEX-5’</td>
<td>5’CCG GAT CCC CGA AGG TCA CCG ACA TTG CGA AC -3’</td>
<td>Alpha 2 giardin</td>
</tr>
<tr>
<td>A2GEX-3’</td>
<td>5’GGG AAT TCT TCA CGC GCC AGA GGG TGC GA AG-3’</td>
<td>Alpha 2 giardin</td>
</tr>
<tr>
<td>A2int-5’</td>
<td>5’-AAG AAC GAC CAC ATG GCC T-3’</td>
<td>Alpha 2 giardin</td>
</tr>
<tr>
<td>A2int-3’</td>
<td>5’-CGG CTT TCA TAC TCC CGT AG-3’</td>
<td>Alpha 2 giardin</td>
</tr>
<tr>
<td>A2ex-5’</td>
<td>5’-CCT CAT GGT GTA CAT GCT GG-3’</td>
<td>Alpha 2 giardin</td>
</tr>
<tr>
<td>A2ex-3’</td>
<td>5’-AAG CAT AGA GTA CGG CCC CT-3’</td>
<td>Alpha 2 giardin</td>
</tr>
<tr>
<td>A xo1-5’</td>
<td>5’-GCA GTG GTG TCT TCT GTG GA-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo1-3’</td>
<td>5’-GGA TGT CCC TCT CCT TCT CC-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo2-5’</td>
<td>5’-GAC ATG GAG GGA ATG CTC AG-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo2-3’</td>
<td>5’-CCT GGA GCT CGT CTT TGA GT-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo3-5’</td>
<td>5’-CGA TAA CCA GAT CGC AGA GC-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo3-3’</td>
<td>5’-GTT GCG TAG GTC CTT CTC CTT GA-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo4-5’</td>
<td>5’-GGC GAG AAC GAC CAC TAC A-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo4-3’</td>
<td>5’-CTC TGC GAT CCT CTC CTT GA-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo5-5’</td>
<td>5’-GAG GCC AAG GAA ATG GAG AC-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>A xo5-3’</td>
<td>5’-CTA CTG GAG GTC CAG GGC-3’</td>
<td>GASP-180</td>
</tr>
<tr>
<td>UPL1-5’</td>
<td>5’-GAA GGT AAA TTA ATG AAA CTT TCC C-3’</td>
<td>UPL-1</td>
</tr>
<tr>
<td>UPL1-3’</td>
<td>5’-CGA GCA AAG AGA GCA GCA TCT CGT-3’</td>
<td>UPL-1</td>
</tr>
<tr>
<td>UDPrs-5’</td>
<td>5’-TAC CAC CTC TTT ATC CGT CC-3’</td>
<td>UPL-1</td>
</tr>
<tr>
<td>UDPrs-3’</td>
<td>5’-CTC CAC AAG GGA ATC CAA GA-3’</td>
<td>UPL-1</td>
</tr>
</tbody>
</table>
6.2.4 Polymerase Chain Reaction

The A2GEX and UPL-1 PCR experiments were conducted using the conditions outlined in Palm et al. (2003). All in house developed PCR reactions were performed in 25 µL reactions using the same base master mix with the only variation being the primers used. Every reaction consisted of 1X PCR buffer, 1.5 mM magnesium chloride, 50 µM of each dNTP, 50 µM of each primer and 1 unit of DNA Taq polymerase (all from Fisher Biotec, Perth, Australia). One microlitre of *G. duodenalis* DNA was added to each of the test reactions and a negative control was used for every PCR run in which no template was added. The base PCR cycling conditions were an initial denaturation at 95°C for 3 mins, followed by 30 cycles of 95°C for 30 s, annealing temperature for 45 s, 72°C for 45 s and a final extension step of 72°C for 5 min. All PCR reactions were performed on a Geneamp 2400 (Applied Biosystems, Foster City, California, U.S.A.). The annealing temperature varied slightly for each PCR. The A2int and A2ex PCR both used an annealing temperature of 50°C, the Axo PCRs all used an annealing temperature of 49°C and the UDPrs PCR used an annealing temperature of 52°C. After the reactions had finished they were either visualised immediately or stored at 4°C.

Reactions were visualised using gel electrophoresis. Agarose gels were prepared at a concentration of 1% w/v in tris acetate EDTA buffer (40 mM Tris, 20 mM glacial acetic acid, 2 mM EDTA) and Sybr Safe was added to dilute 1:25,000. A total of 5 µL of PCR product was mixed with 2 µL of DNA loading buffer (20% w/v sucrose coloured with Orange G) and loaded into the gel. A 100 bp DNA marker (Axygen Biosciences, Union City, U.S.A.) was included with every gel in order to estimate
the size of the PCR products. Gels were run in a tank containing TAE buffer at 80 V for 45 min or until the dye front had run off the gel.

6.2.5 Sequencing

Following visualisation of the PCR products, if a single band was seen the PCR product was purified using the Wizard SV PCR purification kit (Promega). If multiple bands were present the gel was repeated as before except that the remaining 20 µL of the PCR reaction were mixed with 5 µL of DNA loading buffer and loaded into the gel. The band of the correct size was then excised from the gel using a scalpel and the DNA purified as for the PCR product. The concentration of the purified PCR product was estimated by visualising the DNA on a 1% agarose gel as before and comparing the intensity of the purified DNA to the DNA marker.

The sequencing reaction was performed using the BigDye terminator kit version 3.0 (ABI, Foster City) with half reactions in both the forward and reverse direction. Each 10 µL reaction consisted of 4 µL of BigDye mix, 50 mM of primer and 20 ng of DNA from the PCR product. Reactions were run on a thermocycler at 96°C for 2 mins followed by 25 cycles of 96°C for 10 s, annealing temperature for 10 s and 60°C for 4 min. The annealing temperatures used were the same as for the PCR reactions. Sequencing reactions were purified by adding 1 µL of 125 mM EDTA, 1 µL of 3M sodium acetate pH 5.2 and 25 µL of ethanol. The samples were mixed and left at room temperature for 20 mins. Sequencing samples were then centrifuged at 16,000 x g for 30 min and the supernatant was discarded and the pellet left to dry for 10 min. A 125 µL volume of 70% v/v ethanol was added to each sample to rinse the pellet and the sample was centrifuged at 16,000 x g for 5 mins. The supernatant was
discarded and the pellet was dried using a vacuum centrifuge for 10 min. The samples were sequenced at the State Agricultural Biotechnology Centre sequencing facility.

If PCR products failed to give readable sequence after several attempts the PCR products were cloned in order to sequence. Fifty nanograms of PCR product were mixed with 50 ng of pGem-T vector 2 μL of 5X ligation buffer and 1 μL of T4 DNA ligase with the total volume adjusted to 10 μL with nuclease free water (all from Promega, Madison, U.S.A.) and the reactions were left to ligate overnight at 4°C. The next day 2 μL of the ligation reaction was removed to a clean 1.5 mL microcentrifuge tube and mixed with 50 μL of JM109 competent \textit{Escherichia coli} cells (Promega) and incubated on ice for 20 min. Samples were placed at 42°C for 50 secs, then left to cool on ice for 2 min, following which 950 μL of LB broth (Becton Dickson) was added to the sample and the bacteria were left to grow at 37°C for 90 mins. Cells were then concentrated via centrifugation and resuspended in 200 μL of fresh LB broth and 100 μL of this was spread onto duplicate LB agar plates supplemented with 100 μg/mL ampicillin (Sigma), 0.5 mM IPTG and 80 μg/mL X-Gal (both from Fisher Biotec). Cells were incubated overnight at 37°C and examined for the presence of white colonies the next day.

A total of 10 white colonies per sample were picked from plates and used directly in a PCR reaction to test for the presence of the insert and for sub-culturing. Polymerase chain reactions were prepared as above and a small amount of colony was transferred from the growth plates directly to a tube containing the PCR mix using a sterile 200 μL pipette tip. The same tip was then wiped over a section of a
new LB agar plate supplemented with 100 µg/mL ampicillin, 0.5 mM IPTG and 80 µg/mL X-Gal to subculture the colonies tested. The cycling conditions for the PCR were the same as those used before and reactions were visualised as detailed above.

Those colonies containing an insert were then amplified using M13 sequencing primers specific for the vector (m13pUC F 5’-CCC AGT CAC GAC GTT GTA AAA CG-3’, m13pUC R 5’-AGC GGA TAA CAA TTT CAC ACA GG-3’) using the same reagent mix as outlined above. The samples were denatured at 95°C for 3 mins, then underwent 35 cycles of 95°C 30 sec, 52°C 45 sec and 72°C 45 sec, followed by a final extension at 72°C for 7 minutes. The reactions were visualised using agarose gel electrophoresis as outline above. The PCR product was purified using the Wizard SV PCR purification kit (Promega) and the DNA concentration estimated as before. The purified PCR products were sequenced as before.

### 6.2.6 Phylogenetic analysis

Sequencing files were viewed using the FinchTV freeware package (Geospiza, Seattle, U.S.A.). Sequences were checked for the correct base calling and corrected where necessary; if the read was ambiguous the IUPAC degenerative base naming system was used. Consensus sequences were generated from the forward and reverse sequences where possible. Sequences were searched using the NCBI Blast engine with the megablast algorithm, where this did not produce a sequence match the discontinuous megablast algorithm, which allows for a greater divergence between sequences was used. If the sequences showed a high degree of similarity to a *G. duodenalis* sequence, with a low “E-value” and a high “Score”, they were accepted as being analogous.
All sequences generated from the same locus were compared using the Molecular Evolutionary Genetic Analysis (MEGA) version 3.1 freeware package (Kumar et al., 2008). Sequence alignments were performed using the Clustal W algorithm incorporated in the program.

**6.3 Results**

**6.3.1 Alpha 2 giardin**

All three PCR tests used to determine the presence of the alpha-2 giardin gene were only able to amplify a product for the assemblage A isolates (Figures 6.1-6.3). The A2GEX PCR which was developed by Palm et al. (2003) only showed amplicons for assemblage A isolates, however there was a high degree of non-specific amplification (Figure 6.1). The Alph2ex PCR also showed positive results for the assemblage A isolates only with the expected band migrating at 1500 b.p. (Figure 6.2). There were also several bands below 1500 b.p. for both assemblage A and B samples. The third PCR test used, Alph2int, gave bands of the expected size of 500 b.p. for assemblage A isolates and a band at around 1000 b.p. for the assemblage B isolates tested (Figure 6.3). There was no non-specific amplification seen for the Alph2int PCR.
Figure 6.1 A2GEX PCR products run in a 1% agarose TAE gel stained with Sybr Safe
Lane 1 BAH 2c2 (A), Lane 2 BAH 26c11 (A), Lane 3 BAH 40c10 (A), Lane 4 BAH 34c8 (B), Lane 5 BAH 15c1 (B), Lane 6 BAH 12c14 (B), Lane 7 negative control

Figure 6.2 Alph2ex PCR products run in a 1% agarose TAE gel stained with Sybr Safe
Lane 1 BAH 2c2 (A), Lane 2 BAH 26c11 (A), Lane 3 BAH 40c10 (A), Lane 4 BAH 34c8 (B), Lane 5 BAH 15c1 (B), Lane 6 negative control
Only the Alph2int PCR products were sequenced, as they gave a band for both assemblage A and B isolates, however the assemblage B band was double of that expected. The assemblage A sequences generated all gave high scores and low E-values for identities to alpha 2 giardin. Initial sequencing runs of the assemblage B product failed to generate any meaningful data with many mixed nucleotide positions in the sequence. The assemblage B sequences were therefore cloned into the pGEM-T vector and sequenced using primers specific for the vector sequence.

The sequence generated from the cloned assemblage B PCR products did not show homology to any known genetic sequence in the NCBI database using the megablast search algorithm. The search was then changed to use the discontinuous megablast algorithm, which allows for a greater degree of divergence between the sequences searched. This identified the assemblage B band as non-specific amplification of a Wee Kinase. The assemblage B nucleotide sequences only showed 71% sequence identity with the published assemblage A sequence (XM_001708868) over the area tested (Figure 6.4). Investigation of the genome database for the assemblage B GS isolate enabled examination of the full length assemblage B sequence (GenBank...
gene ID GL50581_3347), there was only 70% sequence similarity at the nucleotide level between and the assemblage A and B wee kinase and only 63% similarity at the amino acid level. This is an extremely low level of the
Figure 6.4 Alignment of assemblage B sequences obtained from the 1000bp amplicon of the Alpha2int PCR with the genomic Assemblage A Wee Kinase sequence.

The number before the decimal point denotes the sample ID, 7 is BAH 7c5 and 15 is BAH 15c1. The number after the decimal point represents the bacterial clone which was used for the sequencing. The F or R denotes a forward or reverse sequence.
6.3.2 GASP-180

Five separate PCR reactions were used to try and amplify the whole GASP-180 gene. All PCR reactions were able to amplify a PCR product from assemblage A isolates however only one, the Axo3 PCR, was able to amplify assemblage B sequences (Figure 6.5). Positive samples for the Axo3 PCR were sequenced and were all found to be similar to the published axoneme associated proteins GASP-180 sequence available in the NCBI genetic database. Analysis in MEGA showed several substitutions along the length of the sequence analysed for the assemblage B isolates when compared to the assemblage A sequence (Figure 6.6). Overall there are 104 nucleotide substitutions over the 648 base pairs available for all sequences including the insertion of three nucleotides as a group into the assemblage B sequence.

![Figure 6.5 Axo3 PCR products run in a 1% agarose TAE gel stained with Sybr Safe](image)

Lane 1 DNA Marker; Lane 2 BAH 2c2; Lane 3 BAH 26c11; Lane 4 BAH 34c8; Lane 5 BAH 15c1; Lane 6 BAH 7c5; Lane 7 negative control
Figure 6.6 Alignment of Axo3 PCR amplicon sequence from assemblages A and B with the assemblage A genomic sequence.

2 is BAH 2c2, 34 is BAH 34c8, 15 is BAH 15c1 and 7 is BAH 7c5. F and R denote forward and reverse.
6.3.3 Uridine phosphorylase like protein-1

Two separate PCR tests were used for the analysis of the UPL-1 gene. The initial PCR performed was first developed by Palm et al. (2003), however positive PCR tests were not achieved. The in house designed PCR yielded positive PCR results for both assemblage A and assemblage B samples (Figure 6.7). The positive PCR samples were sequenced and searched against the NCBI nucleotide database using the megablast algorithm with 2 assemblage B and 1 assemblage A samples successfully sequenced. All gave low E values and high scores for homology matches with *Giardia* UPL-1. As before sequences were further analysed using MEGA to align the files using the Clustal W algorithm, of the 699 bp examined of the 1115 bp gene there were no nucleotide differences in the nucleotide sequence between assemblages A and B (Figure 6.8).

![Fig 6.7 UDP PCR products run in a 1% agarose TAE gel stained with Sybr Safe](image)

Lane 1 DNA marker; Lane 2 BAH 2c2; Lane 3 BAH 26c11; Lane 4 BAH 7c5, Lane 5 BAH 15c1; Lane 6 BAH 34c8; Lane 7 negative control
<table>
<thead>
<tr>
<th>Species/Abbrv</th>
<th>Sequence Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. UDP 2 F</td>
<td></td>
</tr>
<tr>
<td>2. UDP 2 R</td>
<td></td>
</tr>
<tr>
<td>3. UDP 7 F</td>
<td></td>
</tr>
<tr>
<td>4. UDP 7 R</td>
<td></td>
</tr>
<tr>
<td>5. UDP 15 F</td>
<td></td>
</tr>
<tr>
<td>6. AFS14364.1</td>
<td>UPE-1 gene</td>
</tr>
</tbody>
</table>

**Figure 6.8 Alignment of UDP PCR amplicon sequences from Assemblages A and B with the assemblage A genomic sequence.**

2 is BAH 2c2, 7 is BAH 7c5 and 15 is BAH 15c1. F and R denote forward and reverse.
6.4 Discussion

6.4.1 Alpha 2 giardin

The protein alpha 2 giardin was found in assemblage A isolates but was absent from assemblage B protein gels of human infective *Giardia duodenalis*. To determine if the protein was truly absent or if it was under differential expression control in assemblage B the genome was examined using PCR analysis. The three separate PCR tests used were targeted to different regions of the alpha 2 giardin gene in order to maximise the likelihood of obtaining a positive result. All three PCR tests failed to amplify the alpha 2 giardin gene from any of the assemblage B isolates tested confirming the protein analysis. There were several bands of the wrong size seen in the A2Gex and Alph2ex PCR assays for both assemblage A and B. This could be due to the relatedness of the alpha giardins and that primers are amplifying other closely related alpha giardins from the genome.

Alpha 2 giardin, like all alpha giardins, is a *Giardia*-specific structural protein related to the annexin class of proteins (Morgan and Fernandez, 1995). Annexins are Ca$^{+2}$ dependent phospholipid binding proteins. Alpha 2 giardin itself is known to associate with the plasma membrane and the flagella of the trophozoite (Weiland *et al.*, 2005). Alpha 2 giardin is also a target of the host immune response with a study by Palm *et al.* (2005) identifying it as an immuno-dominant protein in Western blot analysis. The exact function of alpha 2 giardin is unknown, however it is hypothesised to play a role in anchoring the plasma membrane to the cytoskeleton and may have a role in motility due to its association with the caudal flagella (Weiland *et al.*, 2005).
The absence of alpha 2 giardin in assemblage B is significant for several reasons. Firstly, this represents a novel phenotypic difference between the human infective assemblages of *G. duodenalis* and such a difference has never before been noted between the assemblages. This raises questions as to the speciation of assemblages of *G. duodenalis* which has been a contentious issue as the level of genetic difference between some of the assemblages of *G. duodenalis* is as great as between other genera of parasites. However, there is a lack of phenotypic differences to back up these genetic differences (Mayrhofer *et al.*, 1995a). In this study we have identified two proteins that are variable between the assemblages of *G. duodenalis* which also vary at the genetic level, representing phenotypic differences with a defined genetic basis.

Secondly, if alpha 2 giardin has such an important role in anchoring the plasma membrane to the cytoskeleton, as it has been hypothesised, how do assemblage B isolates cope with the loss of this integral protein? The likeliest explanation is that assemblage B isolates have their own alpha giardin-like protein that performs this role. This could allow for a stronger binding of the cytoskeleton to the plasma membrane increasing the integrity of the trophozoite. A recent study undertaken by Franzen *et al.* (2009) to sequence the assemblage B GS isolate reported the presence of an alpha 2 giardin analogue in assemblage B. The reported assemblage B analogue has 80% sequence similarity with the assemblage A gene sequence and 81% identity when the amino acid sequence is compared, which is rather low. When compared to other alpha giardins the putative assemblage B alpha 2 giardin actually displays more similarity with assemblage A alpha 1 giardin with 88% identity at the
nucleotide level. This indicates that the purported assemblage B alpha 2 giardin may not actually be alpha 2 giardin but rather a second alpha 1 giardin variant or as hypothesised above an assemblage B specific alpha giardin.

6.4.2 Genetic Variation of Wee Kinase

For the Alph2int PCR both assemblage A and assemblage B samples gave a positive result in the form of the expected 500bp band for assemblage A isolates and a 1000bp band for the assemblage B isolates. When sequenced the 500 bp band for assemblage A isolates was found to be alpha 2 giardin as expected. The assemblage B band could not initially be sequenced effectively. The sequences generated were very poorly resolved with multiple mixed base positions. The PCR products were therefore cloned into the pGem-T vector for sequencing of individual PCR products using primers specific to the vector. This method obtained clean sequence reads for all samples.

The assemblage B sequences generated did not match any G. duodenalis genetic sequence within the NCBI genetic database when Blast searched using the megablast algorithm. This algorithm is designed to provide homology information for sequences that have at least 95% sequence similarity. The discontinuous megablast algorithm was then used for homology searches. This allows for a much looser association between sequences tested and is recommended for use in inter-species analysis (www.ncbi.nlm.nih/blast). Using this analysis the assemblage B Alph2int PCR product was identified as non-specific amplification of a Wee Kinase. The assemblage B Wee Kinase genetic sequence showed 70% similarity with the assemblage A sequence generated from the Giardia genome project explaining why
it was not identified using megablast, which requires at least 95% sequence similarity in order to find a match.

The wee kinase protein was first described in the yeast *Schizosaccharomyces pombe* as being involved in mitosis (Russell and Nurse, 1987). The process of mitosis in this yeast relies on the activity of the protein kinase cdc2 (cell division cycle 2). Wee kinase acts to deactivate this protein by phosphorylating specific serine and tyrosine residues within the protein, therefore delaying the start of mitosis (Featherstone and Russell, 1991). A second cdc protein, cdc25, works in opposition to wee kinase by dephosphorylating the cdc2 protein, activating it (Rowley *et al*., 1992). Yeast cells deficient in wee kinase display a reduced size when compared to the wild type strains, referred to as the ‘wee’ (little) phenotype. These cells enter mitosis at a much earlier stage in their life cycle and are therefore limited in their growth potential. Cells that are forced to over express the wee kinase are 2 to 3 times larger than the wild type cells, as mitosis is delayed.

The level of variation in the wee kinase of the assemblages is interesting. With only 70% similarity between assemblages A and B for this locus it is strongly suggestive of species level classification between the assemblages. The functional significance of this variation is unclear. As the wee kinase has a role to play in cell division it could be assumed that variation in the protein could lead to a change in the mean generation time of the isolates as a more efficient wee kinase could delay mitosis and therefore increase the mean generation time of the isolates, and vice versa. However there is no evidence of variation in mean generation time based on the assemblage of isolates (Binz, 1996). Alternatively, assemblages A and B may have variation in the
cdc2 homolog which would require different wee kinase proteins to phosphorylate them, delaying the onset of mitosis.

Investigation of the genome database for the assemblage B GS isolate enabled examination of the full length assemblage B sequence (GenBank gene ID GL50581_3347), there was only 70% sequence similarity at the nucleotide level between and the assemblage A and B wee kinase and only 63% similarity at the amino acid level. This is an extremely low level of similarity between two analogous genes. A conserved domain search of the assemblage A sequence identifies two separate protein koinase domains within the protein sequence, however no such domains are found in the assemblage B sequence. This raises the question of whether the assemblage B Wee Kinase analogue is not functionally active, further testing would be required to test this.

6.4.3 GASP-180

For the analysis of the GASP-180 five individual PCR tests were developed in an attempt to amplify the whole gene for both assemblage A and assemblage B. All primer sets were able to amplify assemblage A sequences however only the Axo3 primer pair gave reliable amplification of both assemblages A and B. Overall there was an 85% sequence similarity between the assemblage A and B sequences tested. This level of diversity may explain why only the one PCR gave reliable results for assemblage B as there was likely variation in the other primer sites, which would not facilitate binding of the primer to the complimentary DNA, therefore a PCR product would not be produced.
As its name suggests the GASP-180 protein associates with the axoneme of the trophozoite. A monoclonal antibody specific to a peptide of GASP-180 bound to the intracellular axonemes (Elmendorf et al., 2005). The axonemes are the internal portion of the flagella. They extend from the basal bodies between the nuclei to the point in the cell surface where the flagella emerge (Elmendorf et al., 2003). The axonemes of the anterior flagella start at the basal bodies, cross and loop around the top of the nuclei along the ventral flange and come out in line with the nuclei with the flagella trailing to the posterior of the cell. The other six axonemes extend through the middle of the cell towards the posterior end. The posterior lateral flagella axonemes diverge and emerge from the cell approximately half way along the length of the trophozoite. The caudal and ventral flagella continue along the midline with the ventral flagella emerging in the ventral groove and the caudal axonemes continuing to the posterior tip of the trophozoite where the flagella emerge (Elmendorf et al., 2003). Due to the involvement of GASP-180 with the axonemes it has been hypothesised to have a role in the motility of the trophozoite (Elmendorf et al., 2005).

The large amount of variation seen at the GASP-180 locus is of interest regarding the functionality of the protein. If GASP-180 has an integral role in motility as hypothesised does the variation affect the movement of trophozoites? The movement of trophozoites is not the only important function of the flagella. One of the theories of attachment of trophozoites is that the constantly beating flagella push the adhesive disc against a surface creating an area of negative pressure underneath the disc, allowing the trophozoite to attach to the surface using suction (Holberton, 1974). Variation in the GASP-180 protein could allow for the flagella to beat at a faster rate.
or with greater force, increasing the strength of the suction-like attachment. Alternatively, the GASP-180 protein may not directly drive movement but may allow for a stronger attachment of the flagella to the axoneme therefore increasing the amount of force able to be generated. Evidence that assemblage B isolates are more likely to cause chronic infection could be due to their ability to attach to the gut surface with greater strength, therefore stopping them from being removed from the host via mechanical disruption (Read et al., 2002, Haque et al., 2005, Sahaqun et al., 2008).

A recent paper studying adhesion deficient clones of the WB isolate of *G. duodenalis* which belongs to assemblage A identified a protein migrating around 200 kDa that was absent from those cells that were unable to adhere to cell monolayers (Hernandez-Sanchez et al., 2008). This protein appears to migrate at the same rate as the GASP-180 protein however MS data was not included in the study. If the protein absent from the adherent deficient clones is the same as the GASP-180, it would support the theory that it is involved in attachment of the trophozoite to a substrate related to the suction type attachment of the trophozoite.

A recent study has found that the flagella themselves are not directly required for the trophozoite to maintain attachment to a surface (House et al., 2011). Trophozoites were transfected with a vector which reduced the expression of one of flagella proteins, subsequently these trophozoites and malformed, non-functional flagella. Although the trophozoites were able to maintain attachment without properly functioning flagella the rate of attachment was significantly reduced (House et al., 2011). This indicates the flagella still have an important role in attachment and if an
isolate is able to move at a faster rate against the peristaltic flow within the small intestine it is more likely to be able to reach the mucosal surface so it can attach. Therefore variation in GASP-180 between the assemblages could still have a role in the more chronic nature of assemblage B infections by increasing the likelihood of assemblage B isolates making their way to the mucosal surface.

An investigation of the available genome sequence for assemblages A and B identifies four GASP-180 genes in both assemblages. These vary in size, with the same sizes for both assemblages; the predicted proteins are 104 kDa, 116 kDa, 174 kDa and 237 kDa. Based on the location of the bands on the gels it would seem that assemblage A isolates are predominantly expressing the 174 kDa variant and assemblage B isolates are expressing the 237 kDa variant. For all of the proteins there is between 85% and 90% similarity for each of the GASP-180 size variants at the amino acid level. Further studies looking at the level of expression of each of the GASP-180 variants for the two assemblages need to be undertaken to determine the dominant GASP-180 type for the assemblages.

**6.4.4 Uridine Phosphorylase-like Protein**

From the 2-DE analysis of assemblages A and B the uridine phosphorylase like protein of assemblage B had a more basic pI than the assemblage A protein. The initial PCR analysis using the UPL1 primers failed to give repeatable positive results, as such a second PCR test was designed using the UPL-1 genetic sequence. This resulted in a positive PCR result for assemblage A and B isolates. These samples were then sequenced and compared to the genomic sequence for UPL-1. Both the assemblage A and assemblage B sequences were found to be identical to the genomic sequence over the 700 bp tested of the 1115 bp gene.
Using the giardia genome database the whole protein amino acid sequence of UPL1 for assemblage A (XP_001707342) and assemblage B (EET01236) was compared. Over the 310 amino acids there was 95% similarity between the two assemblages and 97% positivity, which accounts for amino acid substitution where the two amino acids have similar physio-chemical properties. The area encoded by the 700 bp tested above is homologous for assemblages A and B with all substitution occurring after the first 123 amino acids. The predicted isoelectric points for the assemblage variants differ by 0.4 pH units which is less than the distance observed in the 2D PAGE gels.

The homology of the genetic sequences between the assemblages is unusual given the differences seen at the protein level. The difference at the protein level is therefore likely to be due to variation in the post translation modification of the enzyme. Once a protein is translated at the ribosome it is often modified with different chemical groups attached to specific amino acid positions. These modifications are commonly used to change the activation state of the enzyme within a system, normally by changing the phosphorylation state (Lalle et al., 2006). The nature of the post translational modification of this protein is unknown; however, it would appear to be significant given the change in pI seen between the assemblage A and B isolates.

**6.4.5 Conclusions**

The overall aim of this series of experiments was to determine if differences at the protein level correlated with differences at the genetic level. This was found to be
dependent on the locus investigated. The alpha 2 giardin gene was unable to be amplified from the assemblage B genome using three separate PCR tests, agreeing with the proteomic data where the protein was only present in the assemblage A gels. The GASP-180 protein which showed size variation at the protein level showed only 85% similarity over the 648 bp examined. Finally the uridine phosphorylase like protein showed no variation at the genetic level over the 700 bp examined, whereas at the protein level there was a clear shift in the pI of the protein towards the basic end for assemblage B isolates which is greater than the 0.5 pH unit shift predicted from the full amino acid sequence. This highlights why combined approaches are best used when examining variation. The shift in pI of the UPL-1 protein could only be found using proteomic investigation as it appears to be due to a post translational modification. All other methods for studying variation would fail to identify this change between the assemblages.
Chapter 7 Comparison of the encystation rate of *Giardia duodenalis* isolates from assemblages A and B

7.1 Introduction

All species of *Giardia* have a two staged life cycle, the trophozoite that colonizes the small intestine of the host and the cyst that is passed into the environment with the faeces and is the infective particle. The cyst itself is a partially replicated trophozoite that has undergone DNA replication, but is yet to undergo cytokinesis, encased in a carbohydrate and protein matrix. The major carbohydrate is N-acetylagalactoseamine with four proteins ranging from 29-102 kDa in size.

The process of encystation is complex and the complete pathway is not clearly defined. It was originally thought that the presence of bile was required for encystation as incubating trophozoites in a high bile medium promoted encystation. A study by Lujan *et al.* (1996) investigated the effect of cholesterol deprivation on trophozoites, as *G. duodenalis* requires cholesterol for membrane production but is unable to *de novo* synthesise cholesterol. Lujan *et al.* (1996) found that by depriving trophozoites of cholesterol they were able to trigger encystation. The high concentration of bile commonly used to promote encystation *in vitro* was hypothesised to decrease the amount of free cholesterol, thereby triggering cyst formation (Lujan *et al.*, 1996). The molecular processes that drive encystation are poorly defined.

One protein that has been implicated in the control of encystation is the product of *Giardia lamblia* open reading frame C4 (commonly referred to as GLORF-C4)
(Nores et al., 2009). GLORF-C4 is upregulated during encystation and by decreasing the levels of GLORF-C4 within the trophozoite the onset of encystation can be delayed. From the SDS-PAGE and mass spectrometry data outlined in Chapters 3 and 5, GLORF-C4 was repeatedly identified in assemblage B and E isolates but never from assemblage A. To confirm the protein data observed, the expression of the glorf-c4 gene was determined by performing quantitative reverse transcriptase PCR of mRNA isolated from trophozoites of assemblage A and B. As GLORF-C4 regulates the encystation process the rate of encystment was determined for isolates from assemblages A and B to resolve firstly if GLORF-C$ is constitutively expressed in assemblage B and induced in assemblage A and secondly if this constitutive expression increases the rate of encystation. Along with this, the effect of bile source, sera source and variation in pH on percentage of cells encysted was also tested.

7.2 Materials and Methods

7.2.1 Encystation

Trophozoites were cultured and harvested as outlined in Chapter 2. Trophozoites from a 50mL culture were harvested via centrifugation at 2,500 x g for 10mins and washed twice in PBS. Trophozoites were resuspended in 1mL of sterile BS-I-33 medium (pH 7.6-7.8) containing 1% w/v bovine bile and cultures were left at 37°C to allow trophozoites to encyst. Cultures were measured microscopically to assess encystation. Cells displaying typical cyst morphology were assessed as being fully formed cysts. Samples were also stained with Giardi-a-Glo A300FLK (Waterbourne Inc, New Orleans), which is specific for cyst wall protein (CWP) a marker of
encystation, to ascertain the number of trophozoites undergoing encystment. Cells that stained positively for CWP but lacked typical morphology or did not show uniform staining with Giardi-a-Glo were counted as partially encysted trophozoites.

7.2.2 Effect of Bile Source on Encystation

Trophozoites were induced to encyst as described above except that cells were induced to encyst with either 1% w/v bovine bile or 1% w/v porcine bile and left at 37°C for 96 hours, and reactions were performed in triplicate. After incubation the percentage of encysting cells (number of cysts plus the number of trophozoites staining positively for CWP) was calculated for each condition.

7.2.3 Effect of pH on Encystation

Trophozoites were induced to encyst as described above, however four different pH points were used for the culture medium, 6.6, 7.6 and 8. Trophozoites were incubated at 37°C for 96 hours to allow encystment with the percentage of encysting cells calculated for each condition. All tests were performed in triplicate.

7.2.4 Effect of Sera on Encystation

The type of sera used in the encystation culture medium was also varied. Trophozoites were induced to encyst as above except that the BS-I-33 medium was prepared using either 10% v/v new born calf serum (NBS) or 10% v/v foetal bovine serum (FBS), with 1% w/v bovine bile. The cultures induced with FBS were grown in the presence of FBS prior to encystation induction. Trophozoites were incubated
at 37°C for 96 hours in triplicate to encyst with the percentage of encysting cells calculated for each condition.

### 7.2.5 Encystation Rate

To study the encystation rate, cells were induced to encyst as above then monitored microscopically every 24 hours for a total of 96 hours. Experiments were prepared in triplicate. The number of cysts and partially encysting trophozoites were calculated as a percentage of the total number of cells counted, with the number of cysts and partially encysted cells counted separately.

### 7.2.6 Statistical Analysis

The significance of the results was tested using the SPSS statistics package. A paired sample t-test was used to compare the data with a confidence interval of 95% used for all analysis.

### 7.2.7 RNA Extraction and Conversion to cDNA

Ribonucleic acid was extracted from *G. duodenalis* trophozoites and encysting trophozoites at 8 and 18 hours post induction using the TRI® reagent (Sigma, UK) method. Cells were enumerated and approximately 1x10^7 cells were mixed with an equal volume of TRI® reagent and the sample mixed by repeated pipetting, then allowed to stand at room temperature for 5 minutes. An equal volume of chloroform was added to the cell mixture which was allowed to stand for 15 minutes at room temperature and then centrifuged at 12,000g for 15 min at 4°C. The aqueous phase was transferred to a fresh tube and 0.5mL of isopropanol was added. This was mixed
and allowed to stand for 10 min at room temperature then centrifuged at 12,000g for 10 min. The pelleted RNA was washed in 1mL of 70% v/v ethanol and the RNA pellet was allowed to air dry for 10 minutes after which it was redissovled in 0.2mL of DEPC treated water and the concentration and purity of RNA estimated spectrophotometrically by measuring absorbance at 260 and 280 nm. The RNA solution was stored at -80°C when not in use.

Total mRNA was reverse transcribed using the Enhanced Avian HS RT-PCR kit (Sigma, UK) using the two stage method with oligo-dT as primer. The procedure was carried out as per the manufacturers instruction with a pre-incubation at 70°C for 10 minutes followed by incubation at 45°C for 50 min during which time mRNA was converted to cDNA. The cDNA was stored at -80°C for long term storage.
7.2.8 Quantitative Real Time PCR of glorf-c4 Expression

A relative quantitative real time PCR of glorf-c4 expression was carried out using the glyceraldehydes 3-phosphate dehydrogenase (gapdh) gene as a reference. The primers used for the analysis are displayed in Table 7.1 and were designed using Primer3 software package. For the real time PCR 1μL of cDNA was added to 19μL of SYBR green master mix (Roche, Germany) with 0.25μM of the relevant primers added. The PCR was conducted on a Roche Light Cycler 480 using a standard protocol, with data analysed using the inbuilt software package and results expressed as a ratio between glorf-c4 and gapdh expression. All reactions were conducted in triplicate. Results generated were statistically tested using a Mann-Whitney U test.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH F</td>
<td>5′ GTA CGA TCC GAG CAC GAT GAA-3′</td>
</tr>
<tr>
<td>GAPDH R</td>
<td>5′-GCC CTC CTT GAT ACC AAA CTT CT-3′</td>
</tr>
<tr>
<td>GLORF F</td>
<td>5′-AGC TCA TCA TCG TCC TCT A-3′</td>
</tr>
<tr>
<td>GLORF R</td>
<td>5′-CAA TCT TGT TTG CAT ACG A-3′</td>
</tr>
</tbody>
</table>
7.3 Results

7.3.1 Effect of Bile on Encystation

The percentage encystation achieved with bovine and porcine bile is shown in Figure 7.1. All isolates except BAH 12c14 had a significantly higher percentage of encysting cells for bovine bile when compared to porcine bile (p<0.05). BAH 12c14 was the only isolate that had a significantly higher number of encysting cells for porcine bile when compared to bovine bile (p<0.05).

![Figure 7.1 Percentage Encystation of Trophozoites incubated with 1% w/v Porcine and Bovine bile](image)

Isolates 2c2 and 3c3 are representative of assemblage A and 34c8 and 12c14 are representative of assemblage B. * indicates significant difference p<0.05.
7.3.2 Effect of Variation of pH on Encystation

Four separate pH points were tested to determine if the different assemblages showed variation in the pH at which maximal encystation would occur (Figure 7.3). All isolates achieved the highest percentage of cysts when induced at a pH of 7.6 (p<0.05) and the lowest at a pH of 6.6 (Figure 7.3).

![Figure 7.3 Effect of Variation in pH on Encystation of *Giardia duodenalis* from Assemblages A and B](image)

Isolates 2c16 and 3c3 are from assemblage A and isolates 34c8 and 12c14 are from assemblage B.

7.3.3 Effect of Sera on Encystment

The type of sera used in the culture and encystation was varied to determine their role in encystation. For both assemblage A (BAH 2c2) and assemblage B (BAH34c8) isolates the combination of new born calf serum with a high bile environment (HB NBS) resulted in significantly higher levels of encystment than foetal bovine serum in a high bile environment (HB FBS) (p<0.05, Figure 7.2). A low background level of encystment was still present under normal culture conditions.
conditions (Norm FBS and Norm NBS). Results are presented as the number of cysts, as ascertained by staining with Giardi-a-Glo

![Graph showing encystation rates with error bars]

**Figure 7.2 Effect of New Born Serum and Foetal Bovine Serum on the Percentage of Trophozoites Undergoing Encystation**

Isolate 2c2 is from assemblage A and isolate 34c8 from assemblage B. FBS is foetal bovine serum and NBS is new born calf serum. Norm is a normal bile concentration of 0.05% w/v used for culturing Trophozoites. HB is a high bile concentration of 1% w/v bile. * denotes significant difference p<0.05

**7.3.4 Variation in Encystation Rate**

For the assemblage A isolates both BAH 2c2 and BAH 3c3 had an initial increase in the number of encysting trophozoites post induction with a peak at 48 hours after which time the number of encysting trophozoites present decreased (Figure 7.4).

The number of cysts increased slowly post induction then, between 48 and 72 hours, the number of cysts increased markedly. After 72 hours the number of cysts continued to increase but at a decreased rate.
Figure 7.4 Encystation Rate for Assemblage A Isolates BAH 2c16 and BAH 3c3
 ■ denotes isolate BAH 3c3 and ● isolate BAH 2c16. The broken line plots the number of trophozoites that stain positively for the presence of cyst wall protein. The solid line plots the number of cysts present.

The assemblage B isolates BAH 12c14 and BAH 34c8 both had a steady increase in the number of cysts to 48 hours post induction after which time the number of cysts continued to increase but at a slower rate (Figure 7.5). The number of encysting trophozoites remained low throughout the experiment peaking at 24 hours.
Figure 7.5 Encystation Rate for Assemblage B Isolates BAH 34c8 and BAH 12c14

■ denotes isolate BAH 12c14 and ▲ isolate BAH 34c8. The broken line plots the number of trophozoites that stain positively for the presence of cyst wall protein. The solid line plots the number of cysts present.

7.3.5 Comparison of glorf-c4 Expression in Trophozoites

For the analysis of glorf-c4 expression in trophozoites the data from three separate cultures of each isolate was used for comparison, with each culture assayed for glorf-c4 expression in triplicate. The two isolates investigated from assemblage B both displayed a higher level of glorf-c4 expression compared to gapdh with 7.5 times the level of expression (Figure 7.6). Assemblage A isolates expressed 4 times less glorf-c4 than gapdh. This equates to around a 10 fold difference in expression of glorf-c4 mRNA in assemblage B when compared to assemblage A which was found to be significantly different using the Mann-Whitney U test (p<0.1).
7.3.6 Analysis of glorf-c4 Expression during Encystation

The expression of glorf-c4 during the early stages of the encystation process was also examined. For the assemblage B isolates (34c8 and 12c14) there was no significant increase in the level of glorf-c4 expression over the 18 hours (Figure 7.7). The assemblage A isolates showed a significant increase in the level of glorf-c4 expression up to 8 hours post induction (p<0.1), however any further increases were not significant.
Figure 7.7 Expression of glorf-c4 before and 8 and 18 hours post induction for encystment
Isolates 34c8 and 12c14 are from assemblage B and isolates 3c3 and 2c16 from assemblage A.

7.4 Discussion

The differentiation of a trophozoite into a cyst is a major factor in the transmission of *G. duodenalis* as it is the cyst which is the environmentally resistant infectious stage. Variation in the encystation process has been investigated in one study previously by Kane *et al.* (1991), however this was before the designation of assemblages (Kane *et al.*, 1991). This is the first study to assess the impact of assemblage on encystation. We have not only examined the rate of encystation but also if encystation can be affected by variation in the induction process, in an assemblage dependent manner. All of the major data is summarised in Table 7.2
Table 7.2 Summary of data for the variation in encystment and expression of *glorf c4*

<table>
<thead>
<tr>
<th>Assay</th>
<th>BAH 2c16 (A)</th>
<th>BAH 3c3 (A)</th>
<th>BAH 34c8 (B)</th>
<th>BAH 12c14 (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of bile producing maximum cysts (porcine or bovine)</td>
<td>Bovine</td>
<td>Bovine</td>
<td>Bovine</td>
<td>Porcine</td>
</tr>
<tr>
<td>Optimum pH of encystation medium</td>
<td>7.6</td>
<td>7.6</td>
<td>7.6</td>
<td>7.6</td>
</tr>
<tr>
<td>Time taken for cyst numbers to peak post induction (hrs)</td>
<td>72</td>
<td>72</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td><em>glorf c4</em> expression pre induction*</td>
<td>0.25</td>
<td>0.5</td>
<td>6</td>
<td>4.3</td>
</tr>
<tr>
<td><em>glorf c4</em> expression 8 hrs post induction*</td>
<td>2.5</td>
<td>2</td>
<td>8.2</td>
<td>5.8</td>
</tr>
</tbody>
</table>

*Expression is presented as a ratio compared to gapdh expression i.e. 2.5 is 2.5 times more *glorf c4* than gapdh.

7.4.1 Variation in Encystment Induction

The impact of pH, bile type (porcine and bovine) and serum source (NBCS and FCS) on encystation was tested for assemblages A and B of *Giardia duodenalis*. All isolates had the same optimum pH of 7.6 which is similar to the optimum of 7.8 identified by Kane *et al.* (1991). In all cases an acidic pH of 6.6 produced significantly lower levels of cyst induction. This is not surprising given that as trophozoites move distally from the start of the small intestine the environment becomes increasingly more basic. This has been hypothesised as a factor contributing to the trophozoites undergoing encystation and the results presented here indicate that a more basic environment is conducive to an increased cyst production (Kane *et al.*, 1991); however as the pH continues to increase there is a negative effect on the levels of cyst production.

The source of bile used was varied to ascertain if the assemblages would respond differently to varied sources of bile. Isolates BAH 2c2 (A), BAH 3c3 (A) and BAH 34c8 (B) all had a significantly higher percentage of encysting cells when incubated
with bovine bile when compared to porcine bile. It is possible that as the isolates are
normally grown in medium containing bovine bile they may become conditioned to
it therefore responding better in higher concentrations. Interestingly isolate BAH
12c14 had a significantly higher percentage of cells undergoing encystation when
incubated with porcine bile. As for the other isolates tested BAH 12c14 is regularly
cultured in medium containing bovine bile. This implies that there is a component of
porcine bile absent from bovine bile that is triggering higher levels of encystment in
isolate BAH 12c14.

The effect of bile on the assemblage A isolate WB has been discussed in several
previous studies. Gillin et al. (1988 and 1989) examined the effect of human,
porcine and bovine bile on encystation as well as the effect of different bile salts.
They found that porcine bile routinely produced higher number of cysts than either
bovine or human bile (Gillin et al., 1989). A study by Kane et al. (1991) also using
the WB isolate found that the use of bacteriological grade bovine bile resulted in a
similar number of cysts to the porcine bile used by Gillin, et al. (1989), although at
20 times the concentration. The study by Gillin et al. (1989) also investigated the
composition of the human, porcine and bovine bile to determine if differences in the
bile components could account for variation in the ability to trigger encystation.
They found that porcine bile had a higher proportion of chenodeoxycholate and
hyocholate bile salts whereas the bovine bile was predominantly cholate and
deoxycholate. The human bile contained predominantly cholate and
chenodeoxycholate bile salts (Gillin et al., 1989). In a separate study cholate was
found to produce three times as many cysts as chenodeoxycholate, however th
reason behind this is unknown (Gillin et al., 1988).
Subsequent studies to that by Gillin *et al.* (1989) have found that the encystation process can occur in the absence of bile and that cholesterol starvation is a key trigger in inducing encystment (Lujan *et al.*, 1996). Bile itself is believed to reduce the availability of cholesterol to the trophozoites. However, if bile is only required to reduce the availability of cholesterol then varying the source of bile should not change the ability of an isolate to encyst. This is not the case as demonstrated in the results of this Chapter and also by Gillin *et al.* (1988 and 1989), which indicate that there is another factor in the bile that must contribute in stimulating encystation.

The significantly higher levels of encystation seen for BAH 12c14 incubated with porcine bile may be due to the presence of hyocholate in the porcine bile, which is absent from human bile and only present in low percentages (0.9% of total bile salts) in bovine bile (Gillin *et al.*, 1989, Lubec and Afjehi-Sadat, 2007). This intra-assemblage variation is important as it indicates that the number of cysts released in the faeces is not indicative of the level of infection as other factors are involved in the regulation of the encystation process, which may be strain dependent.

All isolates produced significantly higher numbers of cysts when incubated with NBCS compared to FCS. This is likely due to the isolates being placed under a higher level of stress when grown with NBCS. Foetal calf serum is commonly used for the cultivation of fastidious organisms and is thought to be more nutrient rich and contain growth factors absent from NBCS. The encystation process is the trophozoite’s mechanism for surviving within stressful environments. The extra
nutrients supplied by the FCS likely reduce the amount of nutritional stress the trophozoites are under leading to a decreased rate of encystation.

7.4.2 Variation in Encystation Rate

The progression of encystation over time was studied for two assemblage A (BAH 2c2 and BAH 3c3) and two assemblage B isolates (BAH 34c8 and BAH 12c14). The percentage of formed cysts as well as encysting trophozoites was enumerated for each isolate at 24, 48, 72 and 96 hours post induction. This is the first time that a comparison between genotyped organisms has been made and that the number of encysted and encysting cells has been quantified and compared. The results in Figure 7.4 show that the assemblage A isolates had a steady increase in the number of partially encysted trophozoites up to 48h post induction and subsequently, between 48h and 72h, the number of encysting trophozoites started to decrease and the number of fully formed cysts increased significantly. Between 72h and 96h the number of fully formed cysts continued to increase but at a much slower rate. The assemblage B isolates represented in Figure 7.5 displayed a steady and steep increase in the number of cysts up to 48h post induction, from 48h to 96h the number of cysts produced continued to increase but at a greatly reduced rate.

The results indicate that the assemblage B isolates are undergoing the encystation process approximately 48h earlier than the assemblage A isolates. There is also a decreased number of partially encysting trophozoites present in the assemblage B isolates indicating that the assemblage A isolates have a longer lag phase, in which they are likely producing the proteins necessary for encystation. This would indicate the assemblage B isolates are either more efficient at producing the encystation
specific proteins after induction or that some of the proteins are already being expressed at the trophozoite stage.

### 7.4.3 Variation in glorf-c4 gene expression

From the SDS-PAGE analysis and protein identification, outlined in Chapters 3 and 5 respectively, the protein GLORF-C4 was identified as a component of the assemblage B trophozoite proteome but was absent from the assemblage A proteome. When the amount of mRNA present in the trophozoite stage was tested, Assemblage B isolates exhibited levels of transcription 10 times higher than assemblage A isolates, correlating with the inability to identify the GLORF-C4 protein on assemblage A protein gels. When trophozoites were induced to encyst the assemblage A isolates displayed a significant increase in the level of glorf-c4 expression which is similar to previously reported results (Yang et al., 2002, Yong et al., 2002). The levels of assemblage B transcription did increase post induction but not at levels considered significant. This is further support for the theory that glorf-c4 is constitutively expressed in assemblage B isolates whereas assemblage A isolates only express large amounts of glorf-c4 upon induction for encystation. This difference in the regulation of expression of glorf-c4 appears to affect the rate at which the different assemblages can undergo the encystations process.
7.4.4 General Conclusions

GLORF-C4 has structural similarities to α-crystallin heat shock proteins and is thought to help maintain integrity of cyst proteins prior to incorporation into the cyst wall (Nores et al., 2009). The constitutive expression of GLORF-C4 in assemblage B may decrease the rate of encystment as the trophozoites do not need to synthesise the chaperone proteins prior to production of the proteins required for the cyst wall.

This comparative study of the encystation process of *Giardia duodenalis* has raised some interesting issues. Firstly the bile used in the encystment induction must have a more significant role than just depriving the trophozoites of cholesterol, as variation in the type of bile causes variation in the level of encystment. Secondly and most importantly the encystation rates of assemblage A and B isolates differ markedly, indicating that the two assemblages are using different mechanisms to regulate expression of proteins involved in encystation, such as GLORF-C4. This has key implications for the epidemiology of *G. duodenalis* infections as, according to this data, mixed infection may not be detected for a 24 hour period between the peak of assemblage B and assemblage A cyst production.
8.1 Introduction

*Giardia duodenalis* has the ability to not only infect humans but also a wide variety of animal hosts (Thompson, 2004). Domesticated animals are often seen as potential sources of *G. duodenalis* within the environment and could be seen as a possible zoonotic risk (Geurden and Olson, 2011). Production animals are seen as potential sources of contamination to water sources as they are usually kept in large numbers and have the potential to excrete a large number of cysts into a water system. Cattle are one such production animal that have been implicated as a reservoir for human infective *G. duodenalis* (O’Handley *et al.*, 2000).

Cattle are often infected with *G. duodenalis* from the human infective assemblages (A and B), however there is also an assemblage of *G. duodenalis* which is specific to cattle and other livestock, assemblage E. Genetically, assemblage E appears to share a lineage with assemblage A (Monis *et al.*, 1999) however, as with variation between the human assemblages, little is known about the protein variation between human infective assemblages and assemblage E. A study of the variable surface proteins showed that assemblage E isolates share several VSPs with assemblage A (Ey and Darby, 2002). Understanding how non-human infective assemblages differ from human infective assemblages is important to gain an insight into how isolates are able to infect humans. Proteins that are absent from assemblage E isolates could be required for infection in humans. Such proteins would make perfect drug or vaccine targets as by blocking them it could be possible to prevent infection. Now that we
have an understanding of how the human infective assemblages differ, this chapter
aims to determine the proteins of difference between human infective isolates of *G.
duodenalis* and the assemblage E isolate P15c1 using SDS-PAGE.

### 8.2 Materials and Methods

#### 8.2.1 SDS-PAGE

Samples BAH 2c2 from assemblage A, BAH 34c8 from assemblage B and P15c1
from assemblage E were analysed using large format 10% SDS-PAGE gels as
outlined in Section 3.2.2 of Chapter 3. Gels were stained with modified Coomasie
blue G-250 stain (Candiano et al., 2004) overnight and imaged as per Chapter 3
Section 3.2.4.

#### 8.2.2 Protein Identification by Mass Spectrometry

Protein bands were excised using the protocols stated in Section 5.2.1. Gel plugs
were then processed for mass spectrometry according to Section 5.2.2 and the mass
spectrometry analysis performed as per Section 5.2.3. Finally the mass spectrometry
data was analysed in line with Section 5.2.4. Bands of isolate P15c1 that correlated
with the bands identified for assemblage A and B in Chapter 5 were chosen for
identification by mass spectrometry.

### 8.3 Results

#### 8.3.1 SDS-PAGE

The proteins of assemblage E isolate P15c1 resolved well on the 10% SDS-PAGE
gel. There were several areas of difference evident between the three assemblages
and these are highlighted in Figure 8.1. Bands 1, 2, 3, 4, 5 and 7 all show variation between the assemblages. In most cases (bands 1, 2, 4, 5 and 7) the variation seen is in the number of bands at a particular position. Band 3 appears to vary in size between the three assemblages. There is one band, band 6, that is present in the human infective assemblages and absent in assemblage E.

8.3.2 Protein Identification

Protein bands from P15c1 identified as being variant from either of the human infective assemblages were excised for mass spectrometry, as well as several bands that were common between the three assemblages as a positive control. The individual bands excised are displayed in Figure 8.2. Of the 16 bands excised and submitted for mass spectrometry 15 returned positive identifications for *G. duodenalis* proteins. The 15 bands consisted of 35 proteins, five of which are classed as provisional. There are 5 proteins that are present multiple times giving 27 individual proteins. The proteins identified can be classed into 5 major categories; metabolic proteins (5), structural proteins (11), proteins involved in protein production and structure (6), proteins involved in cell cycling (3) and proteins with no homology with any known proteins (2).
Figure 8.1 SDS-PAGE of Assemblages A and B and Comparing with Assemblage E
Lane 1 molecular mass marker with size in kDa, lane 2 assemblage A isolate BAH 2c2, Lane 3 assemblage B isolate 34c8, lane 4 assemblage E isolate P15c1. Arrows mark areas of difference between the isolates.
Figure 8.2 Protein Bands of Assemblage E Isolate P15c1 that were submitted for mass spectrometry
Table 8.1 Proteins of Assemblage E Isolate P15c1 Identified Using Mass Spectrometry form SDS-PAGE Gels

<table>
<thead>
<tr>
<th>Number</th>
<th>Protein Identity</th>
<th>Score</th>
<th>Peptide Matches</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Pyruvate-flavodoxin oxidoreductase</td>
<td>104</td>
<td>3</td>
<td>XP_001708704</td>
</tr>
<tr>
<td></td>
<td>pyruvate ferrodoxin oxidoreductase</td>
<td>54</td>
<td>2</td>
<td>XP_001708948</td>
</tr>
<tr>
<td>E2</td>
<td>Dynamin</td>
<td>214</td>
<td>6</td>
<td>XP_001704962</td>
</tr>
<tr>
<td></td>
<td>axoneme associated protein</td>
<td>75</td>
<td>2</td>
<td>XP_001709214</td>
</tr>
<tr>
<td></td>
<td>Protein 21.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>50</td>
<td>1</td>
<td>XP_001708485</td>
</tr>
<tr>
<td>E3</td>
<td>alpha tubulin&lt;sup&gt;c&lt;/sup&gt;</td>
<td>53</td>
<td>1</td>
<td>XP_001705720</td>
</tr>
<tr>
<td>E4</td>
<td>Beta tubulin</td>
<td>295</td>
<td>5</td>
<td>XP_001707372</td>
</tr>
<tr>
<td></td>
<td>Alpha tubulin</td>
<td>199</td>
<td>3</td>
<td>XP_001705720</td>
</tr>
<tr>
<td>E5</td>
<td>NEK kinase</td>
<td>178</td>
<td>4</td>
<td>XP_001707030</td>
</tr>
<tr>
<td>E6</td>
<td>Hyp Prot protein disulphide isomerase</td>
<td>191</td>
<td>4</td>
<td>XP_001707782</td>
</tr>
<tr>
<td></td>
<td>translation elongation factor 1 gamma&lt;sup&gt;c&lt;/sup&gt;</td>
<td>57</td>
<td>1</td>
<td>XP_001707321</td>
</tr>
<tr>
<td>E7</td>
<td>NADP-specific glutamate dehydrogenase</td>
<td>527</td>
<td>10</td>
<td>XP_001707257</td>
</tr>
<tr>
<td></td>
<td>NADH oxidase protein</td>
<td>129</td>
<td>2</td>
<td>XP_001709257</td>
</tr>
<tr>
<td></td>
<td>phosphoglycerate kinase</td>
<td>62</td>
<td>2</td>
<td>XP_001706998</td>
</tr>
<tr>
<td>E8</td>
<td>Hyp prot no homology</td>
<td>76</td>
<td>2</td>
<td>XP_001706898</td>
</tr>
<tr>
<td>E9</td>
<td>Hyp prot no homology</td>
<td>165</td>
<td>5</td>
<td>XP_001706436</td>
</tr>
<tr>
<td></td>
<td>Hyp Prot disulfide isomerase</td>
<td>60</td>
<td>2</td>
<td>XP_001707782</td>
</tr>
<tr>
<td>E10</td>
<td>Alpha 11 Giardin</td>
<td>386</td>
<td>6</td>
<td>XP_001705101</td>
</tr>
<tr>
<td>E11</td>
<td>Alpha 11 Giardin</td>
<td>386</td>
<td>6</td>
<td>XP_001705101</td>
</tr>
<tr>
<td></td>
<td>Hyp Prot 14-3-3 homology</td>
<td>82</td>
<td>2</td>
<td>XP_001706755</td>
</tr>
<tr>
<td></td>
<td>Alpha 10 giardin</td>
<td>110</td>
<td>2</td>
<td>XP_001705104</td>
</tr>
<tr>
<td>E12</td>
<td>Beta Giardin</td>
<td>274</td>
<td>3</td>
<td>XP_001705425</td>
</tr>
<tr>
<td></td>
<td>delta giardin</td>
<td>224</td>
<td>4</td>
<td>XP_001707449</td>
</tr>
<tr>
<td></td>
<td>Hyp Prot 14-3-3 homology</td>
<td>111</td>
<td>2</td>
<td>XP_001706755</td>
</tr>
<tr>
<td></td>
<td>Alpha 2 Giardin</td>
<td>110</td>
<td>2</td>
<td>XP_001706958</td>
</tr>
<tr>
<td>E13</td>
<td>Beta Giardin</td>
<td>555</td>
<td>8</td>
<td>XP_001705425</td>
</tr>
<tr>
<td></td>
<td>Hyp Prot 14-3-3 homology</td>
<td>88</td>
<td>2</td>
<td>XP_001706755</td>
</tr>
<tr>
<td>E14</td>
<td>SALP1</td>
<td>162</td>
<td>3</td>
<td>XP_001708306</td>
</tr>
<tr>
<td></td>
<td>Ribosomal protein L7</td>
<td>126</td>
<td>3</td>
<td>XP_001705231</td>
</tr>
<tr>
<td></td>
<td>Beta Giardin&lt;sup&gt;c&lt;/sup&gt;</td>
<td>61</td>
<td>1</td>
<td>XP_001705425</td>
</tr>
<tr>
<td></td>
<td>spindle assembly checkpoint component</td>
<td>57</td>
<td>2</td>
<td>XP_001705618</td>
</tr>
<tr>
<td></td>
<td>GLORF-C4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>54</td>
<td>1</td>
<td>XP_001704917</td>
</tr>
<tr>
<td>E15</td>
<td>Peroxiredoxin 1</td>
<td>324</td>
<td>4</td>
<td>XP_001704199</td>
</tr>
<tr>
<td></td>
<td>Ribosomal Protein L13A</td>
<td>86</td>
<td>3</td>
<td>XP_001704829</td>
</tr>
<tr>
<td>E16</td>
<td>NR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Number of peptides matched to the protein  
<sup>b</sup> Accession number for protein in NCBI database  
<sup>c</sup> Provisional identification as only one peptide match
### 8.4 Discussion

The SDS-PAGE gels of assemblage A, B and E were compared to determine the level of protein variation between the human infective assemblages (A and B) of *G. duodenalis* and an isolate from the livestock specific assemblage (E). The areas of difference between the three assemblages were similar to those seen when the two human infective assemblages were compared in Chapter 3. The protein banding pattern for the assemblage E isolate P15c1 did not have any more similarity with either of the human infective assemblages. This is in contradiction to phylogentic based studies which places assemblage E isolates in a group with assemblage A isolates (Wielinga and Thompson, 2007). A recent study performed whole genome sequence comparison of assemblages A, B and E, their results indicate that assemblage E is equally dissimilar to assemblages A and B (Jerlstrom-Hultqvist et al., 2010). This demonstrates that although assemblage A and E are of the same lineage they are still very different.

#### 8.4.1 Identification of Variant Proteins

As for the assemblage A and B SDS-PAGE gels, proteins were identified for the majority of the protein bands excised, the only band that did not return a result was E16 (Fig 8.3.1). As the areas of variation between assemblages A and B were the same as the areas of variation between A and E and B and E, similar proteins were identified for P15c1. Table 8.4.1 summarises the 12 proteins of difference between assemblages A, B and E. Where a protein is labelled as NI it was not identified for that particular assemblage. This would indicate that it is not present within the
sample and may therefore be missing from the trophozoite proteome of that assemblage.

Table 8.2 Variant Proteins Identified by Mass Spectrometry and Their Position on Gels

<table>
<thead>
<tr>
<th>Protein</th>
<th>Assemblage A</th>
<th>Assemblage B</th>
<th>Assemblage E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamin</td>
<td>A3</td>
<td>B2</td>
<td>E2</td>
</tr>
<tr>
<td>NEK Kinase</td>
<td>A5</td>
<td>NI</td>
<td>E4</td>
</tr>
<tr>
<td>Alpha tubulin</td>
<td>NA</td>
<td>B6</td>
<td>E4</td>
</tr>
<tr>
<td>NADP specific glutamate dehydrogenase</td>
<td>A7</td>
<td>B8, B9</td>
<td>E7</td>
</tr>
<tr>
<td>Phosphoglycerate Kinase</td>
<td>NI</td>
<td>B9</td>
<td>E7</td>
</tr>
<tr>
<td>Hypothetical Protein XP_001706898</td>
<td>A8</td>
<td>NI</td>
<td>E8</td>
</tr>
<tr>
<td>Hypothetical Protein XP_001706436</td>
<td>A8</td>
<td>B10, B11</td>
<td>E9</td>
</tr>
<tr>
<td>Alpha 11 giardin</td>
<td>NI</td>
<td>B13</td>
<td>E10</td>
</tr>
<tr>
<td>Alpha 10 giardin</td>
<td>NI</td>
<td>B13</td>
<td>E11</td>
</tr>
<tr>
<td>14-3-3</td>
<td>A10, A12</td>
<td>B13, B15</td>
<td>E11, E12, E13</td>
</tr>
<tr>
<td>Alpha 2 giardin</td>
<td>A13</td>
<td>NI</td>
<td>E12</td>
</tr>
<tr>
<td>GLORF-C4</td>
<td>NI</td>
<td>B15, B16</td>
<td>E14</td>
</tr>
</tbody>
</table>

Assemblage A and B bands refer to Figure 5.3.1
Assemblage E bands refer to Figure 8.3.2

Dynamin is a guanosine triphosphatase and is involved in protein trafficking through the formation of membrane bound vesicles. In *G. duodenalis* the only known transport system are the encystation specific vesicles, which are primitive golgi, and form prior to encystment (Gaechter *et al.*, 2008). Based on SDS-PAGE analysis the protein bands containing dynamin of assemblages A and B are the same size and the assemblage E band is smaller in molecular mass. This would indicate that the assemblage E dynamin has a deletion when compared to assemblages A and B. Insertions and deletions (indels) are the most common form of polymorphism found in *G. duodenalis* and due to their nature will result in a change to the amino acid sequence and hence the molecular mass of a protein (Jerlstrom-Hultqvist *et al.*, 2010).
The NEK kinase family is a dominant gene family within *G. duodenalis* (Jerlstrom-Hultqvist *et al.*, 2010). There are 180 NEK kinases in assemblage A with only 43 believed to be functionally active based on their sequence (Morrison *et al.*, 2007). A NEK kinase XP_001717030 was identified in both assemblage A and E with the assemblage E variant having a larger molecular mass. The protein was not identified on assemblage B gels. The exact role of NEK kinases is poorly characterised so the biological meaning of this difference is unknown (Sonda *et al.*, 2010).

Glutamate dehydrogenase (GDH) was found in different bands between all three assemblages. The NADP-specific glutamate dehydrogenase of *G. duodenalis* is used to produce glutamate and assimilate ammonia (Yee and Dennis, 1992). The assemblage E isolate had one band containing the GDH protein which was smaller in molecular mass than the one band containing the GDH protein from assemblage A. The assemblage B isolate BAH 34c8 had two bands in which the GDH protein was present. The higher molecular mass band of the two assemblage B bands migrates above the assemblage A band and the smaller molecular mass band migrates alongside the assemblage E band. The genomes of A, B and E all predict one copy of the GDH gene and a sequence analysis of the predicted proteins indicates that there is 98-99% positivity between the assemblages, where positivity is a measure not just of identity, but, if there is variation, whether this is likely to alter the protein. This indicates that the mass change seen on the protein gels is more likely due to variation in the post translational modification applied to the proteins by the different assemblages.
Another case of size variation between the assemblages is with the hypothetical proteins XP_001706898 and XP_001706436. In assemblage A isolate BAH 2c2 these are present in the one band (A8), in isolate BAH 34c8 (B) only the XP_001706436 is present in bands B10 and B11 (Figure 5.3.1), indicating two size variants in the one isolate. The assemblage E isolate contains both XP_001706898 and XP_001706436, however unlike assemblage A where these proteins co-migrate in assemblage E the XP_001706898 is higher in molecular mass than XP_001706436. The two assemblage E bands migrate alongside the dual bands of assemblage B for XP_001706436. Another explanation for the dual banding for assemblage B is that it has it’s own version of the XP_001706898 and XP_001706436 genes which have homology to XP_001706436. This is reinforced by the fact that the bands B10 and B11 had the same peptide hits for the XP_001706898 protein. This may be due to a similarity in this part of the assemblage B orthologs rather than there being two copies of the one gene for assemblage B.

The structural proteins alpha 10 and 11 giardin were both found to vary among the assemblages. Both proteins were not identified from assemblage A trophozoite SDS-PAGE gels, however alpha 11 giardin was seen on 2D-PAGE gels (Chapter 5 Figure 5.3.2 and Table 5.3.3). A reverse transcriptase PCR based study of mRNA levels of several alpha giardins found a high level of expression of alpha 11 giardin when compared to beta giardin in assemblage A isolate WB (Weiland et al., 2005). The same study also found that the WB isolate had negligible levels of expression of alpha 10 giardin which may indicate that alpha 10 giardin is not normally expressed in assemblage A. For the assemblage B isolate alpha 10 and 11 giardin co-migrate as
a single band whereas in assemblage E they resolve into two separate bands with alpha 11 giardin having a larger mass than alpha 10 giardin. The assemblage B band (B16) migrates alongside the assemblage E alpha 11 band (E10) which indicates that assemblage E alpha 10 giardin has had a deletion when compared to the assemblage B isolate reducing the mass of the protein. The alpha giardins are a family of proteins with homology to annexins, which are calcium dependent phospholipid binding proteins (Weiland et al., 2005). They are theorised to bind the plasma membrane to the cytoskeleton. There is evidence of variation between the cytoskeletal proteins of different assemblages of G. duodenalis, as seen with the GASP 180 protein of assemblages A and B (Chapter 3 and 5). Differences in these structural proteins may require variation in the alpha giardins to anchor them to the plasma membrane.

Another alpha giardin previously identified as being an assemblage A specific protein of human infective G. duodenalis, alpha 2 giardin (Chapters 3, 4 and 5), was also identified in assemblage E. Alpha 2 giardin has been localised to the plasma membrane and flagella of assemblage A trophozoites (Weiland et al., 2005). As stated previously in Chapter 6 the loss of alpha 2 giardin in assemblage B is likely compensated for by a new alpha giardin specific for assemblage B. The presence of alpha 2 giardin in assemblage E adds weight to the argument that assemblage E is of the same genetic lineage as assemblage A (Monis et al., 1999). It would be interesting to examine assemblages C and D for the presence of alpha 2 giardin, as these assemblages are of the assemblage B genetic lineage. If they too lack alpha 2 giardin then alpha 2 giardin may have been an addition to the genome after the divergence of assemblages A and B.
The final variation between the three assemblages is GLORF-C4. This has been identified in protein gels of assemblages B (Chapters 3 and 5) and E isolates. Based on mRNA analysis GLORF-C4 is constitutively expressed in assemblage B and is induced as part of the encystation process in assemblage A (Chapter 7). The variation in expression of GLORF-C4 results in assemblage B being able to undergo the encystation process faster than assemblage A (Chapter 7). The presence of the GLORF-C4 protein in the trophozoite stage of the assemblage E isolate P15c1 implies that assemblage E is constitutively expressing GLORF-C4 and will produce fully formed cysts at a faster rate than assemblage A. The expression of GLORF-C4 in assemblage E still needs to be confirmed via mRNA analysis. As previously stated assemblage E is of the same genetic lineage as assemblage A, and thus the expression of GLORF-C4 in assemblage E would indicate that assemblage A has had selective pressure in favour of induced expression of GLORF-C4.

Although there has not been an assemblage specific protein identified for assemblage E or proteins specific for human infective assemblages, significant variation has been identified between the three assemblages. Dynamin is the only protein which appears to have host specific variation with the assemblage E protein having a smaller molecular mass than the human infective G. duodenalis proteins. In future studies the variants identified between the three assemblages should be investigated at the genetic level and for the case of GLORF-C4 at the mRNA level. These differences not only provide interesting information of the protein variation but also likely patterns of divergence between the assemblages namely with alpha 2 giardin as a potential A lineage protein and the potential of GLORF-C4 constitutive expression being selected against in assemblage A after the divergence of assemblage E.
Protein variants between the assemblages could also be used to produce an antigen panel for serotyping of *G. duodenalis* infections. This would be important in persistent infections where there is low or intermittent cyst production and in cases where previous infection status is of interest. For example with the increasing evidence that previous infection with *G. duodenalis* may be involved in the development of irritable bowel syndrome and linking this to a specific genotype (Dizdar *et al.*, 2007).
Chapter 9 General Discussion

9.1 Introduction

*Giardia duodenalis* is a major cause of gastrointestinal disease in both the developing and the developed worlds. Around the world prevalence rates vary from 1.4 (Louisiana, U.S.A.) to 816.9 (Romania) cases per 100,000 individuals, and is among the most common notifiable disease where applicable (Caccio and Sprong, 2011). It is in developing countries where infections with *G. duodenalis* cause greater problems, as malnutrition further complicates infection leading to failure to thrive in children. Cytopathologically infection with *G. duodenalis* leads to a diffuse shortening of the microvilli with a decreased absorptive surface area in the gut. The decreased absorptive area causes an increase in the concentration of nutrients in the lumen eventuating in water entering the lumen from the interstitial space of the small intestine via osmosis (Buret and Cotton, 2011). There is also a change in the tight junctions between individual cells in the small intestine allowing water to flow from the interstitial space into the lumen (Buret *et al.*, 2002). The factors produced by *G. duodenalis* to cause these structural and biochemical changes in the small intestine have not yet been identified.

The species *G. duodenalis* is separated into seven distinct genetic groups (termed assemblages A-G). Assemblages A and B are at least in part zoonotic, as there is evidence of sub-assemblage host specificity (Monis *et al.*, 2003a), whereas the other assemblages only infect specific groups of animals (Thompson and Monis, 2004). The mechanisms behind this host specificity, or lack thereof in assemblages A and B, is unclear. The genes used in the majority of genotyping studies encode for metabolic enzymes (glutamate dehydrogenase and triosephosphate isomerase) and
structural proteins (beta giardin) as well as the 18S rRNA gene (Wielinga and Thompson, 2007). There is one reported case of an assemblage specific gene, *glorf-c4* from assemblage B (Nash and Mowatt, 1992), however a homologue has been sequenced from assemblage A (Yong *et al.*, 2002).

At the protein level there is even less information as to the amount of variation between assemblages. Several studies have investigated protein variation between isolates, however these were performed before the designation of assemblages (Moore *et al.*, 1982, Smith *et al.*, 1982, Nash and Keister, 1985, Wenman *et al.*, 1986, Adam *et al.*, 1988, Aggarwal and Nash, 1988, Capon *et al.*, 1989, Bruderer *et al.*, 1993). The most in-depth studies on protein variation have been targeted isozyme studies targeting specific enzymes which can be stained for using substrate based techniques. Many of these studies where the first to identify the assemblages of *G. duodenalis* (Andrews *et al.*, 1989, Ey *et al.*, 1995b, Monis *et al.*, 1999, Monis *et al.*, 2003b). The isoenzyme studies show that there is protein variation between the assemblages of *G. duodenalis* and that investigations into the global proteome of the assemblages is likely to identify assemblage specific proteins and variants.

In this study SDS-PAGE and 2D-PAGE coupled with MS were used to identify some of the major proteins in assemblages A, B and E and determine some of the major proteins of difference between these assemblages. Eleven proteins that differ between assemblages A and B and twelve proteins of difference between assemblages A, B and E were determined, seven proteins of difference between assemblages A and B also varied with assemblage E. A selection of these differences
were also investigated using PCR based techniques (Chapter 6) and encystation studies (Chapter 7).

9.2 Implications of Protein Variation on Pathogenicity

There is very little known as to how *G. duodenalis* causes disease in the host. The cytological affects have been well documented and are described above. Investigations in *in vitro* models have found that whole trophozoites and sonicates produce similar pathological effects indicating a factor produced by the trophozoites is responsible (Chin *et al.*, 2002). A study by Kaur *et al.* (2001) identified a 48kDa protein which caused an increase in permeability of the small intestine of mice, however the protein was not further characterised.

There is also a large amount of variation seen between isolates in the type of disease caused. Three studies have found that assemblage A isolates are more likely to cause an acute diarrhoeal episode which is self-limiting and the assemblage B isolates are more likely to cause chronic infection, characterised by intermittent diarrhoea and abdominal discomfit (Read *et al.*, 2002, Haque *et al.*, 2005, Sahaqun *et al.*, 2008). In contrast to this a study from the Netherlands found that assemblage A was more associated with intermittent diarrhoeal disease and assemblage B with persistent chronic diarrhoea (Homan and Mank, 2001). There is also anecdotal evidence that assemblage B isolates are harder to treat. This variation in pathogenicity between the assemblages is likely to be attributed to variation in the proteins produced by these assemblages. Although the data presented in this thesis has not identified a particular protein directly responsible for this variation, the assemblage variant proteins identified may have a role to play.
The variable surface proteins (VSPs) are a group of proteins that cover the surface of the trophozoite (Ey and Darby, 2002). Isolates have the ability to change the major VSPs present on their surface, which is believed to aid the parasite in evading the host immune system (Nash, 2002). Even though these proteins are known to be both highly variable they were not detected on any of the SDS-PAGE or 2D-PAGE gel using mass spectrometry. As these are membrane bound proteins it could be that they were not solubilised in the sample buffers, as membrane bound proteins are notoriously difficult to solubilise (Herbert, 1999). Other studies in which VSPs have been investigated have mainly used Western blots to identify the location of VSPs which is a more sensitive way of detecting a specific protein group than relying on visualising a protein band or spot using general protein stains (Nash and Keister, 1985, Palm et al., 2003, Tellez et al., 2005).

9.2.1 Attachment

The cytoskeleton of *Giardia sp.* is important in regards to pathogenicity. Firstly a key virulence factor is the ability of the trophozoite to move within the gut (Elmendorf et al., 2005). Motility is driven by the flagella and other associated proteins within the cytoskeleton. Motility may also be involved in the attachment of the trophozoite to enterocytes, with the flagella maintaining downward pressure of the trophozoite onto the enterocyte keeping it firmly in place (Holberton, 1974). The GASP-180 protein, which associates with the base of the flagella, has been implicated, either directly or indirectly, with motility (Elmendorf et al., 2005). GASP-180 has been localised to the base of the flagella and it has been proposed that
it may be directly involved in motility by driving the flagella or that it is indirectly involved in motility by acting as an anchor protein for the flagella (Elmendorf et al., 2005). Variation within this protein could cause changes to the motility of trophozoites or to the effectiveness of surface attachment. It is believed that the peristaltic movement of the gut helps to clear infection, in effect physically removing the trophozoites from the enterocytes (Elmendorf et al., 2003). The acute versus chronic nature of infection with assemblages A and B maybe due to their ability to evade removal from the small intestine during infection, with assemblage B isolates being able to maintain their attachment to the small intestine under conditions that do not allow assemblage A isolates to retain their adherence.

The differences in the protein sequence and therefore ultimate structure of the assemblage B GASP-180 protein may enable the parasite to stay attached to enterocytes with continual peristaltic movement. If the GASP-180 protein is an integral part of the machinery required for driving the flagella, variation in the protein could alter the beat rate or strength of the flagella movement. Both scenarios would increase the force produced by the flagella, allowing the trophozoite to swim against the movement of liquid in the gut and stay in the parasites preferred environment longer. If the GASP-180 protein is an essential anchor protein connecting the flagella to the cytoskeleton, a stronger attachment would allow the flagella to beat with more force. A recent study identified clones of the G. duodenalis isolate WB (assemblage A) that were unable to attach effectively. The attachment deficient clones were missing a protein band of approximately 200 kDa when compared to wild type trophozoites, however the protein was not identified (Hernandez-Sanchez et al., 2008).
The attachment process is thought to have several components one of which is a ligand mediated attachment (Hansen et al., 2006). Alpha 2 giardin as with several other giardins associates with the plasma membrane and flagella of the trophozoite (Weiland et al., 2005). The alpha giardins are all related to the annexins and are therefore capable of binding glycosaminoglycans which appear on the surface of endothelial cells (Weiland et al., 2003). Different alpha giardins may bind to various ligands with variable binding efficiency. A stronger bond between ligand and the alpha giardin would allow for a stronger attachment of the trophozoite to the small intestine. As the results presented in this thesis indicate, assemblage B isolates lack alpha 2 giardin. It is likely that assemblage B isolates produce another alpha giardin-like protein to compensate for the lack of alpha 2 giardin. The variation in the alpha giardins produced by the assemblages may lead to different ligand interactions or stronger interactions using the same ligand. This may increase the strength of the attachment of the trophozoites to the small intestine making it more difficult for the host to clear the parasite.

A draft genome for the assemblage B isolate GS was recently released (Franzen et al., 2009). The authors make mention of the fact that an alpha 2 giardin-like protein does indeed exist in assemblage B isolates. This assemblage B variant of alpha 2 giardin is only 81% similar to the assemblage A alpha 2 giardin at the amino acid level. The purported assemblage B alpha 2 giardin is in fact more similar to assemblage A alpha 1 giardin than it is to alpha 2 giardin. Overall the alpha giardins displayed between 57% and 95% amino acid similarity between the assemblage A and B variants, highlighting the large amount of variation within the alpha giardins.
More work is needed to understand the role that these changes play in the cell biology and pathology of *G. duodenalis*.

### 9.2.2 Cyst Production

Another key virulence factor of *G. duodenalis* is its ability to produce cysts, therefore continuing the infection cycle. The success of a strain can be measured by its ability to produce viable cysts and the number of these cysts released in the faeces of the host. The encystation process in *G. duodenalis* has been studied in great detail; however the signalling processes in the trophozoite that lead to encystation are not as clearly defined (Morf *et al.*, 2010). A recent study by Nores *et al.* (2008) found the GlORF-C4 protein regulated the speed at which encystation occurred. When over-expressed trophozoites encysted at a faster rate and when expression was blocked encystation still proceeded but at a significantly slower rate. In Chapters 5 and 8 it was discovered that GlORF-C4 was present in assemblage B and E trophozoite samples but absent from assemblage A trophozoites. The encystation rates of assemblages A and B were compared in Chapter 7 with assemblage B isolates encysting at a faster rate than assemblage A isolates.

The data presented in this thesis indicates that assemblage B isolates are constitutively producing GlORF-C4 and the assemblage A isolates only express the gene when encystation is triggered. This variation in the expression profile of GlORF-C4 would lead to the variation in the encystation rates seen in Chapter 7. Since assemblage B isolates are constitutively expressing the GlORF-C4 protein, as indicated by high levels of mRNA and identifiable protein in the trophozoite stage,
they are able to encyst at a faster rate. This is supported by the study of Nores et al. (2008) where the presence of GlORF-C4 increased the rate of encystation.

GlORF-C4 was initially investigated by Nash and Mowatt (1992) as a potential assemblage B specific protein based on northern blot analysis. This was later refuted by Yong et al. (2002) who amplified the GLORF-C4 gene from assemblage A using a polymerase chain reaction assay. The data presented by Nash and Mowatt (1992) could also be interpreted as a difference in expression between assemblages A and B. Nash and Mowatt (1992) assumed that if the mRNA of GLORF-C4 was being expressed in assemblage B but not in A then the gene must be absent from assemblage A. The data could also indicate that the gene is under differential expression in assemblages A and B, with assemblage B isolates producing mRNA transcripts in the trophozoite stage when the assemblage A isolates are not, which is supported by our ability to identify GLORF-C4 protein and glorf-c4 transcripts in assemblage B trophozoites (Chapters 3, 5 and 7).

If an isolate of *G. duodenalis* is able to encyst at a faster rate than others it gains a competitive advantage over other isolates as it is able to disseminate into the environment at a faster rate. This is especially important in cases of symptomatic giardiasis where diarrhoea is commonly seen. The decreased time that contents are in the small intestine will decrease the time available to form viable cysts. Therefore, an isolate that is able to complete encystation in a shorter period of time will still be able to release viable cysts into the environment thus continuing the infection cycle. GlORF-C4 could therefore be used as a vaccine or drug target to decrease the rate of encystation leading to a decreased cyst count per gram of faeces. In models using
parasitic nematodes such as *Ostertagia ostertagi* in cattle, vaccines that reduce faecal egg counts have been trialled to break infection cycles (Geldhof *et al.*, 2002).

### 9.3 Implications of Protein Variation on the Taxonomy of *Giardia duodenalis*

#### 9.3.1 Review of Taxonomy

As outlined previously *G. duodenalis* is divided into seven distinct genetic groups, termed assemblages A-G. Assemblages A and B are the only assemblages to infect humans however they are also zoonotic, assemblages C and D are specific to canids, E is specific for hoofed animals, assemblage F is specific to felids and G is specific to rodents. Originally, assemblage designation was based on isoenzyme studies (Andrews *et al.*, 1989, Ey *et al.*, 1995a, Monis *et al.*, 1999, Monis *et al.*, 2003b, Monis *et al.*, 2003a) Subsequent studies have used a range of genes including, beta giardin, glutamate dehydrogenase, triosephosphate isomerase, GLORF-C4 and heat shock protein 70. These studies have strengthened the assemblage designation based on the isoenzyme data.

The taxonomic classification of *G. duodenalis* is currently a contentious issue (Thompson and Monis, 2011). Genetic diversity is now widely used to assess relatedness between species and genus; however there is not a prescribed level of sequence diversity that can be used to delineate species. The level of genetic difference between assemblages A and B of *G. duodenalis* is as great as that between some other species of protists (Mayrhofer *et al.*, 1995a). The mainstream use of PCR based methods for detecting *Giardia* sp. and the examination of the genetic sequence
amplified have vastly increased the understanding of the genetic structure of *G. duodenalis*.

### 9.3.2 Nucleotide variation vs Protein variation

A major issue with the studies of genetic variation is that the significance of the variation in nucleotide sequence on the biological activity of the isolates is rarely investigated. A gene may contain 10 polymorphic sites within its nucleotide sequences between isolates, however if each of these substitutions is a synonymous mutation, that is the protein code does not change, there will be no difference in the biological activity of the proteins. Many mutations lead to a change in the amino acid sequence although the amino acid substituted has the same properties, for example if a lysine is replaced with an arginine there is likely to be little change to the overall protein as both are basic amino acids. This is extremely important in assessing the variation between organisms as classical taxonomy is based upon phenotypic variations, and as stated before there is yet to be a cut off value given for the amount of genetic variation required to speciate eukaryotes, although using 16S rRNA in prokaryotes similarity less than 97% is considered enough variation to attribute separate species (Gevers *et al.*, 2005).

Two phenotypic differences between assemblages A and B of *Giardia duodenalis* with a genetic basis have been studied; alpha 2 giardin and GASP-180. The alpha 2 giardin protein was found in assemblage A isolates with the genetic sequence for the gene also amplified. The protein alpha 2 giardin was absent from assemblage B and the gene was unable to be amplified using several PCR tests. The protein GASP-180 was found to vary in size between assemblages A and B, with a 648bp fragment of
the gene showing 84% similarity between assemblages A and B. These are the first phenotypic differences between assemblages A and B which have an identifiable genetic basis. This finding supports the continuing calls for the reclassification of *G. duodenalis* into several distinct species (Thompson and Monis, 2011).

**9.3.3 Is *Giardia duodenalis* one species?**

*Giardia* sp. was first described on the basis of host range, with each host group being infected by their own species of *Giardia*. Filice in 1952 performed a thorough morphological examination of the *Giardia* species and grouped the majority of those described from mammals, together under the name *Giardia duodenalis* (Filice, 1952). All isolates within this group displayed a pear shaped trophozoite with a claw-hammer shaped median body (See Chapter 1 for full description). Filice stated that his designations were only temporary until such time as more discriminating characteristics could be discovered. Many see the genetic data currently being discovered within *Giardia* sp. as an opportunity for the reclassification of *Giardia*, specifically the classification of the assemblages as separate species groups.

The protein variation seen within this study between assemblages A and B gives credence to the call for the separation into two distinct species groups. It has been proposed that assemblage A isolates keep the name *G. duodenalis*, assemblage B isolates be termed *Giardia enterica*, assemblage C and D be combined to become *Giardia canis*, assemblage E be termed *Giardia bovis*, assemblage F as *Giardia cati* and assemblage G *Giardia simondi* (Monis *et al*., 2009). The designation of assemblages A and B as separate species was based on the amount of genetic difference characterised in molecular epidemiological studies and isoenzyme
analysis. The phenotypic variation discovered in this study further strengthens this argument by providing more evidence to complement the genetic distance between these isolates.

9.4 Comparison of electrophoretic methods for assessing protein variation in *Giardia duodenalis*

Protein electrophoresis is driven by two major concerns; the number of proteins visible and the resolution of these proteins. Unfortunately these two concerns often impact on each other. To increase the number of proteins visible the amount of sample run on the gel needs to be increased however this can decrease the resolution as more abundant proteins become over represented and in the case of 2D-PAGE these abundant proteins can have a deleterious effects on isoelectric focusing (see chapter 4). An aim of this thesis was to ascertain which method of protein separation, SDS-PAGE or 2D-PAGE is most effective in characterising protein diversity. Ultimately, both methods have their own “pros” and “cons” and the method recommended for use will be governed by the needs of a project.

Other than those used within this study there are other techniques which can be used for global proteomic studies. For comparative studies such as the one undertaken here there are several other techniques which can be employed which may increase the number of proteins identified. A very similar approach to that used here is difference in gel electrophoresis, where two separate samples are labelled with different fluorophores then run on the same gel (Tonge *et al.*, 2001). This technique
reduces gel to gel variation and makes comparison between gels easier (Alban et al., 2003). Unfortunately the technology is in many cases cost prohibited as the fluorophores used are proprietary products and the technique requires sophisticated imaging equipment.

Another technique that allows comparative proteome analysis is the isobaric mass tagged systems. These solely use HPLC coupled with mass spectrometry (HPLC-MS) to identify proteins, utilising mass tag that are attached to the proteins, so that the relative amount of a particular protein between the samples can be quantified (Wiese et al., 2007). Isobaric tagged relative and absolute quantification (iTRAQ) is one of isobaric tagged systems used for comparative studies. These HPLC-MS based systems are able to identify a large number of diverse proteins as they can detect much smaller amounts of proteins and are less affected by issues of protein solubilisation (Aggarwal et al., 2006).
9.4.1 Number of Proteins

Ideally when examining the proteome, the total protein complement within a sample needs to be visualised. From this study SDS-PAGE allowed for a larger number of proteins over a broader mass range to be visualised when compared to 2D-PAGE. From SDS-PAGE proteins below 25kDa were under represented on the gel making it difficult to identify individual bands. This is likely due to the relative higher abundance of the larger molecular mass proteins masking the presence of the less than 25kDa proteins. It could be possible to use mass filters to enrich for proteins below 25kDa in future studies of these lower molecular weight proteins.

In 2D-PAGE the inverse was seen with fewer proteins seen in the higher molecular mass region. This is due to the difference in the protein solubilisation techniques used for SDS-PAGE and 2D-PAGE. As the name suggests SDS-PAGE utilises sodium dodecyl sulphate (SDS), a strong ionic detergent to break inter and intra protein bonds and to give the proteins a negative net charge. As 2D-PAGE first separates proteins on the basis of their isoelectric point, using SDS would alter the proteins pI affecting migration. Instead of SDS, 2D-PAGE makes use of urea, thiourea and zwitterionic detergents to solubilise the proteins. These 2D-PAGE solubilisation buffers are not as effective as the SDS-PAGE buffers at breaking the protein bonds therefore resulting in fewer proteins visible on the gel. Interestingly, 2D-PAGE had clear protein spots in the lower molecular weight range (below 25kDa) which was underrepresented in SDS-PAGE; this could be due to the 2D-PAGE solubilisation buffers enriching certain proteins.
9.4.2 Resolution

As previously stated, proteomics is not just concerned with seeing as many proteins as possible but also with the resolution of those proteins. Proteins must appear as clear bands in the case of SDS-PAGE and spots for 2D-PAGE to complete downstream analysis. For SDS-PAGE three different percentages of polyacrylamide were used in the gels to increase the resolution within specific protein mass ranges. Changing the percentage of polyacrylamide varies the pore size in the gel matrix changing the migration rate of proteins through the gel. A lower percentage of acrylamide increases the pore size making it easier for larger molecules to migrate. The three percentages used, 7.5%, 10% and 12.5% (w/v), increased resolution in different molecular weight ranges allowing for a greater spread of proteins to be visualised and compared between the isolates examined. The majority of SDS-PAGE gel bands examined via MS contained at least two proteins. This could lead to proteins that co-migrate with abundant proteins being masked.

The use of 2D-PAGE increases the resolving power of gel electrophoresis as two properties of the proteins are used to separate them. This decreases the possibility of abundant proteins masking other proteins of the same size as they will likely have different isoelectric points. 2D-PAGE also allows for the identification of post translationally modified proteins. The PTM status of a protein is important in understanding protein regulation; this data can not be gained from standard SDS-PAGE. Due to its ability to separate proteins based on both size and isoelectric point, 2D-PAGE had a greater resolving power.
The question of which electrophoretic technique is best for proteomics is debatable. SDS-PAGE is much better at resolving based on size as the solubilisation techniques employed in SDS-PAGE allow for an investigation of a broader range of proteins as can be seen in Chapter 3. However, the increase in resolution from combining this with isoelectric focusing in 2D-PAGE allows for much more information to be gained on the limited number of proteins present in relation to variation in pI and PTMs. For a global view of protein variation or change, or if examining larger molecular mass proteins which are underrepresented in 2D-PAGE, SDS-PAGE is the best technique, although for a finer view of the protein complement or if examining variation in PTMs 2D-PAGE is best. The answer to which electrophoretic method should be used is really dependant on the aim of the set of experiments, within this thesis both techniques were used to identify protein variants between the assemblages.

9.5 Future Studies

Although *Giardia* has been studied for many years there are still many avenues that need further exploration. Within the field of proteomics there is much to be done. There is yet to be a concerted effort to map the proteome of *G. duodenalis*, as has been accomplished for *Toxoplasma gondii* and *Cryptosporidium parvum* (Sanderson *et al.*, 2008, Xia *et al.*, 2008). This would allow for a better understanding of pathogenicity and metabolism in this parasite.

Many of the proteomic based investigations into variation of *G. duodenalis* where performed prior to readily available protein identification techniques and information on isolate assemblage. Repeating the studies performed by Capon *et al.* (1989), Nash
and Keister (1985), Smith (1982) and Wenman (1986) would allow for the identification of the protein variants highlighted in these studies. This is especially important for the work performed by Capon et al. (1989) as they identified a 35kDa protein that was absent from three assemblage A isolates purified from cats. This may represent a protein required for human infection, supporting the idea of host specific sub-assemblages in assemblage A (Monis et al., 2003b). If the 35kDa protein is indeed required by *G. duodenalis* to infect humans then it would be a target for vaccine and anti-*Giardia* drug development.

Several previous studies have identified differences in the proteins in Western blot based experiments with patient and polyclonal antibodies produced against a variety of isolates (Nash and Keister, 1985, Capon et al., 1989). As these studies were performed before both the designation of assemblage and mass spectrometry based protein identification, the assemblage specific nature of the variation and the identification of the variant immune-dominant proteins have not been investigated. Identifying variable proteins that the host mounts an antibody mediated response to could also be used to produce novel diagnostics capable of identifying the assemblage of the infective isolate.

An extension of this would be to determine the response of the immune system of other mammalian hosts to infection with *G. duodenalis*, and examine if they differ from those seen in humans. There is evidence that dogs and cattle are more likely to display clinical symptoms if they are infected with *G. duodenalis* from their host specific assemblages (Thomas Geurden, personal communication). Determining if dogs and cattle mount an antibody mediated immune response against different
proteins of assemblage A and B isolates than humans, could identify if this difference in clinical presentation between species is due to the proteins targeted by the immune system. Proteins that are only bound by animal derived antibodies could be used as vaccine candidates to block disease in humans.

As previously stated the knowledge of pathogenicity in G. duodenalis is poorly defined. The pathological effects the parasite causes are clearly understood, however the mechanisms the parasite uses to cause these are not well defined. Previous studies indicate that there is a factor produced by the trophozoites that contributes to the disease state, as in in vitro models cell sonicates incubated with human endothelial cells lead to a similar level of disruption as when human endothelial cells are co-cultured with trophozoites. (Chin et al., 2002). By separating the complex whole trophozoite protein mixtures, for example using FPLC, prior to incubation with human endothelial cells it should be possible to narrow the search for proteins involved in the disease process. These protein fractions could then be further analysed to determine what proteins are present with potential cytopathic proteins studied further. Determining if assemblage A and B isolates use the same pathogenic proteins would also be important in understanding their relatedness and their variance in clinical presentation. Any identified pathogenic proteins would be prime candidates for vaccine development as blocking these proteins could stop disease. This could essentially eradicate G. duodenalis as a cause of failure to thrive and malnutrition in the developing and developed worlds.
References

- 4, - 373.


reveals cargo maturation mechanisms associated with golgi-like encystation


Thompson, R. C. A. 2004. The zoonotic significance and molecular epidemiology of
*Giardia* and *giardiasis*. *Veterinary Parasitology*, 126, 15-35.


Thompson, R. C. A. & Monis, P. T. 2004. Variation in *Giardia*: Implications for
taxonomy and epidemiology. *Advances in Parasitology*, 69-137.


Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan,
F., Hawkins, E., Currie, I. & Davison, M. 2001. Validation and development of
fluorescence two-dimensional differential gel electrophoresis proteomics

Torian, B. E., Barnes, R. C., Stephens, R. S. & Stibbs, H. H. 1984. Tubulin and high-

Traub, R. J., Robertson, I. D., Irwin, P., Mencke, N., Monis, P. & Thompson, R. C.
A. 2003. Humans, dogs and parasitic zoonoses - unravelling the relationships

Villegas, E. N., Glassmeyer, S. T., Ware, M. W., Hayes, S. L. & Schaefer, F. W.

Wang, H. & Hanash, S. 2003. Multi-dimensional liquid phase based separations in


Characterisation of alpha-1 giardin: an immunodominant *Giardia lamblia*


