

U. PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO



INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

The High Cysteine Proteins in *Giardia intestinalis*

Contribution to the characterization of the family

Sara Couto de Araújo Campos

MSc Thesis Project

Integrated Master in Bioengineering

September 2014

The High Cysteine Proteins in *Giardia intestinalis*

Contribution to the characterization of the family

Project Supervisor: Marcela Ferella, PhD

Uppsala University – Biomedicinskt Centrum

Department of Cell and Molecular Biology



UPPSALA
UNIVERSITET

Project Co-Supervisor: André Almeida, PhD

ICETA – Instituto de Ciências e Tecnologias Agrárias e
Agro-Alimentares da Universidade do Porto

Universidade do Porto

 PORTO

i c e a
Instituto de Ciências e Tecnologias Agrárias e Agro-Alimentares

Acknowledgments

During this project I counted with several helpful people, who I wish to acknowledge now:

- ✓ Professor Staffan Svärd, for accepting me in his group to develop this project and for always being optimistic, encouraging and supportive during my time abroad;
- ✓ My supervisor Marcela Ferella, for all the knowledge transmitted this semester, for the guidance and trust, the patience, dedication and advices;
- ✓ My work colleagues Elin, Dimitra and Nahuel for making the workplace a better place to be, all the funny moments and *fika* breaks;
- ✓ My Uppsala friends Filipa and Tommie for making the 'being away experience' more pleasant and amusing;
- ✓ My friends from Portugal, for always being there in one way or another, for the support and comprehension;
- ✓ Most of all, to my parents and brother, for the opportunity of studying abroad and always be supportive and provide all the conditions needed for me to pursuit my dreams.

Abstract

Giardia intestinalis is the most common human protozoan pathogen, infecting over one billion people worldwide. It is responsible for a pathology named giardiasis, from which pathogenicity mechanisms and the major host defenses against the infection are not well characterized. While infecting, this protozoan undergoes antigenic variation, a remarkable but poorly understood mechanism that allows the parasite to evade the host's immune defense activity, where only one Surface-specific Variant Protein (VSP) is expressed at a particular time. VSPs are epigenetically regulated and share some structural characteristics, namely CXXC motifs, with a novel *Giardia* protein family, the High Cysteine Proteins (HCPs). Several members from the HCP family showed up-regulation during parasite-host interactions, including 91707. Trophozoites transfected with this protein carrying an HA-tag showed change in cellular localization through different passages, particularly nucleus and, later on, plasma membrane.

Due to the similarities between VSP and HCP protein families there were three main objectives in this project. The first one was to confirm that localization changes in these HCPs were really occurring *in vitro*. If so, and to avoid extra factors associated to episomal protein expression, the second objective was focused on the expression and purification of recombinant HCPs for the production of specific antibodies to evaluate the cellular localization of 91707. Finally, verify the presence of one or more regulatory elements of transcription by miRNAs in some HCPs sequences to evaluate if this protein group could be regulated in the same way as VSPs are.

Regarding localization of 91707-HA in distinct transfectants, the results showed that there is, indeed, change in protein cellular localization, which may indicate that the trophozoites adapted to the overexpression of 91707-HA. The production and purification of 91707 was optimized; however, the final confirmation of the presence of this protein did not give the expected result, whereby the protein sample was sent to mass spectrometry analysis. Finally, qPCR results may indicate the presence of some miRNAs that could target HCPs sequences during parasite-host interactions. Nevertheless, further assays are needed in to verify the mechanism behind HCP regulation.

Resumo

Giardia intestinalis é o protozoário patogénico mais comum em humanos e infecta mais de um bilião de pessoas mundialmente. É responsável pela patologia denominada giardíase, cujos mecanismos de patogenicidade e as principais defesas do hospedeiro contra a infeção ainda não estão totalmente caracterizados. Enquanto infecta, este protozoário sofre variação antigénica, um mecanismo notável mas pouco estudado que permite ao parasita evadir as defesas do hospedeiro, havendo apenas a expressão de uma única *Variant-specific Surface Protein* (VSP) de cada vez. As VSPs são reguladas epigeneticamente e possuem algumas características estruturais comuns, nomeadamente motivos CXXC, com uma outra família de proteínas de *Giardia*, as *High Cysteine Proteins* (HCPs). Vários membros desta família estão sobre-expressos durante interações parasite-hospedeiro, incluindo a proteína 91707. Trofozoítos transfectados com esta proteína e o *tag* de fusão HA mostraram diferenças na localização ao longo das passagens, nomeadamente no núcleo e, mais tarde, na membrana plasmática.

Devido às semelhanças entre as famílias de proteínas VSPs e HCPs, existiram 3 objetivos principais neste projeto. O primeiro foi a confirmação da alteração na localização celular de 91707-HA *in vitro*. Caso se confirmassem, e para evitar factores adicionais associados à expressão episomal da proteína, o segundo objetivo foi focado na expressão e purificação de HCP recombinantes para a produção de anticorpos específico de modo a avaliar a localização celular de 91707. Por fim, verificar a presença de um ou mais elementos de regulação da transcrição por miRNA em algumas sequências de HCPs para avaliar se esta família de proteínas poderia ser regulada de um modo semelhante às VSPs.

Relativamente à localização de 91707-HA em diferentes transfectantes, os resultados mostraram que há, de facto, alteração na localização celular da proteína, o que pode indicar que os transfectantes se adaptaram à sobre-expressão de 91707-HA. A produção e purificação de 91707 foi otimizada; no entanto, a confirmação final da presença da proteína não mostrou o resultado esperado, pelo que uma amostra da mesma foi enviada para análise por espetrometria de massa. Por fim, os resultados de qPCR poderão indicar a presença de alguns miRNAs que tenham como alvo sequências de HCPs durante interações parasite-hospedeiro. Ainda assim, novos testes são necessários para verificar os mecanismos por detrás da regulação de HCPs.

Table of Contents

ACKNOWLEDGMENTS	4
ABSTRACT	5
RESUMO	6
TABLE OF CONTENTS	7
FIGURE INDEX	9
TABLES INDEX	10
1. INTRODUCTION	11
1.1 CELL BIOLOGY OF <i>GIARDIA</i>	11
1.2 LIFE CYCLE OF <i>GIARDIA</i> AND TRANSMISSION OF <i>GIARDIASIS</i>	12
1.3 <i>GIARDIASIS</i> : DISEASE FUNDAMENTALS AND CURRENT STUDIES	14
1.4 PARASITE-HOST INTERACTIONS	16
1.5 THE VARIANT SURFACE PROTEIN FAMILY OF <i>GIARDIA INTESTINALIS</i>	18
1.6 HIGH CYSTEINE PROTEINS IN <i>GIARDIA INTESTINALIS</i>	19
1.7 MIRNA GENE REGULATION IN <i>GIARDIA INTESTINALIS</i>	20
2. THESIS PROJECT BACKGROUND AND OBJECTIVES	23
3. MATERIAL & METHODS	24
3.1 MATERIAL STERILIZATION	24
3.2 BACTERIAL STRAINS AND PLASMIDS	24
3.3 FUSION TAGS	27
3.3.1 HISTIDINE TAG	27
3.3.2 HEMAGGLUTININ TAG	27
3.4 GENERAL METHODS OF MOLECULAR BIOLOGY	28
3.4.1 PRIMERS DESIGN	28
3.4.2 POLYMERASE CHAIN REACTION	28
3.4.3 COLONY-PCR	28
3.4.4 AGAROSE GEL ELECTROPHORESIS	29
3.4.5 DNA EXTRACTION AND PURIFICATION FROM AGAROSE GEL	29
3.4.6 TOPO [®] CLONING REACTION	30
3.4.7 PREPARATION OF COMPETENT CELLS OF <i>E. COLI</i> STRAINS TOP10 AND BL21 ORIGAMI	30
3.4.8 TRANSFORMATION OF COMPETENT <i>E. COLI</i> STRAINS TOP10 AND BL21 ORIGAMI	31
3.4.9 EXTRACTION OF PLASMID DNA FROM <i>E. COLI</i> STRAINS TOP10 AND BL21 ORIGAMI TRANSFORMED CELLS	31
3.4.10 DNA SEQUENCING	32
3.5 CULTURE CONDITIONS FOR <i>GIARDIA INTESTINALIS</i>	32
3.6 DNA EXTRACTION FROM <i>GIARDIA INTESTINALIS</i>	32

3.7	GIARDIA TRANSFECTION	33
3.8	IMMUNOFLUORESCENCE OF GIARDIA TRANSFECTANTS.....	33
3.9	LONG TERM STORAGE OF GIARDIA TRANSFECTANTS.....	34
3.10	LONG TERM STORAGE OF BACTERIAL CLONES	34
3.11	PROTEIN EXPRESSION AND PURIFICATION	35
3.11.1	OPTIMIZATION OF BACTERIAL RECOMBINANT PROTEIN EXPRESSION.....	35
3.11.2	PURIFICATION OF RECOMBINANT PROTEINS AND PROTEIN QUANTIFICATION.....	35
3.12	IMMOBILIZED METAL ION AFFINITY CHROMATOGRAPHY (IMAC).....	37
3.13	BIOINFORMATICS ANALYSIS	38
3.14	STEM LOOP-QPCR	39
4.	RESULTS AND DISCUSSION.....	41
4.1	EVALUATION OF EXPRESSION LEVELS AND LOCALIZATION OF HCPS TRANSFECTANTS.....	41
4.2	RECOMBINANT HIGH CYSTEINE PROTEINS EXPRESSION AND PURIFICATION	47
4.2.1	CLONING OF 91707 AND 7715 FOR RECOMBINANT PROTEIN EXPRESSION.....	47
4.2.2	OPTIMIZATION OF PROTEIN INDUCTION AND EXPRESSION.....	50
4.2.3	EVALUATION OF PROTEIN PURIFICATION BY NGC™ CHROMATOGRAPHY	54
4.3	BIOINFORMATICS ANALYSIS & MIRNA REGULATORY TARGET SEQUENCES IN HCPS.....	60
5.	CONCLUDING REMARKS.....	64
6.	FUTURE STUDIES.....	65
7.	REFERENCES.....	66
8.	APPENDIX.....	69
8.1	PRIMERS DESIGNED AND TESTED ALONG THE PROJECT.....	69
8.2	EXAMPLE OF HCP GROUPING BASED ON SEQUENCE HOMOLOGY USING CLUSTALOMEGA SOFTWARE	71
8.3	PUTATIVE MIRNAS IN GIARDIA INTESTINALIS (ZHANG ET AL., 2009).....	72

Figure Index

FIGURE 1. <i>GIARDIA LAMBLIA</i> STAGES. (A) <i>GIARDIA</i> TROPHOZOITE. (B) <i>GIARDIA</i> CYST [FROM: TOUZ MC, J., 2011].	11
FIGURE 2. LIFE CYCLE OF <i>GIARDIA INTESTINALIS</i> [FROM: ANKARKLEV, J., 2010].	12
FIGURE 3. LOCALIZATION OF IDENTIFIED PROTEINS IN <i>G. LAMBLIA</i> TROPHOZOITES BY POLYCLONAL MOUSE ANTIBODIES (GREEN) RAISED AGAINST PURIFIED RECOMBINANT ADI (A), OCT (B) AND ENOLASE (C). SCALE BAR: 10 μ M [ADAPTED FROM: RINGQVIST <i>ET AL.</i> , 2008].	16
FIGURE 4. VARIANT-SPECIFIC SURFACE PROTEIN STRUCTURE [ADAPTED FROM: ANKARKLEV <i>ET AL.</i> , 2010].	18
FIGURE 5. TRAFFIC OF HCNCP AND CYST PROTEINS DURING GROWTH AND DIFFERENTIATION OF <i>G. INTESTINALIS</i> [ADAPTED FROM: DAVIDS <i>ET AL.</i> , 2006].	19
FIGURE 6. MIRNA- AND SIRNA-MECHNISMS [ADAPTED FROM: HE AND HANNON, 2004].	20
FIGURE 7. MAP OF PEXP5-CT/TOPO [®] (INVITROGEN [™] , 2006).	25
FIGURE 8. SCHEMATIC DESCRIPTION OF STEM LOOP QPCR USING TAQMAN PROBE [ADAPTED FROM: CHEN <i>ET AL.</i> , 2005].	39
FIGURE 9. CELL COUNTING FOR PROTEIN LOCALIZATION OF 91707-HA A <i>GIARDIA</i> TRANSFECTANT.	41
FIGURE 10. CELL COUNTING FOR PROTEIN LOCALIZATION OF 91707-HA B <i>GIARDIA</i> TRANSFECTANT.	42
FIGURE 11. PCR PRODUCTS FROM <i>GIARDIA</i> GENOMIC DNA IN 1% AGAROSE GEL. 1 – GENE RULER [™] 1KB DNA LADDER; 2 AND 3 – 7715 FRAGMENT; 4 AND 5 – EMPTY LANES; 6 AND 7 – 91707 FRAGMENT.	47
FIGURE 12. COLONY-PCR PRODUCTS OF PEXP5-CT/TOPO [®] -7715 AND PEXP5-CT/TOPO [®] -91707 CONSTRUCTS EXTRACTED FROM <i>E. COLI</i> TOP10 TRANSFORMANTS IN 1% AGAROSE GEL.	48
FIGURE 13. COLONY-PCR PRODUCTS OF PEXP5-CT/TOPO [®] -7715 CONSTRUCT EXTRACTED FROM <i>E. COLI</i> TOP10 TRANSFORMANTS IN 1% AGAROSE GEL.	49
FIGURE 14. SDS-PAGE ANALYSIS OF THE EXPRESSION OF HIS TAGGED PROTEINS. A – COOMASSIE GEL; B – WESTERN BLOT.	50
FIGURE 15. SDS-PAGE ANALYSIS OF THE EXPRESSION OF 91707-HIS AT 30°C AND ROOM TEMPERATURE. A – COOMASSIE GEL; B – WESTERN BLOT.	51
FIGURE 16. SDS-PAGE ANALYSIS OF THE EXPRESSION OF 7715-HIS AT 30°C AND ROOM TEMPERATURE. A – COOMASSIE GEL; B – WESTERN BLOT.	52
FIGURE 17. WESTERN BLOT 91707-HIS INDUCTION AT RT: SONICATION OPTIMIZATION. L – PAGE RULER PLUS PRESTAINED PROTEIN LADDER; NI – NON INDUCED; I – INDUCED; PEL – INSOLUBLE FRACTION; SUP – SOLUBLE FRACTION.	53
FIGURE 18. SDS-PAGE ANALYSIS OF THE PURIFICATION OF 91707-HIS BY IMAC. A – COOMASSIE GEL; B – WESTERN BLOT.	54
FIGURE 19. SDS-PAGE ANALYSIS OF THE PURIFICATION OF 7715-HIS BY IMAC. A – COOMASSIE GEL; B – WESTERN BLOT.	55
FIGURE 20 – SDS-PAGE ANALYSIS OF THE PURIFICATION OF 91707-HIS BY IMAC FROM 250ML CULTURE. A – COOMASSIE GEL; B – WESTERN BLOT.	56
FIGURE 21 – SDS-PAGE ANALYSIS OF THE PURIFICATION OF 91707-HIS BY IMAC FROM 1L CULTURE. A – COOMASSIE GEL; B – WESTERN BLOT.	57
FIGURE 22 – CHROMATOGRAM OF THE PURIFICATION OF 91707-HIS FROM 1L OF CULTURE WITH CHROMLAB [™] SOFTWARE.	58
FIGURE 23 – CHROMATOGRAM OF THE SIZE EXCLUSION CHROMATOGRAPHY OF 91707-HIS USING CHROMLAB [™] SOFTWARE.	58
FIGURE 24 – SDS-PAGE ANALYSIS SIZE EXCLUSION FRACTIONS OF 91707-HIS.	59
FIGURE 25 – ANALYSIS OF SAMPLES FROM SLRT-QPCR IN 1% AGAROSE GEL.	62
FIGURE 26 – EXPRESSION LEVEL OF MIR6 MIRNA IN THE BIOLOGICAL REPLICATES A AND B.	63

Tables Index

TABLE 1. GENOTYPE FROM <i>E. COLI</i> TOP10 AND <i>E. COLI</i> BL21 ORIGAMI (LIFE TECHNOLOGIES, <i>GENOTYPES OF INVITROGEN™</i> COMPETENT CELLS, 2014)	24
TABLE 2. COMPOSITION OF THE DIFFERENT BUFFERS USED IN THE PURIFICATION OF THE HIS-TAGGED PROTEINS.....	38
TABLE 3. LOCALIZATION OF 91707-HA IN <i>GIARDIA INTESTINALIS</i> TRANSFECTANTS A AND B.	43
TABLE 3 (CONT.). LOCALIZATION OF 91707-HA IN <i>GIARDIA INTESTINALIS</i> TRANSFECTANTS A AND B.....	44
TABLE 3 (CONT.). LOCALIZATION OF 91707-HA IN <i>GIARDIA INTESTINALIS</i> TRANSFECTANTS A AND B.....	45
TABLE 4. HCPS GROUPING ACCORDING TO THE ALIGNMENT OF THE 3'END AND 3'UTR SEQUENCES.....	60
TABLE 5. IDENTIFICATION OF PUTATIVE MIRNA REGULATION SITES IN HCPS GROUPED SEQUENCES	61
TABLE 6. DESIGNED PRIMERS FOR 91707 AND 7715 PCR AMPLIFICATION AND ANALYZED FEATURES. FW – FORWARD; RV – REVERSE.....	69
TABLE 7. DESIGNED PRIMERS FOR THE MIRNAS TESTED WITH THE HCP SEQUENCES.....	70
TABLE 8. PUTATIVE MIRNAS (ZHANG <i>ET AL.</i> , 2009) AND MIR6 AND MIR10 (LI <i>ET AL.</i> , 2012) FROM <i>GIARDIA INTESTINALIS</i> MATCHED WITH HCP SEQUENCES.....	72

1. Introduction

1.1 Cell Biology of *Giardia*

Giardia intestinalis, also known as *Giardia lamblia* or *Giardia duodenalis*, is an eukaryotic intestinal parasite that belongs to the family Hexamitidae (order *Diplomonadida*) and the genus *Giardia*, being worldwide distributed (Cavalier-Smith T., 2003). This parasite is considered to be the most common intestinal pathogenic protozoa of humans. *G.intestinalis* is responsible for the disease named giardiasis, also known as "Beaver Fever" (Amar et al., 2002). It can also infect various animal species such beavers, cows and domestic dogs and cats. There are eight assemblages (A to H) that have different genetic backgrounds and host specificity, but are morphologically identical. Humans and other animals like dogs and cats are infected by assemblages A and B, the two major genotypes of the protozoa. These two assemblages diverge in approximately 20% at the DNA level and although the available data in this matter is limited, some studies suggest that the infection cause by assemblage A leads to more severe symptoms of the pathology, but B isolates may be more common worldwide (Singh *et al.*, 2009). The other *Giardia* assemblages include: assemblage C and D infecting dogs, E which infects hoofed animals, F, G and H having as hosts cats, rodents and seals, respectively. Assemblage E, isolate P15, is the only non-AB isolate that can be efficiently grown *in vitro* (Ankarklev *et al.*, 2010).

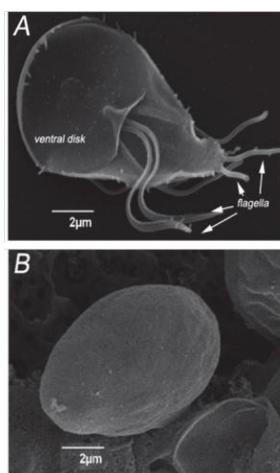


Figure 1. *Giardia lamblia* stages. (A) *Giardia* trophozoite. (B) *Giardia* cyst [From: Touz MC, J., 2011].

Giardia intestinalis' isolates are characterized by the presence of four pair of flagella, two nuclei, lack of aerobic mitochondria and a different genetic code (Tovar, J., 2003). *Giardia* also lacks peroxisomes and Golgi apparatus and owns lysosome-like peripheral vesicles, known as PV's (Carranza & Lujan, 2010). This protozoon is an interesting model for cell biology, since it has undergone a reductive evolution (early-divergent) and presents evidence for lateral gene transfer.

Giardia is a parasite with low infection doses, whereby the ingestion of only ten cysts can cause infection. The disease can be diagnosed by finding cysts or trophozoites, showed in Figure 1, the two main forms of the parasite, in the feces. Both of these life cycle stages

have a very distinct appearance (Roxström-Lindquist *et al.*, 2005).

1.2 Life Cycle of *Giardia* and Transmission of *Giardiasis*

Giardia's life cycle is quite simple and it is shown in Figure 2. The cycle begins when contaminated food or water or fomites with infective cysts are ingested by host, human or animals. The cyst is a non-motile and highly resistant form that allows the parasite to survive under severe and hostile conditions (Palm *et al.*, 2003). After ingestion, occurs in the small intestine a process named excystation. In this process, the cysts undergo a change of environmental conditions, such as pH by the gastric acid in the host stomach, leading to the metabolic activation of the cysts – the opening of the rigid and protective cyst wall and its degradation by cysteine proteases (Ankarklev *et al.*, 2010). The process of excystation lasts about ten minutes and each cyst releases two trophozoites, which will colonize the mucosa of the small intestine (Adam, RD, 2001).

In the intestine the trophozoites multiply by binary fission remaining free or attaching to the host epithelial cells through the ventral disc, a very complex structure composed by several vesicles and microtubules, which allows the parasite to be protected from the peristaltic flow. Trophozoites are motile due to their four pairs of flagellae and are about 12-15 μm long and 5-9 μm wide (Ankarklev *et al.*, 2010).

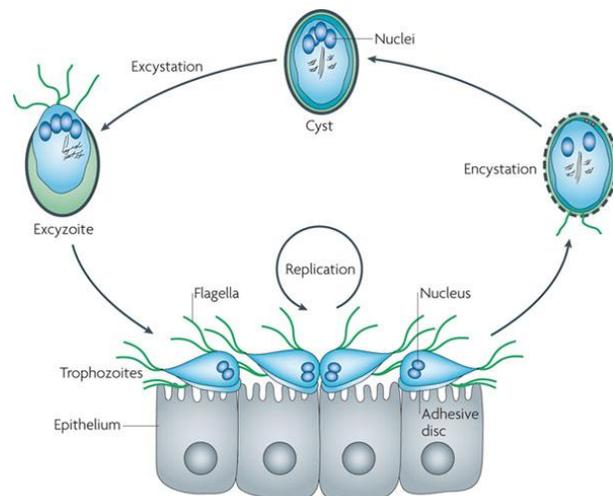


Figure 2. Life cycle of *Giardia intestinalis* [From: Ankarklev, J., 2010].

As the parasites transit towards the colon, another process occurs: encystation. This process is induced by factors inherent to the host, such as high levels of biliary secretions and basic pH, which lead to the transformation of the motile trophozoite into the infective cyst (Lauwaet *et al.*, 2007). Encystation involves the intracellular synthesis and transport of cyst wall components, as cyst wall protein (CWPs), and the assembly of cyst wall filaments. This process comprises two phases: early and late. During the early phase, cyst wall components are synthesized and transported by Golgi-like stack of membranes to encystation-specific vesicles (ESVs) (Adam, RD, 2001). In turn, ESVs transport those components to the nascent cyst wall. There are two important cyst wall proteins, CWP1

and CWP2, which are obviously highly expressed along this phase. These proteins are transported to the cyst wall by ESVs. In the late phase, the cyst wall filaments start to assemble and the trophozoite plasmalemma starts to appear on that region. During this stage, the flagella are internalized and the adhesive disc starts to fragment, which blocks the adhesion of the parasite to the host cells (Ringqvist *et al.*, 2008). When the process of encystation is over, the motility disappears and the cysts become rounded and filamentous. Each cyst contains one trophozoite with four nuclei that starts dividing as the cyst components disassemble. Two trophozoites will be separated when they finish cytokinesis and released from the cyst.

1.3 Giardiasis: Disease Fundamentals and Current Studies

Giardiasis is the disease caused by the parasite *Giardia intestinalis*. *Giardia* infects nearly 2% of adults and 6% to 8% of children in developed countries worldwide. In the United States of America, *Giardia* infection is the most common intestinal parasitic disease affecting humans. In developing countries, nearly 33% of people have had giardiasis (*Centers for Disease Control and Prevention, 2012*).

The severity and aggressiveness of the pathology is also related with the assemblage infecting a particular host. The symptomatic spectrum of the disease is determined by genetic and environment factors of the host. *Giardia* is not an invasive parasite, it does not penetrate the epithelium and does not enter to the blood stream, however it penetrates into the intestinal crypts (Carranza & Lujan, 2010) causing the partial destruction of the epithelial barrier. Parasite attachment to the epithelial cell layer of the small intestine is a critical step in initiating and maintaining the infection. After excystation, trophozoites divide by binary fission and give rise to several populations. These populations differ in their surface proteins or antigens, expressing each parasite a different antigen at a time, which then switches to a different surface protein in order to avoid host immune responses (Carranza & Lujan, 2010). This phenomenon is called antigenic variation and the surface antigens are variant surface proteins or VSPs.

Variant-specific surface proteins, VSPs, are a group of cysteine-rich proteins that represent around 200 genes of *Giardia* assemblages A and B (Ankarklev *et al.*, 2010). Their molecular weight varies from 20 to 200kDa. These proteins contain both a variable and a semi-conserved domain, where numerous CXXC amino acid motifs are present. Besides this, they have a transmembrane domain, since they are present on the surface (including the flagella and the ventral disk), and a unique and highly conserved CRGKA in the carboxy-terminal motif, where ADI (arginine deiminase) specifically binds and *citrullinates* the region. VSPs citrullination is a mechanism that apparently is involved in VSP-switching (Touz *et al.*, 2008).

Antigenic variation phenomenon occurs in *Giardia intestinalis* trophozoites *in vitro* every 6.5 to 13 generations in the two major human-infecting assemblages, although this is dependent on the isolate, the growth conditions and the specific VSPs being expressed (Ankarklev *et al.*, 2010). Only one member of the VSP-family is expressed at a time, except during cellular differentiation and VSP-switching.

In terms of the pathology, human giardiasis causes defects in the villous structure, due to the increase of crypt depth, with different stages of severity. The intestine permeability is also affected. In chronic human giardiasis, there is a loss of epithelial barrier function which leads to the reduction of the absorptive surface area, consequently causing maldigestion and malabsorption of nutrients, sodium, and water (Troeger *et al.*, 2007). Early and recent studies have proposed different disease mechanisms in giardiasis, including apoptosis of enterocytes, loss of epithelial-barrier function, hypersecretion of chlorides (Cl⁻), microvillus shortening, interference with bile and salt metabolism, among others (Ankarklev *et al.*, 2010).

1.4 Parasite-Host Interactions

Giardia is a parasite that nowadays can easily be grown *in vitro* in optimal culture conditions, which allows the accomplishment of a variety of studies. Transcriptomic studies of *Giardia intestinalis* during parasite-host interactions are highly relevant since they can disclose the mechanism behind the infection. An *in vitro* model has been established where *Giardia* cells are seeded over a Caco-2 cell monolayer and allowed to interact for various periods of time after which both cell types are collected and its RNA extracted for microarray studies (Roxström-Lindquist *et al.*, 2005; Ringqvist *et al.*, 2011).

Caco-2 cells, a cell line isolated from colon sarcomas, have been the current accepted *in vitro* model for intestine epithelial cells, since cell-cell contact triggers differentiation into enterocyte (Roxström-Lindquist *et al.*, 2005; Sambuy *et al.*, 2005). There is a group of specific proteins that are exceptionally released during parasite-host interactions, namely acute giardiasis. Among them, there are arginine deiminase (ADI) and ornithine carbamoyl transferase (OCT) (Ringqvist *et al.*, 2011). These proteins are related to arginine metabolism. *Giardia* captures arginine by an arginine-ornithine antiporter, releasing ornithine to the surrounding area (Stadelmann *et al.*, 2012). Intestinal epithelial cells' arginine uptake is inhibited by the release ornithine, but this does not affect *Giardia*. Thus, with this mechanism, the parasite has the ability to starve host cells by an arginine-dependent pathway, reducing cell-host proliferation.

Enolase enzyme is also released during parasite-host interactions and it is involved in *Giardia* excystation. It was shown to be present on the wall of mature cysts and cell membrane, flagella and cytosol of trophozoites, but not in ESVs (Castillo-Romero *et al.*, 2012).

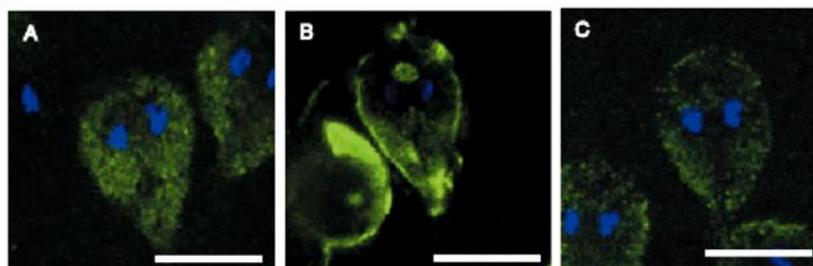
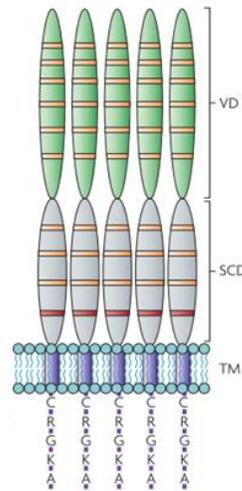


Figure 3. Localization of identified proteins in *G. lamblia* trophozoites by polyclonal mouse antibodies (green) raised against purified recombinant ADI (A), OCT (B) and enolase (C). Scale bar: 10 μm [Adapted from: Ringqvist *et al.*, 2008].

Using recombinant biology technology, protein analysis techniques and immunofluorescence microscopy it was possible to determine the localization of ADI, OCT and Enolase metabolic proteins. ADI and Enolase are mainly expressed in the cytoplasm, showing a punctuated pattern in that area, while OCT localizes in cytoplasm but with a strong accumulation near the plasma membrane seen on Figure 3 (Adapted from Ringqvist *et al.*, 2008).

1.5 The Variant Surface Protein family of *Giardia intestinalis*

The Variant Surface Proteins (VSPs) involved in the antigenic variation mentioned previously, are a group of cysteine-rich proteins represented by more than 200 genes in *Giardia* assemblages A and B (Ankarklev *et al.*, 2010). Their molecular weight varies from 20 to 200kDa. These proteins contain both a variable (N-terminal) and a semi-conserved (C-terminal) domain, where numerous CXXC amino acid motifs are present and therefore their amino acid sequence is



very cysteine rich. These motifs are represented by orange boxes in Figure 4 (Adapted from: Ankarklev *et al.*, 2010). VSPs have a transmembrane domain of type-I integral membrane proteins and a unique and highly conserved CRGKA carboxy-terminal motif, where ADI (arginine deiminase) specifically binds and citrullinates the R residue. VSPs citrullination is a mechanism that apparently is involved in VSP-switching (Touz *et al.*, 2008).

In vivo the switching is triggered by the pressure induced by the adaptative immune system from the host, preventing the clearance of the infection. The name antigenic variation derives from the fact that homologous cell surface molecules that are antigenically different undergo a continuous switching (Prucca *et al.*, 2011). VSP switching

Figure 4. Variant-specific Surface Protein structure [Adapted from: Ankarklev *et al.*, 2010].

occurs *in vitro* spontaneously every 6.5 to 13 generations in the two major human-infecting assemblages, although this is dependent on the isolate, the growth conditions and the specific VSPs being expressed. Only one member of the VSP-family is expressed at a time, except during cellular differentiation and VSP-switching (Ankarklev *et al.*, 2010).

1.6 High Cysteine Proteins in *Giardia intestinalis*

High Cysteine Proteins, HCPs, are a family of proteins rich in cysteine content as its name refers to. In *Giardia intestinalis* there are 86 members of this family with only 62 of them having a transmembrane domain in their amino acid sequence – type I integral proteins (Davids *et al.*, 2006).

There is still very few knowledge regarding this protein family, whereby so far only the study from Davids *et al.*, 2006 contributes to its characterization.

HCPs have 11-16% cysteine content within several CXXC and CXC motifs with a molecular weight range from 50 to 250kDa. Figure 5 shows the localization of a High Cysteine Non-variant Cyst protein, HCNCp (Adapted from Davids *et al.*, 2006). This protein seems to be localized in the nucleus or its periphery in the trophozoite stage, but in the cyst form the HCNCp is found in the cell body of the cyst. This suggests that this particular protein may be important in the trophozoite stage, since it is accumulated in the cyst during late encystation. A set of HCPs also showed upregulation in microarray data in the presence of the histone deacetylases (HDAC) inhibitor FR235222, indicating that some members from this protein family maybe regulated at the chromatin level (Sonda *et al.*, 2010).

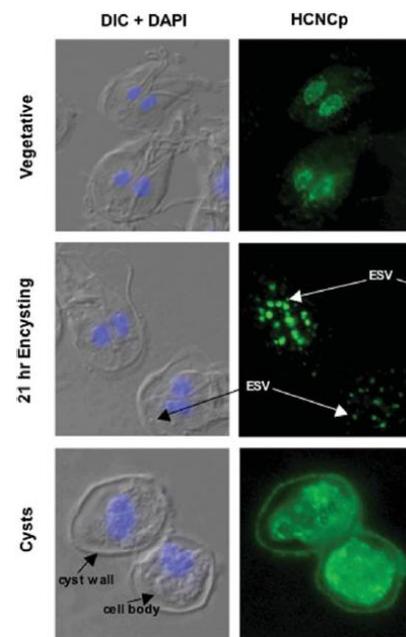


Figure 5. Traffic of HCNCp and Cyst proteins during growth and differentiation of *G. intestinalis* [Adapted from: Davids *et al.*, 2006].

High Cysteine Proteins present a common amino acid motif to the VSPs, CXXC, which is likely to be related with protein folding and response to O₂ levels. All the resemblances between these two peculiar protein families suggest that HCPs could be a subgroup of VSPs. Additionally, it is suggested that HCPs can be regulated the same way VSPs are, by a microRNA mechanism or at the chromatin level (Sonda *et al.*, 2010).

1.7 miRNA Gene Regulation in *Giardia intestinalis*

The control of gene expression is a biological process essential to all organisms. Gene expression can be controlled during transcription or post-transcriptionally by mechanisms such as RNA transport, mRNA stability, regulation of chromatin structure and post-transcriptional modifications among others (Yilmaz & Grotewold, 2010).

MicroRNAs (miRNAs) are an abundant class of small non-coding RNAs which post-transcriptionally regulate the expression of genes based on partial sequence complementarity. Usually this regulation results in repression of gene expression and can either occur by mRNA degradation or repression of translation (Cai *et al.*, 2009). Several studies in different biological fields have shown that this class of RNAs is involved in gene regulation in distinct biological pathways (Bartel, 2009). In 1993, Rosalind Lee identified the first miRNA, *lin-4*, which down regulates the levels of LIN-14 protein, being in turn involved in timed regulation of the larval development stage in *Caenorhabditis elegans* life cycle (Lee *et al.*, 1993). miRNAs are usually 21-25 nucleotides in length and are present in several organisms, as worms, flies, plants and mammals (He & Hannon, 2004). The process of miRNA formation in animals is not completely

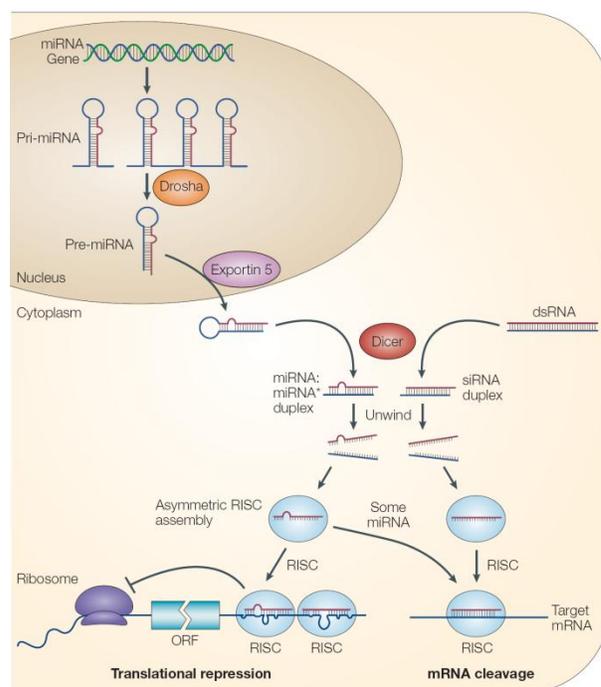


Figure 6. miRNA- and siRNA-mechanisms [Adapted from: He and Hannon, 2004].

clear, but the general mechanism of its biogenesis is known. miRNA transcripts, also known as pri-miRNA, are processed into pre-miRNA by the RNase-III enzyme Drosha, which are precursors with approximately 70 nucleotides, in the nucleus. These precursors are exported to the cytoplasm by the nuclear export receptor Exportin-5 through nuclear pore complexes. Once in the cytoplasm, another RNase-III enzyme, Dicer, cleaves pre-miRNA into 20-25 nucleotide miRNA fragments, termed mature miRNAs, removing the loop region from the sequences as exemplified in Figure 6. As mentioned, both Drosha and

Dicer are RNase-III enzymes with endonuclease activity, meaning cleavage of phosphodiester

bonds within a polynucleotide sequence. These dsRNA-specific enzymes create 2-nucleotide-long 3' overhangs similar to a mini-helix during pri- and pre-miRNA splicing (Kim Narry, 2004). The presence of this structure in pre-mRNA makes it suitable to be recognized by Exportin-5. In order to turn into functional molecules, double-stranded miRNA sequences suddenly convert into single strands and joins the Argonaute protein, forming the RNA-induced silencing complex (RISC), to allow target recognition (He & Hannon, 2004). The remaining strain (miRNA*) is degraded. The strain which comprises RISC usually has the less stable base-paired 5' end in the miRNA/miRNA* duplex (Khvorova *et al.*, 2003). The Argonaute proteins are included in an evolutionary conserved family and contain both a PAZ and Piwi domain. These domains recognize dsRNA ends with a 3' two-base overhang (MacRae *et al.*, 2006). RISC binds to the mRNA and the miRNA present in the complex anneals to it, usually with imperfect base-pair (Krol *et al.*, 2010). The annealing occurs between the seed region near the 5' terminus in the miRNA strain (~ nucleotide 2-7) and the complementary sequence in the mRNA (Cai *et al.*, 2009). Thus protein synthesis is inhibited by either translational repression or mRNA deadenylation with consequent degradation.

In *Giardia* many pathways from fundamental cellular processes seem to be simpler when compared to other eukaryotic organisms. Transcriptional and translational expressions are extremely regulated in higher eukaryotes like *Giardia*. However, this protozoon has few consensus promoters and particularly short 3' and 5' UTRs (untranslated regions), meaning that regulatory sites for translational regulation may be less available (Saraiya *et al.*, 2008).

In *Giardia intestinalis* miRNA mechanisms occur slightly different from what was earlier described, since its RNAi (RNA interference) machinery comprise distinct components. *Giardia* contains a homologous of the ribonuclease protein Dicer, (GIDcr) which three-dimensional structure is resolved by X-ray crystallography. *Giardia* also presents an Argonaute-like protein (GIAgo), but no homolog for Drosha or Exportin-5 (Saraiya *et al.*, 2008). GIAgo function and activity are still poorly understood, however a few studies provided evidence that this protein may be involved in miRNA-mediated translational repression (Kiriakidou *et al.*, 2007; Saraiya *et al.*, 2008).

In 2008, a study project from César Prucca showed that the process of antigenic variation in *Giardia lamblia* is regulated by RNA interference. Several transcripts from the VSPs

are produced even though only one VSP transcript accumulates, which suggested that the mechanism involved in its regulation should be at post-transcriptional level (Prucca *et al*, 2008). The reduction of *rdp* (RNA-dependent RNA polymerase) and *dicer* expression gave rise to a trophozoite population expressing more than one *vsp* in their surface, whereas silencing of *Giardia ago* resulted in the absence of viable clones, meaning that this protein is essential for cell viability. Thus, cells with downregulation of RNAi machinery are suitable for the production of a vaccine since they express several VSPs simultaneously (Prucca *et al*, 2008).

There were identified six miRNAs in *Giardia* – miR2, miR3, miR4, miR5, miR6 and miR10 – and all of them derive from small nucleolar RNAs (snoRNAs), except from miR4, which is derived from a canonical biogenic pathway mediated by GIDcr (Saraiya *et al.*, 2013). From this miRNAs, regions 100nt upstream and 50nt downstream the stop codon were analyzed and only miR3 has not shown to have *vsp* mRNAs as putative target (Saraiya *et al.*, 2014).

2. Thesis Project Background and Objectives

Previous experiments performed at *Giardia*'s research group at ICM were based on RNA sequencing of *Giardia intestinalis* transcriptome at several time points during *in vitro* interaction with intestinal epithelial cells (Caco-2). These experiments showed that among the highly expressed genes were several members of the HCP family. These results were in agreement with previous studies (Roxström-Lindquist *et al.*, 2005; Ringqvist *et al.*, 2011; Maayeh *et al.*, 2012).

As the biological functions of the High Cysteine Protein family is completely unknown, the *Giardia* group decided to begin its characterization in order to understand their role during *in vitro* interactions and therefore during disease. Several HCPs were cloned into expression vectors fused to an HA tag and transfected into *Giardia* parasites in order to follow their localization by antibodies against the tag. Almost all HCPs tested, presented a particular localization except for two HCPs (GL50508_91707 and GL50508_115066 annotated in *GiardiaDB*) which showed different patterns with time: nucleus and, later on, plasma membrane.

Based on this, the first two objectives of this project were:

- a. to confirm that localization changes in these HCPs were really occurring *in vitro*, and
- b. the expression and purification of recombinant HCPs for the production of specific antibodies to avoid the use of transfected cells lines.

As mentioned along the introduction of this thesis, HCPs protein family has several similarities with another protein family present in *Giardia*: variant-specific surface proteins (VSPs). Due to the high resemblances structure-wise between the proteins from these two families, it was hypothesized that HCP regulation could occur in a similar way of VSP regulation: by miRNA regulation. Starting from here, the third and last objective in this project is to:

- c. determine the presence of one or more regulatory elements of transcription by miRNAs in some HCPs sequences.

3. Material & Methods

3.1 Material Sterilization

All the material, culture media and solutions suitable to high temperature treatment methods were autoclaved at 121°C and 1 atmosphere, for 20 minutes. All termolabile solutions were sterilized by filtration using 0.20µm or 0.45µm sterile filters.

3.2 Bacterial Strains and Plasmids

In this project were used two strains of *E. coli* – TOP10 and BL21 origami. *E. coli* TOP10 chemically competent cells were used to manipulate and clone the recombinant plasmid pEX5P-CT/TOPO (Invitrogen™). *E. coli* BL21 Origami chemically competent cells were used in the expression of the high-cysteine proteins GL50803_91707 and GL50803_7715. The genotype of these two strains is described in Table 1.

Table 1 Genotype from *E. coli* TOP10 and *E. coli* BL21 Origami (Life Technologies, *Genotypes of Invitrogen™ competent cells*, 2014)

<i>E. coli</i> Strain	Genotype
TOP10	F – <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ(<i>ara leu</i>) 7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>
BL21 Origami	F – <i>ompT hsdSB</i> (rB–, mB–) <i>gal dcm</i> (DE3)

E. coli TOP10 chemically competent cells are engineered in order to increase the transformation efficiency. This strain allows a stable replication of high-copy number plasmids. They have different features regarding their final genotype: *hsdR* for efficient transformation of unmethylated DNA from PCR amplifications, *mcrA* for efficient transformation of methylated DNA from genomic preparations, *lacZ*ΔM15 for blue/white color screening of recombinant clones, *endA1* for cleaner DNA preparations and better results in downstream applications due to elimination of nonspecific digestion by Endonuclease I and *recA1* for reduced occurrence of nonspecific recombination in cloned DNA.

BL21 Origami cells are ideal for use with T7 promoter-based expression systems. This strain allows protein expression. The DE3 designation indicates the strains contain the DE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lacUV5* promoter. IPTG

(Isopropyl β -D-1-thiogalactopyranoside) is required to induce expression of the T7 RNA polymerase. *ompT* indicates that the *E. coli* lack an outer membrane protease, reducing the degradation of heterologous strains and improving the recovery of intact recombinant proteins.

pEXP5-CT/TOPO[®] (Invitrogen[™]) was the plasmid used to clone the proteins of study. This vector has a bacteriophage T7 promoter for high-level, inducible expression of the recombinant protein of interest in *E. coli*, C-terminal 6xHistidine fusion tag for detection and purification of recombinant proteins, TOPO[®] Cloning site for rapid and efficient cloning of Taq-amplified PCR products, ampicillin resistance gene for selection in *E. coli* and pUC origin for high-copy replication and maintenance of the plasmid in *E. coli*, as shown in the plasmid map in Figure 7. pEXP5-CT/TOPO[®] vector is supplied linearized with single 3-thymidine (T) overhangs for TA Cloning[®] and Topoisomerase I covalently bound to the vector (pEXP5-NT/TOPO[®] and pEXP5CT/TOPO[®] TA Expression Kits, Invitrogen[™], 2006).

During PCR amplification *Taq*-polymerase adds a single deoxyadenosine (A) to the 3' ends of PCR products. Since the linearized vector is supplied linearized with single 3'-T overhangs, the PCR inserts can efficiently be ligated with the vector.

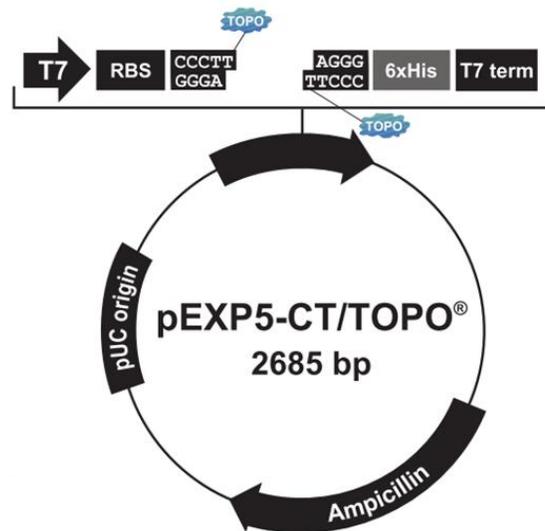


Figure 7. Map of pEXP5-CT/TOPO[®] (Invitrogen[™], 2006).

pAC plasmids carrying 91707 and 7715 genes were available in lab due to previous experiments performed by the group. These constructs were used to transfect *Giardia* expressing the 3xHA-tagged 91707 and 7715 proteins. This plasmid carries a puromycin *N*-acetyltransferase gene as a selection marker for *Giardia* transfectants. The PAC-pBS gene is

flanked by 47 bp of the giardial glutamate dehydrogenase (GDH) promoter and a 123-bp 3' untranslated region (UTR) from GDH inserted between *KpnI* and *HindIII* restriction sites (Jerlström-Hultqvist et al., 2012).

3.3 Fusion Tags

3.3.1 Histidine Tag

The polyhistidine tag is a DNA sequence coding for six to nine histidine residues. This tag is frequently used in vectors for production of recombinant N- or C-terminal tagged proteins. After expression in an adequate *E. coli* strain, his-tagged proteins can be purified since the sequence of histidine residues is able to bind to several types of immobilized metal ions, including nickel, cobalt and copper, under specific buffer conditions. The sequence, position and length of the tag can influence the rate of protein expression, the accessibility for binding the column, the solubility and activity of the protein. Besides this, there are anti-His-tag antibodies commercially available, allowing the detection of tagged proteins in assays such as Western Blot (Loughran & Walls, 2011).

3.3.2 Hemagglutinin Tag

Hemagglutinin (HA) is a surface glycoprotein from Influenza A virus responsible for binding sialic-acid present in the cells' surface of the upper respiratory tract in mammals. The HA tag is derived from the HA-molecule corresponding to amino acids 98-106 (YPYDVPDYA). This tag is widely used in expression vectors in order to track or detect proteins or purify them, with the use of specific antibodies anti-HA tag. HA-tag does not appear to interfere with the bioactivity or the biodistribution of the recombinant protein (Merck Millipore™, 2014).

3.4 General Methods of Molecular Biology

3.4.1 Primers Design

Forward and reverse primers specific to *7715* and *91707* were design in order to amplify these genes from *Giardia* genomic DNA. Both genes sequences were extracted from *GiardiaDB* (GL50803_91707 and GL50803_7715). The signal peptide and transmembrane domain sequences were removed from the original gene sequences to prevent insolubility of the expressed proteins. The signal peptide corresponds to nucleotides 1-57 (19 aa) in both genes and the transmembrane domain corresponds to nucleotides 1917-1983 (22 aa) in *91707* and nucleotides 1890-1956 (22 aa) in *7715*. Primers were analyzed using the online tool OligoAnalyzer 3.1 – Integrated DNA Technologies® (IDT®) and are shown in Table 6 in Appendix 8.1.

3.4.2 Polymerase Chain Reaction

The DNA fragments of *7715* and *91707* were amplified from *Giardia intestinalis* Assemblage A genomic DNA by polymerase chain reaction (PCR). PCR was performed on an Arktik™ Thermal Cycler (Thermo Scientific™). Each PCR reaction contained 20ng of Genomic *Giardia* DNA, 0.6µL of 10mM dNTPs, 1.2µL of primer forward, 1.2µL of primer reverse, 6µL of Buffer High Fidelity 5x with MgCl₂ (Thermo Scientific™), 0.5µL of Phusion Hot Start II High-Fidelity DNA Polymerase (proof reading – Thermo Scientific™) and sterile water up to 30µL.

The reaction comprised the following steps: denaturation and enzyme activation step: 98°C for 5 minutes; 35 cycles of amplification, including a denaturation step at 98°C for 15 seconds, an annealing step at the primer-specific melting temperature (see Table 6 in Appendix 8.1) and a extension step at 72°C for 1 minute and 20 seconds, followed by an extra extension step 72°C for 5 minutes. After this, the tubes were slightly centrifuged and 0.5µL of Taq DNA Polymerase (no proof-reading – Thermo Scientific™) and the reaction continued at 72°C for 20 minutes in order to add the 3'-A overhang using the Terminal Transferase activity of the Taq polymerase.

3.4.3 Colony-PCR

Colony-PCR was performed in order to detect bacterial clones carrying the desired insert in the right orientation in the construct with pEX5P-CT/TOPO, hence detecting the positive

colonies. Each PCR reaction contained 0.4 μ L of 10mM dNTPs, 0.8 μ L of primer T7 forward, 0.8 μ L of gene-specific reverse primer, 2 μ L of Taq Buffer with KCl 10x (Thermo Scientific™), 1.6 μ L of MgCl₂, 0.2 μ L of Taq DNA Polymerase (no proof-reading – Thermo Scientific™) and sterile water up to 20 μ L. The tubes with the PCR mix were placed on ice and a TOP10 *E. coli* colony was added with sterile toothpicks both to each reaction tube and to a gridded numbered plate, in order to track the positive colonies after the reaction assessment in agarose gel electrophoresis.

3.4.4 Agarose Gel Electrophoresis

Electrophoresis in agarose gel is one of the most common techniques to evaluate the size of DNA or RNA fragments. The principle of this technique is based on the migration of nucleic acids fragments from a negative to a positive pole, since DNA is negatively charged in neutral pH. The fragments will migrate according to their size, meaning that small fragments will more easily cross the porous matrix produced by agarose in solution.

The analysis of the DNA fragments was evaluated by electrophoresis in 1% Tris-acetate buffer (TAE) agarose gel stained with SYBR® Safe DNA Stain (Invitrogen™). The ladder used for size quantification of the double –stranded fragments was GeneRuler™ 1kb DNA Ladder (Thermo Scientific™). DNA samples were loaded in the gel together with 1 μ L of 6X DNA Loading Dye (Thermo Scientific™) to help the settling of the sample in the well.

The gels were visualized in Gene Genius Bio-Imaging System (SynGene™) and, in case of gel extraction, the bands were excised on a gel detection tray (Pharmacia Biotech™).

3.4.5 DNA Extraction and Purification from Agarose Gel

The extraction and purification of the interest DNA fragment was performed using GeneJET™ Gel Extraction Kit (Thermo Scientific™), which contains a silica-based membrane technology in the form of a convenient spin column.

The DNA fragment of interest was excised from a 1% TAE agarose gel with a clean scalpel and placed in a microcentrifuge tube previously weighed. The tube was weighed once again after collecting the band of interest. Binding Buffer was added in a proportion of 1:1 (100 μ L/100mg of agarose gel) to the gel slice and the mixture was incubated on a heat-block at 55°C for 10 minutes with slight shaking. After the gel slice was completely dissolved, the mixture was transferred to a GeneJET™ purification column and centrifuged for 1 minute. The

column was then washed with 700µL of Wash Buffer, centrifuged for 1 minute and dried at room temperature. The column was transferred to a clean microcentrifuge tube and incubated at room temperature for 2 minutes with 50µL of elution buffer. After centrifuge for 1 minute, the purification column was discarded and the purified DNA was stored at -20°C.

3.4.6 TOPO® Cloning Reaction

The principle of the TOPO® Cloning Reaction is based on linkage of the Topoisomerase I from *Vaccinia* virus to duplex DNA at specific sites (CCCTT), which leads to the cleavage of the phosphodiester backbone in one strand (Shuman, 1991). The energy released in the cleavage is conserved by formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of topoisomerase I. Thus, the bond between the DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase (Shuman, 1994).

Ligation reaction was performed according to the protocol provided by Invitrogen™ (pEXP5-NT/TOPO® and pEXP5-CT/TOPO® TA Expression Kits). 0.5 to 4µL of fresh PCR product, in a final concentration no higher than 100 ng/reaction, 1µL of Salt Solution (supplied with the kit), sterile water up to a final volume of 5µL and 1µL of TOPO® vector are gently mixed and incubated for 5 minutes at room temperature. TOPO® Cloning Reaction can either be kept overnight at -20°C or used immediately.

3.4.7 Preparation of Competent Cells of *E. coli* strains TOP10 and BL21 Origami

E. coli TOP10 and BL21 Origami competent cells are able to take up foreign DNA during transformation process due to an increase of the permeability of the cell-membrane.

Overnight cultures of both strains were set up and inoculated in 200mL of sterile LB. The flasks were incubated at 37°C with shaking till the cultures reach and Optical Density (550 nm) of 0.4 (exponential growth phase). At this point, the culture was transferred to 50ml falcon tubes and spun down at 3000rpm, for 10 minutes at 4°C. Each pellet was resuspended in 20mL of ice-cold TFBII and centrifuged in the same conditions. The pellet was then resuspended in 2mL of TFBII and placed on ice for 15 minutes. The suspension was aliquoted in microcentrifuge tubes (100µL/tube) which were immediately frozen in liquid nitrogen. The cells were stored at -80°C.

3.4.8 Transformation of Competent *E. coli* strains TOP10 and BL21 Origami

Transformation is the process by which foreign DNA is introduced into a prokaryotic cell. Introducing plasmid DNA in bacteria has an important relevance for several studies, since storage and replication of plasmids are ensured. These plasmids carry a bacterial origin of replication, the gene of interest and an antibiotic resistance gene which works as a selectable marker for positive transformants.

Transformation reaction was performed according to the protocol provided by Invitrogen™ (pEXP5-NT/TOPO® and pEXP5-CT/TOPO® TA Expression Kits). pEXP5-CT/TOPO-7715 and pEXP5-CT/TOPO-91707 were introduced in *E. coli* TOP10 cells and in BL21 Origami cells. 2µL of ligation reaction was added to 100µL aliquots of chemically competent cells and incubated on ice for 20 minutes. Cells were then heat-shocked at 42°C for 45 seconds on a heat-block and immediately placed on ice for 5 minutes. 500µL of LB were added to each tube and the cells were placed at 37°C with shaking for 1 hour. Finally, the cells were slightly spun down and plated in LB plates containing 100µg/ml and incubated o/n at 37°C. Control cells were transformed with pUC19 plasmid.

3.4.9 Extraction of Plasmid DNA from *E. coli* strains TOP10 and BL21 Origami

Transformed Cells

Plasmid DNA from *E. coli* strains TOP10 and BL21 Origami transformed cells was extracted using GeneJET Plasmid Miniprep Kit (Thermo Scientific™). 5-10ml of bacterial culture was spun down at 3500rpm for 15 minutes. The supernatant was discarded and the pellet was gently resuspended in 250µL of resuspension solution and transferred to a microcentrifuge clean tube. 250µL of lysis solution was added and the solution was mixed by inversion, followed by the addition of 350µL of neutralization solution. The solution was mixed by inversion and centrifuged at 13000rpm for 5 minutes at RT. The supernatant was transferred to a GeneJET™ spin column and centrifuged for 1 minute. The column was then washed twice with 500µL of Wash Buffer, centrifuged for 1 minute and dried at room temperature. The column was transferred to a clean microcentrifuge tube and incubated at room temperature for 2 minutes with 50µL of elution buffer. After centrifuge for 2 minutes, the spin column was discarded and the purified plasmid DNA was stored at -20°C.

3.4.10 DNA Sequencing

To verify the correct sequences of the cloned plasmids, samples of the plasmids were sent for sequencing at the Uppsala Genome Center. All the samples contained 450ng of plasmid together with efficient sequencing primers (4pmol/tube) and water to a final volume of 18 μ l.

3.5 Culture conditions for *Giardia intestinalis*

Giardia intestinalis, strain WB clone 6 (GL_50803) trophozoites were cultured *in vitro* in TYDK medium supplemented with 10% heat inactivated Bovine Serum (Gibco®) with a pH of 6.8 and grown at 37°C in an incubator with 5% O₂, according to Keister DB, 1983.

TYDK media consists of 30 mg/ml Peptone (Becton, Dickinson and Company), 55.6mM Glucose (Sigma®), 34.2mM NaCl (Sigma®), 1.14mM L-ascorbic acid (Sigma®), 5.74mM K₂HPO₄ (Sigma®), 4.41mM KH₂PO₄ (Sigma®), 11.4mM L-cysteine hydrochloride monohydrate (Sigma®) and 0.038mM ferric ammonium citrate (Sigma®), into filtered dH₂O. The pH was adjusted to 6.8 by adding 4M NaOH. The medium was sterile filtered with a 0.45 μ m filter system and supplemented with 0.268mg/ml filtered bovine bile (Sigma-Aldrich®) in a final concentration of 1%.

3.6 DNA extraction from *Giardia intestinalis*

Cells from a confluent tube of *G. intestinalis* (1 \times 10⁷ cells) were harvest by centrifuge at 3000rpm for 8 minutes. The supernatant was removed and the cell pellet was washed in 1ml of cold PBS to remove any media components by centrifuge at 2500rpm for 8 minutes once again. The pellet was resuspended in 500 μ l of lysis buffer including 50mM EDTA, 1% SDS and 10mg/ml Proteinase K. The resuspended pellet was later incubated at 56°C for at least 1 hour and mixed occasionally by vortexing. 20 μ l RNase A was added and the mixture was incubated for 20 minutes at room temperature. To the tube were added 275 μ l of Phenol (pH 8), 275 μ l of Chisam (24:1 chloroform:isoamyl alcohol) and mixed carefully by vortexing for 20 seconds. The sample was centrifuged for 10 minutes at 13000rpm. The aqueous phase of the tube (upper phase) was removed and an equal volume of Chisam was added in it. After vortexing for 20 seconds and centrifuging the tube for 10 minutes at 13000rpm the aqueous phase was precipitated with an equal volume of isopropanol and followed by 10 minute incubation at room temperature. The

tube was centrifuged for 10 minutes at 13000rpm, the supernatant was removed and the pellet was washed with 1ml of 70% cold ethanol and then centrifuged for 5 minutes at 13000rpm at 4°C. The supernatant was removed and the pellet was air-dried and dissolved ddH₂O (20-300µl depending on the amount of start material). The final concentration of genomic DNA was measured by NanoDrop®.

3.7 *Giardia* Transfection

A nearly confluent tube of *G. intestinalis* (80-90%) was placed on ice for 15 minutes to allow cell detachment. Cells were collected by centrifuging the tube at 2500rpm for 5 minutes. The pellet was resuspended in 300µl of TYDK and placed into a 4mm gap electroporation Gene Pulser® Cuvette (BioRad®) and placed on ice. To the cuvette, ~20µg the recombinant plasmids pAC carrying 91707 or 7715 were added, mixed with pipette and electroporated immediately. These plasmids were used in previous experiments by the lab team and were available for this experiment. The settings used for the electroporator Gene Pulser (Bio-Rad®) were: pulse in 350V, capacitance in 960µF and resistance in 800Ω.

After the electroporation the cuvettes were incubated on ice for 10 minutes and later transferred to culture tubes with 10 ml of warm TYDK media. The electroporated cells were grown at 37°C for 24 hours before 50µg/ml of puromycin (selection drug) was added. Every week, fresh media and selective drug was given to the transfected cells until they become confluent and stable. In order to avoid contamination during the cell cultivation, 100µg/ml of gentamicin was added to the culture tubes.

3.8 Immunofluorescence of *Giardia* Transfectants

In order to prepare the samples for immunofluorescence assay, one confluent tube of *Giardia* was used and put on ice for 10 minutes and centrifuged at 2500rpm for 5 minutes. The cell pellet was washed in PBS and centrifuged again at 2500rpm for 5 minutes. Next, the cell pellet was resuspended in 1ml of ice-cold HBS-glucose buffer (Hepes Buffered Saline) to wash away media components and centrifuged again at 2500rpm for 5 min. Thereafter, 25µl drops of cells were placed on poly-L-lysine coated "Diagnostic Microscope Slides" (Thermo Scientific®) with 10 wells. The cells were allowed to attach to the surface of the slide at 37°C for 10 minutes in a humidity chamber. *G. intestinalis* trophozoites were fixed with 2% paraformaldehyde (PFA) at 37°C for 30 minutes. The glycine was removed and 20µl of 0.1M Glycine in PBS was added to

the wells to quench any remaining traces of fixative. The fixative was removed and the wells were washed 2 times with PBS before 20µl of 0.2% Triton-X dissolved in PBS were added and incubated for 30 minutes at 37°C to permeabilize the cells. The solution was removed and 20µl of 2% BSA dissolved in PBS and incubated at room temperature during 1h or at 4°C o/n.

After incubation, the blocking solution was removed and 20µl of Anti-HA monoclonal antibody (product no. H 9658, Sigma-Aldrich®) diluted 1:500 times containing the DNA stain 4',6'diamidino-2-phenylidole (DAPI – 200mg/ml) 1% in the blocking buffer were added and incubated for 1 hour at room temperature. Next, the antibody was removed using vacuum suction and the wells were washed 4 times with 2%-BSA-PBS before 20µl of the secondary anti-mouse antibody conjugated to Alexa 488 diluted 1:250 times in the blocking buffer were added and incubated for 1 hour at room temperature. The antibody was removed using vacuum suction and the wells were washed 5 times with PBS. In the final step, 3µl of mounting media Vectashield (Vector Laboratories®) were added and a cover slip was placed over the wells and sealed with nail varnish. The slides were stored at -20°C in darkness. The immunostained cells were examined with a Zeiss Axioplan2 fluorescence microscope and the images were processed using the software Axiovision Rel. 4.8.

3.9 Long term storage of *Giardia* Transfectants

80% confluent tubes of a *Giardia* transfectants were spun down at 2500rpm for 5 minutes. Supernatant was discarded and the cell pellet was resuspended in 1ml of complete TYDK plus 10% DMSO (Dimethyl sulfoxide, Sigma-Aldrich®) and transferred to a well labeled cryo-tube (Nunc®). Stocks were stored at -80°C.

3.10 Long term storage of Bacterial clones

1 ml of transformed *E. coli* (TOP10 or BL21 Origami) o/n culture in LB 80% glycerol (Sigma®) in a final concentration of 15% glycerol. Glycerol prevents cell membrane damage and keeps the cells alive. The mixture is then transferred to a glass vial with metal lid properly labeled and stored at -80°C.

3.11 Protein Expression and Purification

3.11.1 Optimization of Bacterial Recombinant Protein Expression

E. coli BL21 Origami cells carrying either pEXP5-CT/TOPO-7715 or pEXP5-CT/TOPO-91707 construct were incubated o/n at 37°C with shaking in LB with 1% glucose and 0.1% ampicillin. The day after, these cultures were inoculated in a dilution of 1:20 in 500mL of LB with 1% glucose, 0.1% ampicillin and 0.1% MgSO₄ and incubated at 37°C till they reach and OD₆₀₀ of 0.8. At this point, 1ml samples were collected – labeled as non-induced samples – and 1mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) was added to the E-flasks in order to trigger induction. Protein induction was performed in three different conditions: o/n at room temperature, o/n at 4°C, 3 hours at 30°C and 3 hours at 37°C, with shaking. After induction, OD₆₀₀ was measured once again and 1ml samples were collected – labeled as induced samples. The remaining culture was transferred to 50ml falcon tubes and centrifuged at 3500rpm for 20 minutes. Supernatant was discarded after saving a 2ml sample and each pellet was resuspended in 5 ml of ice-cold resuspension buffer containing protease inhibitor cocktail tablets (Roche®). Resuspended pellets were added together in a final volume of 15ml and resuspension buffer was added up to a final volume of 25ml in each falcon. Different sets of buffers were used during the optimization of recombinant proteins expression and purification.

3.11.2 Purification of Recombinant Proteins and Protein Quantification

3.11.2.1 *E. coli* BL21 Origami Cell Lysis

One of the most important steps when purifying recombinant proteins is the preparation of the bacterial lysate. The key is to determine the optimal conditions in order to take full advantage of cell lysis and increase the yield of extracted recombinant protein and, on the other hand, minimize protein oxidation events, proteolysis and contamination with genomic DNA (Structural Genomics Consortium *et al.*, 2008).

Bacteria sonication was performed using a vibracell™ sonicator (Sonics®). The resuspended pellets (in a volume of 25 ml) were sonicated with a 6mm tip, during 10 minutes, with a pulse of 10/10 and amplitude of 40%. 1ml samples were collected – labeled as total sonicated – and the falcon tubes were centrifuged at 4000 rpm during 1 hour at 4°C. The

supernatant was transferred to a clean 50ml falcon and the remaining pellet was resuspended in 1ml of resuspension buffer.

3.11.2.2 SDS-PAGE Gels

SDS-PAGE (Sodium dodecyl sulfate Polyacrylamide gel electrophoresis) is a technique that allows the electrophoretic separation of proteins and nucleic acids according to their mobility. SDS detergent denatures proteins, which lose their tridimensional structure and native properties, and negatively charges the protein surface. Since the proteins charge is no longer a determinant factor for its separation, these molecules can be separated by size, migrating from the negative to the positive electrode. The concentration of SDS in the gel is chosen according to the size of the proteins that one intends to separate: proteins with a higher molecular weight need higher SDS concentration to denature.

Mini-PROTEAN® TGX Stain-Free™ pre-casted gels (Bio-Rad®) were used to separate the study proteins. These gels are suitable to separate protein with a wide range of molecular weight (kDa). Proteins lysates and fraction were collected after cell disruption and protein purification by Immobilized Metal Affinity Chromatography (IMAC). All the samples were treated with 1.2µL of 2-mercaptoethanol to reduce disulfide bonds and, consequently, disrupt protein tertiary structure, and 7µL of 4x Sample Buffer (Bio-Rad®). The samples were boiled at 98°C during 10 minutes. The tubes were slightly spun down and 30-50µL of each samples were loaded in the gel wells. The running occurred at 100V during 1 hour. The gel was disassembled and stained with Coomassie Blue reagent during 1h with shaking or used for western blot method.

3.11.2.3 Western Blots

Western blotting was performed using standard techniques. The proteins separated by SDS-PAGE gel was transferred to a FluoroTrans® PVDF (Polyvinylidene Flouride) membrane (Pall Life Sciences), previously activated in methanol for 5 minutes followed by equilibration in transfer buffer- The membrane was placed on top of the gel stacked between pads and filterpapers. The stack was placed in 1x Transfer buffer and the transfer was performed at 100V for 1 hour at room temperature with cooling and continuous stirring of the buffer for homogenous cooling. The membrane was then blocked using 3 % non-fat dry milk in Tris buffered saline (TBS) containing 0.1 % Polyoxyethylenesorbitan monolaurate, Tween20

(SigmaAldrich®) to reduce unspecific binding of the antibodies. The membrane was blocked for 1 hour at room temperature followed by washing three times for five minutes each in TBS-T. The membrane was incubated with Anti-His₆ antibody (product no. 04905318001, Roche®) as the primary antibody which was diluted 1:6000 in TBS-T for 1 hour at room temperature. Thereafter the membrane was washed as previously followed by a 1 hour incubation with the secondary antibody, Anti-mouse coupled with horseradish peroxidase (HRP) (product no. P0161, Dako®) diluted 1:20 000 in TBS-T. After one hour incubation the membrane was washed as before with TBS-T. Blots were developed using Clarity™ Western ECL Substrate (Bio-Rad®) according to the manufacturer's instructions.

3.12 Immobilized Metal ion Affinity Chromatography (IMAC)

Immobilized metal ion affinity chromatography (IMAC) was the method used to purify the his-tagged proteins. The purification was performed in a NGC™ Chromatography System (BioRad®) with a HisTrap FF column (GE Healthcare®) of 1ml capacity. HisTrap FF column is prepacked with precharged Nickel Sepharose 6 Fast Flow. Nickel Sepharose 6 Fast Flow consists of 90µm beads of highly cross-linked agarose, to which a chelating ligand has been immobilized. This chelating ligand is charged with Ni₂⁺ ions. This column has a high binding capacity, of approximately 40 mg/ml medium.

Histidine has a pK_a of approximately 6.0, meaning that at pH 8.0 this amino acid will be negatively charged. Due to the high content of negatively charged histidines present in the His-tag, there is a high affinity to bind the agarose matrix covered by Ni₂⁺ ions. The procedure starts with the equilibration of the column with binding buffer, which is the same buffer where the lysate is resuspended. This buffer is previously treated with DNase (Thermo Scientific™), RNase (Thermo Scientific™) and protease inhibitors (*cOmplete Protease Inhibitor Cocktail Tablets*, Roche®). The sample is then applied to the system at a flow rate of 1ml/minute to allow the binding of the recombinant proteins. The column is washed with wash buffer to remove the unspecific and weakly bound proteins and components of the mixture. The final step is the recovery of the his-tagged protein by applying the elution buffer with high concentration of Imidazole, since this compound competes for coordination sites on Ni₂⁺ displacing His-tagged protein from the matrix. The composition of all buffers used in the purification of 91707-His and 7715-His can be found in Table 2. The purification profile was analyzed using ChromLab™ Software.

Table 2. Composition of the different buffers used in the purification of the his-tagged proteins.

	Set 1		Set 2
Resuspension /Binding Buffer	HEPES 100mM pH 8; DDT 2mM; NaCl 200mM; Triton 1%; Glycerol 20%; 10mM Imidazol; Proteinase Inhibitors.	Resuspension/Binding Buffer	HEPES 100mM pH8; NaCl 200mM; DTT 1mM; Glycerol 10%; Imidazole 5mM; DNaseI-RNaseA 20ug/ml each.
Wash Buffer	DDT 2mM; NaCl 200mM; HEPES 100mM pH 8; Triton 1%; Glycerol 20%.	Wash Buffer 1	HEPES pH8 100mM; NaCl 300mM; DTT 1mM; Glycerol 10%; Imidazole 5mM.
Elution Buffer	DDT 2mM; NaCl 200mM; HEPES 100mMpH 8; Triton 1%; Glycerol 20%.	Wash Buffer 2	HEPES pH8 100mM; NaCl 500mM; DTT 1mM; Glycerol 10%; Imidazole 10mM.
Desalting Buffer		Elution Buffer	HEPES pH8 100mM; NaCl 500mM; DTT 1mM; Glycerol 10%; Imidazole 500mM.
			HEPES pH8 100mM; NaCl 500mM; DTT 1mM; Glycerol 10%.

3.13 Bioinformatics Analysis

The 3' end and 3'UTR sequences of the 86 annotated HCP genes were downloaded from GiardiaDB.org. 50 nucleotides from the 3' end sequences and 100 nucleotides from the 3'UTR sequences we analyzed. The 86 3'end-3'UTR sequences were aligned using the alignment tool *Clustal Omega* (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). From these alignments, the HCPs were grouped according to the homology between the 3'end-3'UTR sequences. In each HCP group, the 3'end and 3'UTR were align using the same tool. To evaluate the potential presence of miRNA regulation sites in the 3'end and/or 3'UTR of these HCPs, the seed region of 50 putative miRNAs from *Giardia* were align with the HCP groups.

3.14 Stem Loop-qPCR

Stem Loop Quantitative Real-Time PCR is a technique used to detect and quantify the expression levels of miRNAs. This technique comprises two main steps schematically described in Figure 8: the first step is a Stem-Loop reverse transcript reaction where cDNA is produced with

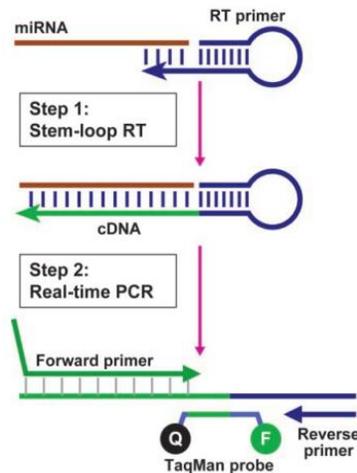


Figure 8. Schematic description of Stem Loop qPCR using TaqMan probe [Adapted from: Chen *et al.*, 2005].

the use of a stem-loop primer which anneals in the 3' end of the miRNA strand. The second step is a Real-Time Polymerase Chain Reaction where the amount of DNA is measured after each cycle via fluorescent dyes. The fluorescent signal is directly proportional to the number of PCR product molecules – amplicons (Chen *et al.*, 2005).

Two SLRT-qPCR reactions were performed: the first one with all sets of primers present in Table 7 in Appendix 8.1 and total RNA from *Giardia* trophozoites; the second with mir6 and mir6-control sets of primers and total RNA extracted from trophozoites during interaction with intestinal epithelial cells (IEC) in different time points. All RNA samples had been extracted and prepared prior to my arrival in the project.

Each SLRT reaction contained 50ng of DNaseI-treated total RNA, 0.5µL of 10mM dNTP mix, 2µL 10X first-strand Buffer, 2µL 50mM MgCl₂, 2µL 0.1M DTT, 0.1µL RNaseOUT, 0.25µL of SuperScript® III Reverse Transcriptase (all from Invitrogen™) and H₂O up to 19µL. 1µL of the correspondent 1mM SLRT-primer previously denaturated at 65°C for 5 minutes is added to the reaction tube. Reaction without SuperScript® III Reverse Transcriptase (-RT) and without template (-T) were also prepared. The reactions comprised the following steps: 16°C for 30 minutes, 60 cycles of 30°C for 30 seconds, 42°C for 30 seconds and 50°C for 1 second, and 85°C for 5 minutes. The reaction was ready for qPCR.

Each qPCR reaction SuperScript® III Reverse Transcriptase (+RT qPCR) contained 1µL of 10mM Fw primer, 1µL of 10mM Rv primer, 10 µL of SYBR® Green Master Mix (BioRad®), 6µL of sterile dH₂O and 2µL of RT product. The 96-well plate was sealed and placed in a StepOnePlus™ Real-Time PCR System and analyzed with StepOne™ Software from Life Technologies™. The reactions comprised the following steps: 95°C for 5 minutes, 40 cycles of 95°C for 5 seconds and 60°C for 10 seconds; for melting curve analysis, samples were denaturated at 95°C and cooled to

65°C at 0.2°C/second. Fluorescence signals were collected at 530nm continuously from 65°C to 95°C at 0.2°C/second.

4. Results and Discussion

4.1 Evaluation of expression levels and localization of HCPs transfectants

As mentioned in section X – Project Thesis Background and Objectives, the *Giardia intestinalis* transfectant 91707-HA showed a change in localization of the overexpressed tagged protein. Transfectants 91707-HA presented different protein localization when defrosted closer to or further from the transfection date, termed 91707-A and 91707-B respectively. Starting from here, these transfectants were followed by immunofluorescence assay (IFA) in two different approaches. Firstly, the cells were followed during 3 days and immunofluorescence assay was performed two times per day (within 8h). Considering the generation time of *Giardia* around 8h, during this period 9 generations were followed. A few days after this, it was decided to follow these transfectants for a longer period, covering more generation of the parasite. Thus, the same transfectants were followed by IFA, but this time only one time per day during 13 days. IFA was performed in the first 6 days, in day 8, day 10 and day 13 and a total of 37 generations were followed. After these two assays, the cells were counted and evaluated according to the cellular localization of the protein. The results from these counting are showed in the graphs from Figures 9 and 10. These values were obtained by counting two slides from each transfectant and generation.

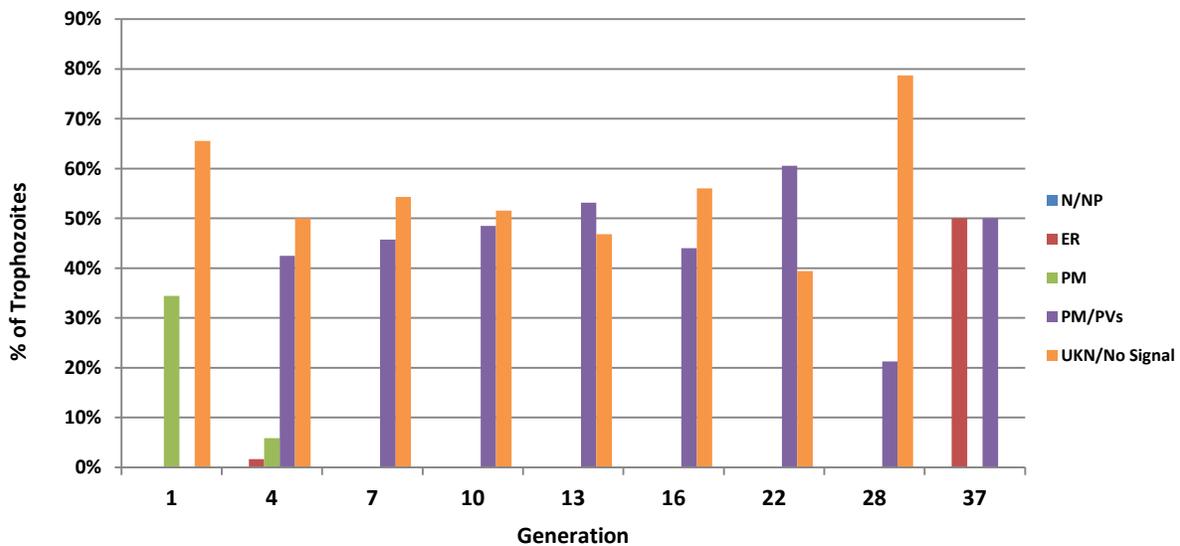


Figure 9. Cell Counting for Protein Localization of 91707-HA A *Giardia* transfectant.

N/NP – Nucleus/Nuclear Periphery; ER – Endoplasmic Reticulum; PM – Plasma Membrane; PM/PVs – Plasma Membrane/Peripheral Vacuoles; UKN/No Signal – Unknown localization/No signal.

Looking at the cell counting results shown in Figure 9 it can be concluded that the main patterns of cell localization for 91707-HA protein are plasma membrane and peripheral vacuoles (PM/PVs) in A transfectant. In generations 1 and 4 there are cells that show exclusively plasma membrane localizations, which was never detected alone again along the 37 generations followed. In generation 37 there were very few cells in the slides and the results obtained for this generation did not reflect the real scenario of the culture at this stage, whereby the cell counting was not valid.

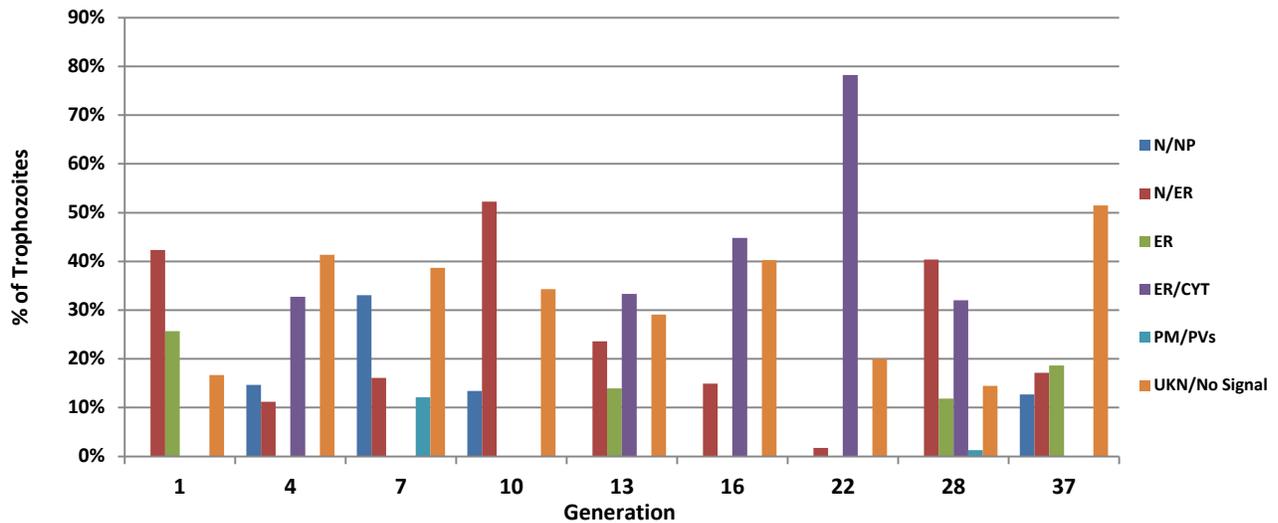


Figure 10. Cell Counting for Protein Localization of 91707-HA B *Giardia* transfectant.

N/NP – Nucleus/Nuclear Periphery; N/ER – Nucleus and Endoplasmic Reticulum; ER – Endoplasmic Reticulum; ER/CYT – ER and Cytoplasm; PM – Plasma Membrane; PM/PVs – Plasma Membrane/Peripheral Vacuoles; UKN/No Signal – Unknown localization/No signal.

Regarding the cell counting results for 91707-HA B transfected trophozoites present in the graph on Figure 10, it can be seen that the predominant localizations for the recombinant protein are nucleus (N/NP) and endoplasmic reticulum (ER), in opposition to transfectant A.

In Table 3 is present an example of IFA pictures of the localization of 91707-HA in transfectants A and B along the several followed generations.

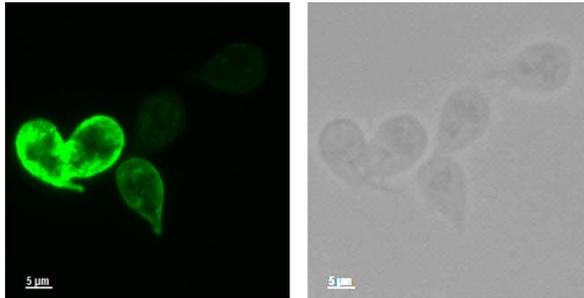
Table 3. Localization of 91707-HA in *Giardia intestinalis* transfectants A and B.

Generation

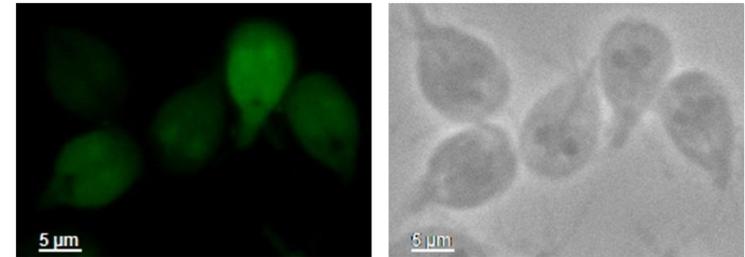
91707-HA A

91707-HA B

1

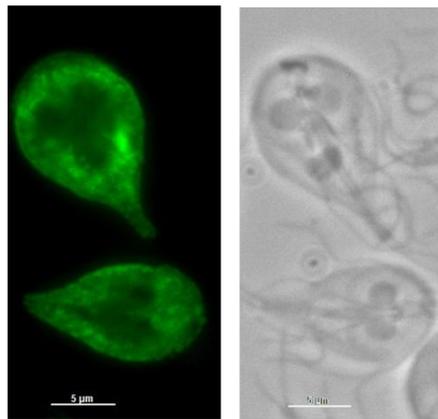


Plasma Membrane

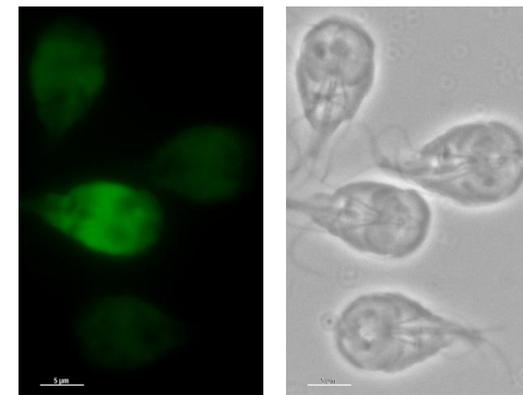


Endoplasmic Reticulum/Nucleus

4

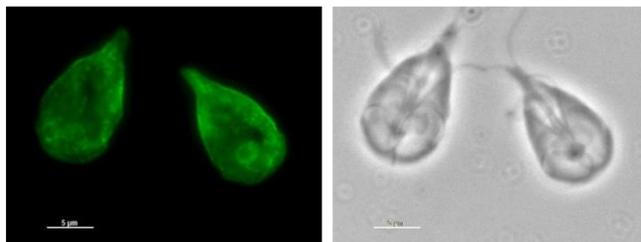


Endoplasmic Reticulum/PVs

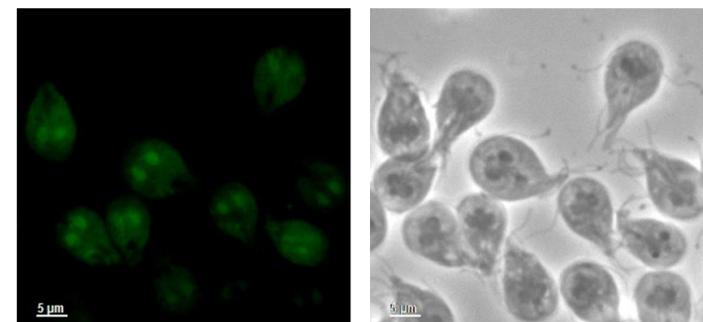


Endoplasmic Reticulum

7



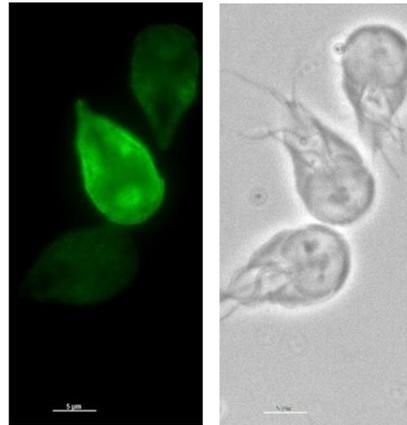
Plasma Membrane



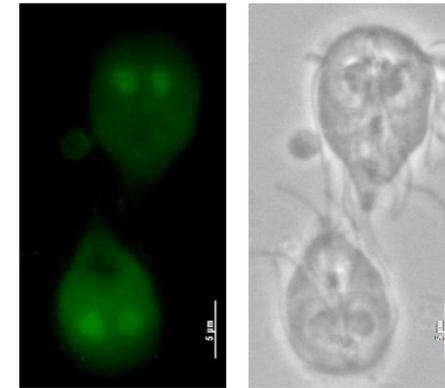
Endoplasmic Reticulum/Nucleus

Table 4 (cont.). Localization of 91707-HA in *Giardia intestinalis* transfectants A and B.

10

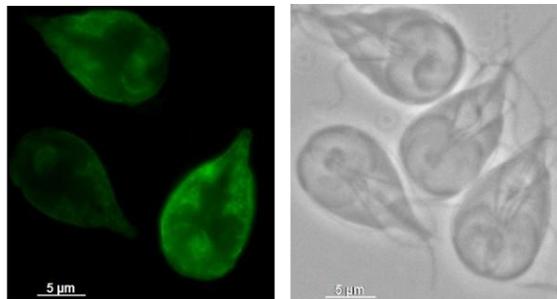


Plasma Membrane/PVs

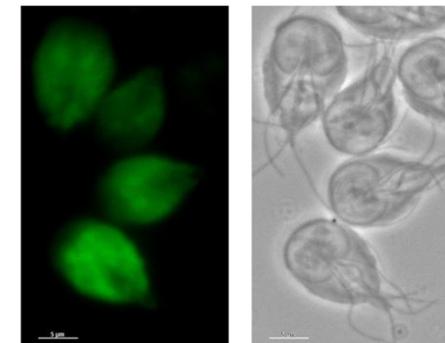


Endoplasmic Recitulum/Nucleus

13

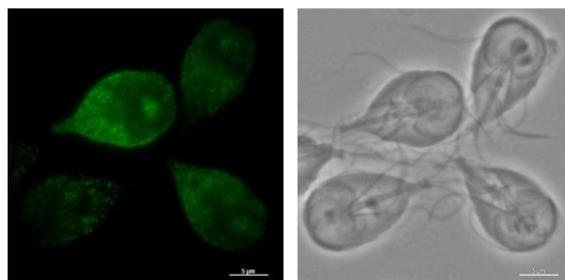


Plasma Membrane/PVs

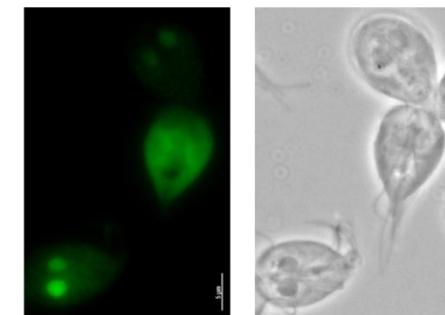


Endoplasmic Recitulum/Nucleus

16



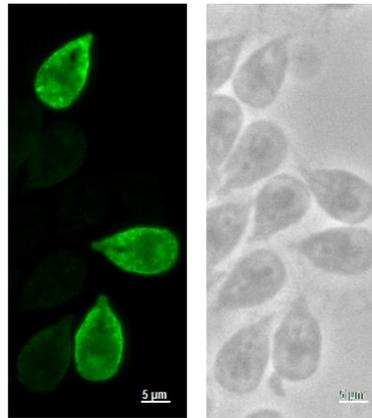
Plasma Membrane/PVs



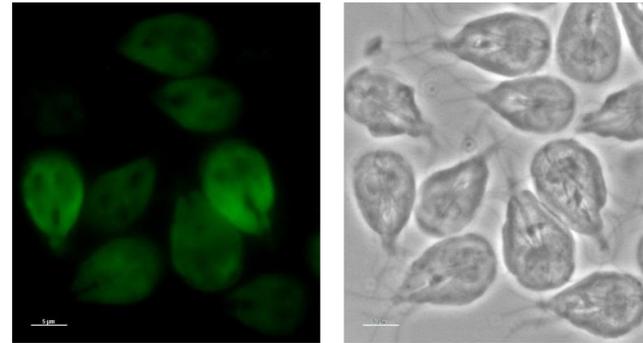
Endoplasmic Recitulum/Nucleus

Table 5 (cont.). Localization of 91707-HA in *Giardia intestinalis* transfectants A and B.

22

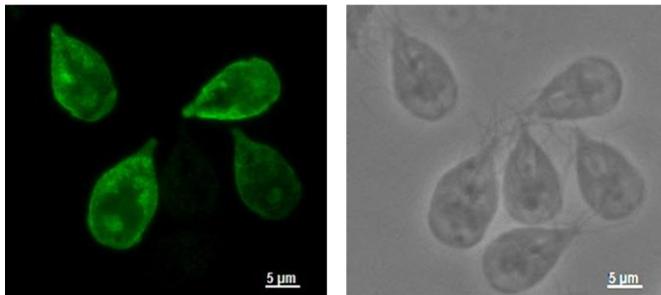


Plasma Membrane/PVs

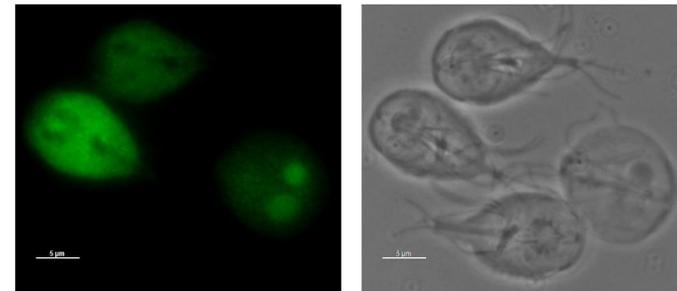


Endoplasmic Recitulum

28

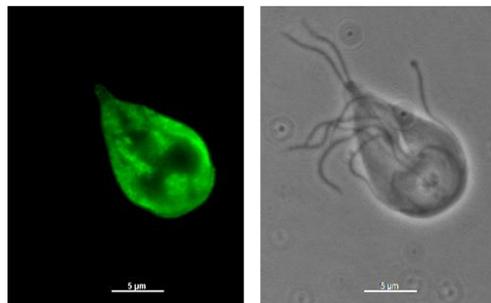


Plasma Membrane/PVs

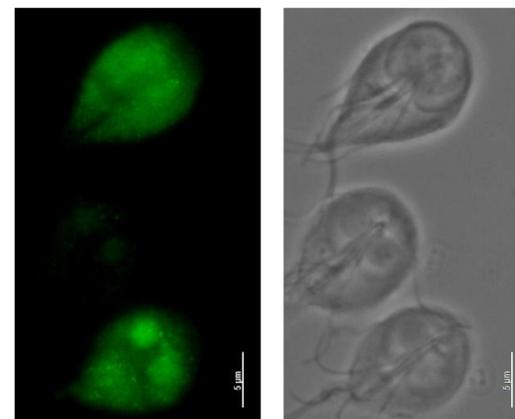


Endoplasmic Recitulum/Nucleus

37



Plasma Membrane/PVs



Endoplasmic Recitulum/Nucleus

The pattern of cellular localization for 91707-HA in transfectant A is the plasma membrane. This localization is believed to be the correct one for this protein, based on the results with the freshly transfected cells obtained from early experiments in the laboratory. Transfectant A is referred as the culture passage closer to the date of transfection. On the other hand, there is different scenario regarding transfectant B. This transfectant has been passed for a longer period of time, using a small amount of initial culture each time. Transfectant B showed mainly ER and nucleus localization for 91707-HA to later change into plasma membrane. One hypothesis to explain this phenomenon is the adaptation of the trophozoites to the overexpression of 91707-HA. This means that somehow, along several generations, this transfectant is trying to regulate the expression of the HCP to avoid the high loading of the recombinant protein to the plasma membrane. It is to remember that the tagged protein is expressed episomally while the cell still carries the other four chromosomal copies of the gene. What it is observed in transfectant B could be the actual biosynthesis and transport of the 91707 HCP, being synthesized in the nuclear periphery and delivered into the ER for transport to the plasma membrane, as its final cellular localization. If this is true, there is no signal of 91707-HA in the plasma membrane when the nuclear periphery is synthesizing the protein. This means that there is a protein turnover not by internalization and degradation but probably by secretion of 91707 to the media for then synthesize more of the protein. To prove this a specific antibody is needed to be able to detect also the 91707 protein generated from the chromosomal copies of the gene not carrying the tag. The chromosome versions of the protein might be present in the plasma membrane when the tagged protein is being synthesized. If this is the situation, then it means the cell is expressing 91707 from different genomic loci and became able to regulate the episomal copy in the same manner as the chromosomal ones.

In both transfectants, the cell counting revealed a high number of cells with no signal detected when compared to the total number of counted cells. These cells in fact might be expressing the untagged 91707 protein translated from the chromosomal genes, similar to the above explanation.

4.2 Recombinant High Cysteine Proteins Expression and Purification

Two HCPs GL50508_91707 and GL50508_7715 were selected for production and purification to produce an AB (tool) and be used in functional assays with *G. intestinalis* trophozoites. From now on, the proteins will be mentioned as 7715-His and 91707-His.

4.2.1 Cloning of 91707 and 7715 for Recombinant Protein Expression

The first step of this part of the project was the construction of the vectors carrying 91707 or 7715. pEXP5-CT/TOPO® (Invitrogen™) was the chosen vector for this cloning containing a Histidine tag on its sequence, which will be fused to the c-terminal region of the protein once expressed. Both 91707 and 7715 gene sequences did not contain the signal peptide and transmembrane domain fragments. Therefore, an ATG codon was added in the 5' region of the forward primer. PCR reactions were performed in order to amplify the desired gene sequences from *Giardia* genomic DNA. The expected band sizes for the genes without the referred fragments are 1806bp for 91707 and 1836bp for 7715. PCR products were obtained according to the description in Section 3.4.2 and its sizes could be compared to the expected ones by visualization in 1xTAE agarose gel. In Figure 11 are shown the PCR products from the amplification reaction.

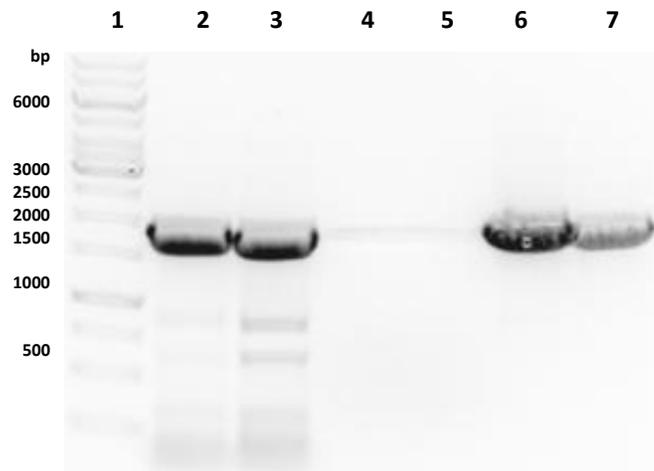


Figure 11. PCR products from *Giardia* genomic DNA in 1% agarose gel. 1 – GeneRuler™1kb DNA Ladder; 2 and 3 – 7715 fragment; 4 and 5 – Empty lanes; 6 and 7 – 91707 fragment.

Lanes 2 and 3 show the PCR products of 7715 fragment. PCR products in these lanes are localized between the 1.5kb and 2.0kb bands showed by the ladder (lane 1), meaning that the amplified products have the expected size. There can also be seen some extra bands lower in

these two lanes, meaning that some unspecific amplification may have occurred. However, these accessory bands will not interfere with the amplified fragment, since the latter will be excised by GeneJET™ Gel Extraction Kit (Section 3.4.5).

Lanes 6 and 7 show the PCR products of *91707* fragment. Once again, the bands are found slightly below the 2.0kb band showed in the ladder lane, meaning that these fragments have the expected size. The four bands obtained from the PCR reaction show high intensity. The concentration of DNA obtained after performing GeneJET™ Gel Extraction Kit was 74ng/μL for *7715* and 89ng/μL for *91707*. These numbers represent a good recovery yield in gel extraction recovery.

After gel-purify *91707* and *7715* fragments obtained by PCR, these DNA sequences were cloned into pEXP5-CT/TOPO® vector according to the description in Section 3.4.6 and *E. coli* TOP10 strain was transformed with the recombinant constructs, as mentioned in Section 3.4.8. To analyze the transformants, colony-PCR was performed as mentioned in Section 3.4.3 in order to evaluate the orientation of the gene in the vector. This is possible since the two primers used in the PCR reaction anneal with different sequences: one of the primer is specific for the vector and the other for the insert. 10 colonies from each construct transformation were randomly selected for colony-PCR. The 20 reactions were detected and analyzed in 1% agarose gel, as is shown in Figure 12.

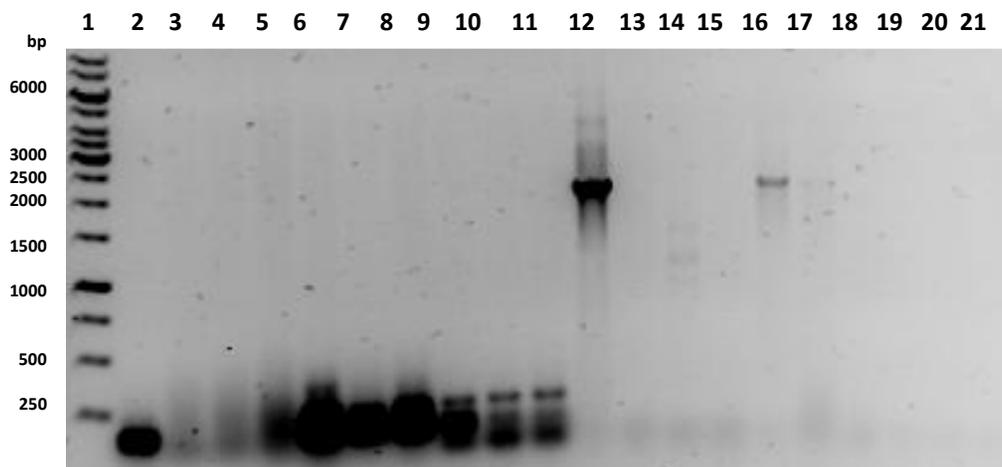


Figure 12. Colony-PCR products of pEXP5-CT/TOPO®-*7715* and pEXP5-CT/TOPO®-*91707* constructs extracted from *E. coli* TOP10 transformants in 1% agarose gel. 1 – GeneRuler™1kb DNA Ladder; 2-11 – *7715*; 12-21 – *91707*.

In the lanes corresponding to 7715 fragment (lanes 2 to 11), no bands with the expected size, ~ 1.8kb, can be detected. Essentially, in all lanes are visible intense bands of primer dimer with a size lower than 250bp. Primer dimer occurs when primer molecules hybridize and DNA polymerase amplifies this dimer, leading to the inhibition of the amplification of the target DNA sequence.

In the lanes corresponding to 91707 fragment (lanes 12 to 21), bands with an approximate size of 2.0kb are visible. The extension with the T7 forward primer used in the amplification adds 76 nucleotides from the vector to the fragment, increasing the expected size to ~ 1.9kb, in agreement to what is seen in Figure 12. In lane 12 a clear and intense band is visible while in lanes 16 and 17 the bands are faint but they are present in the correct size position. These three clones, 1, 5 and 6 respectively, were confirmed to be positive ones by Sanger-sequencing (Uppsala Genome Center).

Since no positive clones were detected for pEXP5-CT/TOPO[®]-7715 construct, a new TOPO[®] ligation reaction and *E. coli* TOP10 transformation were performed and 20 new colonies were tested by colony-PCR. The results from the colony-PCR were analyzed in 1xTAE agarose gel, as shown in Figure 13.

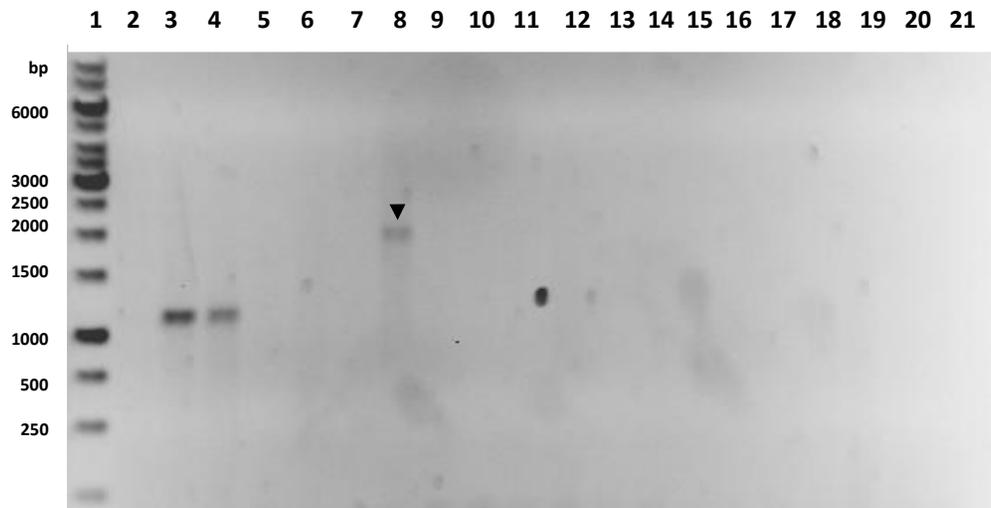


Figure 13. Colony-PCR products of pEXP5-CT/TOPO[®]-7715 construct extracted from *E. coli* TOP10 transformants in 1% agarose gel.

1 – GeneRuler™1kb DNA Ladder; 2-21 – 7715.

In lanes 3 and 4 is visible an artifact band slightly above the 1.0kb band showed by the ladder. In lane 8, a very faint band can be detected around 2.0kb, shown by the black arrow. This clone, termed clone 7, was confirmed to be positive by Sanger-sequencing.

After confirming that the putative positive clones from pEXP5-CT/TOPO[®]-7715 and pEXP5-CT/TOPO[®]-91707 were, in fact, positive, these constructs were transformed into the expression *E.coli* strain BL21 Origami, as described in Section 3.4.8. We assumed that all colonies present in the plates were positive, given that they acquired the resistance to ampicillin only by internalizing the recombinant TOPO[®] vector.

4.2.2 Optimization of protein induction and expression

E. coli BL21 Origami transformants with pEXP5-CT/TOPO[®]-7715 and pEXP5-CT/TOPO[®]-91707 were induced to express the respective proteins according the conditions described in Section 3.11.1. These inductions were assessed in 10% SDS-PAGE followed by Western Blot or Coomassie Blue staining. These recombinant proteins, 7715-His and 91707-His, have a molecular weight of 66.2 kDa and 65.6 kDa, respectively.

Different conditions were tested to induce the expression of proteins 91707-His and 7715-His. In Figure 14 are represented the results from the first induction of proteins expression, which occurred at 37°C during 3h and 1mM IPTG.

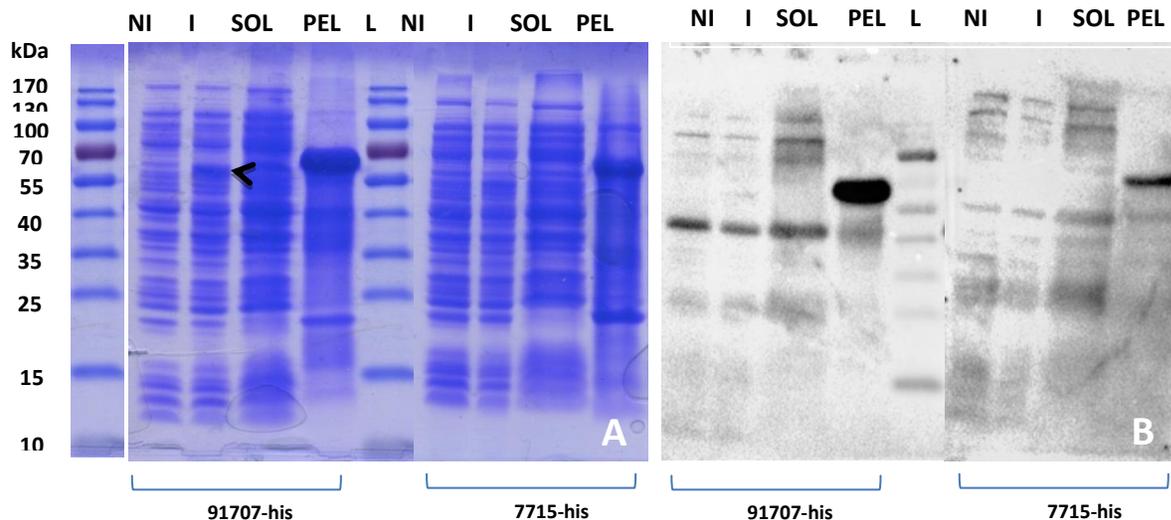


Figure 14. SDS-PAGE analysis of the expression of His tagged proteins. **A** – Coomassie gel; **B** – Western blot.

NI – Non Induced; I – Induced; SOL – Soluble fraction; PEL – Insoluble fraction; L – PageRuler Prestained Protein Ladder.

In the Coomassie gel from Figure 14-A the lanes corresponding to 91707-His samples a band between the 55 kDa and 70 kDa in the induced sample is present as indicated by the black arrow. This meant that the induction of 91707-His might have worked. However, in the western blot corresponding to the induced sample (Figure 14-B) the His-tagged protein does not seem to appear. In the soluble fraction, there is a faint band in the correct molecular weight position (~60kDa), which is not confirmed by the western blot. On the other hand, in the insoluble fraction there is present a strong and distinct band in the same position that also shows up in the western blot, meaning that 91707 was expressed in the inclusion bodies. Inclusion bodies consist in aggregates of denaturated protein in the form of particles. Usually, the aggregated components belong to the organism of which the expressed recombinant protein is native from. The rest of the bands observed in the Blot represent background given by the His-tag antibody, since histidine is present in several proteins. Additionally, the overexposure of the membrane in the detector increases the detected signal.

One explanation for 91707-His not giving a signal in the induced samples by Western Blot could be that the protein is not highly expressed after induction and, therefore, not detected by the anti-his tag antibody. However, the presence of a highly intense band with the expected molecular weight in the insoluble fraction supports the conclusion that there was an induction; although, the protein is insoluble.

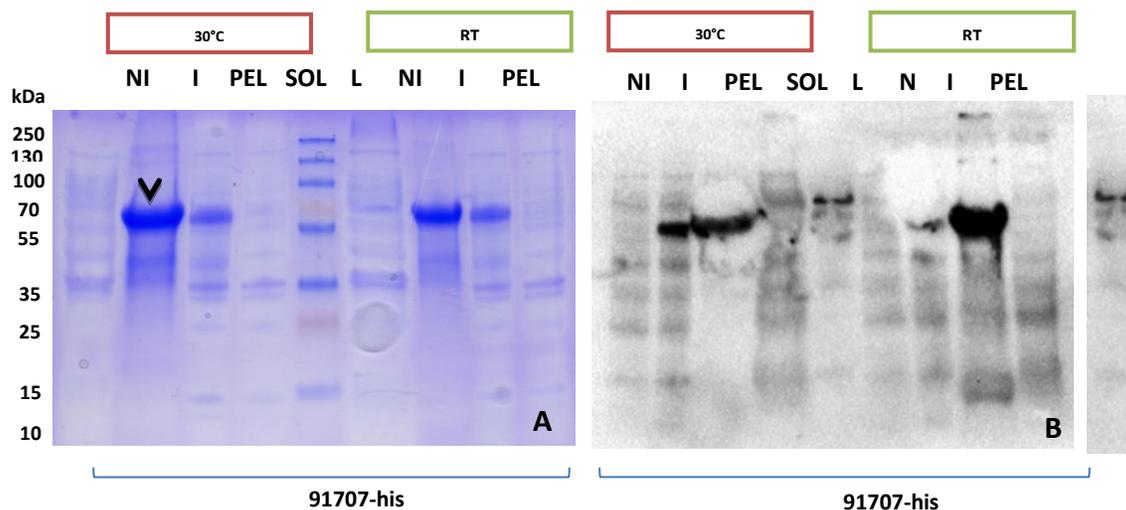


Figure 15. SDS-PAGE analysis of the expression of 91707-His at 30°C and room temperature. A – Coomassie Gel; B – Western Blot.

NI – Non Induced; I – Induced; SOL – Soluble fraction; PEL – Insoluble fraction; L – PageRuler Plus Prestained Protein Ladder.

Regarding the lanes corresponding to 7715-His samples, does not seem to be a difference between the non-induced and the induced samples, both in Coomassie gel and western blot. Once again, there is a distinct band around 60 kDa present in the insoluble fraction, which is confirmed to be the His-tagged protein in the western blot. Thus, the expression of this protein also led to the formation inclusion bodies. In the non-induced, induced and soluble fraction samples there are present some accessory bands which represent unspecific antibody binding.

The induction of 91707-His and 7715-His expression was performed in different conditions from the ones adopted in the first induction in order to assess if decreasing the temperature would avoid the formation of inclusion bodies. The Coomassie gels and Western Blots from these inductions are represented in Figures 15 and 16.

In Figure 15-A is shown the Coomassie gel from the induction of 91707-His at 30°C and room temperature o/n. The expression of the protein in different conditions was successful. A distinct band, indicated by the black arrow, with a molecular weight around 60kDa can be seen in the lane corresponding to the induced sample. This is both verified in the Coomassie gel and in the Western Blot, in opposition to what happened with the first induction. However, the protein is still being expressed with the formation of inclusion bodies in both conditions, similarly to what was verified in the induction at 37°C. Apparently, the decrease of temperature did not prevent the formation of inclusion bodies.

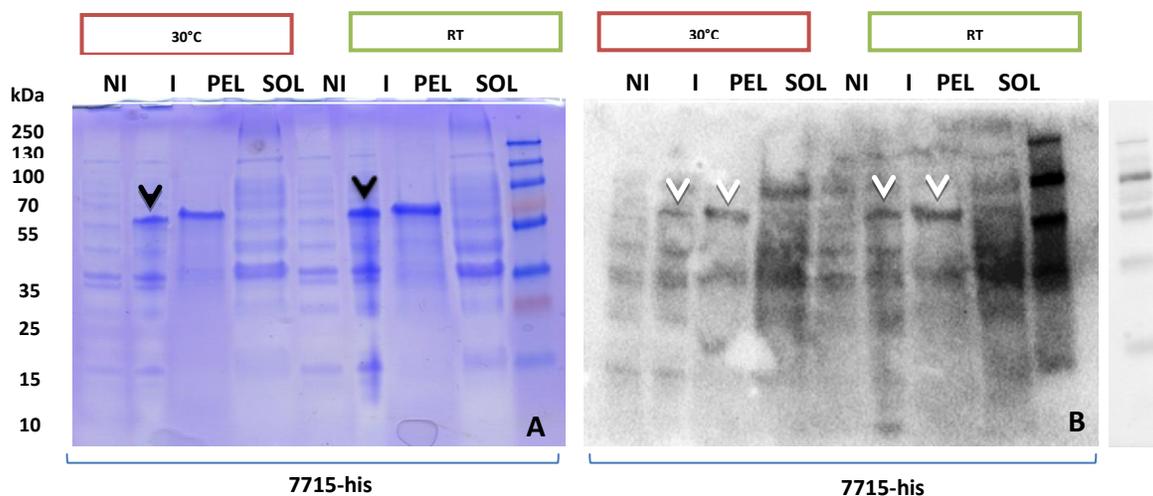


Figure 16. SDS-PAGE analysis of the expression of 7715-His at 30°C and room temperature. A – Coomassie Gel; B – Western Blot.

NI – Non Induced; I – Induced; SOL – Soluble fraction; PEL – Insoluble fraction; L – PageRuler Plus Prestained Protein Ladder.

In the Coomassie gel from Figure 16-A is shown the expression of 7715-His induced at 30°C and room temperature. In the induced samples from both conditions, there is a band with a molecular weight around 60kDa. In the soluble fraction from the 30°C induction it is present a band in the correct molecular weight position, potentially corresponding to the His-tagged protein. The results observed in the Coomassie Gel are confirmed by Western Blot, shown in Figure 16-B. The interest bands in the induced and insoluble samples are indicated with white arrow. Once again, the protein seems to be expressed on its insoluble form.

Overall, the expression of 91707-His and 7715-His was higher at 30°C and RT when compared to the one at 37°C. Unfortunately, the insolubility issue of both proteins was not solved with the decrease of the temperature during the inductions. Facing this problem, the first approach to solubilize the proteins was performed using 8M urea. However, the solubilization of the protein was not complete and a high amount of the protein fractions was still insoluble. Additionally, the use of high concentration of Urea could interfere with the process of purification of the his-tagged proteins. The second approach to solubilize the protein was the improvement of cell sonication. In the results from one of the several protein inductions performed, part of the protein appeared in the soluble form, as it can be seen in Figure 17. Starting from here, the initiative to optimize cell sonication in order to obtain the proteins in their soluble form was thought of.

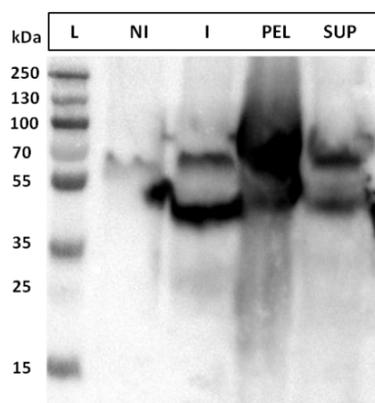


Figure 17. Western Blot 91707-His induction at RT: sonication optimization. L – PageRuler Plus Prestained Protein Ladder; NI – Non Induced; I – Induced; PEL – Insoluble fraction; SUP – Soluble Fraction.

4.2.3 Evaluation of protein purification by NGC™ Chromatography

The purification of 91707-His and 7715-His was made when the proteins were finally produced on their soluble form. The results from the first purification of the his-tagged proteins are shown in Figures 18 and 19. In the Coomassie gel from Figure 18, in the lanes corresponding to the flowthrough (FT) and column washes (5 and 20) there is high content of protein that is not binding to the column matrix. In the lanes corresponding to proteins elution fraction, 46 to 55, it is present a very thin band with the expected molecular weight for 91707-his, indicated by the black arrow. In the Western Blot, Figure 18-B, there is a distinct band with the expected molecular weight in each lane from protein fraction, indicating that 91707-his was purified. However, similar bands to these ones are also faintly present in the flowthrough and column washes fraction, meaning that this buffer should be modified in order to wash unspecific unbound proteins without removing the interest his-tagged protein. Additionally, the wash volume should be increase to make sure that a higher amount of unspecific bound protein and components are removed before the elution of 91707-His.

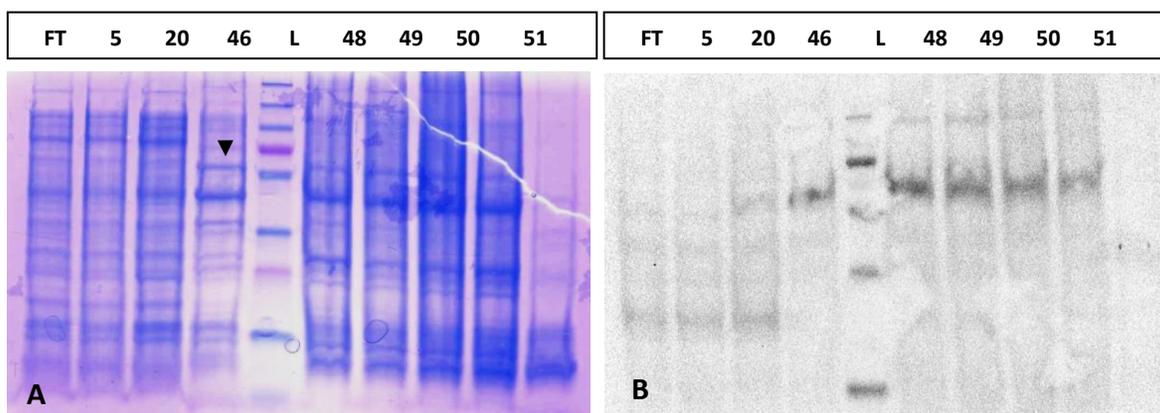


Figure 18. SDS-PAGE analysis of the purification of 91707-His by IMAC. A – Coomassie Gel; B – Western Blot.

FT – Flowthrough; 5 and 20 – Wash Fractions; 46-55 – Protein Elution Fractions; L – PageRuler Plus Prestained Protein Ladder.

In Figure 19-A is shown the Coomassie gel from the purification of 7715-His by IMAC. It is worth to mention that the induced BL21 Origami culture for the expression of this protein used for this purification did not show to be positive. Thus, this purification was performed in a blind way, in case the induction turned out to be positive due to a misevaluation. As expected, both Coomassie gel and Western Blot did not reveal the presence of 7715-His in the eluted protein fractions or, actually, in any other fractions. Do to all the issues associated to this gene that were

faced during the production of 7715-His, it was decided to move forward and optimize exclusively the purification of 91707-His.

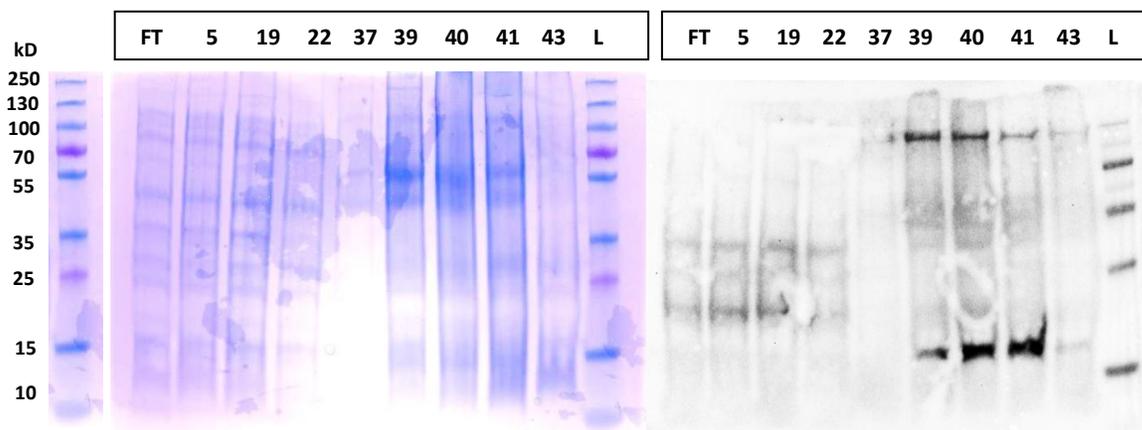


Figure 19. SDS-PAGE analysis of the purification of 7715-His by IMAC. A – Coomassie Gel; B – Western Blot.

FT – Flowthrough; 5, 19 and 22 – Wash Fractions; 37-43 – Protein Elution Fractions; L – PageRuler Plus Prestained Protein Ladder.

At this point, more bacteria culture was induced and processed in order to perform a new purification of 91707-His and evaluate the purification profile with the new set of buffers. The composition of these buffers can be found in Table 2 (Section 3.12). Triton was a component excluded from the primary set of buffers. Triton is a detergent usually used in particular protein purifications. It helps cell lysis and solubilization of proteins by disrupting lipid membranes, not affecting the biological activity of the protein. Since one managed to have 91707-His in its soluble form, the main function of the detergent was no longer needed. Besides this, the detergent will also be present in the protein fractions after elution, helping the protein to be stable. However, to be sent for antibody production, the protein should be stable in a buffer without components that may harm the animal during immunization. Thus, the Triton would have to be removed from these buffers, which could lead to the precipitation of the protein due to loss of stability. For this new purification two different wash buffers were prepared, with increasing concentration of NaCl, to assess if more unspecific bound proteins could be removed without eluting the his-tagged protein. The concentration of Imidazole was also increased within these buffers in order to remove the unspecific protein histidine bound.

In Figure 20 it is present the result from the purification of 91707-His when 250ml of culture were processed. In the lanes corresponding to the clear protein lysate and Flowthrough in the Coomassie gel there is a high protein content, which results in a huge background showed in the Western Blot. When looking at the wash fractions with buffers 1 and 2, the ones corresponding to washes with buffer 2 contain two faint bands, being the lowest one in the molecular weight corresponding to 91707-His, around 60 kDa, meaning that some of the protein could had been eluted. However, there is no detection in the Western Blot from either one of these bands, which indicates that the increase in concentration of NaCl and Imidazole were sufficient to elute unspecific bound proteins without washing out the his-tagged protein. Additionally, there are no bands present in the fraction from the wash with buffer 1, meaning that the concentration of these two components in this buffer is not adequate to properly wash the column. Finally, in the protein fractions samples there are present two bands with similar molecular weight seen in the Coomassie gel, which are equally detected in the Western Blot. The presence of these two bands may indicate protein degradation, indicating that the histidine tag is present in an incomplete 91707, binding to the Nickel ions present in the column and being eluted with high concentration of Imidazole.

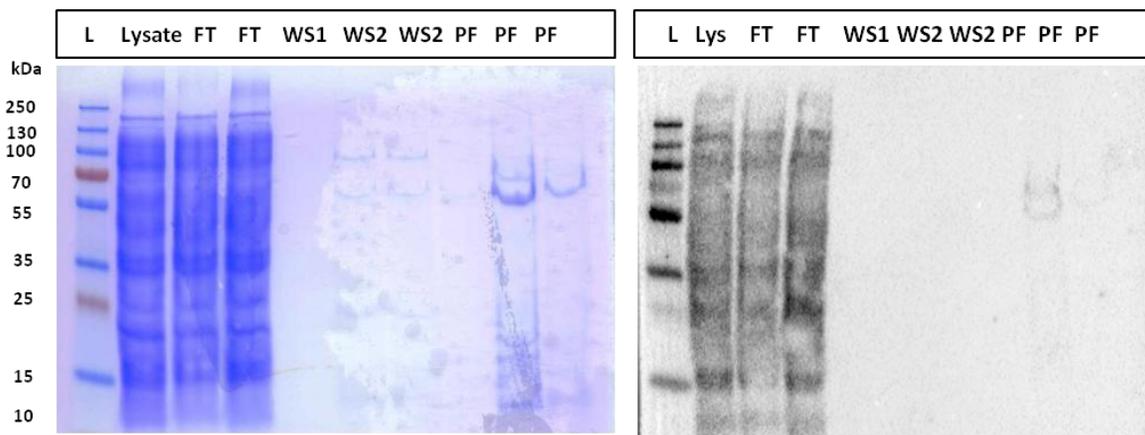


Figure 20. SDS-PAGE analysis of the purification of 91707-His by IMAC from 250ml culture. A – Coomassie Gel; B – Western Blot.

L – PageRuler Plus Prestained Protein Ladder; Lysate – Clear Protein Lysate; FT – Flowthrough; WS1 – Wash Fractions 1; WS2 – Wash Fractions 2; PF – Protein Fractions.

After inducing another 1L of bacteria culture to express 91707-His, a new purification was made, using the new buffers chosen: resuspension buffer, wash buffer 2 and elution buffer, all without Triton. The results from this purification are shown in Figure 21. By analyzing the Coomassie gel and the Western Blot, the induction of expression of 91707-His seems to be positive, since one band around 60kDa can be seen in the Induced fraction in the Coomassie gel, which in turn gives a strong and clear signal in the Western Blot. The clear protein lysate and Flowthrough present a high protein content, which results in a huge background in the Western Blot, being imperceptible a signal from the his-tagged the protein in these fractions. The wash fraction shows the elution of unspecific bound proteins but not a his-tag signal in the Western blot, indicating that 91707-His was not eluted during the column wash. Concerning protein fraction samples it is present a strong band in both lanes with the correct molecular weight. Some accessory bands with distinct molecular weights are also present, which resemble the protein pattern present in the wash fraction lane. This means that the wash volume used in this purification should have been increased. In the protein fraction present in the Western Blot, there are present two faint bands, one of them with the expected molecular weight for 91707-His. However, the bands in the Coomassie gel which apparently correspond to these ones are strong. Thus, it was expected a stronger signal in the Western Blot for these same bands.

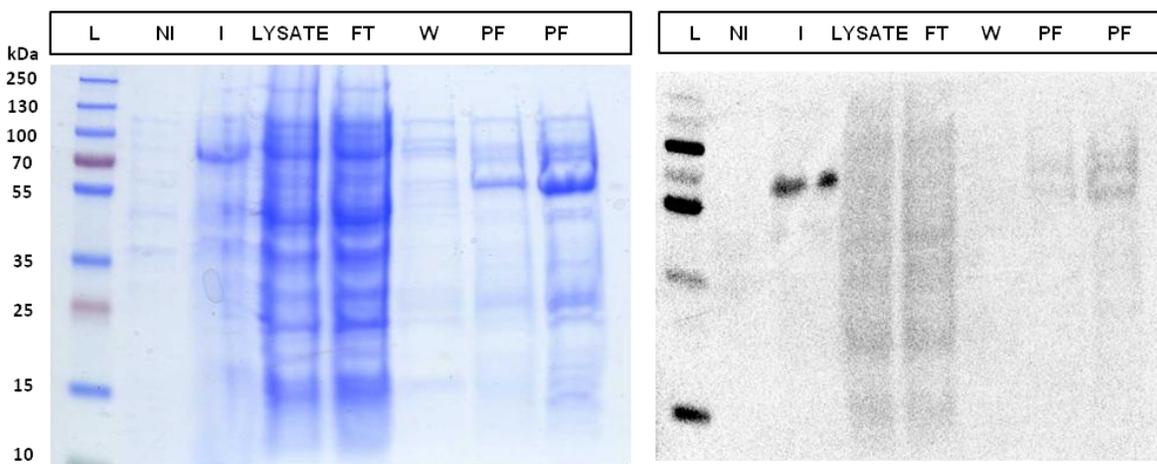


Figure 21. SDS-PAGE analysis of the purification of 91707-His by IMAC from 1L culture. A – Coomassie Gel; B – Western Blot.

L – PageRuler Plus Prestained Protein Ladder; NI – Non-Induced Fraction; I – Induced Fraction; Lysate – Clear Protein Lysate; FT – Flowthrough; W – Wash Fractions; PF – Protein Fractions.

Figure 22 represents the chromatogram from this purification obtained with the analyzer software ChromLab™. The green line represents the protein content distributed by the fractions collected along the chromatography and it is detected in an absorbance peak of 280nm due to the aromatic ring present in some amino acids, such as histidine and tryptophan. The red line represents the conductivity in the fraction, which increases exactly when the wash buffer (2) is injected due to an increase of NaCl concentration. The conductivity level is unaffected from this point on since the concentration of NaCl is the same in the wash and elution buffers.

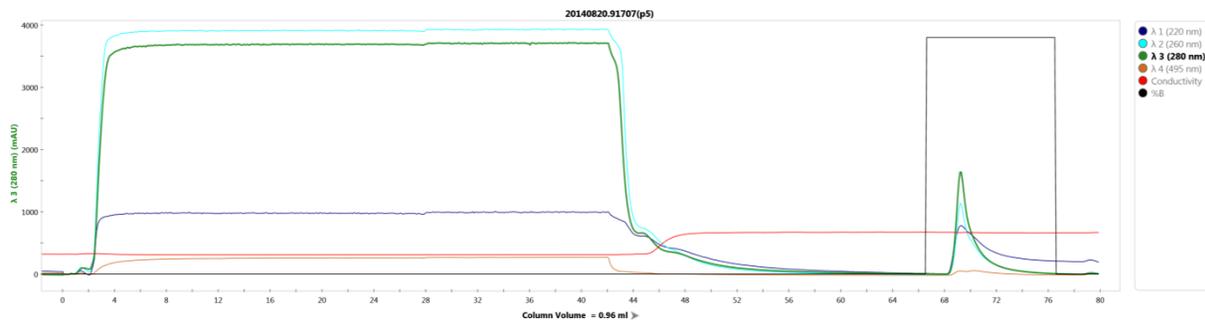


Figure 22. Chromatogram of the purification of 91707-His from 1L of culture with ChromLab™ software.

A new Western Blot was performed with the same samples to reevaluate the purification in case some mistake had been made during the process. The results were exactly the same. Thus, the proteins fractions were desalted, meaning 91707-His was placed in a new buffer without Imidazole. The high amount of Imidazole present in the eluted fractions may harm the rabbit during the boosts. However, the removal of the Imidazole may lead to protein precipitation since this alkaloid stabilizes the molecule.

The desalted fractions were then processed in a size exclusion chromatography (SEC) to exclude proteins with a different size of 91707-His. The chromatogram from SEC is present in Figure 23. The three peaks represent the fractions with higher protein content.

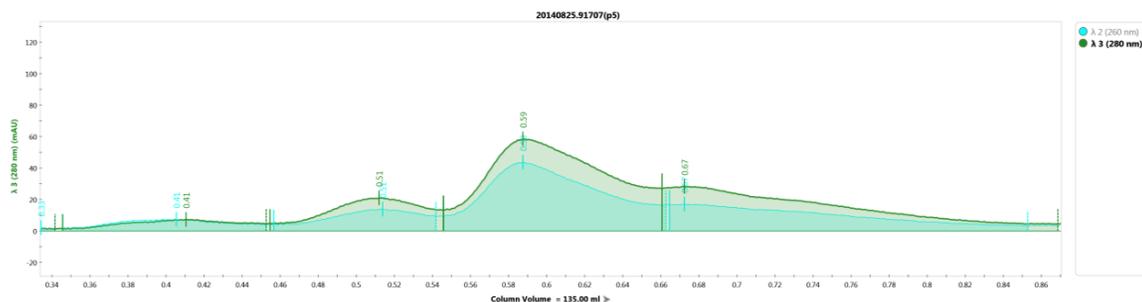


Figure 23. Chromatogram of the size exclusion chromatography of 91707-His using ChromLab™ software.

These three fractions corresponding to the three peaks were evaluated by SDS-PAGE followed by Western Blot and its result can be seen in Figure 24. In all fractions was observed more than one band. In fraction 5G, two bands with close molecular weight are observed. Both bands show an apparent correct molecular weight for 91707-His. However, this may indicate protein degradations along the purification. Therefore, a protein sample will be send to the Mass Spectrometry facilities in Uppsala University to evaluate if 91707-His is, in fact, present in the SEC fraction, before trying to repeat the whole procedure to obtain the protein.



Figure 24. SDS-PAGE analysis Size Exclusion Fractions of 91707-his.

4.3 Bioinformatics Analysis & miRNA regulatory target sequences in HCPs

In this part of the project a strategy was design and followed in order to evaluate if some HCPs could be regulated by microRNAs. For this purpose I searched for shared sequence homology across the 3'end and 3'UTR regions of the whole HCP family which could indicate the presence of binding sites for a regulatory element.

150 nucleotides from the 3'end and 3'UTR from 82 HCP sequences were aligned using the online tool *Clustal Omega*. According to the sequence homology, 33 HCPs were grouped as shown in Table 3. In Appendix 8.2 can be found an example of the alignment of the sequence in Group 1.

Table 4. HCPs grouping according to the alignment of the 3'end and 3'UTR sequences.

HCP SUBGROUP	GENES	
1	GL50803_113531	GL50803_114626
	GL50803_137727	GL50803_113836
	GL50803_114161	
2	GL50803_113987	GL50803_27717
	GL50803_103943	GL50803_112126
	GL50803_87706	GL50803_114891
3	GL50803_115158	GL50803_10659
	GL50803_101589	GL50803_113512
	GL50803_14783	GL50803_114470
	GL50803_32701	GL50803_112673
	GL50803_112135	
4	GL50803_16936	GL50803_14791
	GL50803_17380	GL50803_112584
	GL50803_113801	GL50803_114852
	GL50803_114888	GL50803_137672
5	GL50803_7715	GL50803_15008
	GL50803_91707	GL50803_112604
	GL50803_102180	

After grouping the selected HCPs, the seed regions of 50 putative miRNA published in Zhang *et al.*, (2009) and mir6 and mir10, two confirmed miRNA in *Giardia* (Li *et al.*, 2012), were matched with the 3'end and 3'UTR sequences present in the five groups. Table 4 shows in which HCP sequence the seed region of a determined miRNA is present.

Table 5. Identification of putative miRNA regulation sites in HCPs grouped sequences

Groups	Protein	Targeting miRNA
Group 1	GL50803_113531	Gl-mir-23,
	GL50803_114626	Gl-mir-23,
	GL50803_137727	-
	GL50803_113836	Gl-mir-34, Gl-mir-46
	GL50803_114161	Gl-mir-34,
Group 2	GL50803_113987	Gl-mir-41
	GL50803_27717	Gl-mir-26, Gl-mir-41
	GL50803_103943	Gl-mir-26, Gl-mir-41
	GL50803_112126	Gl-mir-3, Gl-mir-5, Gl-mir-20, Gl-mir-41
	GL50803_87706	Gl-mir-5, Gl-mir-34, Gl-mir-41
	GL50803_114891	Gl-mir-5, Gl-mir-34, Gl-mir-41
Group 3	GL50803_115158	Gl-mir-35,
	GL50803_10659	Gl-mir-5, Gl-mir-18, Gl-mir-45, mir10
	GL50803_101589	Gl-mir-5, Gl-mir-20, Gl-mir-32
	GL50803_113512	Gl-mir-30,
	GL50803_14783	Gl-mir-5, Gl-mir-30, Gl-mir-49, Gl-mir-50, mir10*
	GL50803_114470	Gl-mir-5, Gl-mir-30
	GL50803_32701	Gl-mir-11, Gl-mir-18
	GL50803_112673	Gl-mir-2, Gl-mir-7, Gl-mir-13, Gl-mir-14
	GL50803_112135	Gl-mir-2, Gl-mir-7, Gl-mir-15
Group 4	GL50803_16936	Gl-mir-2, Gl-mir-39, Gl-mir-49,
	GL50803_14791	Gl-mir-23,
	GL50803_17380	Gl-mir-7, Gl-mir-10
	GL50803_112584	Gl-mir-43
	GL50803_113801	Gl-mir-3,
	GL50803_114852	-
	GL50803_114888	-
	GL50803_137672	-
Group 5	GL50803_7715	Gl-mir-5
	GL50803_15008	Gl-mir-34
	GL50803_91707	Gl-mir-1, Gl-mir-5, Gl-mir-22, Gl-mir-36
	GL50803_112604	Gl-mir-16
	GL50803_102180	Gl-mir-26, Gl-mir-31

After assessing that some HCPs sequences contain putative regulation target sites for miRNA, it was decided to test if some of these miRNAs, particularly the most frequent ones and tested in *Giardia*, could be detected. To do so, Stem-Loop RT-qPCR (SLRT-qPCR) was performed using specific primers to these miRNAs. In the first SLRT-qPCR assay was used total RNA from *Giardia intestinalis* to test the primers. Primers designed to target and amplify miRNA are extremely short, which can lead to an annealing in unspecific regions. The primers designed for SLRT-qPCR can be found in Annex 8.1. In the end, the samples were evaluated in 1% agarose gel,

shown in Figure 25. NTC (non-template control) sample for mir2 was selected because it showed detection in the qPCR curve analysis. Since this is a non-template control, it should not have given an amplification product.

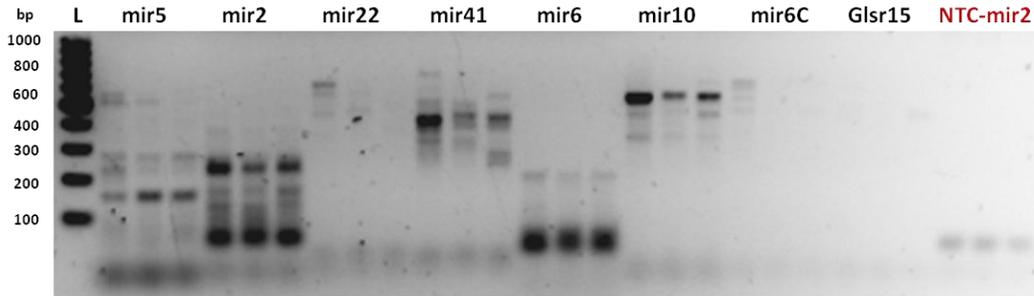


Figure 25. Analysis of samples from SLRT-qPCR in 1% agarose gel.
L – GeneRuler 100 bp DNA Ladder (thermo Scientific™).

Samples from mir5, mir22, mir41 and mir10 showed several PCR products with a base-pair number higher than 100bp. The miRNAs that are being targeted have an average length of 25nt plus the bases added by the Stem-loop primer, whereby smaller products should have been detected. Due to the short length of the primers, it might have happened non-specific annealing, which lead to the amplification of the snoRNA from where these miRNAs are derived, rather than the respective miRNAs or other completely unrelated transcripts. Some of mir2 PCR products have a length below 100bp, which may indicate a correct amplification of the target miRNA with this set of primers. However, since a PCR product was present in the NTC for this miRNA, it was presumed that a contamination might have happened. Finally, mir6 samples showed PCR products of the expected sizes. Additionally, the control for this miRNA, mir6C, did not show any amplification, indicating that this set of primers has worked. Even though mir6 is not present in Table 5, this miRNA was found in some HCP sequences when longer 3'UTR were considered. Thus, a new SLRT-qPCR was performed with these two sets of primers. This time, the template used in the reaction was total RNA extracted from trophozoites during interaction with intestinal epithelial cells (IEC) in different time points. Total RNA from two biological replicates were used.

The results obtained from the qPCR reaction were evaluated according to the following formulas:

$$(1) \Delta C_T = C_{T_{tested\ sample}} - C_{T_{control\ sample}}$$

$$(2) \text{Fold Change} = 2^{-(\Delta C_T)}$$

Formula (2) for comparative analysis was adjusted since $\Delta\Delta C_T$ could not be calculated, due to the absence of a reference gene, a gene which expression is constant during the whole experiment which represents the background of the method. Based on this a relative quantification of the change in expression levels is obtained as compared to the use of a standard curve method (Absolute Quantification) for which we did not have enough material.

When analyzing the graphs Figure 26, both in replicates A and B there is a decrease in the expression level of mir6 up to 3h of interaction when compared to samples at time 0 (A0 and B0). These samples consist in total RNA collected from trophozoites before the interaction with IEC. In replicate A, when looking at the sample from 4.5h of interaction, the expression of mir6 seems to increase to a level very close to the control one. However, in replicate B there is also an increase of mir6 expression when comparing to earlier time points from interaction samples, but it does not reach the level verified in the control sample.

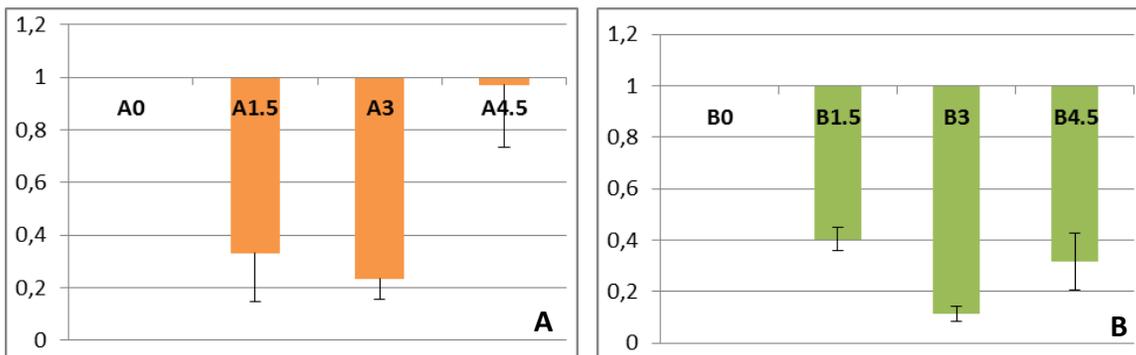


Figure 26. Expression level of mir6 miRNA in the biological replicates A and B.

A0/B0 – Control sample; A1.5/B1.5 – Interaction during 1.5h; A3/B3 – Interaction during 3; A4.5/B4.5 – Interaction during 4.5h.

During parasite-host interactions a set of HCPs showed very high fold up-regulation (unpublished data), including 7715 and 91707. If mir6 is, in fact, regulating the expression of some HCPs, it makes sense that the level of this miRNA decreases throughout interactions. During interactions the trophozoites are obviously under different conditions than the ones they face when cultures *in vitro*, for instance exposure to IEC secretion products and higher O₂ concentration. The levels of O₂ also vary a lot in the intestine, whereby HCPs may play a role to protect the parasite from the oxygen stress.

5. Concluding Remarks

The High Cysteine Protein family is a novel protein group in *Giardia intestinalis* which is very poorly studied. Due to the similarities between HCP and VSP protein family there were three main objectives in this project. The first one was to confirm that localization changes in these HCPs were really occurring *in vitro*. If so, and to avoid extra factors associated to episomal protein expression, the second objective is focused on the expression and purification of recombinant HCPs for the production of specific antibodies to evaluate the cellular localization of 91707. Finally, verify the presence of one or more regulatory elements of transcription by miRNAs in some HCPs sequences to evaluate if this protein group could be regulated in the same way as VSPs are.

Along the different parts of the project there were some limitations which delayed its progression. One of the limitations of cellular localization of 91707-HA by immunofluorescence assay was the episomal protein expression, which may lead to a mislocalization of the tagged protein. Besides this, the use of antibodies with a confirmed target and localization in *Giardia* should have been used, since protein co-localization usually results in a more accurate cellular localization. One limitation of the expression of 91707-His in *Escherichia coli* was the production of the recombinant protein on its insoluble form. This is a very common problem associated to the expression of recombinant proteins in this host. Additionally, the purification of 91707-His by IMAC resulted in a low yield of pure recovered protein, given that there was a very dubious detection of the recombinant protein in Western Blots. Regarding the determination of target miRNA regulation sequences, many of the miRNAs tested by SLRT-qPCR were not detected, whereby only mir6 was tested and indicated a potential regulation of the HCP family.

This research project was very broad and complete regarding both concepts and techniques and it lead to new ideas for the characterization of this protein family. It is important to remember that this study is part of a bigger project and there are several other studies to perform in order to follow up the notions revealed here.

6. Future Studies

Since High Cysteine Proteins in *Giardia intestinalis* are a poorly studied protein family, there are several studies that can be made in order to achieve more insights about this particular protein group. Starting from the results obtained in this project, the production and usage of the specific antibody against 91707 is the most immediate future step. The application of this antibody would be extremely important to evaluate the protein synthesis and transport in the cell in non-transfected cell lines. Additionally, a putative secretion of 91707 could be detected both in normal growth conditions and during interaction with host cells. Besides this, the trophozoites could be subjected to stress conditions, such as treatment with H₂O₂, to produce cellular redox stress and, thus, evaluate changes at the protein level both by Western Blots and Immunofluorescence assays. Further on, the purified recombinant protein could also be used to assess its binding *in vitro* to intestinal epithelial cells, if this is what justifies the secretion of these proteins during parasite-host interactions.

Regarding the putative miRNA regulation of HCPs, more work is needed to evaluate if these groups of proteins or, at least, some members, are actually targets of these miRNAs in the same way VSPs are. The overexpression of particular miRNAs *in vitro* such as mir6 and observation of the transcript levels of HCPs is one step to follow up this regulation; however, it will still not completely proven the regulation at this level. Further on, regulation at the chromatin level should be explored as well.

7. References

- Adam, RD (2001), Biology of *Giardia lamblia*, 14(3):447–475.
- Ankarklev J, Jerlström-Hultqvist J, Ringqvist E, Troell K & Svärd SG (2010), Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nature Reviews Microbiology*, 8:413–422.
- Bartel DP (2009), MicroRNAs: target recognition and regulatory functions, *Cell*, 136(2):215–33.
- Cai Y, Yu X, Hu S & Yu J (2009), A brief review on the mechanisms of miRNA regulation, *Genomics. Proteomics Bioinformatics*, 7, 147–54.
- Carranza, PG and Lujan, HD (2010), New insights regarding the biology of *Giardia lamblia*, *Microbes and Infection*, 12(1):71–80.
- Castillo-Romero A, Davids BJ, Lauwaet T, Gillin FD (2012), Importance of enolase in *Giardia lamblia* differentiation, *Molecular and Biochemical Parasitology*, 184(2):122–125
- Farthing, Michael JG (1997), The Molecular Pathogenesis of Giardiasis, *Journal of Pediatric Gastroenterology & Nutrition*, 24(1):79–88.
- Chen C, Ridzon D, Broomer A, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ & Guegler KJ (2005), Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33, e179.
- Davids BJ, Reiner DS, Birkeland SR, Preheim SP, Cipriano MJ, McArthur AG & GillinFD (2006), A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein, *PLoS One*, 1(1):e44.
- He L & Hannon GJ (2004), MicroRNAs: small RNAs with a big role in gene regulation, *Nature Reviews, Genetics*, 5, 522–31.
- Jerlström-Hultqvist J, Stadelmann B, Birkestedt S, Hellman U & Svärd SG (2012), Plasmid vectors for proteomic analyses in *Giardia*: purification of virulence factors and analysis of the proteasome, *Eukaryotic Cell*, 11, 864–73.
- Keister DB (1983), Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 77:487–488.
- Khvorova A, Reynolds A & Jayasena SD (2003), Functional siRNAs and miRNAs exhibit strand bias, *Cell*, 115, 209–16.
- Kim VN (2004), MicroRNA precursors in motion: exportin-5 mediates their nuclear export. *Trends in Cell Biology*, 14, 156–9.

- Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT & Mourelatos Z (2007), An mRNA m7G cap binding-like motif within human Ago2 represses translation, *Cell*, 129, 1141–51.
- Krol J, Loedige I & Filipowicz W (2010), The widespread regulation of microRNA biogenesis, function and decay, *Nature. Reviews, Genetics*, 11, 597–610.
- MacRae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD & Doudna J (2006), Structural basis for double-stranded RNA processing by Dicer, *Science*, 311, 195–8.
- Lauwaet T, Davids BJ, Reiner DS & Gillin FD (2007), Encystation of *Giardia lamblia*: a model for other parasites, *Current Opinion in Microbiology*, 10(6):554-559.
- Lee RC, Feinbaum RL & Ambros V (1993), The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*, *Cell*, 75, 843–54.
- Li W, Saraiya A & Wang CC (2012), The profile of snoRNA-derived microRNAs that regulate expression of variant surface proteins in *Giardia lamblia*. *Cell, Microbiology*. **14**, 1455–73.
- Loughran, ST & Walls D (2011), Protein Chromatography, 681, 311–335.
- Ma'ayeh SY & Brook-Carter PT (2012), Representational difference analysis identifies specific genes in the interaction of *Giardia duodenalis* with the murine intestinal epithelial cell line, IEC-6, *International Journal for Parasitology*, 42: 501–9.
- Palm JE, Weiland ME, Griffiths WJ, Ljungström I & Svärd SG (2003), Identification of immunoreactive proteins during acute human giardiasis, *The Journal of Infectious Diseases*, 187(12):1849-1859.
- Prucca CG, Slavin I, Quiroga R, Elías EV, Rivero FD, Saura A, Carranza PG, & Luján HD (2008), Antigenic variation in *Giardia lamblia* is regulated by RNA interference, *Nature*, 456, 750–4.
- Prucca CG, Rivero FD & Luján HD (2011), Regulation of antigenic variation in *Giardia lamblia*, *Annual Review of Microbiology*, 65:611-30.
- Ringqvist E, Palm JE, Skarin H, Hehl AB, Weiland M, Davids BJ, Reiner DS, Griffiths WJ, Eckmann L, Gillin FD & Svärd SG (2008), Release of metabolic enzymes by *Giardia* in response to interaction with intestinal epithelial cells, *Molecular & Biochemical Parasitology*, 159(2):85-91.
- Ringqvist E, Avesson L, Söderbom F & Svärd SG (2011), Transcriptional changes in *Giardia* during host-parasite interactions, *International Journal for Parasitology*, 41(3-4):277-285.
- Roxström-Lindquist K, Ringqvist E, Palm D & Svärd SG (2005), *Giardia lamblia*-induced changes in gene expression in differentiated Caco-2 human intestinal epithelial cells, *Infection & Immunity*, 73(12):8204-8208.

- Sambuy Y, De Angelis I, Ranaldi G, Scarino ML, Stammati A & Zucco F (2005), The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics, *Cell Biology and Toxicology*, 21(1):1-26.
- Saraiya A & Wang CC (2008), snoRNA, a novel precursor of microRNA in *Giardia lamblia*, *PLoS Pathogens*, 4, e1000224.
- Saraiya A, Li W & Wang CC (2014), Transition of a microRNA from repressing to activating translation depending on the extent of base pairing with the target, *PLoS One*, 8, e55672.
- Saraiya A, Li W, Wu J, Chang CH & Wang CC (2014), The microRNAs in an ancient protist repress the variant-specific surface protein expression by targeting the entire coding sequence, *PLoS Pathogens*, 10, e1003791.
- Shuman S (1991), Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific, *Proceedings of the National Academy of Sciences of the United States of America*, 88, 10104–8.
- Shuman S (1994), Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase, *The Journal of Biological Chemistry*, 269, 32678–84 (1994).
- Sonda S, Morf L, Bottova I, Baetschmann H, Rehrauer H, Caflisch A, Hakimi MA & Hehl AB (2010), Epigenetic mechanisms regulate stage differentiation in the minimized protozoan *Giardia lamblia*, *Molecular Microbiology*, 76(1):48–67.
- Structural Genomics Consortium *et al.* (2008), Protein production and purification, *Nature Methods*, 5, 135–146.
- Touz MC, Rópolo AS, Rivero MR, Vranich CV, Conrad JT, Svard SG & Nash TE (2008), Arginine deiminase has multiple regulatory roles in the biology of *Giardia lamblia*, *Journal of Cell Science*, 121(Pt 17):2930-2938.
- Troeger H, Epple HJ, Schneider T, Wahnschaffe U, Ullrich R, Burchard GD, Jelinek T, Zeitz M, Fromm M & Schulzke JD (2007), Effect of chronic *Giardia lamblia* infection on epithelial transport and barrier function in human duodenum, *Gut*, 56(3):328–35.
- Yilmaz A & Grotewold E (2010), Components and mechanisms of regulation of gene expression, *Methods in Molecular Biology*, 674:23-32.

8. Appendix

8.1 Primers designed and tested along the project

Table 6. Designed primers for *91707* and *7715* PCR amplification and analyzed features. FW – Forward; RV – Reverse.

Gene	Primer	Number of nucleotides	GC Content	Melting Temp.
7715	RV 5'- GATAATCATTGTCGTAACCGGCACCC -3'	26	50.0 %	59.6 °C
	FW 5'- ATGCCCTGCGCAACTCGCCCTAT -3'	24	62.5 %	67.2 °C
91707	RV 5'- GAGAATGTACCCCTCCATG -3'	19	52.6 %	52.1°C
	FW 5'- ATGGATAGTTCGACTCCGTATGTAAAC -3'	27	40.7 %	56.1°C

Table 7. Designed primers for the miRNAs tested with the HCP sequences.

Transcript	Primer	Number of nucleotides
mir2	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-GCACTG-3'	40
	FW 5'-GAGTCGGGCAGGA-3'	13
mir5	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-CGGAAAG-3'	41
	FW 5'-GAGTCGGAGCAGA-3'	13
mir6	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-CAGAAT-3'	40
	FW 5'-GAGTCGGGACGCG-3'	13
mir6-C	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-TCGTCA-3'	40
	FW 5'-GAGTAGGTGATGA-3'	13
mir10	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-GGTCAG-3'	40
	FW 5'-GAGTCGGTGAGGA-3'	13
mir22	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-TGTACA-3'	40
	FW 5'-GAGTCGGTGTAAAC-3'	13
mir41	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-AGAGGC-3'	40
	FW 5'-GAGTCAGCCCGTT-3'	13
GlsR15	SLRT 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-GGAGTC-3'	40
	FW 5'-GAGTCGGGATTCA-3'	13
Universal Primer	RV 5'-GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCAC-3'	34
	SLRT 5'-GTGCAGGGTCCGAGGT-3'	16

8.2 Example of HCP Grouping based on sequence homology using *ClustalOmega* software

```

GL50803_113531 -----CATCTCATGCCGGTTCTAGTG--TCCAGCT---CTAGTTCGCCGCCATCCATG
GL50803_114626 -----CATCTCATGCCGGTTCTAGTG--TCCAAC---CTAGTTCGCCGCCATCCATG
GL50803_137727 -----TGTCTGGTGCCAACCTCCAGCA--TCCAAC---TTGATACCGTTATTGTCTGT
GL50803_113836 -----CGTCCAACGTCAACTCCAGCA--TCCAGCT---CTAGCCCCTGCCCGCAAGG
GL50803_114161 -----TGTCTGGTGCCAACCTCCAGCA--TCCAGCT---CTAGCCCCTGCCCGCAAGG

GL50803_113531 -----G-CGCCCCTACTGAGCCG-GT-----
GL50803_114626 -----G-CGCCCCTACTGAGCCG-GT-----
GL50803_137727 -----G-TGCCCCTGCTGAGTCA-GC-----
GL50803_113836 -----G-TGCCCCTGCTGAGCCA-GG-----
GL50803_114161 -----G-TGCCCCTGCTGAGTCA-GC-----

GL50803_113531 CAAGATA-----TTGCAGCATGCACG-AGTCGGTCTCCAA-AACT---GTAATCTG
GL50803_114626 CAAGATA-----TTACAGCATGCACG-AGTTGGCCCTCCAA-AACT---GTAATCTG
GL50803_137727 GCGGAGC-----ATTGGTTTTGGTGA-TGC-AGTGGCGTAT-AAAT---TTA--TGGG
GL50803_113836 TGTCGGT-----ATTGGTAAGTCGA-----CGGCGCGG-ATCT---GCA--TACG
GL50803_114161 TGTCGGT-----ATTGGTAAGTCGA-----CGGCGCGG-ATCT---GCA--TACG

GL50803_113531 GG-TATAAGACTCATTCTGA-CTCGAGTAGAGGA-ATCA-----T-----
GL50803_114626 GG-TATAAGACTCACTCAG-CTCGAGTAGAGGA-ATCA-----T-----
GL50803_137727 TT-CGGAATTGAAATTG-----AGGGTATGG-AAACC---ACCTATCA-----
GL50803_113836 AA-TGTAACATGATTTAC-TTCGAGGTTAGTT-TGTCA---ACTGCTGTC-----
GL50803_114161 AA-TGTAATTGAAACGTA-AGAGACGGGCCGG-TTAGT---GCTTAGAAC-----

GL50803_113531 -----
GL50803_114626 -----
GL50803_137727 -----
GL50803_113836 -----
GL50803_114161 -----

```

8.3 Putative miRNAs in *Giardia intestinalis* (Zhang *et al.*, 2009)

Table 8. Putative miRNAs (Zhang *et al.*, 2009) and mir6 and mir10 (Li *et al.*, 2012) from *Giardia intestinalis* matched with HCP sequences.

miRNA	Sequence 5'-3'	Seed Region (2-7 nt)	DNA seq.	DNA seq. Reverse Complement
<i>Gl-mir-1</i>	<i>cucaucuauacacagugucgc</i>	<i>ucaucu</i>	<i>tcatct</i>	<i>agatga</i>
<i>Gl-mir-2</i>	<i>gcaggaguugcagcgcagugc</i>	<i>caggag</i>	<i>caggag</i>	<i>ctctg</i>
<i>Gl-mir-3</i>	<i>cucuaccuccugacugcauug</i>	<i>ucuacc</i>	<i>tctacc</i>	<i>ggtaga</i>
<i>Gl-mir-4</i>	<i>aucgcacaccuagccuugug</i>	<i>ucgcac</i>	<i>tcgcac</i>	<i>gtgcga</i>
<i>Gl-mir-5</i>	<i>agcagagaagugucuuccg</i>	<i>gcagag</i>	<i>gcagag</i>	<i>ctctgc</i>
<i>Gl-mir-6</i>	<i>aggucggacucgcgccucc</i>	<i>ggucgg</i>	<i>ggtcgg</i>	<i>ccgacc</i>
<i>Gl-mir-7</i>	<i>ggaccaggccgucgucgaagu</i>	<i>gaccag</i>	<i>gaccag</i>	<i>ctgtgc</i>
<i>Gl-mir-8</i>	<i>uagcacggcgaucugugucug</i>	<i>agcacg</i>	<i>agcacg</i>	<i>ctgtct</i>
<i>Gl-mir-9</i>	<i>aggacguagauagacaucu</i>	<i>ggacgu</i>	<i>ggacgt</i>	<i>acgtcc</i>
<i>Gl-mir-10</i>	<i>gaggucuucuuccaaggcaau</i>	<i>aggucu</i>	<i>aggctt</i>	<i>agacct</i>
<i>Gl-mir-11</i>	<i>cuauaagaagagauaucuugu</i>	<i>uauaag</i>	<i>tataag</i>	<i>cttata</i>
<i>Gl-mir-12</i>	<i>uagcuuauagguaguugcaccu</i>	<i>agcuua</i>	<i>agctta</i>	<i>taagct</i>
<i>Gl-mir-13</i>	<i>cgugcgcacccccuugguau</i>	<i>gugcgc</i>	<i>gtgcgc</i>	<i>gcgcac</i>
<i>Gl-mir-14</i>	<i>gacgaagaucgccccacga</i>	<i>acgaag</i>	<i>acgaag</i>	<i>cttcgt</i>
<i>Gl-mir-15</i>	<i>cauccauguagacaucuugcg</i>	<i>auccau</i>	<i>atccat</i>	<i>atgfat</i>
<i>Gl-mir-16</i>	<i>ugggagcgugucacaggcag</i>	<i>gggagc</i>	<i>gggagc</i>	<i>gctccc</i>
<i>Gl-mir-17</i>	<i>agugcaggaccacgcgguu</i>	<i>gugcga</i>	<i>gtgcga</i>	<i>tcgcac</i>
<i>Gl-mir-18</i>	<i>ccccucugggacucgugagu</i>	<i>cccucu</i>	<i>ccctct</i>	<i>agaggg</i>
<i>Gl-mir-19</i>	<i>guccgauuaccaaccuuuga</i>	<i>uccgau</i>	<i>tccgat</i>	<i>atcgga</i>
<i>Gl-mir-20</i>	<i>ugagcagauucacagguaua</i>	<i>gagcag</i>	<i>gagcag</i>	<i>ctgtct</i>
<i>Gl-mir-21</i>	<i>ugaguacuguuuugcaccuu</i>	<i>gaguac</i>	<i>gagtac</i>	<i>gtactc</i>
<i>Gl-mir-22</i>	<i>uguacagucacauacauuaca</i>	<i>guuaca</i>	<i>gtaaca</i>	<i>gtttac</i>
<i>Gl-mir-23</i>	<i>cgugcauggaauaguguguau</i>	<i>gugcau</i>	<i>gtgcat</i>	<i>atgcac</i>
<i>Gl-mir-24</i>	<i>ugucuuuugaagacuuugug</i>	<i>gucuuu</i>	<i>gtcttt</i>	<i>aaagac</i>
<i>Gl-mir-25</i>	<i>agaucuggaucgccuucagcu</i>	<i>gaucug</i>	<i>gatctg</i>	<i>cagatc</i>
<i>Gl-mir-26</i>	<i>cgugcggcccgggggcagg</i>	<i>gugcgg</i>	<i>gtgcgg</i>	<i>ccgcac</i>
<i>Gl-mir-27</i>	<i>uuggcuacugcagucagcucu</i>	<i>uggcua</i>	<i>tggcta</i>	<i>tagcca</i>
<i>Gl-mir-28</i>	<i>ugagaucuuagagccguuug</i>	<i>gagauc</i>	<i>gagatc</i>	<i>gatctc</i>
<i>Gl-mir-29</i>	<i>cccucaaaucuccugauuauug</i>	<i>ccuaca</i>	<i>ccttca</i>	<i>tgaagg</i>
<i>Gl-mir-30</i>	<i>cagcuguguguccgcccag</i>	<i>agcug</i>	<i>agcgtg</i>	<i>cacgct</i>
<i>Gl-mir-31</i>	<i>cgaggagacguuggagaauug</i>	<i>gaggag</i>	<i>gaggag</i>	<i>ctcttc</i>
<i>Gl-mir-32</i>	<i>guagcaagaauccuucugcu</i>	<i>uagcaa</i>	<i>tagcaa</i>	<i>ttgcta</i>
<i>Gl-mir-33</i>	<i>aucgaagcacucuggagcu</i>	<i>ucgaaa</i>	<i>tcgaaa</i>	<i>tttcga</i>
<i>Gl-mir-34</i>	<i>cgccgucacggagucgccc</i>	<i>gccguc</i>	<i>gccgtc</i>	<i>gacggc</i>
<i>Gl-mir-35</i>	<i>aguguaucaaaaccuggcac</i>	<i>guguau</i>	<i>gtgtat</i>	<i>atacac</i>
<i>Gl-mir-36</i>	<i>ugccagguuuuagauacacua</i>	<i>gccagg</i>	<i>gccagg</i>	<i>ctctgc</i>
<i>Gl-mir-37</i>	<i>auccauguugagcgcugguu</i>	<i>uccaug</i>	<i>tccatg</i>	<i>catgga</i>
<i>Gl-mir-38</i>	<i>ucccugaagguacuuggcuua</i>	<i>cccuga</i>	<i>ccctga</i>	<i>tcaggg</i>
<i>Gl-mir-39</i>	<i>cccagagugcugucucacagc</i>	<i>ccagag</i>	<i>ccagag</i>	<i>ctctgg</i>
<i>Gl-mir-40</i>	<i>uccgguagguuuuuuuuuuu</i>	<i>ccgguu</i>	<i>ccggtu</i>	<i>taccgg</i>
<i>Gl-mir-41</i>	<i>cccguuuuuggggucucucu</i>	<i>ccguuu</i>	<i>ccgttt</i>	<i>aaacgg</i>
<i>Gl-mir-42</i>	<i>ccagucuaacauagacacagu</i>	<i>caguca</i>	<i>cagtca</i>	<i>tgactg</i>
<i>Gl-mir-43</i>	<i>gaugggugguuauagagccug</i>	<i>augggu</i>	<i>atgggt</i>	<i>accctat</i>
<i>Gl-mir-44</i>	<i>aagcugcuuuuugugaaguc</i>	<i>agcugc</i>	<i>agctgc</i>	<i>gcagct</i>
<i>Gl-mir-45</i>	<i>uggauucagggccguguuuu</i>	<i>ggauuc</i>	<i>ggattc</i>	<i>gaatcc</i>
<i>Gl-mir-46</i>	<i>gagcagugcugccuuugaua</i>	<i>agcagu</i>	<i>agcagt</i>	<i>actgct</i>
<i>Gl-mir-47</i>	<i>uucgaaauaggagccaagagu</i>	<i>ucgaaa</i>	<i>tcgaaa</i>	<i>tttcga</i>
<i>Gl-mir-48</i>	<i>ggcgucacagccagcagcucu</i>	<i>gcguc</i>	<i>gcgtcc</i>	<i>ggacgc</i>
<i>Gl-mir-49</i>	<i>aggugcauggacggcgccgc</i>	<i>ggugca</i>	<i>gggtca</i>	<i>tgacc</i>
<i>Gl-mir-50</i>	<i>cacgagagucgacgaccggg</i>	<i>acgaga</i>	<i>acgaga</i>	<i>tctctg</i>
<i>mir6*</i>	<i>gacgcgtgacgaagttgtctattctg</i>	<i>acgcgt</i>	<i>acgcgt</i>	<i>acgcgt</i>
<i>mir10*</i>	<i>tgaggaagaaccgcttctgtctgacc</i>	<i>gaggaa</i>	<i>gaggaa</i>	<i>ttcttc</i>

*mir6** and *mir10** are tested and confirmed miRNAs (Li *et al.*, 2012).

8.4 Composition of Reagents and Solutions used in the project

Reagent/Solution	Composition	Use
Tris-acetate buffer (TAE)	40mM Tris, 20mM Acetic Acid, 1mM EDTA.	Agarose Gels
TFBI	30mM $\text{KC}_2\text{H}_3\text{O}_2$, 100mM RbCl, 10mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 50mM $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 15% glycerol – pH 5.8.	Competent Cells
TFBII	10mM MOPS, 10mM RbCl, 75mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 15% glycerol – pH 6.6.	Competent Cells
TYDK	30 mg/ml Peptone, 55.6mM Glucose, 34.2mM NaCl, 1.14mM L-ascorbic acid, 5.74mM K_2HPO_4 , 4.41mM KH_2PO_4 , 11.4mM L-cysteine hydrochloride monohydrate, 0.038mM ferric ammonium citrate – pH 6.8	<i>Giardia intestinalis</i> trophozoites media
HBS-glucose buffer	130mM NaCl, 5mM KCl, 10mM glucose, 1mM MgCl_2 – pH 7.4	IF Assay
Running Buffer	25 mM Tris, 192 mM glycine, 0.1% SDS	SDS_PAGE
Transfer Buffer	25 mM Tris, 192 mM glycine, 20% methanol	Westren Blot
Resuspension/Binding Buffer A	HEPES 100mM pH 8; DDT 2mM; NaCl 200mM; Triton 1%; Glycerol 20%; 10mM Imidazol; Proteinase Inhibitors.	
Wash Buffer A	DDT 2mM; NaCl 200mM; HEPES 100mM pH 8; Triton 1%; Glycerol 20%.	
Elution Buffer A	DDT 2mM; NaCl 200mM; HEPES 100mMpH 8; Triton 1%; Glycerol 20%, 500mM Imidazole.	
Resuspension/Binding Buffer B	HEPES 100mM pH8; NaCl 200mM; DTT 1mM; Glycerol 10%; Imidazole 5mM; DNaseI-RNaseA 20ug/ml each.	Protein Purification by IMAC
Wash Buffer 1 B	HEPES pH8 100mM; NaCl 300mM; DTT 1mM; Glycerol 10%; Imidazole 5mM.	
Wash Buffer 2 B	HEPES pH8 100mM; NaCl 500mM; DTT 1mM; Glycerol 10%; Imidazole 10mM.	
Elution Buffer B	HEPES pH8 100mM; NaCl 500mM; DTT 1mM; Glycerol 10%; Imidazole 500mM.	
Desalting Buffer	HEPES pH8 100mM; NaCl 500mM; DTT 1mM; Glycerol 10%.	

