LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Investigation of the oxidative folding pathway in *Trypanosoma brucei*

Thesis submitted by Michael I. Millerick in fulfilment of a Ph.D

Supervisor Dr Derek P. Nolan.

School of Biochemistry and Immunology

Trinity College Dublin

2013

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Mich

LLEGE
2013
UBLIN
Thesis 10148

Acknowledgements

This project could not have come about without the help and support provided by the exceptional people I had the pleasure to work with.

First and foremost, I would like to express my very great appreciation to Dr. Derek Nolan for his guidance and support throughout this project. The advice and wisdom he imparted was key to my success, both inside and outside the lab, and I am forever in his debt.

For tireless support in the lab I would like to thank Dr. Nóirín Nic a' Bháird, Dr. Joyce Rubotham and Dr. Diana Spitznagel who helped me master the techniques that we central to the completion of this project.

For moral support and advice I would like to thank Jennie McFarland, Orla O'Hanrahan, Connor Breen, Andy Knox, Rob O'Byrne who were always there to talk about the finer points of trypanosomes.

I would also like to give special thanks to the academic staff, especially Dr. Vincent Kelly for help with molecular cloning, Dr. Tim Mantle for protein purification advice and Dr. H. Paul Voorheis for additional trypanosome advice.

I would also like to thank the School of Biochemistry and Immunology for the smooth running of the department and for providing an excellent work environment.

And lastly I would like to thank the Welcome Trust for their support, without whom none of this would be possible.

Dedications

I would like to dedicate this thesis to my family who supported me whole heartedly in this endeavor and who have always been there for me in whatever I do,

And I would also like to dedicate this to Dr. Sinead Healy-Evans who gave me unconditional support and help at every twist and turn of the process.

Abstract

In eukaryotic organisms the native structure of proteins destined to be secreted, such as immunoglobulins, or mounted surface proteins rely on the formation of disulphide bridges between cysteine residues for structural integrity. These bonds are thought to be formed in the endoplasmic reticulum via a pathway commonly referred to as the oxidative folding pathway. Many studies to date have been carried out to investigate this pathway *in vitro* although due to the complexity of the oxidative folding pathway in higher organisms, *in vivo* investigations into this pathway are still lacking conclusive evidence demonstrating aspects of the function of the oxidative folding pathway. This thesis presents the first characterization of the oxidative folding pathway of *Trypanosoma brucei*, a parasitic protozoan whose oxidative folding machinery appears to be substantially less complex than that of other eukaryotes.

It was found that *T. brucei* possesses six PDIs (TbPDI-1 to 6) and a single Ero 1 (TbEro 1). Additionally it was demonstrated that the transcript levels of two of these putative enzymes were expressed three fold in the bloodstream stage of the life cycle of *T. brucei*. During this lifecycle stage the highly immunogenic Variable Surface Glycoprotein (VSG) is expressed, which represents 99% of the surface protein. This protein allows African trypanosomes to evade the host humoral response by the periodic switching of the surface coat from one unique VSG to an entirely different VSG, allowing the parasite to survive extracellularly. As such the inhibition of VSG folding presents an ideal drug target. In this study it was demonstrated that both TbPDI-2 and TbEro 1 are essential for normal growth in bloodstream and procyclic form cells and both are required for the efficient folding of VSG in bloodstream form cells. The two proteins were also found to co-localize and interact. Additionally, it was also shown that TbEro 1 is a resident of the endoplasmic reticulum. These findings taken together indicate that TbPDI-2 and TbEro 1 form the major oxidative folding pathway in *T. brucei* and that VSG is the principal substrate of this pathway. Not only is this pathway essential in bloodstream form cells but also in the insect stage of the lifecycle, the procyclic form and presents a novel drug target for the treatment of trypanosomiasis.

Table of Contents

Declaration 4

Abstract 5

1 Introduction 15

1.1 Trypanosome Biology 15

1.1.1 Trypanosoma History, Taxonomy and Epidemiology 15

1.1.2 Life Cycle of *Trypanosoma brucei* 20

1.1.3 Cell cycle of *Trypanosoma brucei* 24

1.1.4 Genomic Structure of *Trypanosoma brucei* 26

1.1.5 Gene Expression 28

1.1.6 Variable Surface Glycoprotein 31

1.2 The Endoplasmic Reticulum 38

1.2.1 Form and Function 38

1.2.2 Translocation of proteins to the endoplasmic reticulum 42

1.2.3 Secretion and Retention in the Endoplasmic Reticulum 45

1.3 Oxidative Folding Pathway 46

1.3.1 Protein Disulphide Isomerase 46

1.3.2 Structure of PDI 50

1.3.3 Localization of PDI 51

1.3.4 Reaction mechanism of PDI 54

1.3.5 In vitro assays for PDI activity 56

1.3.6 Thioredoxin superfamily in *protozoan* parasites 58

1.3.7 PDIs of *Giardia lamblia* 59

1.3.8 PDIs of *Leishmania major* 60

1.3.9 PDIs of *Trypanosoma brucei* 61

1.3.10 Endoplasmic reticulum oxidoreductin 1 64

1.3.11 Sulfhydryl oxidase family 64
1.3.12 Catalytic Properties 65
1.3.13 Structure of Ero 1 69
1.4 Reverse genetics 71
1.4.1 RNA interference 71
1.4.2 Mechanism of RNA interference 72
1.4.3 RNAi in *Trypanosoma brucei* 73
1.4.4 Gene Knockout 76
2 Materials and Methods 79
2.1 Preparation of reagents 79
2.2 Agarose Gel Electrophoresis of DNA 79
2.3 Polymerase Chain Reaction 80
2.4 Purification of PCR products 80
2.5 T-vector ligation 80
2.6 Restriction enzyme digestion of DNA 80
2.7 Gel purification of DNA 81
2.8 Ligation of digested PCR product into digested vector 81
2.9 Preparation of competent *Escherichia coli* DH5α 81
2.10 Transformation of competent bacterial cells 82
2.11 Selection of transformed colonies. 82
2.12 Generation of glycerol stocks. 82
2.13 Purification of plasmid DNA. 82
2.14 Transfection of bloodstream form *Trypanosoma brucei* using the AMAXA Nucleofector® apparatus 83
2.15 Transfection of procyclic form *Trypanosoma brucei* using the AMAXA Nucleofector® apparatus 83
2.16 Cell counting 84
2.17 Cell culture of bloodstream form cells 85

2.18 Detergent lysis of cultured cells 86
2.19 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) 86
2.20 Semi-dry Western Blotting 86
2.21 Generation of Anti-TbEro 1 antibody 87
2.22 Immunoblotting 88
2.23 RNA interference. 91
2.24 Generation of RNAi constructs 91
2.25 Generation of trypanosome expression constructs 94
2.26 Generation of recombinant TbEro 1 expression construct 94
2.27 Generation of TbPDI-2 knockout constructs 95
2.28 Homologous recombination 106
2.29 Preparation of trypanosomes for storage in liquid nitrogen 106
2.30 RNA isolation 106
2.31 Quantitative Real Time Polymerase Chain Reaction. 107
2.32 Cell fixation. 108
2.33 Immunofluorescence and confocal microscopy 109
2.34 Analysis of the rate of surface protein production using pulse-chase 109
2.35 DNA Sequencing 110
2.36 Pulse chase labelling of surface proteins in Trypanosomes 110
2.37 Co-immunoprecipitation of labeled proteins 111
2.38 Digitonin subcellular fractionation 111
2.39 Digestion of TbEro 1 HA with glycosidases 112
2.40 Purification of bacterial recombinant TbEro1 112
3 Bioinformatics analysis of oxidative folding pathway of Trypanosoma brucei.. 114
3.1 Protein disulphide isomerases of *Trypanosoma brucei* 114
3.1.1 Identification of possible PDIs of *Trypanosoma brucei* 115
3.1.2 Primary sequence analysis of candidate PDIs 118
3.1.3 Phylogenetic analysis of candidate PDIs 120

3.1.4 Multiple sequence alignment of TbPDI-1 & closely related proteins 122

3.1.5 Multiple sequence alignment of TbPDI-2 & closely related proteins 126

3.1.6 Multiple sequence alignment of TbPDI-3 & closely related proteins 132

3.1.7 Multiple sequence alignment of TbPDI-4 & closely related proteins 136

3.1.8 Multiple sequence alignment of TbPDI-5 & closely related proteins 140

3.1.9 Multiple sequence alignment of TbPDI-6 & closely related proteins 144

3.2 Endoplasmic reticulum oxidoreductin of *Trypanosoma brucei* 148

3.2.1 Identification of possible Ero 1s of *Trypanosoma brucei* 149

3.2.2 Primary sequence analysis of candidate TbEro 1 149

3.2.3 Multiple sequence alignment of TbEro 1 & closely related proteins 151

3.3 Relative quantity of candidate PDIs in bloodstream versus procyclic cells 158

3.4 Discussion 160

3.4.1 TbPDI candidate proteins 160

3.4.2 Single domain PDIs of *Trypanosoma brucei* 161

3.4.3 TbPDI-3 (Tb927.7.1300) 163

3.4.4 TbPDI-2 (Tb927.10.8230) 164

3.4.5 TbEro 1 (Tb927.8.4890) 166

4 Analysis of TbPDI-2 *in vivo* 168

4.1 Generation of constructs 169

4.1.1 Generation of knockout constructs 169

4.1.2 Generation of ectopic expression construct 178

4.2 Generation and analysis of bloodstream form (TbPDI-2 +/-) cell lines 179

4.3 Attempted generation of bloodstream form (TbPDI-2 -/-) null mutant cell line 194

4.4 Attempted generation of bloodstream form (TbPDI-2 -/- TbPDI-2 p1791) cell line 200

4.5 Generation and analysis of bloodstream form (TbPDI-2 +/- TbPDI-2 RNAi) cell line 201

4.6 Generation and analysis of procyclic form TbPDI-2 knockout cell line 218

4.6.1 Generation and analysis of procyclic form TbPDI-2 single allele knockout cell lines 218

4.6.2 Generation and analysis of procyclic form (TbPDI-2 -/-) cell line 219

4.7 Generation and analysis of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line 230

4.8 Discussion 240

5 Analysis of TbEro 1 250

5.1 Generation and analysis of bloodstream form TbEro 1 knockdown cell line 251

5.2 Generation and analysis of a bloodstream form cell line that expresses an epitope tagged TbEro 1 252

5.3 Generation and analysis of procyclic form (TbEro 1 RNAi) cell line 290

5.4 Generation and analysis of a procyclic form cell line that expresses an epitope tagged TbEro 1 291

5.5 Generation and purification of recombinant TbEro 1 GST fusion protein ... 308

5.6 Discussion 316

6 Conclusions and Future Work 326

7 Appendix 334

7.1 Genomic sequences of enzymes of the oxidative folding pathway in *Trypanosoma brucei* and the multiple sequence alignment of them using T. Coffee 334

7.2 Analysis of hydrophobic regions of the candidate TbPDIs and TbEro 1 of *Trypanosoma brucei* 344

7.2.1 TbPDI-1 (Tb927.4.2450) 344

7.2.2 TbPDI-2 (Tb927.10.8230) 346

7.2.3 TbPDI-3 (Tb927.7.1300) 348

7.2.4 TbPDI-4 (Tb927.5.1020) 350

7.2.5 TbPDI-5 (Tb927.7.5790) 352

7.2.6 TbPDI-6 (Tb927.7.5780) 354
7.2.7 TbEro 1 (Tb927.8.4890) 356
7.3 Primer sequences 359
8 References 1

Abbreviations.

ATP	Adenosine Triphosphate
BiP	Immunoglobulin heavy chain binding protein
BLAST	Basic Local Alignment Search Tool
bp	Nucleotide base pare
BSA	Bovine Serum Albumin
BSF	Bloodstream form
CHAPS	3-[(3-cholamidopropyl)dimethylammonia]-1-propanesulfonate
DAPI	4',6-diamidino-2-phenylindole
dH_2O	Distilled, deionised water
DMSO	Dimethylsulphoxide
dsRNA	Double Stranded Ribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid
Ero 1	Endoplasmic reticulum oxidoreductin
FP	Forward primer
GPI	Glycophosphatidyl Inositol
GPI-PLC	Glycosylphosphatidylinositol-specific phospholipase C
GSH	Glutathione
GSSG	Oxidized glutathione
GST	Glutathione S-transferase
HA	haemagglutinin
IPTG	Isopropyl β-D-1-thiogalactopyranoside

Mb	Mega base
MCS	Multiple Cloning Site
MITat	Molteno Institute Trypanozoon antigen type
NBT/BCIP	5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrzolium
ORF	Open reading frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCF	Procyclic form
PCR	Polymerase Chain Reaction
PDI	Protein Disulphide Isomerase
PMSI	Phenylmethylsulfonylfluoride
Pol I-III	Polymerase I-III
RNAi	Ribonucleic Acid Interference
RNase A	Ribonuclease A
SDS	Sodium Dodecyl Sulphate
SNP	Single-nucleotide polymorphism
SRP	Signal recognition particle
(TbPDI-2 +/-)	TbPDI-2 single allele knockout blasticidin/puromycin resistant blast/puro)
TBS	Tris Buffered Saline
TEMED	(N,N,N,N',N')-tetra-methyl-ethylene-diamine
Tes	Transferrin
Tet	Tetracyclin
TL	Tomato Lectin

TLCK	N_a-Tosyl-L-lysine chloromethyl ketone
Tris	2-amino-2-hydroxymethylpropane-1,3-diol
TSB	Trypanosome Separation Buffer
Tween	Polyethylene glycol sorbitan monolaurate
Tx-like	Thioredoxin-like
VSG	Variant Surface Glycoprotein
VSP	Variable surface protein
w/v	Weight per volume

1 Introduction

1.1 Trypanosome Biology

1.1.1 Trypanosoma History, Taxonomy and Epidemiology

Trypanosomes have existed for approximately 300 million years, diverging at an early stage of the eukaryotic lineage and evolving with their hosts as microscopic unicellular protozoa. They are parasites of a wide range of hosts including insects, plants, birds, bats, fish, amphibians and mammals. As both host and parasite evolved together it is unlikely these parasites were originally pathogenic. It was only when a change of host occurred, that they became so, explaining the small amount of pathogenic Trypanosomes (Gibson, 2012).

The genus *Trypanosoma* was first described by one of the founders of medical microbiology, David Gruby, who identified *T. sanguinis*, a frog Trypanosome. Later Trypanosomes were linked to the disease Surra, a wasting disease of horses and camels, caused by *T. evansi*, but it was not until 1895 that the connection between host, parasite and vector was made. This finding was reported by a medical officer David Bruce, who said that the trypanosoma, now known as *T. brucei*, which caused Nagana in cattle, was being spread by the tsetse fly, the vector of the parasite. This discovery led to a new era of trypanosomiasis research.

Trypanosomes are generally associated with disease in Africa (Sleeping sickness in humans and Nagana in cattle) and South America (Chagas' disease). A classification of African trypanosomes is shown in Table 1.1. Of the pathogenic Trypanosomes, *T. brucei brucei* is considered to be the most virulent. It is infectious to most mammals but not humans or baboons, due to their susceptibility to lysis by human apolipoprotein L1 (Vanhamme et al., 2003), making *T. brucei brucei* a preferable species for study. This subspecies causes Nagana in cattle, a wasting disease which results in a reduction in cattle size and ability to work. Trypanosomes are distributed throughout the sub-Saharan region of Africa, and are spread by their insect vector the tsetse fly. The areas of infection are defined by the habitat of the tsetse fly (Figure 1.1). The trypanosome *T. b. brucei* has a very specific vector parasite relationship due to its complex life cycle.

Figure 1.1 Predicted Areas of Suitability for Tsetse in Africa

Areas of suitability for tsetse fly were predicted using data collected from satalite images showing certain ecological conditions, specifically vegetation cover, rainfall, temperature and elevation.

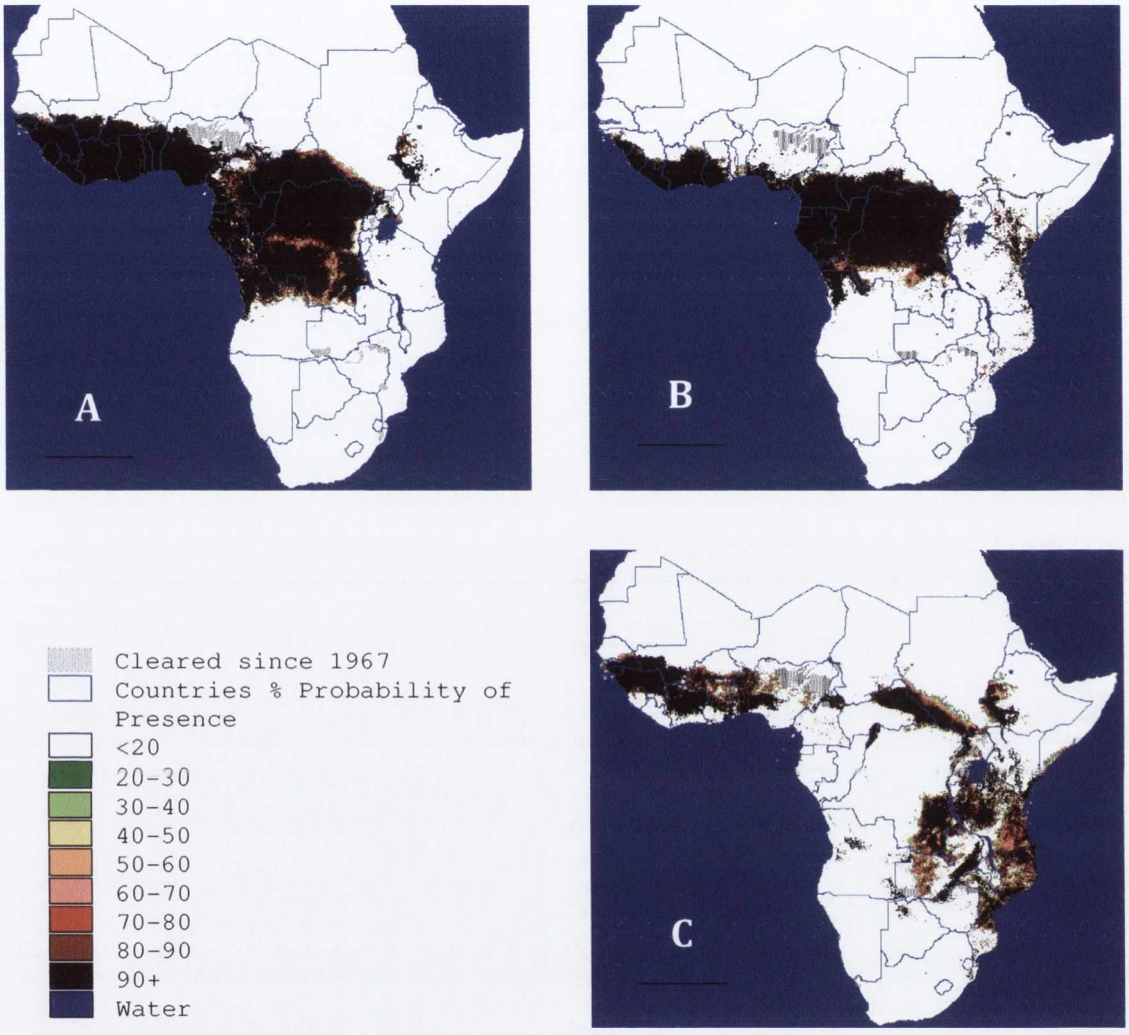
Panel A. Predicted areas of suitability for riverine tsetse: *Palpalis* group.

Panel B. Predicted areas of suitability for forest tsetse: *Fusca* group.

Panel C. Predicted areas of suitability for savanna tsetse: *Morsitans* group, which includes the most effective vector of sleeping sickness, *Glossina morsitans*.

Adapted from (Cecchi et al., 2008).

Figure 1.1 Predicted Areas of Suitability for Tsetse in Africa



Kingdom	Protista	Eukaryotes, generally unicellular.
Phylum	Protozoa	"Animal-like" Protists, nutrition mainly by ingestion.
Subphylum	Sarcomastigophora	
Class	Zoomastigophora	Use flagella for mobility and feeding. Mostly unicellular.
Order	Kinetoplastida	Flagellated protozoa. Kinetoplast present.
Suborder	Trypanosomatina	
Family	Trypanosomatidae	Uniflagellates, complex lifestyles.
Genus	Trypanosoma	Parasites of vertibrates with intermediate invertebrate hosts.
Section	Salivaria	Undergo development in the forgut of their vectors
Subgenus	Trypanozoon	Indistinguishable morphologically, but differ biologically.
Species	T. (t.)brucei	Intermediate vector is the tsetse fly.
Subspecies	T. (t.)brucei brucei	Causative agent of Nagana throughout tropical Africa.
	T. (t.)brucei gambiense	Causative agent of chronic sleeping sickness in West and central Africa.
	T. (t.)brucei rhodiense	Causative agent of acute sleeping sickness in East and South Africa.

Table 1.1 Classification of the African trypanosome group *T. brucei*

Classification of the African trypanosome group *T. brucei* including a description of each groups characteristics (Hoare, 1966).

1.1.2 Life Cycle of *Trypanosoma brucei*

Trypanosomes display different morphology throughout their varied and complex life cycle which occurs in their mammalian hosts and tsetse fly insect vector. These hosts provide two differing environments which trypanosomes must adapt to in order to grow, proliferate and complete the life cycle (Figure 1.2).

During a natural infection trypanosomes are introduced into the bloodstream of their mammalian host via insect bite. In the bloodstream of the mammalian host the trypanosome shows a wide and continuous variation, ranging between three different forms: long slender, intermediate and short stumpy form trypanosomes (Vickerman, 1962, 1970; Vickerman and Luckins, 1968). When in the bloodstream form trypanosomes the *trypomastigote* morphology is observed with the kinetoplast, a single mitochondrion like organelle, located at the posterior end of the parasite with the flagellum emerging from the posterior end of the cell and folding back along the body of the trypanosome. Bloodstream trypanosomes are covered with a variable surface glycoprotein (VSG) which allows the parasite to evade the host humoral immune response. While in the glucose rich environment of the mammalian blood, the majority of ATP is produced via glycolysis which occurs in the glycosomes, an organelle common to *kinetoplastida*(Opperdoes and Borst, 1977). With this readily available source of ATP there is less need for oxidative phosphorylation and as such the mitochondria of slender bloodstream form trypanosomes is acristate and does not engage in oxidative phosphorilation.

The long slender form trypanosomes undergo asexual replication via longitudinal binary fission and monomorphic cell lines, containing long slender morphology only, have been developed for the purposes of growth in tissue culture. These cell lines were employed throughout this study. It is important to note the difference between monomorphic cell lines used during this study and the pleomorphic (many cell morphologies) forms which occur during a natural infection and to consider this differentiation when drawing conclusions from monomorphic studies.

The short stumpy form trypanosomes do not divide and it is believed that they are preadapted to the environment of the tsetse fly (Matthews and Gull, 1997). The environment of the tsetse fly gut is not as rich in glucose and requires a greater use of amino acids as an energy source. As such, cristae form in the mitochondrion of the

short stumpy form in preparation. The VSG coat which is used to evade the host immune system is no longer needed in the tsetse fly and is replaced with the invariant surface protein procyclin (Overath et al., 1986; Roditi and Pearson, 1990; Roditi et al., 1989). The formation of short stumpy form trypanosomes is believed to be triggered by the increased density of trypanosomes in the bloodstream of the host mammal (Hamm et al., 1990; Vassella et al., 1997). Once the short stumpy form trypanosomes are taken up during a blood meal, the drop in temperature from 37 ^{0}C to 27 ^{0}C is believed to trigger transformation to the procyclic form trypanosomes (Matthews and Gull, 1997; Vickerman, 1985). This can be mimicked in culture.

Over a period of 10 to 15 days the procyclic form trypanosomes divide rapidly in the tsetse fly midgut. After this period some procyclic cells transform to the proventricular form trypanosomes which move to the fly's salivary gland (Van Den Abbeele et al., 1999). The proventricular cells then transform to epimastigotes which attach themselves to the epithelial cells of the salivary glands via their flagella (Vickerman 1985). Once there the cells mature to metacyclic form trypanosomes. Mature metacyclic form trypanosomes, like the short stumpy bloodstream form trypanosomes, are believed to be preadapted for a change in environment. The mature cells lose their procyclin coat and regain their VSG coat, their mitochondria lose their cristae once more and the cells detach from the epithelial cells of the tsetse fly so that they are free to be injected into a mammalian host during a blood meal (Vickerman, 1973) thus completing the life cycle of the trypanosome.

Figure 1.2 Schematic representation of developmental cycle in mammal and in the tsetse fly vector

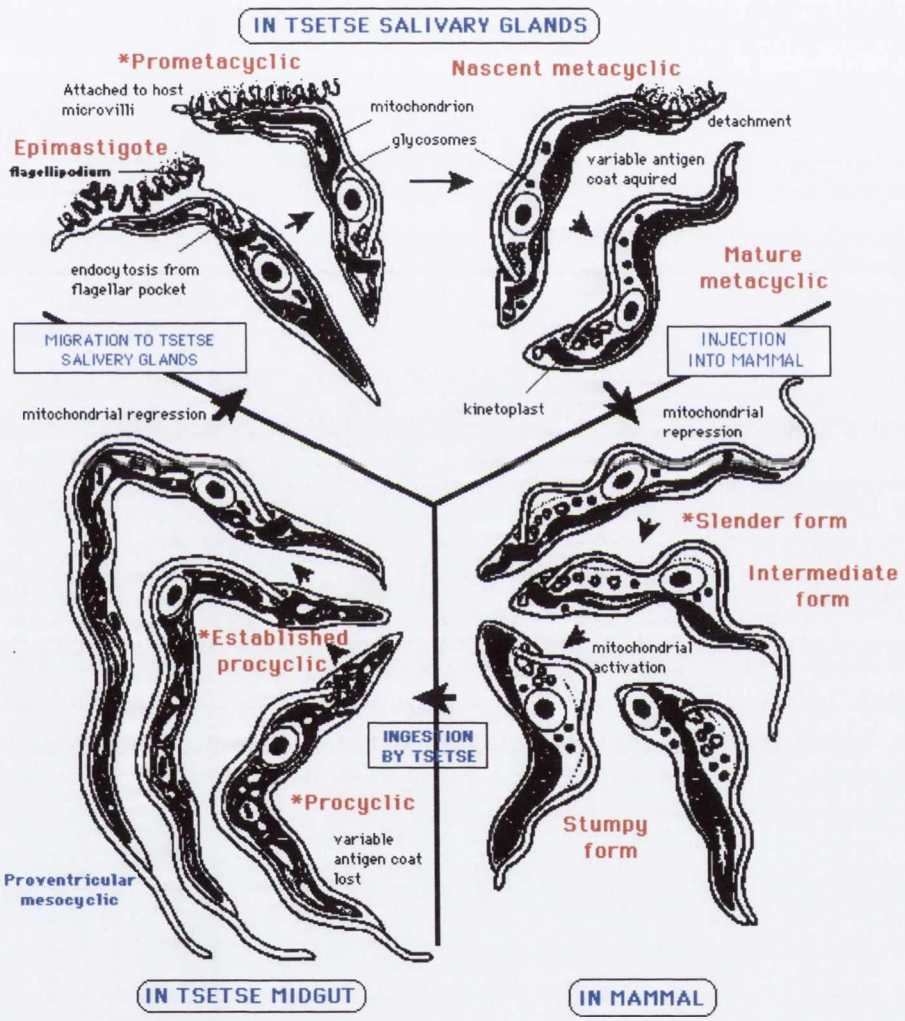
During a natural infection, mature metacyclic form trypanosomes enter the mammalian host while the tsetse fly takes a blood meal. In the bloodstream, metacyclic form trypanosomes transform into trypomastigotes varying in morphology from long slender, intermediate to short stumpy cells, all of which are protected from the host humoral response by the variant surface glycoprotein. Long slender cells divide rapidly by binary fission. Short stumpy form cells which arise from dividing long slender cells are preadapted to transform into procyclic cells once taken up during a blood meal.

Once a blood meal occurs short stumpy trypomastigotes are taken into the tsetse fly's midgut. They then transform to procyclic form cells, where VSG is lost and replaced with the invariant surface protein *procyclin*. Cells rapidly divide; some become proventricular mesocyclic form trypanosomes which migrate to the salivary glands of the tsetse.

In tsetse salivary glands, epimastigote cells go through various changes to form mature metacyclic form trypanosomes, which contain the VSG coating and other metabolic changes to infect new mammalian host during blood meal. This completes the lifecycle of the parasite.

Adapted from (Vickerman, 1985).

Figure 1.2 Schematic representation of developmental cycle in mammal and in the tsetse fly vector.



1.1.3 Cell cycle of *Trypanosoma brucei*

The eukaryotic cell cycle typically consists of four phases: G_0/G_1 phase (Gap0/Gap1), S phase (synthesis), G_2 phase (Gap2 or interphase) and M phase (mitosis).

In the G_1 phase, the cell is growing and preparing the replication of DNA by expressing proteins required for DNA replication. Once the cell is ready to replicate the genome of the cell, the cell enters S phase, where DNA is replicated. Once S phase is complete the cell enters the G_2 phase, where the biosynthesis process is increased in preparation for cell division. Following this phase is the M phase where DNA is partitioned into two daughter cells. In mammalian cells, M phase is tightly coupled with cytokinesis which begins before chromosome segregation is complete. Each step in the cell cycle requires the completion of the preceding step; cells which temporarily stop dividing are said to have entered the G_0 phase (Hammarton, 2007).

In *T. brucei*, the G_1 phase is characterized by the cells having a single nucleus, kinetoplast, flagellum and Golgi apparatus (Figure 1.3) (McKean, 2003). Once the cell enters G1 phase, the elongation and maturation of the pro-basal body to a basal body occurs. Nucleation of a new flagellum occurs followed by the replication of the Golgi body (Hammarton, 2007). During S phase kinetoplast duplication occurs as the cell then enters G2 phase. During G2 phase, segregation of the two kinetoplast occurs. The nucleus then undergoes mitosis once the cell enters M phase. M phase ends as the cell completes cytokinesis into two daughter cells.

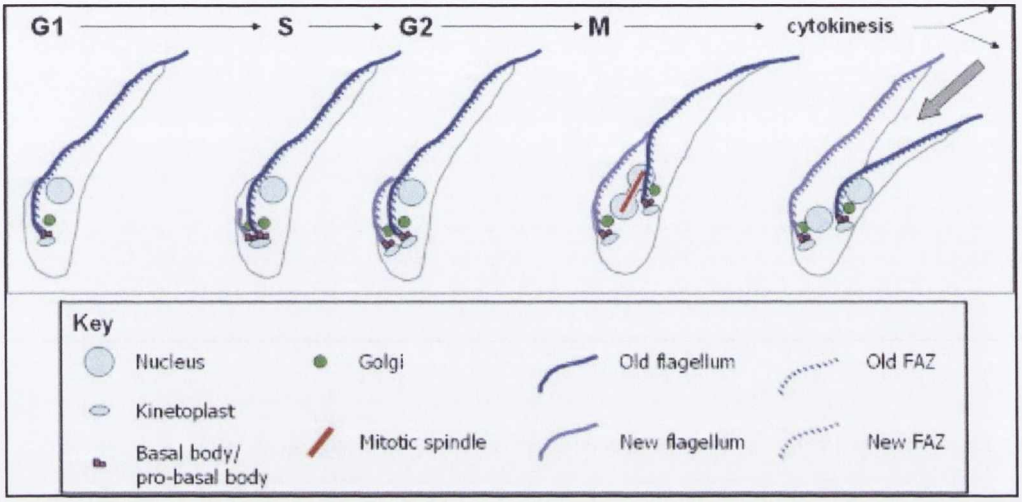


Figure 1.3 Cell cycle of *Trypanosoma brucei*

A schematic representation of the cell cycle phases and the morphological events which occur throughout is shown. Adapted from (Hammarton, 2007).

1.1.4 Genomic Structure of *Trypanosoma brucei*

The majority of eukaryotic genes are located in the nucleus of the cell and the remainder of their genes is located in the mitochondria, chloroplasts or in the case of trypanosomes, the kinetoplast. The genomes of most eukaryotes are orders of magnitude larger than those of bacteria and archaea (Table 1.2). Within the nucleus the DNA is organized into discrete units called chromosomes and each chromosome is made up of a material called chromatin, a complex of proteins and DNA. The number of chromosomes per nucleus is specific to each species. The mitochondria or chloroplast contain their own genome which encodes proteins specific to that organelle. This DNA is contained within the matrix of the organelle and is circular like that of the bacterial genome.

The genome of *T. brucei* consists of a 35 megabase genome, containing 11,412 predicted genes (data taken from Tritrypdb.org). Some genes have multiple copies which exist as tandem repeats or are located at different sites. Examples of these are the genes which encode variable surface glycoproteins (VSG) which are repeated to a great degree. The genome of *T. brucei* consists of mostly polycistronic transcription units which largely contain no introns; there have only been two genes found to be exceptions of this rule (Perelman and Boothroyd, 1990; Siegel et al., 2010). Groups of genes are orientated in the same direction and these genes are separated intervening sequences (Pays, 2005). The nuclear genome can be subdivided into three chromosome sizes; mega chromosomes, intermediate chromosomes and mini chromosomes. Megabase chromosomes are 1 Mb-6Mb in size and contain most of the genes transcribed in the cell including housekeeping genes which are transcribed processively by RNA polymerase II (Gottesdiener et al., 1990). Intermediate chromosomes are 200kb-700kb in size, and mini chromosomes, which are the most unique, are 20kb to 150kb. There are ~100 mini chromosomes which consist of linear DNA containing 177bp repeats, which are followed by silent telomere-linked VSG genes. It is believed that these silent VSG genes make up the library of genes used in antigenic variation, a mechanism through which the VSG is periodically switched to avoid the host humoral response (Weiden et al., 1991). The kinetoplast, which is found in the species of the order *Kinetoplastida*, is a single mitochondrion like organelle and contains ~30% of the total trypanosome DNA. It is organized into a network of maxi circles, 25-50 copies containing 20kb, and mini circles, 5,000-10,000

copies ~1kb in size. It has been shown that the kinetoplast is required for differentiation from bloodstream to procyclic forms (Stuart, 1983).

Table 1.2 Comparison of genome size between species

Species	Genome Size (1000s of nucleotide pairs per haploid genome)	Estimated number of genes coding for proteins
Bacteria		
Escherichia coli	4639	4289
Archaea		
Archaeoglobus fulgidus	2178	2493
Eukaryotes		
Saccharomyces cerevisiae (budding yeast)	12,069	~6300
Arabidopsis thaliana (Thale cress)	142,000	~26,000
Homo sapiens	~3,200,000	~24,000
Trypanosoma brucei	35,000	11,412 (source data: Sanger sequencing project)

Table 1.2 Comparison of genome size between species

For eukaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Adapted from Alberts Cell.

1.1.5 Gene Expression

Within a cell thousands of different protein and RNA molecules arise from the information contained in the DNA sequence of a cells genome. Typically only a fraction of these genes are expressed at any one time; for example the different cells observed in multicellular organisms are able to undertake very specific tasks due to the expression of certain sets of genes. Cells may also express different genes in response to changes in the surrounding environment. To manage the expression of specific genes there are many different levels at which expression may be controlled.

Figure 1.4 Five steps at which eukaryotic gene expression can be controlled

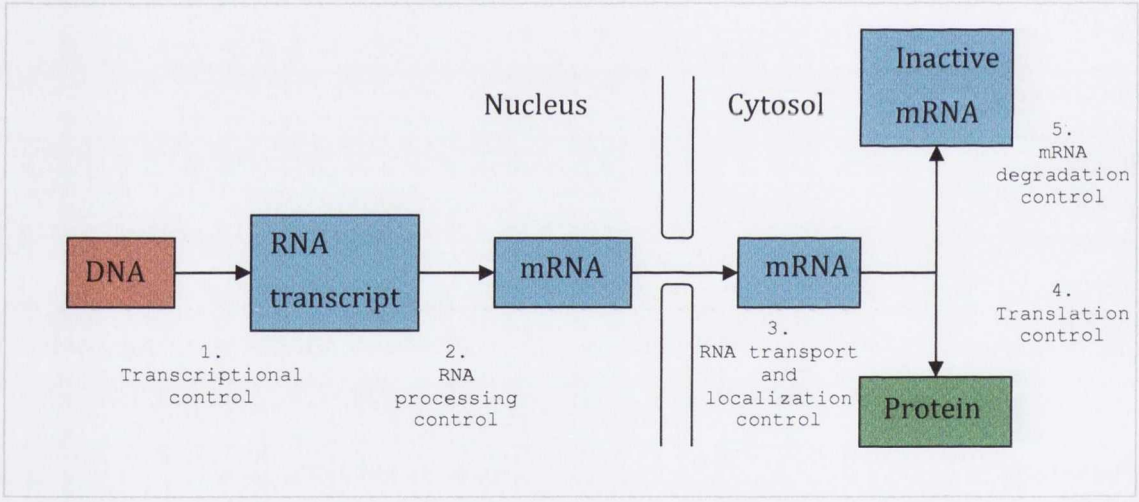


Figure 1.4 Five steps at which eukaryotic gene expression can be controlled

1. Transcriptional control is the control of when and how often a given gene is transcribed.
2. RNA processing control is the control of the splicing and processing of RNA transcripts.
3. Selecting which completed mRNAs are exported from the nucleus to the cytosol and their subsequent localization in the cytosol may be controlled.
4. Once localized in the cytoplasm the mRNAs may be translated by ribosomes which are influenced by translational control.
5. The mRNA may be selectively destabilized via mRNA degradation control.

Transcriptional control is the first level of control of gene expression and in most eukaryotes is the most significant. Within the DNA of a cell lie short stretches of specific sequences which are recognized by gene regulatory proteins. These proteins switch the transcription of individual genes on and off. In prokaryotes, gene regulatory proteins bind to DNA sequences close to the RNA polymerase start site and either activate or repress transcription of the gene. In eukaryotic organisms, expression is typically controlled by many gene regulatory proteins. These regulatory proteins may bind to sequences that can be thousands of nucleotide pairs away from the promoter that directs transcription of the gene. These activators and repressors may influence transcription by a wide variety of mechanisms. These include altering chromatin structure and controlling the assembly of the general transcription factors, mediator, and RNA polymerase at the promoter.

In trypanosomatids, regulation of gene expression through epigenetic mechanisms is significantly different to typical eukaryotes. It has been observed that there are fewer histone post translational modifications when compared to other eukaryotes (Janzen et al., 2006; Mandava et al., 2007). This finding suggests that the majority of control occurs at some other stage of expression.

Generally eukaryotes possess three distinct classes of nuclear RNA polymerases (Pol). Pol I is involved in the production of ribosomal RNAs, Pol III is responsible for small essential RNAs for example transfer RNAs and finally Pol II is responsible for the generation of messenger RNA from protein coding genes. Promoter regions for Pol II typically are usually 40 bases in length and contain functional sub regions called promoter elements which include the TATA box, the initiator and the downstream promoter element. The core promoter elements direct the recruitment and assembly of Pol II and general transcription factors (Juven-Gershon et al., 2008; Lee and Young, 2000).

Unlike the majority of eukaryotic organisms, the protein coding genes of trypanosomatids are organized into polycistronic units (mRNA that encodes for multiple different polypeptides) which are transcribed by Pol II (Clayton, 2002; Martínez-Calvillo et al., 2003; Palenchar and Bellofatto, 2006). Genes from a trypanosomatid polycistronic unit generally code for functionally unrelated proteins unlike the polycistronic units found in bacteria (Campbell et al., 2003). All genes which are part of a polycistronic unit will of course be transcribed at the same level.

As such the level of protein translation must be controlled after transcription of the primary transcript (Clayton and Shapira, 2007). The first major post transcriptional control point of protein expression occurs when mature nuclear mRNAs are generated from primary transcripts. This process occurs by coupled trans-splicing and polyadenylation reactions(LeBowitz et al., 1993; Matthews et al., 1994). Trans-splicing occurs when two exons are spliced together and does not typically occur in higher eukaryotes. Signals contained within the intergenic regions of the polycistronic transcripts direct the splicing machinery to the correct sites at which to splice. The process of trans-splicing adds a common 39-nucleotide miniexon or spliced leader to the 5' end of the mRNAs(Parsons et al., 1984).

The second major control point is regulated by the 3' untranslated region (UTR) of each mRNA which affects the stability of the transcript (Drozdz and Clayton, 1999; Lopez-Estrano et al., 1998). A number of life cycle stage specific proteins have been identified which use this mechanism to affect protein expression, including procyclin, a protein that is expressed on the cell surface in the place of VSG. This expression occurs once *T. brucei* is transferred from mammalian host to insect host. The temperature change from 37 ^{0}C to 22 ^{0}C stabilizes the 3' UTR of the procyclin transcript and allows expression of procyclin (Engstler and Boshart, 2004; Vanhamme et al., 2000).

There is growing evidence to support the view that the mRNA levels in *T. brucei* which code for a protein do not always correlate with the actual levels of protein which are expressed. This evidence suggests that translational control and post translational control in trypanosomatids plays an important role in protein expression (Gale et al., 1994; Mayho et al., 2006; Nardelli et al., 2007). Recently many RNA binding proteins have been identified which may be involved in mRNA stability and translational control (Delhi et al., 2011; Ouna et al., 2012; Wurst et al., 2012), although the mechanism of such control is not yet fully understood. It may be difficult to speculate on the functions of the various controls of RNA degradation as many of these mechanisms appear to be unique to trypanosomes. It is important, from a practical standpoint, to note that high mRNA levels do not always reflect high cellular protein levels as the mRNA may not be translated. However, lower RNA levels usually reflect lower protein levels as there is less transcript to be translated.

1.1.6 Variable Surface Glycoprotein

Bloodstream form trypanosomes are constantly under attack by the humoral immune response while in their mammalian hosts. To evade this immune response, trypanosomes have evolved to physically shield the cell surface and other invariant surface proteins with a densely packed surface coat, which consists of a highly immunogenic protein termed the variable surface glycoprotein (VSG). The surface of the trypanosome is entirely covered with a monolayer of ~10^7 identical VSG, which are periodically switched, allowing the escape of a fraction of the trypanosomes to this specific immune response (Cross, 1978; Doyle, 1977; Vickerman, 1978). As the number of bloodstream form trypanosomes increases in the host, the immune system mounts antibodies to these surface antigens. Cells bearing these antigens are cleared from the host but cells which switched VSG escape to cause a new wave of parasitaemia. This process results in the characteristic waves of fever seen in patients infected with trypanosomes.

1.1.6.1 Genetic organization and expression of VSG

Trypanosomes contain a large library of possible VSG genes and pseudo genes stored in telomeric and subtelomeric regions in a range of chromosomes which are routinely transcribed by RNA Pol I at VSG expression sites (ES) of which there are ~20 (Glover and Horn, 2006; Marcello and Barry, 2007). The analyzed genome strain, *T. brucei 927,* shows ~1600 possible VSG genes arranged as tandem arrays and it is possible that different strains contain larger archives (Masterson et al., 1988). Activation requires recombination into VSG expression sites found at the telomeres of some megabase and intermediate chromosomes (Becker et al., 2004). VSG gene expression may be switched via a number of different mechanisms: transcriptional switching, telomeric VSG conversion, array conversion and segmental conversion. The first, transcriptional switching, simply inactivates one ES and expression occurs at another. The second, telomeric VSG conversion, occurs when a full VSG ORF and upstream and downstream regions are recombined into an active ES. Thirdly, Array conversion differs from the previous mechanism as only the VSG ORF is recombined into an active ES from a VSG pseudo gene. Lastly, segmental conversion occurs whereby a sequence is recombined from a number of normal non-functional VSGs to generate a novel mosaic VSG in the ES (Stockdale et al., 2008).

1.1.6.2 Structure of VSG

Mature VSGs fold to form a dumbbell shaped structure (Figure 1.5) that is attached to the surface membrane through a GPI anchor (Blum et al., 1993; Carrington and Boothroyd, 1996; Carrington et al., 1991). The VSG contain 400-500 amino acids, the majority having between 420 and 460 residues, and contain two domains. The *N*-terminal domain consists of ~350 residues followed by the *C*-terminal domain which contains one or two smaller subdomains, 40-80 amino acids in length (Allen and Gurnett, 1983; Carrington et al., 1991). The identity between different *N*-terminal domains may be as little as ~20% although there appears to be a conserved pattern of cysteine residues. Three patterns have been observed and classified as domain type A, B and C (Carrington et al., 1991) and since the *T. brucei* sequencing project these three subdivision have been further categorized into further divisions of type A (A g1, A g2, A g3 and A g4), B (B g1, B g2 and B g3) and C (C1 to 5) (Marcello and Barry 2007). Although little primary structure is conserved between VSG *N*-terminal domains, the secondary structure appears to be common. The backbone consists of two long antiparallel α helices that form a coiled coil. It has been inferred that the exposed epitopes are located in the loops at the extracellular end of the *N*-terminal domain (Masterson et al., 1988; Miller et al., 1984a, 1984b).

The *C*-terminal domain of the VSG has ~40% identity between different VSGs and because of its position within the mature protein the *C*-terminal domain appears to not be involved in antigenic variation. Previously it was thought that four *C*-terminal domain cysteine patterns exist although the genome project has revealed there are six types. The *C*-terminal domain types 2, 4 and 5 contain one domain and 4 cysteine residues and the remaining types 1, 3 and 6 contain two smaller subdomains, each containing 4 cysteines making a total of 8 cysteine residues. Of the VSG analysed so far different *N*- and *C*- terminal domains may be observed in any combination and there appears to be no restriction. The *C*-terminal domain contains the signal sequence for the addition of a glycosylphosphatidylinsitol (GPI) anchor (Ferguson, 1999).

Of the VSG *N*-terminal domains to have their structures solved by X-ray crystallography (both of which correspond to type A *N*-terminal domains) it is interesting to note the similarity in tertiary structure in spite of only a 20% sequence similarity.

1.1.6.3 Significance of disulphide bonds in VSG

The similarity in primary structure seems to lie mostly with the positioning of the cysteine residues between the two different *N*-terminal and *C*-terminal domains of the VSG (Figure 1.6). This positioning suggests the cysteine residues, and the stabilizing disulphide bridges which are formed from these residues, play a very important role in the higher structure of VSG. The formation of correct disulphide bridges thus presents a significant step in the formation of mature VSG. With each VSG containing 16-20 cysteine residues (always even numbers) and thus a maximum of 8-10 disulphide bridges to be formed correctly, the possibility of misfolding a VSG seems likely. For instance, VSG 117 has 6 disulphide bridges which, in theory, could be arranged into a total of 10,395 disulphide combinations (Anfinsen and Scheraga, 1975) but only one of these is found in the mature VSG 117 on the cell surface (Allen and Gurnett, 1983). Thus to assist in the correct folding of VSG there must be an enzymatic pathway which improves the accuracy and speed of VSG production.

As VSG is a surface protein, it must be folded within the endoplasmic reticulum (Ferguson, 1999). Within the endoplasmic reticulum reside the enzymes of the oxidative folding pathway, enzymes which have been shown to be involved in the correct formation of disulphide bridges (Darby and Creighton 1995). The VSG, due to its high level of disulphide bridges, presents an ideal endogenous substrate for the study of this the oxidative pathway, of which little is known *in vivo*. It has been shown that when VSG production is disrupted at the transcript level, the cell cycle is halted at the precytokinesis stage and cells do not undergo reinitiation of S phase of the cell cycle (Sheader et al., 2005). It is thought this halting of the cell cycle is due to a yet uncharacterized mechanism which "senses" the lack of mature VSG on the cell surface and triggers a cell wide halting of protein production (Smith et al., 2009). If various enzymes of the oxidative folding pathway are disrupted, it is likely that the disturbance will be reported by this characteristic growth arrest.

Figure 1.5 Structure of VSG domains

Panel A. Solution structure if an VSG type an *N*-terminal domain dimer from the VSG MITat1.2as solved using X-ray crystallography (Freymann et al., 1990). *N* and *C*-terminals are indicated, cysteine residues are rendered using a green space filling model. The two separate domains are shown in different colours, α-helices are represented as cylindrical tubes and β-pleated sheets are represented by arrows.

Panel B. A second solution structure of a VSG type an *N*-terminal domain dimer from the VSG ILTat1.42 as solved using X-ray crystallography (Chattopadhyay et al., 2005). The structure is rendered as in the previous diagram. Note the structural similarities between MITat1.2 and ILTat1.42 despite little primary sequence similarity.

Panel C. Solution structure, as solved using NMR spectroscopy, of a VSG type 2 *C*-terminal domain from the VSG MITat1.2 (Jones et al., 2008). *N*- and *C*- terminus are indicated and cysteine residues which form two disulphide bonds are displayed using a space filling green model. The two *N*-terminal and *C*-terminal domains are joined by an additional 15 residue protease-sensitive linker which has been omitted. There are also an additional 16 residues containing an *N*-linked glycosylation site between the C-terminus of the model and the ethanolamine of the glycosylphosphatidylinsitol anchor which are not shown.

Figure 1.5 Structure of VSG domains

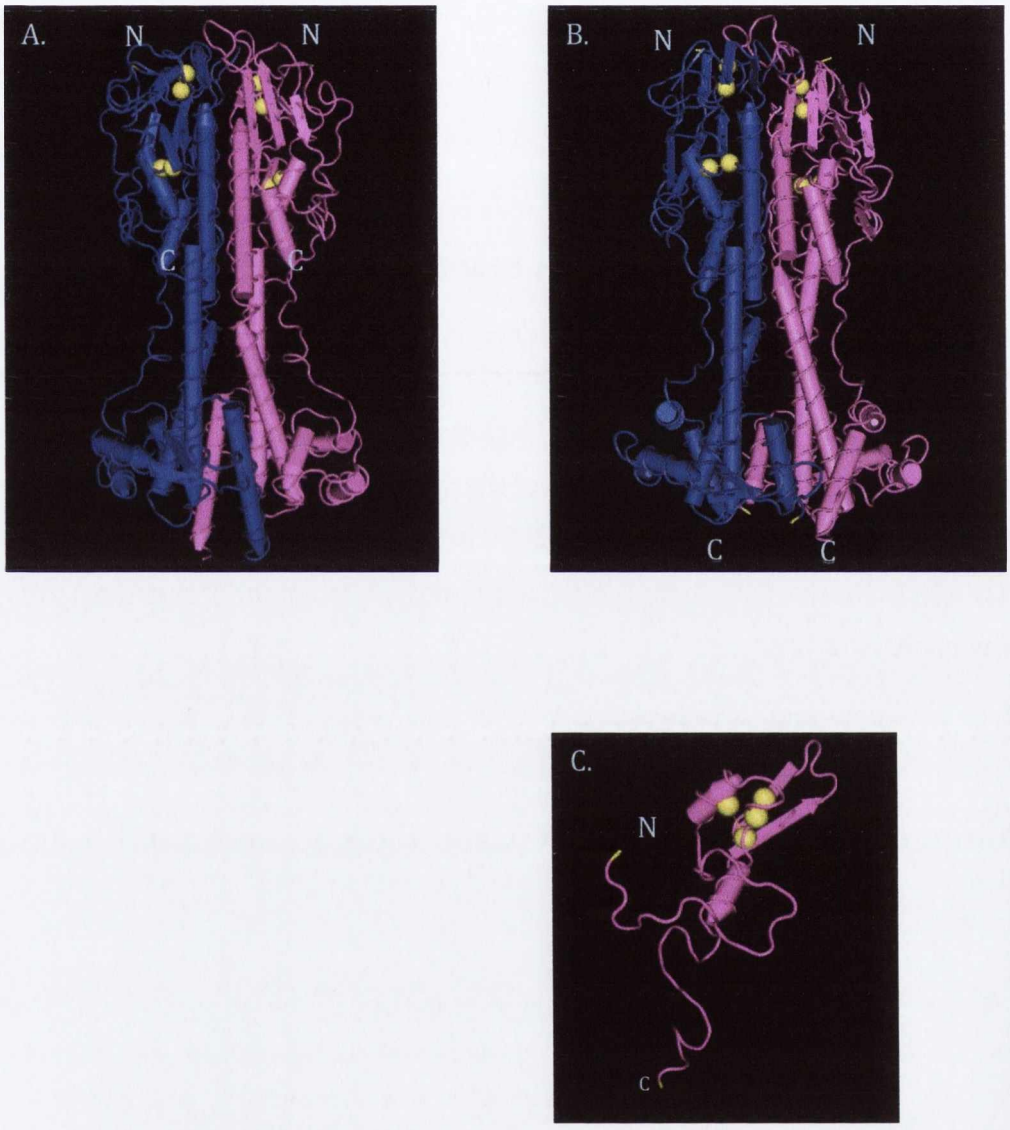


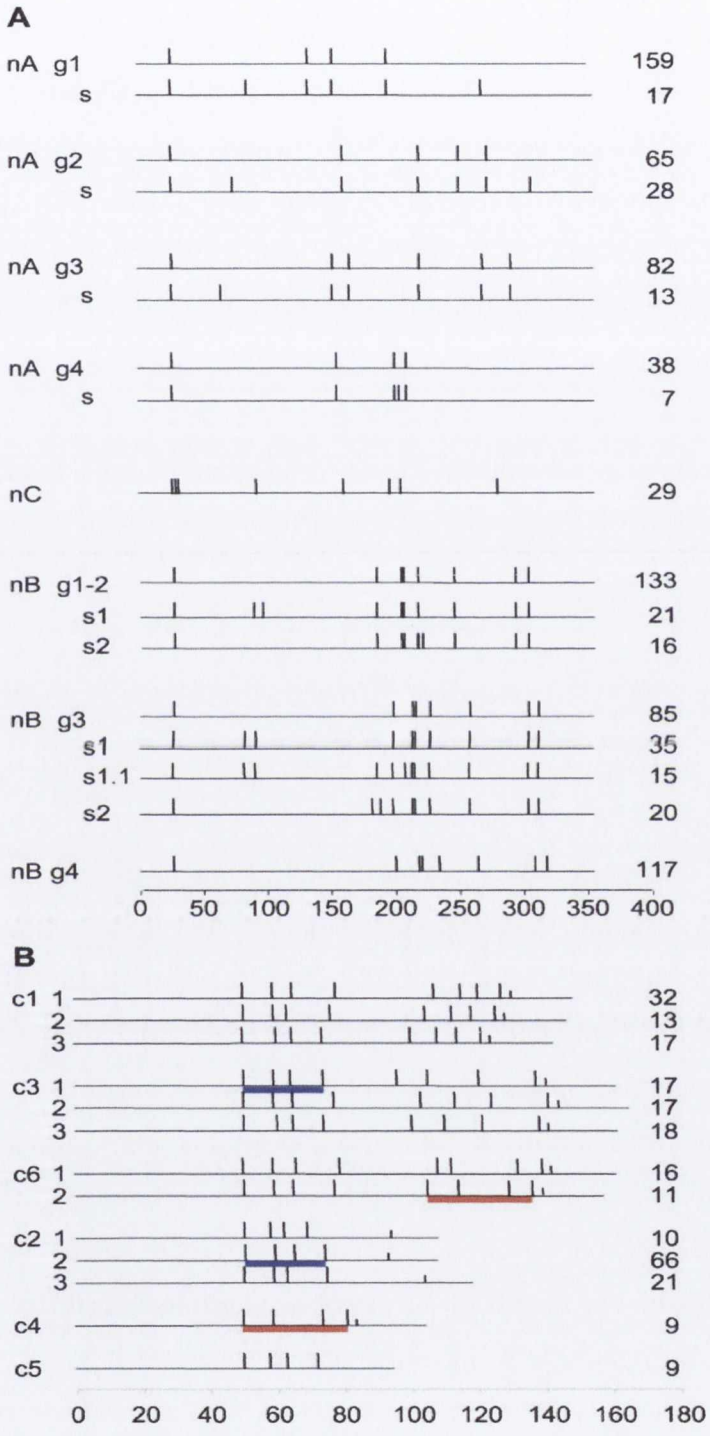
Figure 1.6 Cysteine pattern of VSG domain types

The schematic represents an analysis of the silent VSG archive of *T. brucei* (Marcello and Barry, 2007). The *N*- and *C*- terminal domain sequences were analysed and categorised, further expanding on the classification system previously set (Carrington et al., 1991).

(A) Schematic of VSG *N*-terminal domain. The horizontal line represents the amino acid sequence of the VSG proteins. Cysteine residue positions are shown as vertical bars, and the number of sequences in each category is shown to the right. Group of VSG cysteine pattern is indicated by g 1-4 and subgroup by s1-2. Scale bars (number of residues) are shown at the bottom.

(B) Examples of *C*-terminal domain cysteine patterns. *C*-terminal patterns are labelled c1-5. The horizontal line represents the amino acid sequence of the VSG proteins. Cysteine residue positions are shown as vertical bars, and the number of sequences in each category is shown to the right. Cysteine sub-domains shared between different domain types are indicated by underlines of the same colour. Two to three examples of domain type subgroups are given for the main domain types. Half-length vertical bars represent putative GPI signal sequences. Scale bars (number of residues) are shown at the bottom.

Figure 1.6 Cysteine pattern of VSG domain types



1.2 The Endoplasmic Reticulum

1.2.1 Form and Function

The endoplasmic reticulum is an organelle which is present in all eukaryotic organisms. The membrane of the endoplasmic reticulum typically constitutes more than half the total cell membrane of an average cell and forms a labyrinth of tubules and sacs which interconnect with the nuclear membrane (Figure 1.7). The membrane forms a continuous sheet with the nuclear outer membrane which encloses the lumen of the endoplasmic reticulum. The endoplasmic reticulum plays an important role in lipid and protein production and is used as a Ca^{2+} store in some cell types which may be used in cell signaling. The endoplasmic reticulum is the site at which all transmembrane proteins are produced. Lipids for most of the cells organelles are also produced here, specifically: the endoplasmic reticulum, the Golgi body, lysosomes, endosomes, secretory vesicles and the plasma membrane. Also the majority of the lipids for mitochondria and peroxisomal membranes are made in the endoplasmic reticulum. Almost all proteins to be secreted to the cell exterior must pass through the endoplasmic reticulum (Anelli et al., 2012).

The endoplasmic reticulum provides an environment where tertiary structures of proteins, including disulphide bond formation and carbohydrate addition, may be executed correctly (Liu *et al.* 1979; Braakman *et al.* 1991). These modifications provide long term protein stability for proteins of the secretory pathway. The endoplasmic reticulum is enriched with many chaperones and folding enzymes which enable the formation of a mature native protein structures. There are also enzymes present for quality control of proteins which have been mis-folded and must be either refolded by an isomerase or processed along the endoplasmic reticulum associated degradation pathway. The endoplasmic reticulum is also relatively oxidizing compared to the cytosol; due in part to the presence of reduced/oxidized glutathione which, the general consensus in the literature states, is present at a concentration of ~9 mM when measured in rat liver (Bass et al., 2004). The ratio of reduced glutathione (GSH) to oxidized glutathione (GSSG) was found to be 3:1 in the endoplasmic reticulum (Hwang *et al.* 1992; Bass *et al.* 2004) in contrast with the cytosol which had a ratio ranging from 30:1 to 100:1 (Hwang *et al.* 1992). These oxidizing conditions present in the endoplasmic reticulum allow the formation of

protein disulfide bonds to be thermodynamically favorable (Hwang et al., 1992). In the yeast *S. cerevisiae* it has been shown that altering of the reducing conditions of the endoplasmic reticulum by the use of reducing agents such as dithiothreitol (DTT) causes the buildup of mis-folded proteins in the endoplasmic reticulum, activating the unfolded protein response (Ron and Walter, 2007; Zhang and Kaufman, 2006). This pathway induces genes which attempt to return the folding capacity of the endoplasmic reticulum to normal levels. Although *Trypanosoma brucei* has been shown to be sensitive to DTT concentration levels, this pathway was reported not to exist (Koumandou et al., 2008). It appears that *T. brucei* is transcriptionally inflexible when it comes to the induction of genes in response to chemical stress. The endoplasmic reticulum contains other reducing /oxidizing proteins including protein disulphide isomerase (PDI) and endoplasmic reticulum oxidoreductin 1 (Ero 1) which are now considered to be part of the major pathway of disulphide bond formation in eukaryotes and the topic of this study.

Figure 1.7 biosynthetic secretory and endocytic pathway of the cell

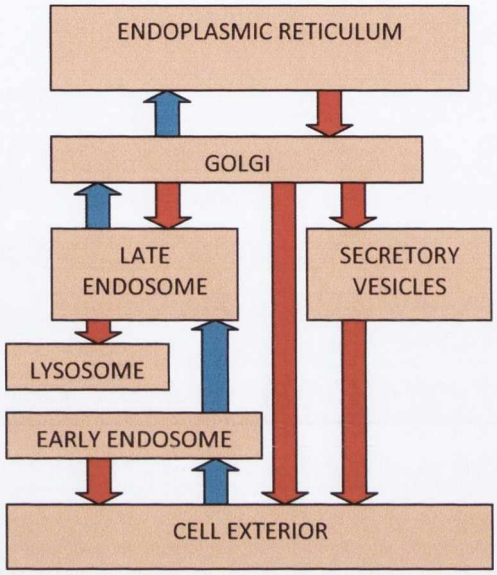
Panel A. In this diagram the biosynthetic secretory and endocytic pathway is represented. The organelles are shown as boxes and their pathways to and from other organelles are indicated by arrows. As can be seen all proteins which are destined for the cell exterior must first pass through the endoplasmic reticulum.

Panel B. The various organelles of the biosynthetic secretory and endocytic pathway are shown in this diagram and here we can see the role of the endoplasmic reticulum in terms of vesicular transport. As can be seen the nuclear envelope and endoplasmic reticulum form one continuous enclosed space known as the lumen of the endoplasmic reticulum.

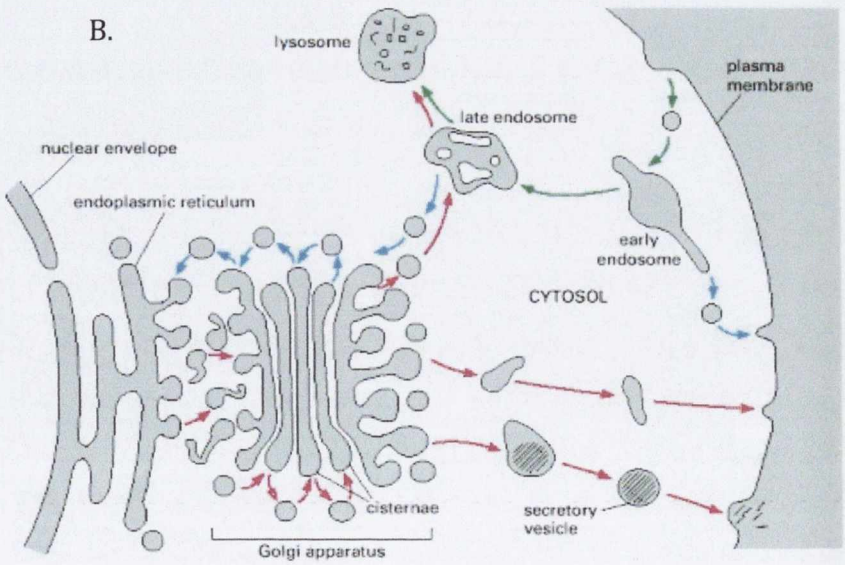
Diagram modified from Alberts Molecular Biology of the Cell.

Figure 1.7 biosynthetic secretory and endocytic pathway of the cell

A.



B.



1.2.2 Translocation of proteins to the endoplasmic reticulum

The first step of protein processing in the endoplasmic reticulum occurs once mRNA leaves the nucleus and translation takes place. Translocation of mRNA ribosome complexes to the endoplasmic reticulum occurs in eukaryotes via the signal recognition particle (SRP) pathway (Shan and Walter, 2005). Proteins which contain the signal peptide sequence for translocation into the endoplasmic reticulum are recognized by the SRP protein which chaperones the ribosome and mRNA complex to the endoplasmic reticulum. The ribosome-nascent chain-SRP complex is bound to the membrane via the SRP receptor; and translation of the protein through the membrane of the endoplasmic reticulum occurs via the translocation channel. The transmembrane protein Sec61 is responsible for forming this channel in eukaryotes (Rapoport, 2007). The complex formed between Sec61 and ribosome-nascent chain-SRP also has the function of inserting hydrophobic regions of membrane proteins laterally into the membrane of the endoplasmic reticulum. It has been shown in yeast that a second method of translocation can occur post translationally(Deshaies et al., 1991). This method requires additional proteins to form a complex with Sec61. These include Sec62, 63, 71 and 72 which in conjunction with the endoplasmic reticulum resident chaperone BiP actively translocate proteins, post transcriptionally, across the membrane of the endoplasmic reticulum which is driven by the hydrolysis of ATP to ADP. In *T. brucei* very little is known about the mechanisms of protein translocation across the endoplasmic reticulum membrane. The SRP pathway consists of the proteins SRP5, SRP68, SRP19 and SRP72. Published RNAi knockdown experiments have shown that the knockdown of individual proteins of this pathway do not affect the translocation of all proteins across the endoplasmic reticulum, suggesting that *T. brucei* also possesses an SRP independent pathway (Lustig et al., 2007). Investigation into this independent pathway, using RNAi techniques, has identified a Sec71 like protein which was shown to be required for the translocation of GPI-anchored proteins. A Sec63 like protein was also identified and knocked down via RNAi, which severely impaired the translocation of all signal peptide containing proteins to the endoplasmic reticulum and blocked polytopic membrane protein biogenesis, suggesting SEC63 like protein was involved in both pathways of translocation across the endoplasmic reticulum (Goldshmidt et al., 2008). These recent studies show *T. brucei* contains both an SRP pathway and an SRP independent

pathway, although both pathways translocate different protein types to the pathways found in yeast and mammals and have yet to be fully understood.

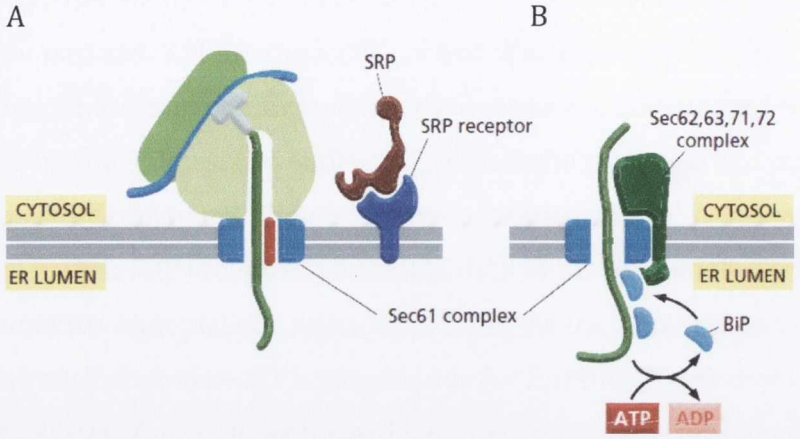


Figure 1.8 Translocation of proteins to the endoplasmic reticulum

Panel A. The components of co-translational translocation are represented in this diagram. The sec 61 protein forms a pore through which the ribosome may translate the protein directly into the lumen of the endoplasmic reticulum. Previously the ribosome is brought to the membrane of the endoplasmic reticulum by the signal recognition particle (SRP) and is bound to the SRP receptor. At the point of translation shown in the diagram, the peptide sequence for translocation to the endoplasmic reticulum is shown in red and has already been cleaved from the emerging protein.

Panel B. A second mechanism of post translational translocation is shown here. An additional complex composed of the Sec62, Sec63, Sec71, and Sec72 proteins attaches to the Sec61 translocator and BiP chaperone molecules attach to the emerging protein and act as a molecular ratchet. This mechanism requires energy in the form of ATP.

1.2.3 Secretion and Retention in the Endoplasmic Reticulum

Secretory and surface proteins which are bound for transport from the endoplasmic reticulum contain a signal peptide which marks them for their destination (McConnell et al., 1981). The peptide is attached to the *N*-terminus of the polypeptide, which is removed once the protein has been processed for transport. These proteins which are packaged in transport vesicles must travel from the endoplasmic reticulum to the golgi apparatus before reaching their final destination. For native proteins of the endoplasmic reticulum, which are to be retained or transported back to the endoplasmic reticulum, a specific signal peptide is located at the *C*-terminus (Munro and Pelham, 1987). The most common of these endoplasmic reticulum retention motif is KDEL(lys-asp-glu-leu) however there are other variations of this motif (Andres et al., 1990).

1.3 Oxidative Folding Pathway

There are many enzymes and chaperones responsible for the proper folding of proteins in the eukaryotic cell. This study is concerned with the folding enzymes which reside in the endoplasmic reticulum and make up a pathway which is known as the oxidative folding pathway. The aim of this introduction is to discuss the group of enzymes thought to be responsible for the majority of this oxidative folding, the form and function of these enzymes and the questions which have yet to be answered regarding their function.

1.3.1 Protein Disulphide Isomerase

1.3.1.1 Protein Disulphide Isomerase Enzyme Family

Disulphide bonds are covalent linkages which are formed between side chains of cysteine residues contained with a protein. Their main function is to stabilize protein structures but they have also been shown to be used in regulatory control. In the early 1960s the need for a protein to catalyze disulphide isomerization was hypothesized. This need was due to the possible number of incorrect combinations of cysteine residues that could be formed during protein folding. The group of Brunó Straub (Venetianer and Straub, 1963) discovered an enzyme purified from the endoplasmic reticulum of both pigeon and chicken pancreas that was able to reduce and oxidize ribonuclease. The group of Anfinsen and co-workers independently discovered an enzyme in rat liver which also accelerated the reactivation of ribonuclease and which they termed Protein Disulphide Isomerase (PDI) (Anfinsen and Haber, 1961). During the intervening 50 years many PDIs have been described. They are considered to have essential activities in eukaryotes because of their role for the maturation of proteins in the endoplasmic reticulum.

The PDI enzyme family consists of a group of enzymes which are thiol-disulfide oxidoreductases. They are characterized by their size, localization in the endoplasmic reticulum, catalytic activity and their thioredoxin-like (Tx-like) active site. This site contains the characteristic CXXC motif which catalyzes the exchange of disulphide bonds (Noiva and Lennarz, 1992). Conserved Tx-like active sites are seen in all members of the thioredoxin super family (Table 1.3). Other members of this family include bacterial Dsb proteins and thioredoxin. Thioredoxin is smaller than PDI, with

a molecular weight of 12kDa, and is found in many cellular compartments where it has many functions involving its reducing properties (Holmgren, 1989). It is believed that PDI and thioredoxin share a common ancestry, a theory which is supported by their sequence similarity and distribution of introns (Sahrawy et al., 1996). Thioredoxin and PDI families differ in other respects besides size; PDIs have a higher redox potential than thioredoxin (under the same conditions)(Hatahet and Ruddock, 2009). Protein disulphide isomerases also contain extra domains which do not contain the CXXC motif. A series of *in vitro* studies have shown that these extra domains are required for other aspects of PDI function, the most important of which is their ability to bind proteins for the purposes of isomerization and chaperone activity. This additional ability implies a more specific and more efficient enzyme-substrate interaction than thioredoxin (Danon, 2002).

The majority of PDIs described contain one or more domains with Tx-like redox active sites. A classification system, based upon the number and position of the Tx-like sites, has been devised (Figure 1.9). Class 2 PDIs have two active sites separated by 14-37 residues while class 1 PDIs have active sites separated by 224-247 residues. Class 3 PDIs are believed to be divergent from class 1 PDIs due to this larger distance between active sites but these contain three Tx-like redox active sites (McArthur et al., 2001). The first PDI characterized was a class 2 PDI and is considered the classical example of a PDI.

Table 1.3 Alignment of thioredoxin super family active sites found in various species

Protein	Active site sequence	
S. cerevisiae(ISBYSS) PDI	APWCGHCK	APWCGHCK
H. Sapiens(Z49835) PDI	APWCGHCK	APWCGHCK
G. lamblia(U64730) PDI-1	APWCGHCK	
G. lamblia (U65017) PDI-2	APWCGHCK	
G. lamblia (AF16411) PDI-3	APWCGHCK	
E. coli (X54539) Thioredoxin	AEWCGPCK	
E. coli (P24991) Dsb A	SFFCPHCY	
E. coli (U28375) Dsb C	DITCGYCH	

Table 1.3 Alignment of thioredoxin super family active sites found in various species

Examples of characterized PDI, thioredoxin and Dsb enzymes were chosen from various species and their active sites were aligned manually. Redox active cysteine residues are shown in green. GenBank/EBI accession numbers are included.

Figure 1.9 Classification of protein disulfide isomerase proteins (after Kanai et al. 1998)

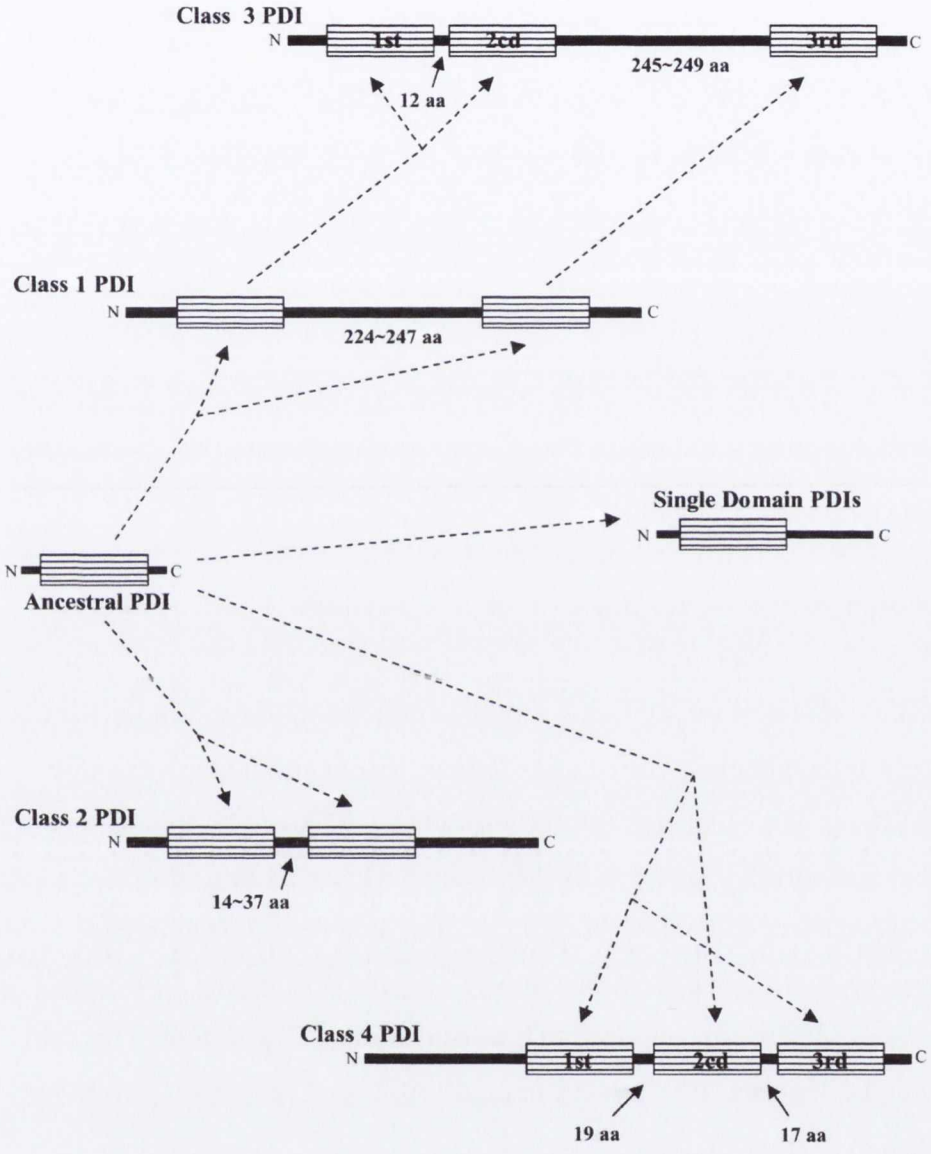


Figure 1.9 Classification of protein disulfide isomerase proteins (after Kanai et al. 1998)

The above diagram shows the possible evolution of the various classes of PDI enzymes. Lined boxes represent thioredoxin-like domains. Interdomain spacing is labeled. (Diagram taken from McArthur et al. 2001)

1.3.2 Structure of PDI

The structure of many single domain PDIs have been solved. The secondary structure is common to all PDIs and is also observed in the single domain members of the thioredoxin superfamily (thioredoxins and glutaredoxins). They contain the following arrangement: β1-α1-β2-α2-β3-α3-β4-β5-α4 (Gruber et al., 2006; Pan and Bardwell, 2006), with the active site cysteine residues positioned at the *N*-terminus of the second helix. A hydrophobic region is found close to the active site implying a substrate binding site (Kemmink et al., 1996; Nordstrand et al., 1999).

Class 1 Members of the PDI family, the classical PDIs, share a common domain structure which was first devised by the use of sequence-based criteria, proteolysis of native PDI and characterization of recombinant proteins (Freedman et al., 1998). The crystal structure of Pdi1p, a multidomain PDI of *S. cerevisiae*, was later solved in 2006 confirming this domain model (Tian et al., 2006). The model contains four structural domains a, b, b', a' with a linker region located between b' and a' and a *C*-terminal extension. In the literature, if the domain is catalytic it is referred to as an a-type domain, and b-type if non catalytic. In some more recently described PDIs the active site of the a-type domains may not be redox active, but due to sequence similarity are still referred to as a-type domains. The b-type domains do not contain any cysteine residues and as such are not redox active but they still share sequence similarity to thioredoxin. When separated, a-type domains are capable of catalyzing simple oxidation or reduction reactions but without b-type domains these enzymes are unable to perform simple isomerization reactions(Darby and Creighton, 1995). It is apparent that b-type domains are required to bind and stabilize the unfolded protein substrate(Darby and Creighton, 1995; Darby et al., 1998) this theory is supported by the structure of Pdi1p (Tian *et al.* 2006).

1.3.3 Localization of PDI

The PDIs are principally located in the endoplasmic reticulum as soluble proteins, as they contain the KDEL *C*-terminal endoplasmic reticulum retention sequence or have a variation of this. There are however exceptions to this localization; for example the *H. sapiens* PDI known as ERp44 localizes between the endoplasmic reticulum and the Golgi apparatus (Anelli et al., 2007). For additional mammalian subcellular localizations of mammalian PDIs refer to Table 1.4. The protein TbPDI-1 (systematic gene ID: Tb927.4.2450) of *T. brucei* was also reported to not contain the usual KDEL *C*-terminal endoplasmic reticulum retention sequence. Indirect immunofluorescent antibody staining of fixed cells suggested that TbPDI-1 may be localized in the highly active endocytic pathway in bloodstream trypanosomes (Rubotham *et al.* 2005). This finding is discussed in more detail in section 1.3.6.

Table 1.4 Mammalian PDI family members and their possible subcellular locations as determined experimentally

Protein	SwissProt accession	Synonyms	Cellular localizations
PDIA1	P07237	PDI	ER, cell surface, nucleus, cytosol, secreted
PDIA2	Q13087	PDIp	ER
PDIA3	P30101	ERp60; ER-60; GRp58; ER-58; ERp58; ERp61; Q2	ER, nucleus, plasma membrane rafts, cell surface, cytosol, secreted
PDIA4	P13667	CaBP2	ER, cell surface, secreted
PDIA5	Q14554		ER
PDIA6	Q15084	P5; CaBP1	ER
PDIA7	Q8N807	PDILT; ERp65	ER
ERp18	O95881		ER
ERp27	Q96DN0		ER
ERp29	P30040	ERp31; ERp28	ER
ERp44	Q9BS26		ER / golgi apparatus
ERp46	Q8NBS9		ER
ERdj5	Q81XB1		ER
TMX1	Q9H3N1		ER
TMX2	Q9Y320		ER
TMX3	Q96JJ7		ER
TMX4	Q9H1E5		ER

Table 1.4 Mammalian PDI family members and their possible subcellular locations as determined experimentally

The above table shows a list of mammalian PDI family members, with their synonyms and subcellular localizations. Each of the above cellular localizations was shown experimentally.

Table adapted from (Turano et al., 2002).

Figure 1.10 Domain Structure and 3d Structure of a classical class 1 protein disulphide isomerase

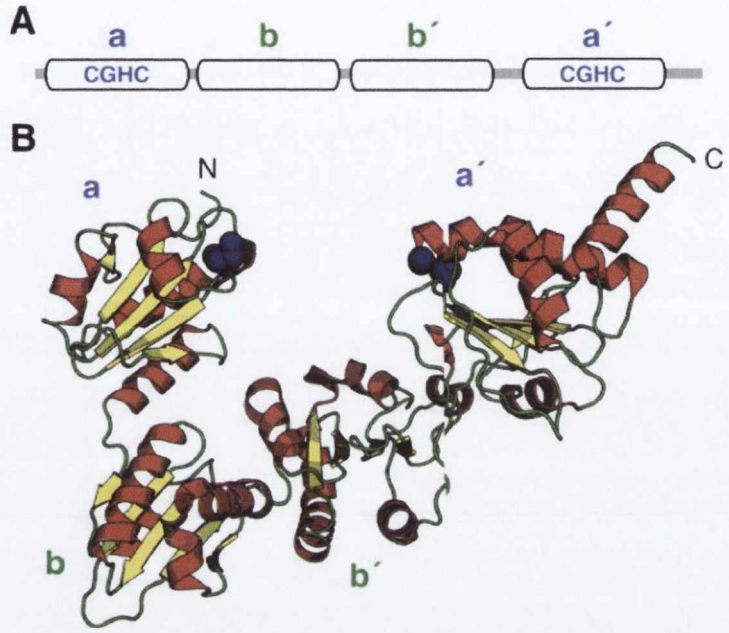


Figure 1.10 Domain Structure and 3d Structure of a classical class 1 protein disulphide isomerase

Panel A. A schematic representation of the PDI domain structure. Thioredoxin-like domains are shown as rounded rectangles with purple lettering denoting catalytic domains (a and a') and green lettering non-catalytic domains (b and b'). The active-site sequences are shown.

Panel B. The 3d structure of PDI solved by x-ray crystallography (Tian et al.,2006). The different domains are indicated. The α-helices are represented as red ribbons and the β-sheets are shown as yellow arrows. The active-site cysteines are depicted as purple spheres.

Diagram adapted from Appenzeller-Herzog and Ellgaard, 2008

1.3.4 Reaction mechanism of PDI

The reaction mechanism of PDI utilizes its thiol groups to catalyze thiol-disulfide reactions which result in the reduction, oxidation or isomerization of disulphide bonds (Creighton, 1984; Walker et al., 1996). This thiol group reaction mechanism is common to all enzymes containing the CXXC active site which includes all members of the thiol-disulphide oxidoreductases family. The two cysteine residues of this active site may be in the disulphide (S-S) state or the $(SH)_2$ state. The mechanism involves the formation of mixed disulphides (Figure 1.11). The first step of the reaction involves the deprotonation of a thiol group on the acceptor protein, resulting in a thiolate anion (S^-). The electrons of the anion form a bond between the sulpher atoms of the two thiol active site, forming a mixed disulphide. The other thiol group of the acceptor protein is then deprotonated and the electrons of the resulting thiolate anion form a new disulphide bond. The remaining thiolate anion of the donor protein receives a proton forming a thiol group completing the disulphide bond exchange. The redox activity of the PDI enzymes depends on the oxidation state of the CXXC active site. If the active site contains a disulphide bond (oxidized state), electrons may be accepted from a protein substrate (oxidation of the substrate). The opposite holds true if the active site is in the reduced state. This allows PDI to accept electrons from cysteine residues from a substrate protein and form disulphide bonds within the substrate protein (oxidation of the substrate), pass on electrons to break incorrectly formed disulphide bonds within the substrate protein (reduction of the substrate) and ultimately form correctly folded proteins which may occur *in vivo* through rounds of oxidation and reduction (isomerization) although this hypothesis has not been conclusively proven (Schwaller et al., 2003; Walker and Gilbert, 1997). In an isomerase reaction the net number of electrons remains the same in the protein being folded.

Figure 1.11 Mechanism of thiol–disulphide exchange reactions

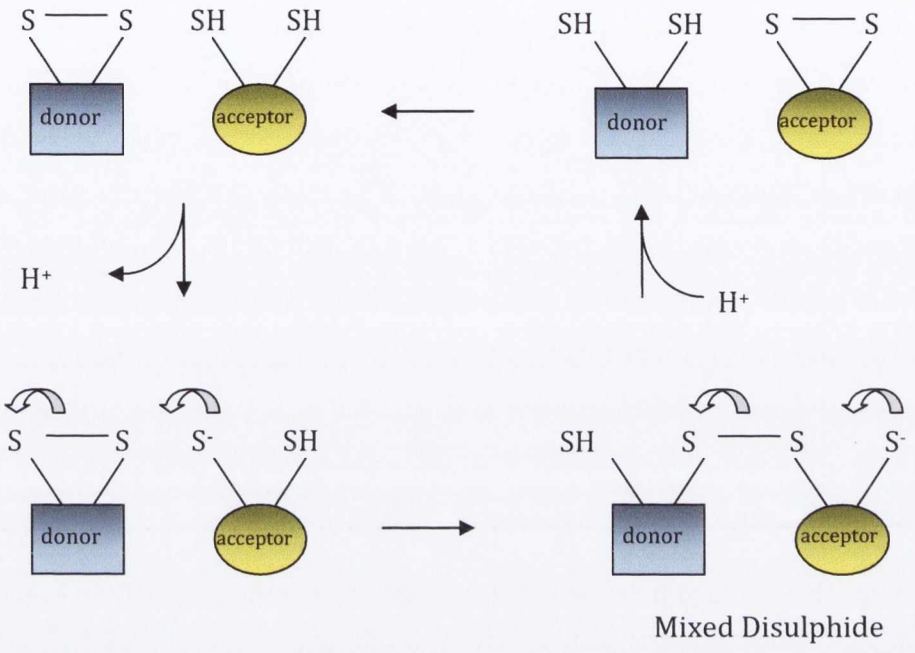


Figure 1.11 Mechanism of thiol–disulphide exchange reactions

Thiol-disulphide exchange reaction between disulphide donor and acceptor proteins. Donors are shown in blue and acceptors in yellow. Thereaction proceeds in the direction of the arrows indicated.

1.3.5 In vitro assays for PDI activity

As isomerases, PDIs are able to catalyze the exchange of electrons using the mechanism of thiol–disulphide exchange. Although PDIs are not classified as an oxidase it is important to note that during thiol–disulphide exchange reactions they do catalyze oxidation reactions and reduction reactions, and have been shown to do so *in vitro*. Many in vitro assays have been developed to investigate these oxidation, reduction and isomerization reactions all with their advantages and disadvantages. All are prone to some form of artifact and so conclusions are usually drawn from a number of results. The consensus in the literature is that natural substrates should be used for the investigation of each individual PDI but often a compromise is met where by a readily available substrate is used for convenience. It is also worth noting that the natural substrates of the PDIs are rarely known so substitutes are used in their place.

A commonly reported substrate used for measuring of the activity of a putative PDI is ribonuclease A (RNase A). The three assays which use RNase A are based on a gain of function; once the RNase is returned to its native conformation the activity of the enzyme returns and digests RNA. The activity of RNase A is measured using a dual-wavelength spectrophotometric assay of ribonuclease activity (Lambert and Freedman, 1983). RNase is a single domain protein which contains four disulphide bonds which offers 105 different disulphide arrangements. To assay the isomerase activity of a putative PDI, scrambled RNase may be used. The use of this substrate is not ideal though, as conformations other than the native conformation of RNase do possess some activity. In spite of this artifact this has been the assay of choice reported in the literature for the measure of isomerase activity (Hatahet and Ruddock 2009). Reduced RNase is used in much the same way to measure the ability of a PDI to oxidize a substrate. Lastly the most popular way to measure the ability of a PDI to reduce a substrate is to use insulin as a substrate (Matsuo et al., 2001; Nørgaard et al., 2001). In this assay, insulin reduction by PDI results in the aggregation of the B chains of insulin which causes an apparent increase in absorbance at ~600nm. This method is however not quantitative and insulin is usually not a natural substrate of the PDI being assayed. The two quantitative assays for PDI described above are difficult to carry out and so are not used routinely to measure quantities of PDI, they are usually only used to indicate a PDI ability to carry out certain reactions. As shown in Table 1.5

only a limited range of PDI candidates have been subjected to biochemical assays measuring thiol isomerase, reduction and oxidation activity.

Table 1.5 Examples of assayed PDI enzymes

Species	Designated name	Accession (UNIPROT)	Class	Oxidation	Reduction	Isomerization	Synonyms
H. sapiens	PDIA1	P07237	Class 1	Yes	Yes	Yes	PDI
	PDIA2	Q13087	Class 1	Yes	Yes	Yes	PDIp
	PDIA3	P30101	Class 1	Yes	Yes	Yes	ERp57
	PDIA4	P13667	Class 3	Yes	Yes	Yes	ERp72
	PDIA5	Q14554	Class 4	Yes	Yes	Yes	PDIr
	PDIA6	Q15084	Class 2	Yes	Yes	Yes	PDI P5
T. brucei	TbPDI-1	Q584E2	Single domain		Yes	Yes	
	TbPDI-2	Q38AE1	Class 1		No	Yes	
G. lamblia	gPDI-1	O97451	Single domain	No	No	Yes (low)	
	gPDI-2	O97452	Single domain	No	Yes	Yes (low)	
	gPDI-3	Q9U015	Single domain	Yes	Yes	Yes	
L. major/ donovani	LmPDI	Q8L8E1	Class 1			Yes	
	LdPDI	Q8L7Z7	Single domain			Yes (low)	

Table 1.5 Examples of assayed PDI enzymes

The above table gives examples of PDI enzymes which have been assayed using previously described methods. The list includes all mammalian PDI which show oxidation, reduction and isomerisation reactions: PDIA1 (Anfinsen and Haber, 1961), PDIA2 (Desilva et al., 1996), PDIA3 (Frickel et al., 2004), PDIA4 (Alanen et al., 2003), PDIA5 (Horibe et al., 2004) and PDIA6 (Alanen *et al.* 2003). Also included are the examples of unicellular eukaryotes that have been characterised including the PDIs of *T. brucei*: TbPDI-1 and TbPDI-2 (Rubotham *et al.* 2005), also the PDIs of G. lamblia (Knodler et al., 1999) and Leishmania: LmPDI (Ben Achour et al., 2002) and LdPDI (Padilla 2002).

1.3.6 Thioredoxin superfamily in *protozoan* parasites

It was originally thought that certain *protozoan* parasites did not poses a thioredoxin system, this was largely due to the failure to detect thioredoxin reductases. It was later discovered that they indeed had a thioredoxin system which was unique (Fairlamb et al., 1985). The glutathione reductase system is replaced by the homologous trypanothione reductase system. Since this discovery many members of the thioredoxin super family have been identified after the genome sequencing projects of *T. brucei, T. cruzi, L. major.* A classical thioredoxin has been characterized in *T. brucei* (Reckenfelderbäumer et al., 2000). Additionally knockdown experiments have been carried out on thioredoxin and have shown the protein to be essential in culture (Comini *et al.* 2007). A gene encoding a class 1 PDI was also identified in *T. brucei* (Hsu et al., 1989) and later its product, which was designated the name TbPDI-2, was characterized along with one other single domain PDI which was designated the name TbPDI-1 (Rubotham *et al.* 2005). Little is known about the role of many PDIs in lower eukaryotes to date, but many unusual features have been observed including potential roles for PDIs in secretory pathways, folding of surface proteins, and developmentally controlled PDIs (section 1.1.6). These examples may help in the understanding of other PDIs and/or may be part of a very unique subset of the thioredoxin super family.

1.3.7 PDIs of *Giardia lamblia*

The protozoan eukaryote *Giardia lamblia* was discovered to contain genes encoding three small PDI (Knodler *et al.* 1999). These enzymes contain only one redox active site. It is important to note that single domain PDIs do not occur often in metazoans but small single domain PDIs appear to be a feature of unicellular eukaryotes (McArthur et al., 2001). Two other PDIs were identified following the *G. lamblia* sequencing project (McArthur *et al.* 2001). All five PDIs contain single redox active sites with the PDI known as gPDI3 being very small with a predicted protein size of only 116 amino acids (Table 1.6). As observed for African trypanosomes, the surface of *G. lamblia* is also covered with a protein coat known as the Variant Surface Protein (VSP). Significantly the VSP contains a very high content of cysteine residues (12-16%). In all the VSP proteins studied, all cysteine residues form intramolecular disulphide bridges (Gillin et al., 1996). It has been postulated that the *G. lamblia* PDIs play a role in the correct folding and maturation of these proteins.

Table 1.6 Characteristics of PDIs from *G. lamblia*.

PDI	Genebank accession	Redox active site	Signal peptide	Transmembrane region	*C*-terminal sequence
1	AAD09365	CGHC	Yes	No	NHAK
2	AAD09366	CGHC	Yes	Yes	KRKK
3	AAF20171	CGHC	Yes	No	KQRL
4	AF295634	CGHC	No	No	DKEL
5	AAF89535	CGHC	Yes	Yes	TSKK

Table 1.6 Characteristics of PDIs from *G. lamblia*.

Various characteristics of PDIs from *G. lamblia*. Genebank accession numbers are shown.

1.3.8 PDIs of *Leishmania major*

The first fully characterized trypanosomatid PDI (GenBank: AAN75008.1) was from *Leishmania major* (Achour *et al.* 2002). This PDI, known as LmPDI, was found to be a class 1 PDI which contains two catalytic domains with the classical CGHC active site motif. The class 1 PDI possessed isomerase activity but was not tested for the ability to reduce or oxidize a substrate *in vitro*. Cells were grown in the presence of zinc bacitracin, DTNB and pCMBA, which were known inhibiters of PDI. This caused a strong negative effect on cell growth. It is important to note these inhibitors may have had an effect on other cell functions including the function of other PDIs as such the role of any one PDI cannot be determined using this approach. Interestingly the PDI was expressed at a higher level in more virulent strains of *L. major* suggesting the protein was linked with virulence. When the protein sequence of LmPDI was compared between virulent and less virulent strains of *L. major* they were found to be identical indicating the increase in virulence observed was caused by the increased level of LmPDI transcription. Additionally it was found that LmPDI was highly expressed in the amastigote stage of the parasite which multiplies in the bloodstream of the host further supporting the theory that LmPDI is linked with virulence. A single domain PDI was found in *Leishmania donovani* (Padilla 2002). The single domain PDI when expressed as a recombinant protein was found to have a mass of 15 kDa and possessed only one thioredoxin like domain which contained the classical CGHC active site motif. Despite possessing only one domain the recombinant protein was able to catalyzed isomerization reactions *in vitro*. The protein LdPDI was found to be expressed equally in the proliferative lifecycle stages of the species - the promastigote and axenic amastigote - and was localized in the endoplasmic reticulum by immunofluorescence microscopy through the co-localization with BiP - a known resident of the endoplasmic reticulum. Disruption of the activity of the PDI, through the over expression of an inactive form, resulted in reduced production of secretory proteins by ~20% (Padilla, 2002). The collective evidence to date suggests PDIs play a role in the natural pathogenicity of *Leishmania*.

1.3.9 PDIs of *Trypanosoma brucei*

The first detailed characterization of two PDIs (TbPDI-1 and TbPDI-2) in *T. brucei* was performed following their identification from a group of proteins which were purified using tomato lectin affinity chromatography (Rubotham et al., 2005). Tomato lectin binds with high specificity to *N*-glycans containing linear poly-*N*-acetyllactosamine, which are associated only with proteins, including receptors for host ligands and invariant proteins, from the flagellar pocket / endocytic pathway (Nolan et al., 1999).

Of the enzymes characterized, TbPDI-1 is a single domain PDI with an active site common to eukaryotic PDIs (FYAPW**CGHC**K). The predicted polypeptide contains 413 residues with a mass of 45 kDa and a pI of 9.0. When expressed as a recombinant protein it was found to possess both thiol–reduction and isomeriziation activity *in vitro*. The protein did not contain the common KDEL-type endoplasmic reticulum retention motif found on most PDIs.

The gene which encoded TbPDI-2 was originally identified as BS2 which is expressed in greater amounts in the bloodstream lifecycle stage (this was not quantified). The protein TbPDI-2 is a 497 residue polypeptide with a predicted mass of 55kDa and a pI of 5.3 (Hsu *et al.* 1989). When TbPDI-2 was expressed as a recombinant protein it was found to possess thiol–isomerization activity but not thiol–reduction activity *in vitro*. Interestingly when antibodies were raised against recombinant TbPDI-2 and used to detect TbPDI-2 in a Western blot of bloodstream form cell lysates, the protein was found to have an apparent mass of ~80 kDa. This difference in mass was attributed to glycosylation of the protein.

The protein TbPDI-2 was also found to be a class 1 PDI with two redox-active sites, the *C*-terminal active site being typical of other eukaryotes and the *N*-terminal active site being atypical (CGYC). As with other PDIs described, TbPDI-2 contained an *N*-terminal signal sequence and a *C*-terminal KQDL endoplasmic reticulum retention sequence. Unusually TbPDI-2 also contained multiple consensus sites for *N*-glycosylation (a total of 12). Genes for closely related proteins have been identified in *T. congolense* and *T. vivax* using a OMNIBLAST search of data from the Sanger pathogen sequencing unit (Rubotham *et al.* 2005). These related proteins also have atypical *N*-terminal redox centers and poses multiple *N*-glycosylation sites. These PDI

N-glycosylation sites appear to only be a feature of African trypanosomes and interestingly the yeast class 1 known as PDIp (Mizunaga et al., 1990). Other unique features of trypanosome PDIs are the following:

1) The *T. brucei* PDIs TbPDI-1 and 2 are developmentally regulated, with expression being upregulated in the bloodstream stage of the life cycle and down regulated in the procyclic form (Rubotham *et al.* 2005). In metazoan organisms expression of PDI generally appears to be constitutive although there are exceptions to this rule - PDIA2 is expressed only in the pancreas showing tissue specificity (Desilva et al., 1996).

2) The protein TbPDI-2 is post-translationally modified by extensive *N*-glycosylation. This glycosylation is not essential for enzymatic activity *in vitro* as isomerase activity was shown when the protein was expressed in *E. coli*. It is however unusual for enzymes to contain such a high amount of glycosylation sites (12 putative sites) and appears to be a feature of the African trypanosomes as *T. congolense* (8 putative sites) and *T. vivax* (15 putative sites) contain a large amount of putative *N*-glycosylation sites (see Figure 3.3 for a direct comparison of protein sequences). TbPDI-1 was only moderately glycosylated (2-3kDa added) which is common to other PDIs. In the case of TbPDI-2, *N*-glycosylation was significant with the difference in apparent molecular mass (~80kDa) and predicted molecular mass being ~25kDa. Digestion with *N*-glycopetidase F, which cleaves most *N*-glycans, resulted in a shift in the apparent molecular mass to the predicted 55kDa. It has been hypothesized that these glycans of TbPDI-2 are likely to be complex and that they contain linear repeats of poly-*N*-acetyllactosamine, the biosynthesis of which are thought to occur exclusively in the Golgi complex. This extent of *N*-glycosylation is very unusual making TbPDI-2 a very unique class 1 PDI.

3) Both PDIs appear to be located in unusual subcellular locations. Typically PDI enzymes were thought to reside solely within the lumen of the endoplasmic reticulum where they were responsible for the maturation of secretory proteins. In the case of TbPDI-1 and 2 there was evidence of at least partial localization with elements of the endocytic pathway when visualized using standard epifluorescence although to conclusively show TbPDI-1 and TbPDI-2 reside outside of the endoplasmic reticulum localization studies using electron microscopy need to be carried out as the resolution of immunofluorescence microscopy is not high enough to determine their locations (Rubotham *et al.* 2005). It is not uncommon to find PDIs in other subcellular

locations and until recently it was accepted that this was due to artifacts of the techniques used. Now, however, it is widely accepted that PDIs do serve functions outside of the endoplasmic reticulum as they are being found in complexes which serve yet uncharacterized function. The human PDIs - PDIA1, PDIA4, PDIA6 and the endoplasmic reticulum chaperone BiP have been found in a complex which occludes the cytosolic face of the translocation protein Sec61p (Stockton et al., 2003)which has been shown to play a role in shedding tumour associated ligands (Gumireddy et al., 2007; Jinushi et al., 2008; Kaiser et al., 2007) and tumour metastasis and also plays a role in platelet function (Jordan et al., 2005).

1.3.10 Endoplasmic reticulum oxidoreductin 1

The protein PDI was discovered in the early 1960s and experimental evidence *in vitro* showed that it was capable of oxidation, reduction and isomerization action depending on the oxidative state of the redox active site. To mimic *in vivo* conditions oxidized glutathione (GSSG) was used to reoxidize PDI during *in vitro* assays, and prior to 1998 it was believed that this reoxidation occurred *in vivo*. Later this view was revised by the discovery and characterization of a new subfamily of sulphhydryl oxidases known as endoplasmic reticulum oxidoreductin 1 or Ero 1. The first of these proteins was an Ero 1 enzyme characterized in a temperature sensitive cell line of the yeast *S. cerevisiae,* now termed Ero 1p (Frand and Kaiser 1998; Pollard *et al.* 1998). Curiously the open reading frame encoding a trypanosome Ero 1 protein was discovered in *T. brucei* seven years earlier during the sequencing of a gene encoding a replicative DNA polymerase of trypanosomes although the function of Ero 1 at the time was not known (Leegwater et al., 1991). It was believed at the time of characterization of the first yeast Ero 1 that the enzyme was the sole provider of reducing equivalence to the endoplasmic reticulum and that GSSG was not involved in the reoxidation of PDI although now both Ero 1 and GSSG are now considered to be part of a larger system of the oxidative pathway (Hatahet and Ruddock, 2009).

1.3.11 Sulfhydryl oxidase family

In addition to Ero 1, there exist other sulfhydryl oxidase family members capable of reducing proteins. Analysis of the genome of *T. brucei* revealed the presence of three putative flavo-linked sulfhydryl oxidases: the first is homologous to yeast Ero1p; the second known as TbALR shows similarities to the yeast Erv1p and human ALR, which have been shown to play a role in the oxidative proteins folding in the mitochondrial intermembrane space , and the third is a quiescin sulfhydryl oxidase (Kodali and Thorpe, 2010). The quiescin sulfhydryl oxidase designated the name TbQSOX has been shown to have the ability *in vitro* to oxidize reduced RNAse. Other QSOX enzyme family members have been found in yeast and one of particular interest to the oxidative folding pathway is the protein Erv2p. The protein Erv2p has been shown to have the ability to oxidize PDI *in vitro* and is thought to provide an alternative pathway for the production of oxidative equivalence in the endoplasmic reticulum (Wang *et al.* 2007). While yeast Erv2p is an endoplasmic reticulum resident, the

trypanosomal TbQSOX contains a putative transmembrane spanning region and has been found in the plasma membrane fraction by proteomic analysis of *T. brucei* (Bridges et al., 2008), suggesting TbQSOX does not play a role in the oxidative folding pathway of the endoplasmic reticulum. This possibility is likely as other members of the QSOX family have been shown to have a wide variety of subcellular locations, including the mitochondrial intermembrane space, unlike the Ero 1 family members which have only been reported to reside in the endoplasmic reticulum (Fass, 2008). To date there has been no characterization of Ero 1 in any parasitic protozoa.

Examples of Ero 1 are contained in a wide variety of species where they are considered to be essential enzymes. However the essential requirement for the activity of this enzyme has yet to be conclusively demonstrated through the generation of a conditional null mutant cell line. Members of the family have been characterized in *Homo sapiens* (Cabibbo et al. 2000; Pagani et al. 2000), *Caenorhabditis elegans* (Harding et al., 2003), *Arabadopsis thaliana* (Dixon et al. 2003), and *Schizosaccharomyces pombe* (Kettner et al., 2004). It is best understood in the yeast species *S. cerevisiae*, where the crystal structure of Ero1p – the yeast homologue of Ero 1– has been solved (Tian *et al.* 2006). Due to similarities between the Ero 1 in *S. cerevisiae* and the Ero 1 inferred from the sequence annotation of *T. brucei* this species will be used primarily as the focus of this introduction to Ero 1.

1.3.12 Catalytic Properties

The catalytic action of Ero 1 is considered to be the main producer of oxidizing equivalence in the endoplasmic reticulum (Gross et al., 2006; Tu and Weissman, 2002; Tu et al., 2000), and the consensus view in the literature is that Ero 1 interacts with PDI in *S. cerevisiae.* It is thought that disulphide bonds in Ero 1 are formed by the reduction of molecular oxygen to hydrogen peroxide. This reaction occurs quite efficiently *in vitro* although the use of oxygen as the sole terminal electron acceptor is still under question as the overproduction of reactive oxygen species has obvious lethal consequences. As disulphide bond formation has not been shown to induce oxidative stress *in vivo* it is very likely that hydrogen peroxide is disposed of very quickly by some molecular mechanism or electrons are accepted via some other mechanism and hydrogen peroxide is not the terminal electron acceptor. Both hypotheses have yet to be investigated conclusively *in vivo* (Hatahet and Ruddock

2009). Other electron acceptors have been used *in vitro* such as flavins and certain exogenous redox metalloproteins for example cytochrome c and cytochrome b5 (Gross *et al.* 2006). Electrons are accepted by Ero 1 from a soluble disulfide-carrier of the PDI family via the formation of mixed disulphides (Frand and Kaiser 1999) as previously described. It has been shown that Ero1p shows a preference for the oxidation of the a' domain active site of Pdi1p *in vitro* (Kulp et al., 2006). This preference was also the case for Ero 1α-PDI interactions in humans (Tu and Weissman 2002). It is thought that *in vivo* PDIs then accept electrons from their substrates. The catalytic study of PDI and Ero 1 interacting to oxidize protein substrates was first characterized in *S. cerevisiae in vitro* (Tu *et al.* 2000).

It has been hypothesized that Ero1p contains a mechanism to regulate the reduction of the terminal electron acceptor. The regulation of the activity of Ero 1 is likely to be highly controlled as the over production of reactive oxygen species is potentially harmful. It has been proposed that the GSSG / GSH ratio plays a role in the regulation of Ero 1 and quite possibly functions as a store of disulphide bonds for later use by PDI, by being directly oxidized by Ero 1 (Hatahet and Ruddock, 2009) but as there is very little *in vivo* data to support this theory.

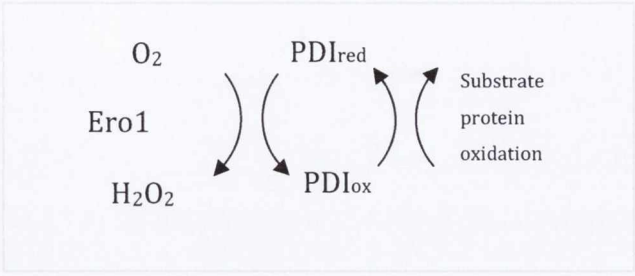


Figure 1.12 Schematic representation of the catalytic role of Ero 1

The enzyme Ero 1 reduces molecular oxygen to hydrogen peroxide and accepts electrons from reduced PDI. The enzyme PDI is then ready to oxidize the substrate protein. This has been shown in yeast *in vitro*. (Kulp *et al.* 2006)

If reduced glutathione and PDI levels are high, a conformational change in Ero1p results in an increase of activity (Figure 1.13). The levels of oxidized glutathione and PDI are thus restored. Conversely, if oxidized glutathione and PDI reach a level above normal, Ero1p undergoes a conformational change to the inactive form of the enzyme which slows the reduction of the terminal acceptor. It has been shown that glutathione is not required for cell viability but it is thought that without glutathione endoplasmic reticulum homeostasis is more unstable. This mechanism has been hypothesized base on in vitro experiments and the crystal structures of Ero1p (Sevier et al., 2007).

Figure 1.13 Schematic diagram of regulation of Ero1p

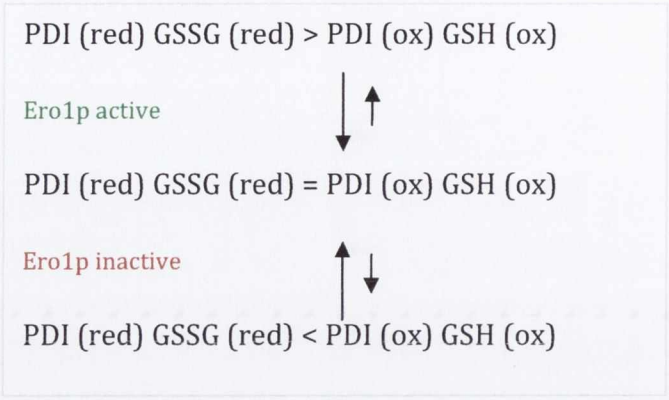


Figure 1.13 Schematic diagram of regulation of Ero1p

If PDI and glutathione are in the reduced state above normal levels, then the activity of Ero1p increases. This increase of activity allows Ero1p to return PDI and glutathione to a normal ratio of reduced to oxidized. Conversely, if the level of oxidized PDI and glutathione increases above normal levels, then the activity of Ero1p falls. This decrease of activity allows the level of oxidized PDI and glutathione to return to normal levels through the oxidation of substrate proteins.

1.3.13 Structure of Ero 1

The yeast Ero 1 protein (Ero1p) is a cysteine-rich enzyme which is peripherally associated with the membrane of the endoplasmic reticulum (Gross et al., 2004). Only one Ero 1-like protein has been found in the genome of *S. cerevisiae* as the case in the genome of *T. brucei* (Chapter 3). However, mammals possess two distinct Ero 1 family members: ERO1α and ERO1β, with the latter being a tissue specific protein found in the pancreas (Zito *et al.* 2010). An additional difference observed between mammalian and yeast Ero 1 proteins are their association with the membrane of the endoplasmic reticulum. Mammalian Ero 1 proteins have been shown to be luminal proteins of the endoplasmic reticulum whereas the yeast Ero1p has been shown to be a peripherally associated with the membrane of the endoplasmic reticulum through the use of a *C*-terminal region (Pagani *et. al.* 2001).

A temperature sensitive cell line of *S. cerevisiae* was found to produce less Ero1p when incubated at 38 ^{0}C. Analysis of this cell line suggested that a lower expression of Ero1p caused the accumulation of misfolded proteins in the endoplasmic reticulum. However, as other defects may have existed in the genome of this cell line and so it was not certain that a lack of Ero1p expression was responsible for the accumulation of misfolded proteins (Frand and Kaiser 1998; Pollard *et al.* 1998).

The structure of Ero 1p consists of a single domain which contains 10α helices, with the first and last helix packed side by side. Five β stretches are present, 3 of which form a β-sheet in the center of the protein. In the sequence of Ero1p there are 11 cysteine residues, forming 5 possible pairs for the formation of disulfide bonds. Of these 5 pairs there exist 3 conserved pairs of cysteine residues, which was revealed through a multiple sequence alignment, with distinct functions (Frand and Kaiser 2000). The cysteine pairs located at 353 and 355 have been shown to form the active site. It has been shown that at this active site molecular oxygen is reduced to hydrogen peroxide. This reaction has been shown *in vitro* but has yet to be shown *in vivo* (Hatahet and Ruddock 2009). It is at this active site where flavin adenine dinucleotide (FAD) is bound in between helices α2 and α3.

The cysteine pair formed from the residues at 100 and 105 is positioned on a flexible loop, which was deduced from the variation observed from different crystal forms solved. These cysteines are believed to accept electrons from the protein target of

Ero 1p, the PDIs. The cysteine pair at 90 and 349 tethers this flexible loop or releases it under certain oxidizing / reducing states of the endoplasmic reticulum. It is believed that these are regulatory cysteine groups which allow transfer of electrons from the active site positioned at the cysteine pair at 353 and 355 or slow it, thus maintaining endoplasmic reticulum homeostasis (Sevier *et al.* 2007).

Figure 1.14 Three dimensional structure of Ero1p

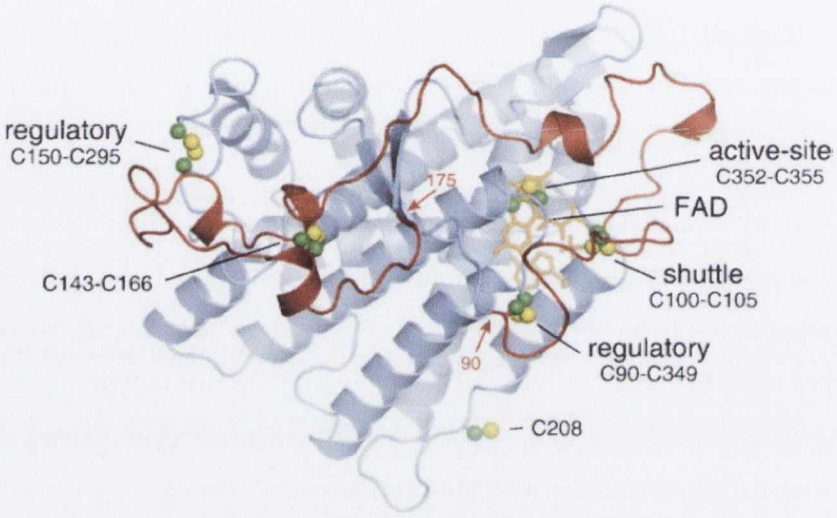


Figure 1.15 Three dimensional structure of Ero1p

Cysteine side chains are represented as green and yellow spheres and the FAD cofactor is drawn as orange sticks. The polypeptide domain (residues 90–175) containing the shuttle cysteine pair is coloured red. Regulatory cysteine residues can be seen on the red region (Cys90 and Cys150) and on the helical core (Cys295 and Cys349). Diagram taken from Sevier et al. 2007.

1.4 Reverse genetics

The approach to understanding molecular mechanisms of biological systems known as *reverse genetics* proceeds by first identifying a gene of interest, obtained by DNA sequencing and through manipulation of the expression of that gene reveal the phenotype which arises as a result of that gene. Manipulation of expression of a gene is achieved through different techniques. The techniques to be used in this project will be discussed in this section.

1.4.1 RNA interference

The mechanism of RNA interference (RNAi) is defined as the mechanism through which gene-specific, double-stranded RNA (dsRNA) triggers degradation of homologous transcripts. This effect knocks down the gene in question at the post transcriptional level. Understood only recently, the phenomenon of RNAi was first discovered by accident when trying to produce a petunia with petals of a deeper purple. An extra gene encoding the flower's pigmentation was introduced with startling results; the petals of the flower contained no pigmentation at all. Both introduced and endogenous genes had been "turned off" (Napoli et al., 1990). In later studies, antisense RNA of particular genes of interest was introduced into the nematode *Caenorhabditis elegans* (Guo and Kemphues, 1995; Montgomery and Fire, 1998). This transfection resulted in the same effect as above, knocking the gene down. It was subsequently discovered to be caused by the formation of double stranded RNA, which was later reported by Fire et.al (Fire *et al.* 1998) earning both Fire and Mello the 2006 Nobel Prize in Physiology or Medicine. The use of RNAi to manipulate gene expression provides a powerful tool for the study of species which contain this pathway. The ability to reduce dramatically the expression of specific proteins in cell culture and *in vivo* has significant applications, especially in those cases when classical genetics cannot be applied.

1.4.2 Mechanism of RNA interference

In eukaryotic organisms it was originally thought that the mechanism of RNAi exists to stop the production of viral RNA and other parasitic genes of the cell. It was subsequently shown to be involved not only as a defense mechanism but also as a mechanism of control for the expression of proteins in general. The enzyme Dicer, an endoribonuclease of the RNase III family, recognizes long double stranded RNA which is then cleaved into 20-25 nucleotide fragments (Bernstein et al., 2001). These small interfering RNA (siRNA) molecules contain a 2 base overhang at the 3' end of the fragment (Castanotto and Rossi, 2009; Siomi and Siomi, 2009; Vermeulen, 2005; Zamore et al., 2000). The siRNAs are separated into single stranded fragments, one of which is known as the guide strand. This strand is incorporated into the RNA-induced silencing complex (RISC) (Ahlquist, 2002). Once the RISC is formed, target mRNAs are sought out. This action is facilitated by the argonaute protein a component of RISC which displays the 5' region of the RNA so that it may optimally base pare with complimentary RNA strand (Bohmert et al., 1998; Montgomery and Fire, 1998; Tabara et al., 1999). Once an mRNA is recognized by the RISC one of two outcomes may occur. If the base pairing is extensive argonaute cleaves the mRNA and it is degraded by exonucleases. If the base pairing is less extensive the mRNA is not cleaved by argonaute but translation is repressed and the mRNA is destabilized. The poly-A tail of the mRNA is shortened and the mRNA are processed in cytosolic processing bodies (P-bodies) where the mRNAs are sequestered from ribosomes and eventually degraded (Brennecke et al., 2005). Although other forms of RNAi exist including micro RNAs which are present in animals and plants as a form of post transcriptional gene control, this project is concerned only with the use of the pathway described above and so other details of microRNAs will not be discussed. The mechanism of RNAi in *T. brucei* has yet to be fully understood but the following has been shown in the literature (Figure 1.15). The identification of an argonaute gene encoding the protein now known as TbAgo1 was straight forward and was discovered using database mining techniques (Durand-Dubief and Bastin, 2003; Shi et al., 2004) but the discovery of a Dicer protein lagged behind as no such gene could be found using the same techniques. Later a Dicer protein was discovered through identification of RNaseIII like domains within possible Dicer like genes. These possible candidates were subjected to further bioinformatical analysis, localization

studies and RNAi knockdown to determine possible functions to identify the Dicer protein now known as TbDcl1 (Shi et al., 2006). Both proteins were found to be predominantly cytoplasmic (Shi *et al.* 2004; Shi *et al.* 2006).

1.4.3 RNAi in *Trypanosoma brucei*

The use of RNAi in *T. brucei* is well established. It was first reported in 1998 (Ngô et al., 1998) when dsRNA was used to knockdown α-tubulin resulting in the so called FAT cell phenotype (rounding of cells). These dsRNAs were generated *in vitro* and introduced into the cell. The modern and more effective approach of using heritable and inducible RNAi has been used extensively since then, making RNAi easier to achieve.

The vector used in this study is the p2T7-177 (Wickstead et al., 2002), which has two tetracycline-inducible promoters arranged as an inverted repeat to allow transcription from both strands of DNA. The p2T7-177 plasmid contains several repeats of the 177 bp repeats of the transcriptionally silent minichromosomes to allow integration of the construct to this region following linearization by *Not I* (Wickstead *et al.* 2002). This region is not transcribed until the activation of the tetracycline operator which allows transcription by the T7 polymerase. The vector also contained a resistance gene to allow selection of successfully transfected cells with the drug Phleomycin.

Cell lines used in this study contained the tetracycline operator and T7 polymerase genes stably transfected into their genome and maintained by selection drugs. In the case of the bloodstream form parental cell line, strain 328.114 was used which maintained selection with G418 disulphate (Wirtz and Clayton, 1995). In the case of procyclic form cells, strain 29.13 cell line, which also expresses the operator and T7 polymerase, was used which was maintained using both G418 disulphate and hygromycin (Wirtz and Clayton 1995; Wirtz *et al.* 1999).

To date many *T. brucei* proteins have been subjected to knockdown by RNAi revealing a wide range of phenotypes. Two PDIs, TbPDI-1 and 2 have been subjected to RNAi but knockdown of these activities had no effect on the growth or morphology of bloodstream forms in culture. Both individual and combined RNAi of TbPDI-1 and 2

showed only a slight growth phenotype. Although both Northern and Western blots reveal a significant loss of the transcript and protein respectively in the induced, it was not possible to calculate the precise level of knockdown using these techniques.

Typically, when quantitative measurements have been made at the protein level, for example by assaying the enzyme activity, knockdown by RNAi usually results in loss of 50 - 90% of the target activity. In the case of *T. brucei* there is a good correlation between loss of mRNA and level of activity remaining with 24 - 48 hours post induction. For example, in a recently published study of the alanine aminotransferase of *T. brucei*, loss of 70% of RNA transcript resulted in a loss of ~70% activity (Spitznagel *et al.* 2009). However, in the case of many enzymes loss of up to 90% of the transcript and activity produced no obvious phenotype. In these situations it is not possible to conclude whether there is an essential requirement or not for the activity as frequently residual levels of activity of less than 10% are sufficient to support normal function and growth. Interestingly in the case of alanine aminotransferase phenotypic effects were observed only when the activity was reduced to less than 5% of control levels.

However, it is important to note that both TbPDI-1 and 2 could not be quantified by assaying their activity. No such assay exists for PDIs in general. Instead the level of knockdown must be quantified by the use of other techniques for example quantitative real time PCR. This technique can be used to quantify the relative level of transcript RNA of a gene of interest between two cell lines. This technique is used throughout this study to compare the expression of target genes between cell lines quantitatively.

Trypanosomes have been known to 'escape' from the effects of inducible constructs. After a period of several days or weeks of cells growing in culture, recombination of the inverted repeats which flank the gene of interest may damage the double stranded RNA system *in situ* (Ullu et al., 2004). This escape causes the trypanosome to behave as if no induction has taken place but still maintains resistance to selection antibiotics (Chen et al., 2003).

Figure 1.15 A schematic representation of the RNAi pathway in *T.brucei*

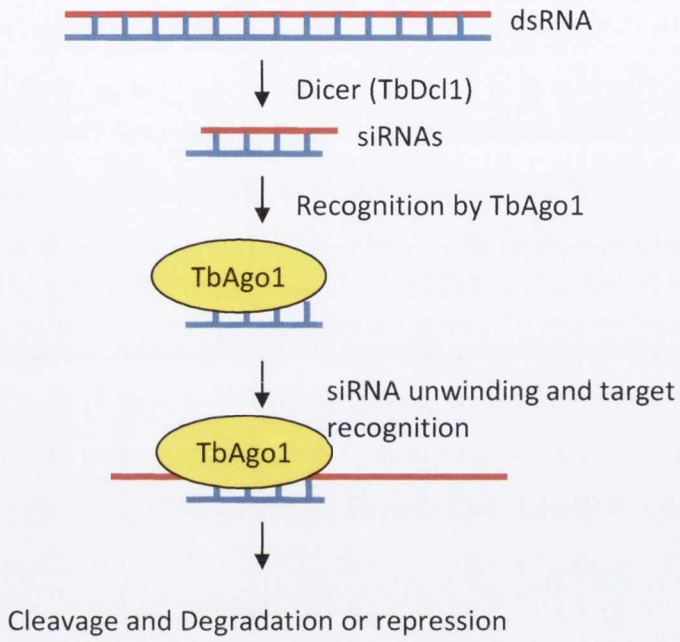


Figure 1.16 A schematic representation of the RNAi pathway in *T.brucei*

dsRNA present in the cytoplasm of the cell is recognized by the enzyme Dicer (TbDcl1) and cut into 20-25 nucleotide dsRNA complexes known as small interfering RNAs (siRNA). These then bind to the trypanosome argonaut protein TbAgo1 which then seeks out complimentary target mRNA using the siRNA as a template. The target is then cleaved and degraded or repressed. Adapted from Ullu et al. 2004.

1.4.4 Gene Knockout

The generation of null mutants in trypanosomes can be achieved through targeted homologous recombination, which occurs with high efficiency in trypanosomes. The target gene is disrupted/replaced with a resistance gene. In these knockout experiments, trypanosomes are transfected with selection markers flanked by regions immediately upstream/downstream of the target gene. Although generation of true null mutant cells is rare in *T. brucei*, this approach has been employed to generate null mutants of genes encoding glycosylphosphatidylinositol-specific phospholipase C (GPI-PLC) in procyclic form trypanosomes (Webb *et al.* 1994) and a gene known as Mre11 found in bloodstream form trypanosomes (Robinson, 2002). The enzyme GPI-PLC stimulates endocytosis of transferrin in bloodstream (Subramanya et al., 2009) and has been shown to cleave GPI at the endoplasmic reticulum (Subramanya and Mensa-Wilmot, 2006). The protein Mre11 was found to have a critical role in maintaining genomic integrity and a growth phenotype was observed as a result of the gene knockout. Successful generation of a null mutant implies that the translated gene product is non-essential in the life cycle stage under consideration. However, it does not necessarily follow that the gene is not essential in other life cycle stages or *in vivo*. However, in the case of the GPI-PLC this has been found to be the case. Recently, the combination of a single gene knockout approach to generate a heterozygote (+/-) cell line and performance of RNAi in the cell line has proved a useful approach. This improved knockdown was used to generate a knockdown of the aforementioned alanine aminotransferase to less than 5% of control levels (Spitznagel *et al.* 2009).

If the enzyme cannot be knocked out as it is essential in culture, an ectopic copy of the gene may be inserted into the genome. This insertion can be achieved using an inducible expression vector, for example p1791, which integrates into the non-transcribed spacer of the *rDNA* locus (Wirtz *et al.* 1999). This vector once transfected incorporates the phleomycin resistance gene which is transcribed by the T7 polymerase in the conditional cell line. As is the case with the RNAi vector, this gene is under the control of the tetracycline-inducible promoter and expression of the target gene is induced in the presence of the drug tetracycline. Once induced, the ectopic copy of the target gene will be transcribed, driven by the EP procyclic promoter using endogenous Pol II. This ectopic gene transcription allows the knockout of the second

gene locus within the diploid genome. Subsequently, removal of the tetracycline in principle leads to silencing of expression of the rescue copy. This technique has been reported in the literature conclusively showing the essential nature of the genes investigated (Estévez et al., 1999; Martin and Smith, 2006a; Shi et al., 2007). There are many advantages to inserting inducible genes into transcriptionally silent regions of the trypanosome genome but a disadvantage which may be significant is the phenomenon of leaky transcription. Although the regions chosen are transcriptionally silent it has been reported that some transcription may take place which may have consequences for use during conditional gene knockouts as the production of even small amounts of target gene may be sufficient to rescue the null mutant (Wirtz *et al.* 1999; Wickstead *et al.* 2002). Despite these disadvantages gene knockdown / knockout strategies remain the tools of choice for the study of trypanosomes.

Aims

In silico analysis

1. To screen the genome of *T. brucei* to identify potential PDIs and Ero genes and identify the possible major oxidative pathway of the endoplasmic reticulum.

In vivo analysis

2. To assess the relative expression of these genes in bloodstream versus procyclic form cells using qRT-PCR to identify the PDIs and Eros most important enzymes in the role of VSG folding.

3. To generate a null mutant cell line of TbPDI-2 by replacement of both gene copies with selection markers. Also introduce an ectopic copy of TbPDI-2 to aid in the generation of a null mutant. This was performed in both bloodstream form and procyclic form cell lines.

4. To generate a combination of single gene knockout and RNAi to enhance knockdown of TbPDI-2 and to analyze any phenotype observed to ultimately establish a link between TbPDI-2 and VSG production. This was performed in both bloodstream form and procyclic form cell lines.

5. To generate a TbEro 1 RNAi cell line to determine if this protein is essential or required for growth in culture and to analyze any phenotype observed again to establish a link between TbEro 1 and VSG production. This was performed in both bloodstream form and procyclic form cell lines.

6. To tag TbEro 1 to characterize the protein using immunnolocalization, immunoprecipetation studies and Western blot to establish a link between TbPDI-2 and TbEro 1's interaction *in vivo*.

In vitro analysis

7. To generate a recombinant TbEro 1 for further *in vitro* studies.

2 Materials and Methods

2.1 Preparation of reagents

All reagents described were weighed using a Mettler (midel KTT) top loading balance (Model R20). Solutions were made up using distilled deionised water unless otherwise stated. Gilson Pipetteman pipettes were used to dispense liquid volumes from 2 µl to 5 ml. The hydrogen ion concentration of solution was measured using an Amagrass glass electrode, which was connected to an EDT RE357 microprocessor pH meter. The hydrogen ion concentration of solutions was adjusted as necessary using HCl (12 M) and NaOH (5 M). Standard buffer solutions of pH 4, 7 and 10 from Sigma were used to calibrate the pH meter. A MSE-O-1582-A centrifuge, a Sorvall RC5B refrigeration centrifuge and a bench Hermle BHG microspin centrifuge were used in centrifugation of samples.

2.2 Agarose Gel Electrophoresis of DNA

Agarose powder (1 % w/v) was mixed with TAE electrophoresis buffer (Tris acetate (40 mM), EDTA (2 mM), pH 8), and heated in a microwave until completely melted. The solution was allowed to cool to ~ 50 °C and poured into a casting tray containing a sample comb to give a well depth of ~ 5/6 mm. The gel was allowed to solidify at room temperature. The comb was removed and the gel placed horizontally into an electrophoresis chamber (Medical Supply Company). TAE was added until just covering the gel. Samples for loading to the gel were prepared in loading buffer (20 % Ficoll 400, 0.1 M EDTA pH 8, l % SDS, 0.025 % bromophenol blue). After loading the gel was allowed to run at 100 V for ~ l h. The electrophoresed gel was placed in a solution of ethidium bromide (0.5 ug/ml) in TAE and left to soak for 10 min to stain the DNA. The gel was then soaked in TAE for a further 10 min to remove background stain. Stained samples were visualised using a UV transilluminator (MSC). DNA to be removed was cut using a scalpel and kept in a minifuge tube on ice for later purification.

2.3 Polymerase Chain Reaction

Primers were designed with standard criteria using sequence data from target DNA to be amplified. Restriction enzyme sites were incorporated into the ends of each primer. The majority of PCR reactions were carried out using the proofreading polymerase Herculase II supplied by Agilent Technologies. PCRs carried out to confirm correct transfection into trypanosome genomic DNA were carried out using GoTq DNA polymerase supplied by Promega. Polymerase chain reaction (PCR) mixtures were made up in thin wall PCR tubes (Molecular bioproducts) with components as described in the supplied protocols. After all additions had been made, samples were centrifuged briefly then placed in a PCR Sprint machine (MSC). PCR products were confirmed by gel electrophoresis.

2.4 Purification of PCR products

PCR products were purified by spin column purification using the 'peqGold miniprep kit I' from Peqlab.

2.5 T-vector ligation

All PCR products which were to be digested were first ligated into pGEM®-T easy vector using the easy vector system from Promega. In the case of all PCR products amplified by proofreading polymerases, an additional incubation with a Taq polymerase was required for the addition of the polyA tails. These additions were required for the ligation to the T-Vector. PCR products were processed using a PCR-cleanup kit and the GoTaq polymerase kit was set up using supplied protocols with the following exceptions: no primers were added, additional template was added and the mixture was incubated for 10 minutes at 72°C. PCR products were cleaned up as before and supplied protocols were followed for the ligation to the T-Vector.

2.6 Restriction enzyme digestion of DNA

DNA samples were singly or doubly digested in sterile 500ul minifuge tubes (Starstedt) using 1-10 U of each restriction enzyme in appropriate buffer (1/10

dilution 10 X stock supplied with the enzyme) for 3 h or overnight at 37 °C to ensure complete digestion.

2.7 Gel purification of DNA

Restriction enzyme digested vector and PCR products in pGEM-T vector were gel purified to remove unwanted uncut DNA and restriction enzymes. Low melt agarose (1 % w/v) was used to make the gel as previously described, and the gel was electrophoresed at 4°C. After staining with ethidium bromide (0.5 ug/ml) both vector and insert bands were excised from the gel using a sterile scalpel and placed in separate minifuge tubes. The DNA was then extracted from the agarose using the Peqlab gel purification system.

2.8 Ligation of digested PCR product into digested vector

Ligation of digested DNA was carried out using the 2xrapid DNA ligation system by promega. Supplied protocols were followed.

2.9 Preparation of competent *Escherichia coli* DH5α

DH5α *E. coli* were grown on antibiotic free Luria-Bertani (LB) agar (10% w/v tryptone, 5 % w/v yeast extract, 5% w/v NaCl, l0 % w/v agar, 70% w/v dH20, autoclaved as standard) overnight at 37 °C in a bacterial incubator. LB broth (10 % w/v tryptone, 5 % w/v yeast, l0 % w/v NaCl, 75 % v/v dH20, autoclaved as standard) was inoculated with resulting colonies and incubated at 18 °C, 200 rpm in a rotary shaker (New Brunswick Scientific). Samples were taken periodically from cultures and culture was monitored at 600 nm in a spectrophotometer until readings were between an optical density of 0.4 — 0.6. At this point cells were centrifuged at 3000 rpm, 4 °C for 15min. The supernatant was discarded and cells were resuspended in 1/3 volume of transformation buffer (10 mM Pipes, 15 mM $CaCl_2$, 250 mM KCl pH 6.8, 55 mM $MgCl_2$ filter sterilised) at 0 °C. Cells were gently resuspended at 4 °C and incubated for 10 min at 0 °C. Cells were again centrifuged and resuspended in 1/12.5 volume of transformation buffer at 0 °C gently. DMSO was added to a final concentration of 7 %, the suspension was mixed, and incubated for a further 10 min

at 0 °C. Cells were distributed in 1 ml aliquots in sterile minifuge tubes and stored at -80°C for several months or until use.

2.10 Transformation of competent bacterial cells

Competent cells were thawed on ice for 10 mins. Following this period 100 ng of plasmid DNA was added to 50 - 100 µl competent cells and left on ice for 30 min. The cells were heat shocked for 2 min at 42 °C, after which they were incubated at 0 °C for 2 min. Cells were transferred into 1 ml LB broth at 37 °C and incubated at 37 °C for 1 hour 200 rpm in a rotary shaker. After the incubation period, 100 µl was taken and aseptically spread on sterile LB agar plates containing ampicillin (50 µg/ml). Plates were incubated overnight at 37 °C in a bacterial incubator.

2.11 Selection of transformed colonies.

In order to select bacterial colonies containing vector into which desired DNA fragments had been ligated, random colonies were selected from those which had grown on LB Amp + plates. Colonies were transferred to 5 ml LB broth containing ampicillin (50 µg/ml) and incubated at 37 °C overnight at 150-250 rpm in a rotary shaker. DNA was isolated from cell cultures (section 2.34) and subjected to enzymatic digestion of restriction enzyme sites at which DNA inserts were ligated. Digested DNA was loaded onto 1 % agarose gels to check for the presence of the insert.

2.12 Generation of glycerol stocks.

Selected transformant cells (800 µl in LB broth) were mixed with a final concentration of 80 % glycerol and stored in sterile cryotubes (Greiner bio-one) at -80° C.

2.13 Purification of plasmid DNA.

Plasmid DNA was purified using two kits: peqGOLD Plasmid Miniprep Kit I and Pure Yield ™ Plasmid Midiprep System from Promega.

2.14 Transfection of bloodstream form *Trypanosoma brucei* using the AMAXA Nucleofector® apparatus

Bloodstream form cells to be transfected were grown in HMI9 including appropriate selection drugs and harvested (3 x 10^7 cells) by centrifugation at 1500 rpm for 10 minutes at room temperature. HMI9 medium was aspirated off and the pellet was resuspended in 100 µl of Amaxa Human T-cell solution (4^0C). The DNA to be transfected (5µg of DNA in 5µl) was added to an Amaxa cuvette and the cell suspension was added. The cells were immediately transfected in the Amaxa electroporator. The cell-DNA mixture was added to pre-warmed HMI9 medium (30 ml). This suspension was gently mixed and cells were incubated at 37°C for 6 hours. After this recovery time the appropriate selection drugs were added. The cells were monitored over a period of 24 hours. Once the majority of cells had died the remaining cells were divided into two flasks; the first flask (flask A) contained 27 ml and the second flask (flask B) contained 3 ml of the cell suspension. The volume of media in both flasks was brought to 36 ml with the addition of warmed HMI9 and additional selection antibiotics were added. The two flasks were plated out into two 24 well plates and incubated over a period of 5-8 days, until positive wells were observed. Positive cells were diluted further into a 24 well plate to ensure the population of cells was derived from a single clone.

2.15 Transfection of procyclic form *Trypanosoma brucei* using the AMAXA Nucleofector® apparatus

Procyclic form cells to be transfected were grown in SDM79 including appropriate selection drugs and harvested (3 x 10^7 cells) by centrifugation at 1500 rpm for 10 minutes at room temperature. SDM79 medium was aspirated off and the pellet was resuspended in 100 µl of Amaxa Human T-cell solution (4^0C). The DNA to be transfected (5µg of DNA in 5µl) was added to an Amaxa cuvette and the cell suspension was added. The cells were immediately transfected in the Amaxa electroporator. The cell-DNA mixture was added to pre-warmed SDM79 medium (30 ml). This suspension was gently mixed and cells were incubated at 27°C for 6 hours. After this recovery time the appropriate selection drugs were added. The cells were monitored over a period of 24 hours. Once the majority of cells had died the

remaining cells were divided into two flasks; the first flask (flask A) contained 27 ml and the second flask (flask B) contained 3 ml of the cell suspension. The volume of media in both flasks was brought to 36 ml with the addition of warmed SDM79 and additional selection antibiotics were added. The two flasks were plated out into two 24 well plates and incubated over a period of 5-12 days, until positive wells were observed.

2.16 Cell counting

Cells were counted and density was calculated using an improved Neubaur haemocytometer with a silvered stage (Gelman-Hawksley, Lancing, Sussex, U.K.) and a light microscope (Olympus CK2).

2.17 Cell culture of bloodstream form cells

The trypanosomes employed for RNAi , single knockout and conditional knockout experiments were monomorphic bloodstream trypomastigotes forms of *Trypanosoma brucei brucei*. This modified cell line expressing MITat (Molteno Institute Trypanozoon antigen type) 1.2 VSG, from strain 328.114 (Wirtz et al., 1999) also expresses a T7 polymerase and tetracycline operator which is maintained by selection drugs (G418 2.5µg/ml). Cultured trypanosomes were grown in HMI-9 medium with 10% heat-inactivated foetal calf serum (FCS) (Hirumi and Hirumi, 1989) and selection drugs. Cells were grown at 37 °C, 5 % CO_2 in a water-jacket CO_2 incubator (Thermo Forma). T-25 and T-75 flasks (Greiner bio-one) were used for the culture of cells. Synchronous growth of cells was achieved when the density of cells in the culture flasks was maintained below 1 x 10^6 cells/ml. Cell growth was monitored for a period of more than 10 days. Growth curves for each clone were constructed based on the mean value of three growth curves. A straight trend line was calculated for each cell line and the doubling time was calculated from the trend line using the equation seen in Equation 2.1.

$$T_d = (t_2 - t_1) \times \frac{\log(2)}{\log(\frac{q_2}{q_1})}$$

Equation 2.1 Equation used to examine the doubling time of cell growth

The equation above was employed to calculate the doubling time of cell growth. Assuming a constant rate of growth, the doubling time (T_d) was calculated using cell count 1 (q_1) taken at time 1 (t_1) and cell count 2 (q_2) taken at time 2 (t_2). (Mather and Roberts, 1998).

2.18 Detergent lysis of cultured cells

Cultured cells were harvested by centrifugation at 800 g / 5 min for bloodstream form cells, or 1, 000 g / 5 min for procyclic form cells. Cell pellets were washed once in PSG buffer and resuspended in PSG to give a cell concentration of 2 x 10^8 cells/ml (4 x 10^6 cells / 20µl). SDS gel samples were prepared by mixing cells with an equal volume of sample buffer [Tris 0.5 M pH 6.8, 10 % glycerol, 0.02 % bromophenol blue in ethanol, 2.5 % (w/v) SDS, 2-mercaptoethanol (50 ul/ml)]. All samples were boiled for 5 min prior to loading on to the gel. The minimum loading per lane was 5 x 10^6 cells.

2.19 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed using a modification of the method of Laemmli (1970). Gel components are given in table 2.1. Ammonium persulphate was prepared immediately prior to use. The separating gel was loaded and left to polymerise with a layer of water saturated butanol on top (60 % dH2O:40 % butanol). This was removed before the stacking gel was loaded. The stacking gel remained the same irrespective of the composition of the separating gel. Samples for gel electrophoresis were prepared in an equal volume of sample buffer (0.5 M Tris buffer pH 6.8, 10 % glycerol, 0.02 % bromophenol blue in ethanol, 2.5 % (w/v) SDS, 2-mercaptoethanol (50 µl/ml)). All samples were boiled for 5 min prior to loading. A vertical slab apparatus (Atto minigel a system) based on the original design of Studier (1973), connected to a BIO-RAD power pack 3000 was used for electrophoresis. Running buffer contained glycine (300 mM), Tris base(80 mM), and SDS (0.1 %). Gels were run at a constant current of 25 mA until the dye front reached the base of the gel.

2.20 Semi-dry Western Blotting

Transfer of protein from SDS polyacrylamide gels to either Polyvinylidene (Difluoride (PVDF) or nitrocellulose membrane was carried out according to the method of Matsudaira (1987). The electrophoresed gel, PVDF/nitrocellulose membranes (pre-soaked in 100 % methanol for PVDF) and sections of chromatography paper cut to size were soaked in transfer buffer (CAPS 10 mM, 10 % methanol, SDS 1.32 mM, pH 11) for 10 min. The gel was overlaid onto the PVDF/nitrocellulose sheet ensuring no

air bubbles were present. Pre-soaked chromatography paper was placed on either side of the gel-PVDF/ nitrocellulose layer (6 sheets on each side). The entire assembly was placed in a Hoeter semi-dry blotter and proteins were transferred to the membrane at 100 mAmp for 1 h. Following transfer, proteins present on the membrane were viewed by staining with Ponceau Red. Molecular weight markers were marked on the membrane and it was washed (4 x 5 min) with dH20, to remove the stain.

Component	12.5 % Separating Gel	Component	5 % Stacking Gel.
30% acrylamide	12.5 ml	30% acrylamide	1.67 ml
1 % bis-acrylamide	3.1 ml	1 % bis-acrylamide	1.3 ml
1.5 M Tris pH 8.7	7.5 ml	1 M Tris pH 6.8	1.25 ml
dH_2O	8 ml	dH_2O	4.4 ml
20% SDS	150 µl	20% SDS	50 µl
10 % Ammonium persulphate solution.	100 µl	10 % Ammonium persulphate solution.	50 µl
TEMED	10 µl	TEMED	5 µl

Table 2.1 Components of SDS-PAGE gel

The components used to product the 12.5 % SDS gels use throughout this study.

2.21 Generation of Anti-TbEro 1 antibody

A specific peptide A: CRAWGKLETAALAT from TbEro 1 was synthesized and coupled, via the *N*-terminal cysteine residue, to a carrier protein (KLH) and used to immunize rabbits. At various intervals the immune response was determined by an ELISA using the peptide antigen. A peptide affinity resin was generated by attaching the peptide to Sulpho-Link resin (Pierce) via the *N*-terminal cysteine after treatment of the peptide with immobilized TCEP disulfide reducing gel (Pierce). The resin (approximately 1 ml) was placed into a column and washed with 10 volumes of column buffer (Tris-Cl 20 mM, NaCl 100 mM, pH 7.5). Rabbit immune serum was dialysed at 4°C overnight against column buffer and then cycled over the affinity resin in a closed loop at a flow rate of 0.2 ml /min overnight. The following day the column was washed with coupling buffer to remove nonspecific antibodies. Specifically bound antibodies were

eluted by washing the resin with 10 column volumes of glycine (100 mM, pH 2.8) followed by 10 volumes of coupling buffer. Fractions (1.8 ml) were collected into tubes containing 0.2 ml of Tris-Cl (1 M, pH 7.5). Those fractions containing antibodies eluted from the resin at the low pH step were dialysed against Hepes buffer (10 mM, NaCl 150 mM, pH 7.5). The antibodies were stored at -80°C.

2.22 Immunoblotting

The PVDF/nitrocellulose membrane was incubated in blocking buffer (5 % (w/v) marvel, 0.05 % (w/v) Tween 20, in TBS) overnight at 4 °C or at room temperature for 1 h 30 min, with gentle agitation (Stuart Scientific rocker). The membrane was washed (3 x 10 min) in TBS, 0.05 % Tween 20 (TBST), before incubation with primary antibody in 3 % (w/v) marvel/TB ST at room temperature for 1h with gentle agitation. The membrane was washed (3 x 10 min) and incubated with an alkaline phosphatase conjugated secondary antibody in 3 % (w/v) marvel/TBST at room temperature for 1 h with gentle agitation. The PVDF/nitrocellulose membrane was washed again (3 x 15 min). Secondary antibody which had bound to the membrane was detected using the 5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium (BCIP/NBT) substrate solution (Roche). The membrane was incubated in this solution with gentle agitation until the colour had developed to the desired degree. The reaction was arrested by incubation in dH2O and the immunoblot was scanned using an Epson scanner to obtain a permanent record of the result. For antibodies used and their dilutions see table 2.2.

Antibody target protein	Animal	Dilution
TbPDI-2	Rabbit	1/5000
TbPDI-1	Mouse	1/5000
TbEro 1	Rabbit	1/5000
HA	Mouse	1/5000
Ectophosphatse	Rabbit	1/5000
Actin	Rabbit	1/5000
BiP	Rat	1/5000

Table 2.2 Antibodies used for immunoblotting

The antibodies and the concentrations used for Western blotting throughout this study.

Name of primer	Sequence	Purpose
TbEro 1 qRT-PCR FP 1	CCG CTC GAG TAA TGA TAC GGC TGG TGT AGA AAA CAG	TbEro 1 RNAi
TbEro 1 qRT-PCR RP 1	AAG CTT CCA CCA CAT TCT CAG CGG AAC	TbEro 1 RNAi
TbEro 1 qRT-PCR FP 2	GGA CTA GTA ATG ATA CGG CTG GTG TAG AAA ACA G	TbEro 1 RNAi
TbEro 1 qRT-PCR RP 2	CGG GAT CCA CCA CAT TCT CAG CGG AAC	TbEro 1 RNAi
TbPDI-2 p1791 FP	GCG AAG CTT ATG CGC GCT ATT TTT CTT GTT GC	TbPDI-2 expression construct
TbPDI-2 p1791 RP	CGG GAT CCC TAC AGA TCT TGT TTA TCA ACA TTG CTC	TbPDI-2 expression construct
TbEro 1 p2620 FP	AAG CTT ATG TTG AAG ATG CGA TTA TTG ATA GTA GTT CCT	TbEro 1 expression construct
TbEro 1 p2620 RP	GGA TCC TCA ACT CAT AGC CGT GTA ATT GAA TTT TTC G	TbEro 1 expression construct
TbEro 1 pGST-Parallel 3 FP	GGA TCC ATG TTG AAG ATG CGA TTA TTG ATA GTA GTT CCT	TbEro 1 recombinant expression construct
TbEro 1 pGST-Parallel 3 RP	GCA CTA GTG AAT TCT CAA CTC ATA GCC GTG TAA TTG AAT TTT TCG	TbEro 1 recombinant expression construct
TbPDI-2 ORF FP	GGA TCC ACA ACT TTA CCC	TbPDI-2 ORF and flanking regions
TbPDI-2 ORF RP	TCT AGA GGC GCG CAG CAC CCC C	TbPDI-2 ORF and flanking regions
TbPDI-2 FlankR FP	ACG CGT CGT GAC GCT TTT TAT TAA CTT	Upstream and downstream regions TbPDI-2
TbPDI-2 FlankR RP	ACT AGT CAA GAT TAA GAA AAC GTT TGG CCC	Upstream and downstream regions TbPDI-2
Blast FP	ACG CGT TGC CCA CCA TCA GCC ATC TCA TTG	Blasticidin R. gene
Blast RP	ACT AGT ATG CAG ATC GAG AAG CAC CTG TCG	Blasticidin R. gene
Puro FP	ACG CGT ATG ACC GAG TAC AAG CCC ACG GTG C	Puromycin R. gene
Puro RP	ACT AGT TCA GGC ACC GGG CTT GCG GGT CA	Puromycin R. gene

Table 2.3 List of cloning primers

The sequences of the primers and their purpose are shown.

2.23 RNA interference.

The knockdown of target genes was achieved using the p2T7-177 vectors (Figure 2.1). The p2T7 vectors were generated by Wickstead et al. (Wickstead *et al.* 2002). Each vector contained a selection marker for bacteria (ampicillin resistance) and one for transformed trypanosomes (phleomycin for p2T7-177 and hygromycin for p2T7-Ver3). Sections of the target genes were subcloned into the vectors then the vector was purified and linearized using the restriction enzyme *Not I*. The p2T7-177 vectors were stably inserted into the transcriptionally silent minichromosomes of *T. brucei*. Transfected cells showing resistance were transferred into 5mls of HMI-9 and allowed to grow. Growth was monitored and once in the log phase of growth, $2x10^6$ cells were frozen as described previously. The remaining cells were subcloned into two new flasks of 5mls of HMI-9 containing selection antibiotics. Tetracycline (1μg/ml) was added to one of the flasks to initiate RNAi. Growth was monitored over a period of longer than 10 days.

2.24 Generation of RNAi constructs

The RNAi of TbPDI-2 transcript was achieved using the p2T7-Ver3 RNAi vector and p2T7-177 vector (Figure 2.2). These vectors were previously made and provided by Joyce Rubotham. A 543 bp fragment from the TbPDI-2 open reading frame (annotated the BS2 gene) was subcloned into the p2T7-Ver3 RNAi vector and the p2T7-177 vector using the restriction enzymes *BamHI* and *XhoI*.

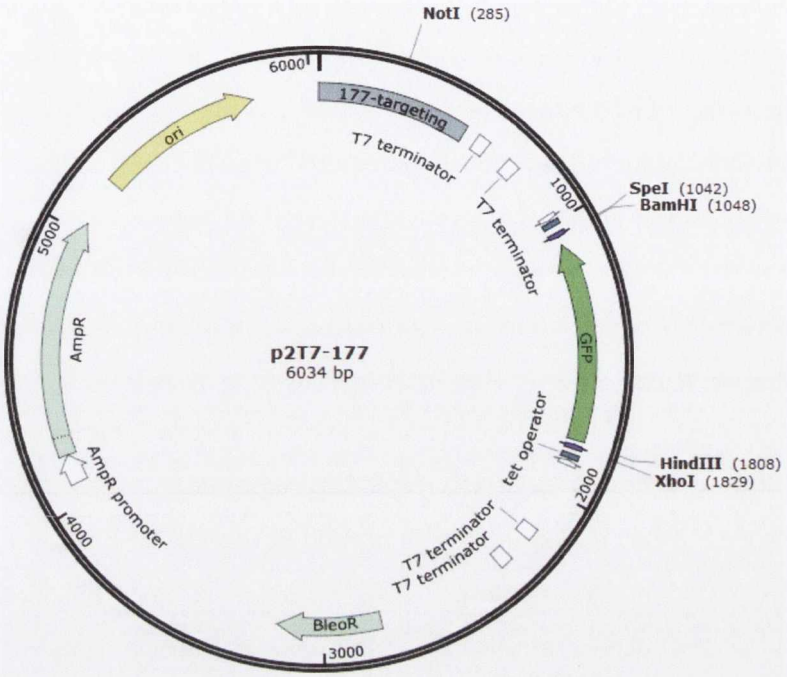
The RNAi of TbEro 1 transcript was achieved using the p2T7-177 RNAi vector (Figure 2.3). The vector was engineered to contain two complementary 594 bp fragments of the TbEro 1 ORF, which once transcribed formed a dsRNA stem loop. The dsRNA stem loop also contained a 717 bp GFP gene as a stuffer. To achieve this arrangement, a fragment of the TbEro1 ORF (Tb927.8.4890) was amplified first with a set primers containing *XhoI* (FP CCG CTC GAG TAA TGA TAC GGC TGG TGT AGA AAA CAG) and *HindIII* (RP AAG CTT CCA CCA CAT TCT CAG CGG AAC) and then with a second set of primers containing *BamHI* (FP GGA CTA GTA ATG ATA CGG CTG GTG TAG AAA ACA G) and *SpeI* (RP CGG GAT CCA CCA CAT TCT CAG CGG AAC). These fragments were subcloned into the p2T7-177 vector and the fragments were orientated in a complementary fashion due to the restriction sites chosen. A list of primers used can be seen in Table 2.3.

Figure 2.1 Schematic diagram of RNAi vectors p2T7-177 and p2T7-Ver3

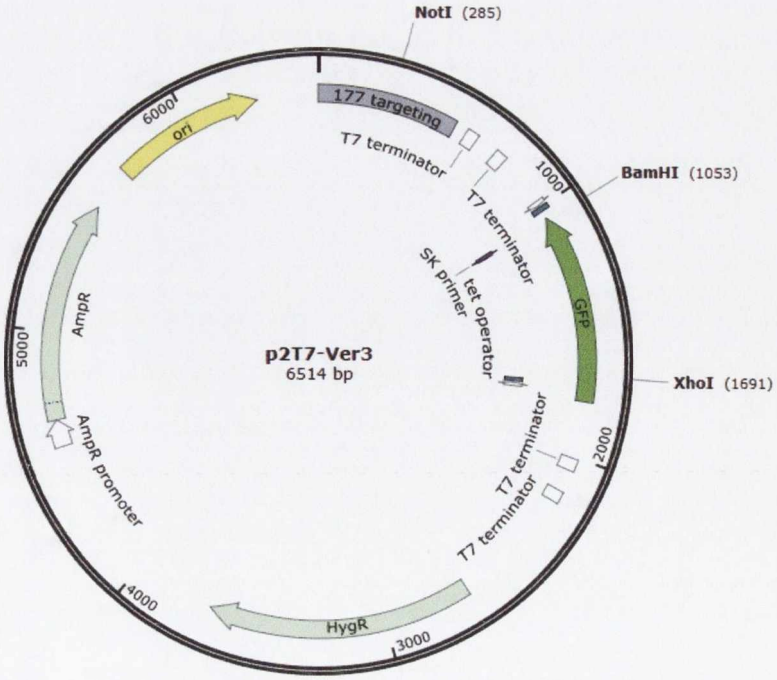
The p2T7 vectors were generated by Wickstead *et al.* (Wickstead *et al.* 2002). The vectors contained multiple cloning sites either side of a green fluorescent protein gene (GFP) to allow the ligation of a target gene fragment. When digested with *NotI* and transfected, the vectors incorporated into the trypanosomes 177 base pare repeats (177-targeting). The vectors contained two opposing T7 promoters under the control of the tet operator. Once induced, the vectors were transcribed in both directions by T7 polymerase. The vectors also contained an ampicillin resistance for cloning in *E. coli*. Panel A: the p2T7-177 vector contained a phleomycin resistance gene (BleoR). Panel B: the p2T7-Ver3 vector contained a hygromycin resistance gene (HygR).

Figure 2.1 Schematic diagram of RNAi vectors p2T7-177 and p2T7-Ver3

A.



B.



2.25 Generation of trypanosome expression constructs

Conditional expression of TbPDI-2 in trypanosomes was achieved using the p1791 expression vector (Figure 2.4), which was provided by the Mark Carrington lab, Department of Biochemistry, University of Cambridge. This expression vector was based on the pLEW100 (Wirtz et al., 1999) but has an enhanced yellow fluorescent protein gene (740 bp) in the place of the luciferase ORF. The BS2 (TbPDI-2 ORF) gene was amplified using primers which contained *HindIII* (FP GCG AAG CTT ATG CGC GCT ATT TTT CTT GTT GC) and *BamHI* (RP CGG GAT CCC TAC AGA TCT TGT TTA TCA ACA TTG CTC) restriction sites. This allowed the subcloning of the BS2 gene into the expression vector. The vector was linearized with *NotI* and transfected into trypanosome cells. Once transfected, the vector incorporated into the trypanosome genome at the 177 base pair repeats. The vector also contained a phleomycin resistance gene (BleoR) which allowed the selection of stably transfected cells. The vector contained a T7 promoter under the control of the tet operator, which preceded the TbPDI-2 ORF. Once induced the TbPDI-2 ORF was translated.

Conditional expression of tagged TbEro 1 in trypanosomes was achieved using the p2620 expression vector (Figure 2.5), which was also provided by the Mark Carrington lab. This expression vector was also based on the pLEW100 but had an additional 12 HA epitopes introduced after the luciferase ORF which can be cleaved by TEV protease. The TbEro 1 ORF was amplified using primers which contained *HindIII* (FP AAG CTT ATG TTG AAG ATG CGA TTA TTG ATA GTA GTT CCT) and *BamHI* (RP GGA TCC TCA ACT CAT AGC CGT GTA ATT GAA TTT TTC G). This allowed the subcloning of the TbEro 1 gene into the expression vector. Once induced with tetracycline the vector produced TbEro1 with the HA tag at the C-terminus.

2.26 Generation of recombinant TbEro 1 expression construct

The pGST-Parallel 3 expression vector was generated by Sheffield *et. al.* (Sheffield et al., 1999) and was employed to produce recombinant TbEro 1 in BL21 cells (Figure 2.6). The TbEro 1 ORF was amplified using primers which contained *BamHI* (FP GGA TCC ATG TTG AAG ATG CGA TTA TTG ATA GTA GTT CCT) and *SpeI* (RP GCA CTA GTG AAT TCT CAA CTC ATA GCC GTG TAA TTG AAT TTT TCG) restriction sites. After digestion, the TbEro1 ORF was subcloned into the pGST-Parallel 3 expression vector.

2.27 Generation of TbPDI-2 knockout constructs

To knockout the two endogenous TbPDI-2 alleles, two constructs were generated (Figure 2.7). These constructs were designed to replace the TbPDI-2 ORF with the resistance genes for blasticidin resistance (bsr) and puromycin resistance (PuroR). The resistance genes were flanked by regions found upstream and downstream of the TbPDI-2 ORF to allow homologous recombination to take place. To generate these constructs, several steps were taken: 1) The TbPDI-2 ORF and flanking regions were subcloned into pUC19 (Figure 2.8) using primers which contained the restriction sites *BamHI* (FP GGA TCC ACA ACT TTA CCC) and *XbaI* (RP TCT AGA GGC GCG CAG CAC CCC C). 2) The vector was then used as a template to amplify all but the TbPDI-2 ORF. The primers used contained the restriction sites *MluI* (FP ACG CGT CGT GAC GCT TTT TAT TAA CTT) and *SpeI* (RP ACT AGT CAA GAT TAA GAA AAC GTT TGG CCC). Finally the blasticidin resistance gene (FP *MluI* ACG CGT TGC CCA CCA TCA GCC ATC TCA TTG, RP *SpeI* ACT AGT ATG CAG ATC GAG AAG CAC CTG TCG) and puromycin resistance gene (FP *MluI* ACG CGT ATG ACC GAG TAC AAG CCC ACG GTG C, RP *SpeI* ACT AGT TCA GGC ACC GGG CTT GCG GGT CA) gene ORFs were then amplified using primers which contained the same restrictions sites *MluI* and *SpeI* and were ligated to the pUC19 and TbPDI-2 flanking regions. To knockout the TbPDI-2 alleles, the constructs were digested with *BamHI* and *XbaI*; releasing the resistance genes flanked by TbPDI-2 upstream and downstream regions.

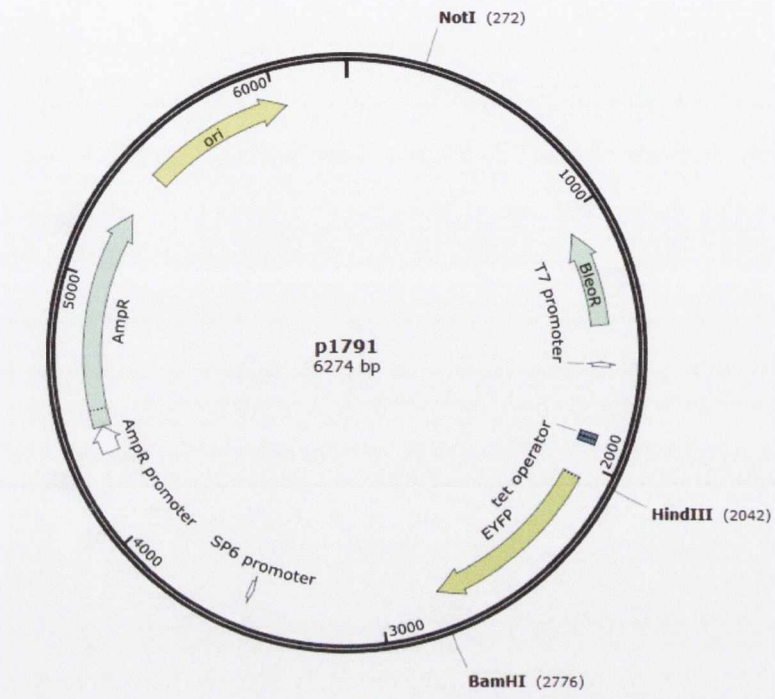
Figure 2.4 Schematic diagram of TbPDI-2 p1791 expression vector

Panel A. The p1791 expression vector was generated and provided by the Mark Carrington lab, University of Cambridge. The vector contained an ampicillin resistance gene (AmpR) for cloning and amplification in *E. coli*.

Panel B. The TbPDI-2 ORF (BS2) was subcloned into the p1791 vector using the restriction sites *HindIII* and *BamHI*. The vector was linearized with *NotI* and transfected into trypanosome cells. Once transfected, the vector incorporated into the trypanosome genome at the 177 base pair repeats (177-targeting). The vector also contained a phleomycin resistance gene (BleoR) which allowed the selection of stably transfected cells. The vector contained a T7 promoter under the control of the tet operator, which preceded the TbPDI-2 ORF. Once induced the TbPDI-2 ORF was translated. The locations of the small nucleotide polymorphisms (snp) are indicated.

Figure 2.4 Schematic diagram of TbPDI-2 p1791 expression vector

A.



B.

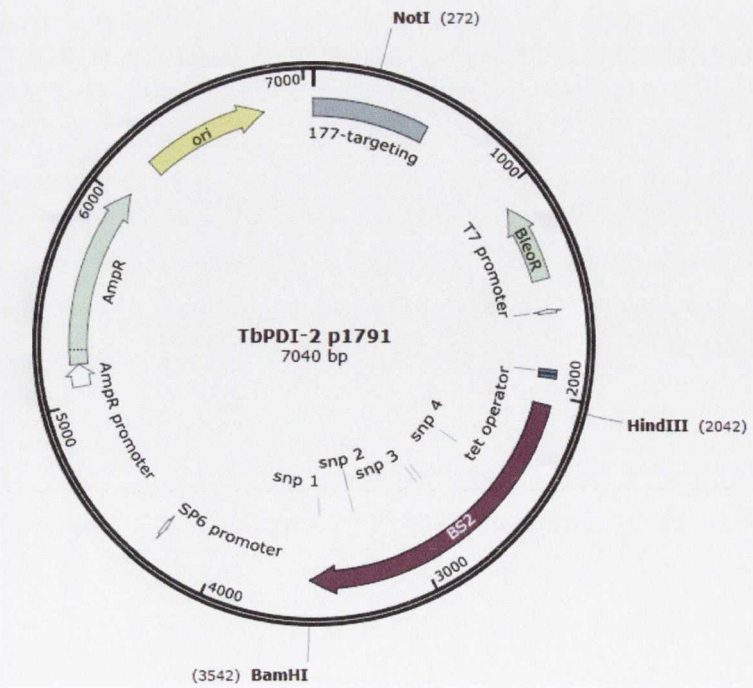


Figure 2.5 Schematic diagram of TbEro 1 p2620 tagging vector

The p2620 expression vector was generated and provided by the Mark Carrington lab, University of Cambridge. The vector contained an ampicillin resistance gene (AmpR) for cloning and amplification in *E. coli*. The TbEro 1 ORF (TbEro 1) was subcloned into the p2620 vector using the restriction sites *HindIII* and *BamHI*. The vector was linearized with *NotI*. Once transfected, the vector incorporated into the trypanosome genome at the 177 base pair repeats (177-targeting). The vector also contained a phleomycin resistance gene (BleoR) which allowed the selection of stably transfected cells. The vector contained a T7 promoter under the control of the tet operator before the TbEro 1 ORF and 12 HA tags (HA) after the TbEro 1 ORF. Once induced the TbEro 1 ORF was translated with 12 HA epitopes attached to the C-terminus of the protein.

Figure 2.5 Schematic diagram of TbEro1 p2620 tagging vector

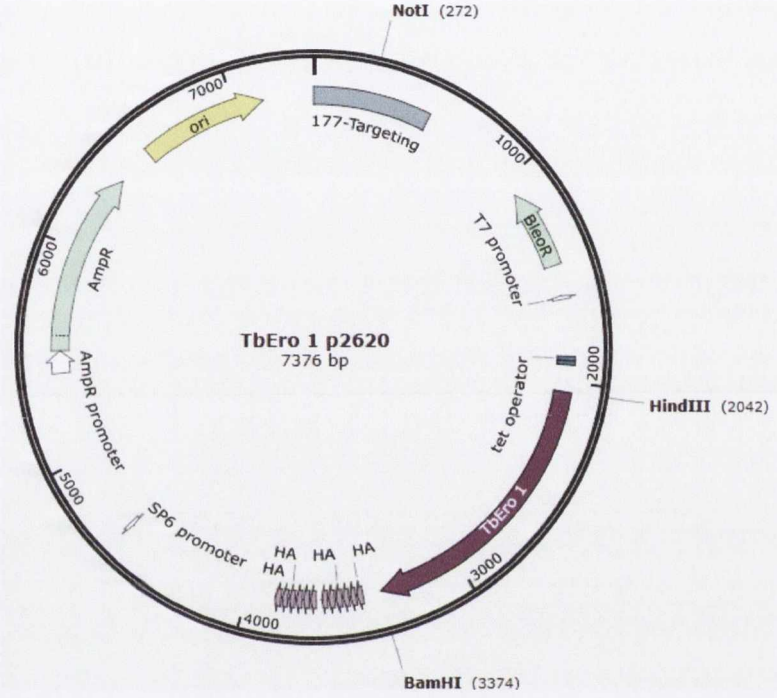


Figure 2.6 Schematic diagram of TbEro 1 pGST-Parallel 3 vector

The TbEro 1 ORF was subcloned into the pGST-Parallel 3 vector using the restriction sites *BamHI* and *XbaI*. The vector contained an ampicillin resistance gene (AmpR) for cloning and amplification in *E. coli*. The vector also contained a gene for the lac repressor protein (lacI). The TbEro 1 ORF promoter (tac promoter) was under the control of the lac repressor. In the presence of IPTG, the TbEro 1 ORF was transcribed. The TbEro 1 is also tagged at the N-terminus by a preceding GST tag (GST), allowing the purification of the protein.

Figure 2.6 Schematic diagram of TbEor 1 pGST-Parallel 3 vector

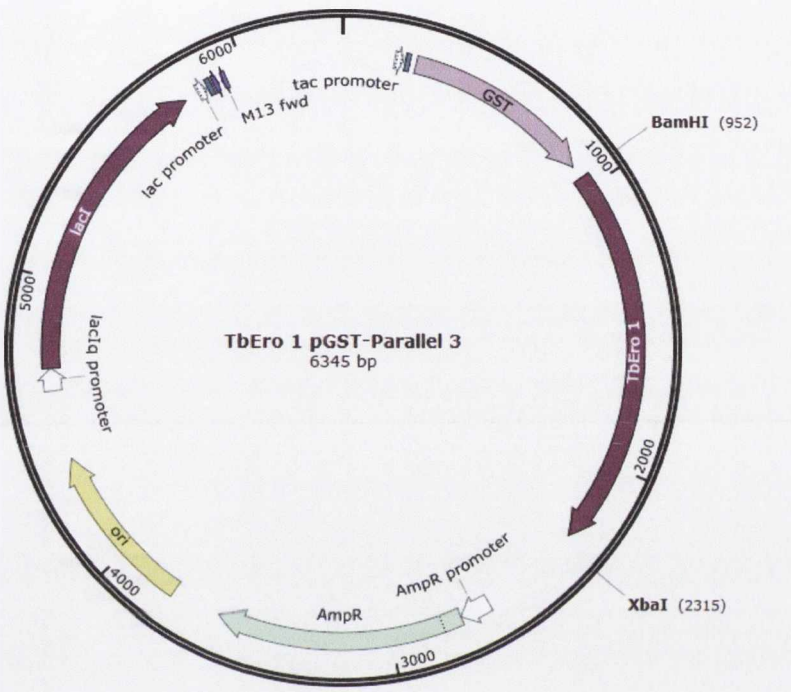
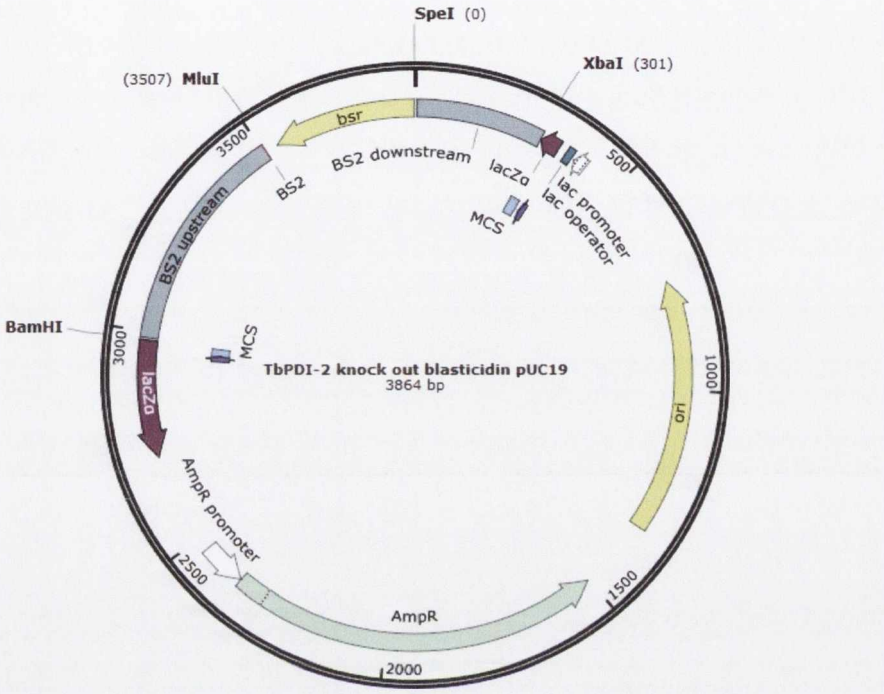


Figure 2.7 Schematic diagram of TbPDI-2 knock out blasticidin and puromycin in pUC19 vector

The schematic diagram of the completed vectors TbPDI-2 knock out blasticidin pUC19 and TbPDI-2 knock out blasticidin pUC19 are shown. Panel A: the plasmid was designed so that after digestion with *BamHI* and *XbaI*, a linearized region of DNA containing the blasticidin resistance gene (bsr) flanked by upstream and downstream regions of the TbPDI-2 ORF (BS2 upstream, BS2 downstream) was released. Panel B: purocmycin resistance gene (PuroR) was subcloned in the place of the bsr gene using the restriction enzymes *MluI* and *SpeI*. This provided an alternative selection drug to knock out the second TbPDI-2 allel.

Figure 2.7 Schematic diagram of TbPDI-2 knock out blasticidin and puromycin in pUC19 vector

A.



B.

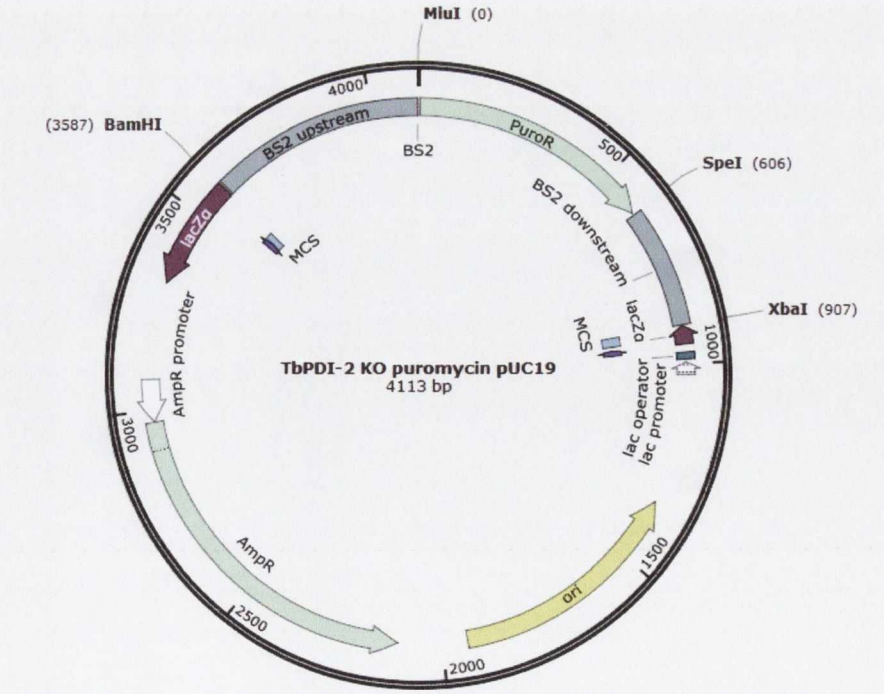
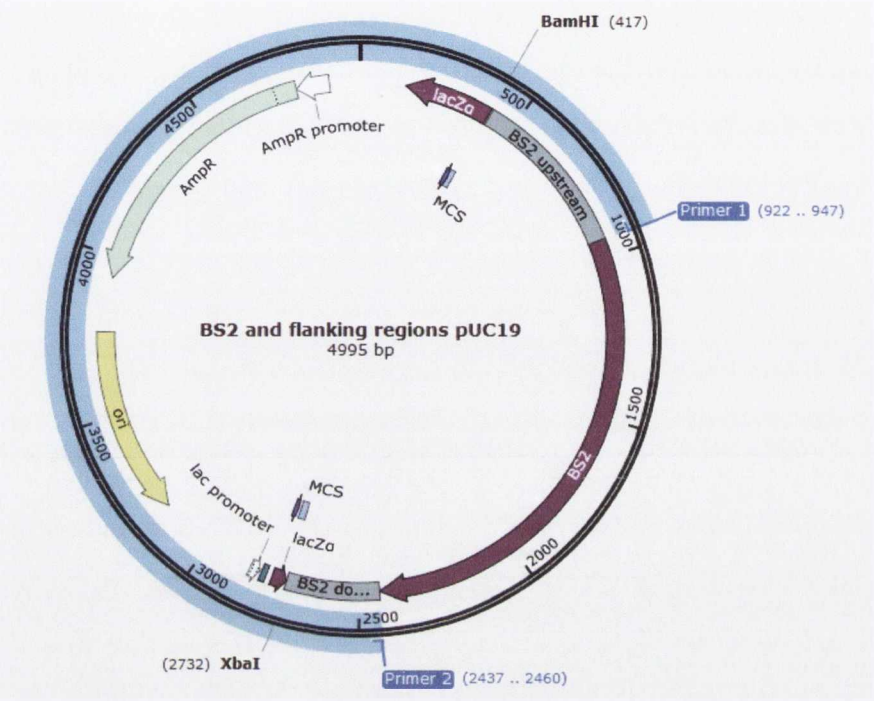


Figure 2.8 Schematic diagram of TbPDI-2 and flanking regions in pUC19 vector

The TbPDI-2 ORF (BS2), 520 bp upstream of BS2 and 295 bp downstream of BS2 were subcloned into the cloning vector pUC19. This was carried out using the restriction enzymes *BamHI* and *XbaI*. The vector plasmid was then used as a template for a PCR using primers (Primer 1 and Primer 2) designed to amplify all but the TbPDI-2 ORF (region of PCR amplification shown as highlighted). This amplification product was subsequently ligated to the blasticidin resistance gene (bsr) or puromycin resistance gene (PuroR).

Figure 2.8 Schematic diagram of TbPDI-2 and flanking regions in pUC19 vector



2.28 Homologous recombination

The aim was to introduce a resistance (blasticidin and puromycin resistance genes) flanked by nucleotide sequences corresponding to regions of 300-400 nucleotides immediately upstream and downstream of the open reading frame of TbPDI-2. This strategy allowed for replacement of the TbPDI-2 gene with the appropriate resistance gene by homologous recombination with the flanking regions. Since trypanosomes are diploid the process involved two rounds of replacement / selection to generate true TbPDI-2 null mutants.

A fragment of 2.3 kb containing the entire open reading frame of TbPDI-2 (1,494 bp) was amplified from genomic DNA using primers base designed around a *BamH I* site located 526 bp upstream of the TbPDI-2 initiation codon and an *Xba I* site located ~300 bp downstream of the stop codon. This fragment was subcloned into the *Bam H I/ Xba I* sites of pUC19. This construct was used as a template to amplify a fragment containing the pUC19 plasmid and regions flanking the TbPDI-2 open reading frame. This involved the use of a forward primer designed to correspond to a region immediately downstream of the TbPDI-2 stop codon and a reverse primer corresponding to the region immediately upstream of the start codon.

2.29 Preparation of trypanosomes for storage in liquid nitrogen.

Cells at ~1 x 10^6/ml were mixed 1:1 with 2X freezing buffer (30 % glycerol, 20 % foetal calf serum, in HMI-9) and transferred to cryovials. Samples were placed in an isulated container at -80°C for 5-6 hr, to allow slow freezing of cells. Samples were then transferred to liquid nitrogen for long term storage.

2.30 RNA isolation

Total RNA was isolated from culture cells in the logarithmic phase of growth using the Absolutely RNA® kit from Stratagene.

2.31 Quantitative Real Time Polymerase Chain Reaction.

Relative quantification of the transcript was performed using the Brilliant® II Sybr green qRT-PCR one step kit and the MX3000P from Stratagene. Primers sets were designed by the Beacon Designer 7 software and were optimized for the RT and amplification steps. Each pair was also assessed to ensure > than 95% efficiency during the amplification stages. Actin was chosen as the reference gene for normalization. In 14 separate experiments, carried out in triplicate, examining the expression of actin in procyclic form (Ct = 21.3 ± 0.8, n =14) and in bloodstream form (Ct = 21.6 ± 0.5, n = 14) it was established that actin is expressed at the same level in both lifecycle forms and was therefore suitable for use as a reference gene. The relative quantitation of genes of interest in bloodstream and procyclic form trypanosomes was determined by normalization of the expression level of these genes to the expression level of actin. Standard curves for all target genes were constructed to ensure amplification/detection efficiencies were reproducible and similar. Reactions were carried out in triplicate and the relative quantities of the unknown genes of interest were calculated relative to the average Cts of the calibrator wells. These relative quantities were then averaged. The error bars included reflect the deviation in the replicates, derived originally from the error of each fluorescence measurement at each cycle, and include estimates of the imprecision of the Normalisers. A list of the primer sequences can be seen in table 2.4.

Name of primer	Sequence	Target gene
QRT-PDI1-FP1	ACTTCCAGCCATTGTTGTC	TbPDI-1 (Tb927.4.2450)
QRT-PDI1-RP1	GCCTCATCGTAAGCATCG	TbPDI-1 (Tb927.4.2450)
QRT-PDI2-FP1	GCAATAGCACTTCACTTTCTAC	TbPDI-2 (Tb927.10.8230)
QRT-PDI2-RP1	TTCTGTCAACTTCTCTACTTCC	TbPDI-2 (Tb927.10.8230)
QRT-PDI3-FP1	GATTCAGCAAGCCGTCTCC	TbPDI-3 (Tb927.7.1300)
QRT-PDI3-RP1	ACCATTCTCAAGGACATTAGTAGC	TbPDI-3 (Tb927.7.1300)
QRT-TbPDI4-FP1	GTGCTTGTTGATTCTGAC	TbPDI-4 (Tb927.5.1020)
QRT-TbPDI4-RP1	GAAGGCTATGAAGGATGG	TbPDI-4 (Tb927.5.1020)
QRT-PDI5-FP1	TTTGATAAGGTGGCGTTG	TbPDI-5 (Tb927.7.5790)
QRT-PDI5-RP1	CTTGTCAGCATCCAATCG	TbPDI-5 (Tb927.7.5790)
QRT-Ero1FP1	GCCTTCTTAACAACATAACG	TbEro 1 (Tb927.8.4890)
QRT-Ero1RP1	TGGAAACATCACTCATATCG	TbEro 1 (Tb927.8.4890)

Table2.4 List of qRT-PCR primers

The sequences of the primers and the target genes which were analyzed using qRT-PCR primers are shown.

2.32 Cell fixation.

Paraformaldehyde solution (6 %) was prepared freshly on the day of the experiment or frozen as aliquots at —20 °C. Paraformaldehyde was added to 5 volumes of distilled deionised water and titrated with NaOH (5 M) until fully dissolved. 10 volumes of PBS were added to the fixative and the pH was adjusted to 7.5 with HCl 5M). Final volume was adjusted with distilled water as required such that the final paraformaldehyde concentration was 6 %. Bloodstream or procyclic form trypanosomes were harvested from culture and washed, or bloodstream forms isolated from rats, and resuspended in PSG at a concentration of 5 x 107 cells / ml. An equal volume of 6 % parafomaldehyde solution was added. Bloodstream form cells were incubated for 10 min, and procyclic forms for 30 min, on ice. Following this incubation the cells were washed twice with PBS buffer containing sodium azide (15 mM) and stored at 4 °C.

2.33 Immunofluorescence and confocal microscopy

Trypanosomes were fixed as previously described. Slides were prepared by washing with polylysine solution (1/10 dilution). These were allowed to dry. The slides were then washed with 70% ethanol. Cells were added and allowed to settle for 15mins. The slide was washed in PBS then blocked with blocking solution (PBS, BSA 5%, and Triton X100 0.15%) for 1 hour. Slides were not allowed to dry out during this period. Cells were then treated with primary antibody solutions for 3 hours (see table 2.4), followed by 3 washes with sterile PBS, then secondary antibody solutions for 3 hours(Alexa 488 green or Alexa 568 red). Mounting solution was added (PBS, glycerol 30%, p-phenylenediamine 1 mg/ml, and DAPI 0.1 μg/ml) and coverslips were then fixed.

Antibody target protein	Animal	Dilution
TbPDI-2	Rabbit	1/100
HA	Mouse	1/200

Table 2.4 Immunofluorescence antibody dilutions

The dilutions for antibodies used during immunnofluorescence microscopy.

2.34 Analysis of the rate of surface protein production using pulse-chase

Cells were cultured in HMI9 medium containing selection drugs and RNAi was induced. Cells were collected by centrifugation ($3x10^8$ cells), washed and resuspended ($2x10^7$ cells/ml) in cysteine and methionine free labelling medium, then incubated for 1 hour at 37°C. A 10 minute pulse labelling was performed by the addition of EasyTag™ EXPRESS^{35}S Protein Labeling Mix, [^{35}S] (0.1 mCi/ml). The medium was diluted with prewarmed HMI9 (1/10) and a sample was immediately removed as the first fraction at time 0. Subsequent fractions were taken at time at various times (10, 20, 30, 50 and 90 minutes). The fractions were diluted (1/10) into ice cold isotonic phosphate wash buffer [NaH2PO4 (3 mM), Na2HPO4 (57 mM), NaCl (44 mM), KCl (5 mM), glucose (10 mM) sucrose (70 mM) pH 8.0], washed twice and

resuspended in 0.9 ml of buffer. Surface proteins were then biotinylated on ice for 10 minutes by the addition of 0.1 ml of sulfo-NHS-biotin (3 mg/ml in wash buffer). Biotinylation was stopped by the addition of 9 ml of cold wash buffer containing glycine (1 mM). The cells were washed a further 2 times and then resuspended in lysis buffer [Tris-Cl (25 mM, pH 7.5), NaCl (150 mM), CHAPS (1%) and protease inhibitors]. The cells were incubated on ice for 1 hour, centrifuged at 18,000 x g for 30 minutes at 4°C. Biotinylated proteins were precipitated from the supernatant using streptavidin agarose. Labelled proteins were analyzed by SDS PAGE autoradiography.

2.35 DNA Sequencing

All vectors constructed throughout this study were sequenced by MWG biotech. The cycle sequencing technology (dideoxy chain termination / cycle sequencing) on ABI 3730XL sequencing machines was employed for all sequences obtained.

Sequencing was carried out using either standard T7 or M13 primers, designed and supplied by MWG biotech, which prime to either the T7 promoter or specific sections of the M13 bacteriophage, or by using specific primers which were designed upstream and downstream of the region of interest.

2.36 Pulse chase labelling of surface proteins in Trypanosomes

Trypanosomes cell lines, which were to be investigated, were grown in HMI9 for the optimum amount of time to induce knockdown in the case of RNAi cell lines. All cell lines investigated were collected and processed simultaneously. Cells were harvested (3×10^8 cells), centrifuged for 5 minutes at 1800 x g and washed with cysteine and methionine free media. The cells were then resuspended in 6.5 mls of prewarmed (37°C) cysteine and methionine free media and incubated for 30 minutes. An aliquot of 59µl of EasyTag(TM) EXPRE35S35S Protein Labelling Mix, [^{35}S] was added to the media (final concentration 0.1 mCi/ml) and the culture was incubated for 10 minutes (pulse). Prewarmed media was added to the flask (32.5 ml) and a 6ml sample was immediately taken representing chase time 0. Additional samples were taken at time chase time 10, 20, 30, 50 and 90 minutes. The samples, once taken, were immediately added to ice cold PSG with added glucose and sucrose (46 ml) and were spun down at

1800 x g at 4°C for 5 minutes. The supernatant was removed and the cells were washed again as before. The cells were then resuspended in 1ml of PSG with NHS biotin (0.3mgs/ml) and incubated on ice for 10 minutes to label the surface proteins of the cells. Ice cold PSG glucose, sucrose and glycine was added (10 ml) to quench the remaining biotin and cells were spun down at 1800 x g at 4°C for 5 minutes. The cells were washed once more in the same wash buffer. The supernatant was removed and the cells were resuspended in to a total volume of 100 µl in lysis buffer (Tris 25 mM, NaCl 150 mM and CHAPS 1% pH 7.5 and a cocktail of protease inhibitors). The cells were incubated on ice for 1 hour and centrifuged at 18,000 x g for 30 minutes at 4°C. Biotinylated proteins were immunoprecipetated from the supernatant using streptavidin agarose. Labelled proteins were analyzed by SDS PAGE autoradiography.

2.37 Co-immunoprecipitation of labeled proteins

Cell lines expressing TbEro 1 p2620 were induced and harvested ($5x10^7$ cells). The cells were washed twice with 1 ml of wash buffer (Na_2HPO_4 8mM, KH_2PO_4 1.5mM, NaCl 137 mM, KCl 2.7mM pH 7.5). After washing, cells were resuspended to a total volume of 400 µl of buffer with 1% CHAPS and a cocktail of protease inhibitors included. Cells were allowed to lyse at 4°C for 15 minutes. After incubation cells were spun down at 18,000 x g for 15 minutes and the supernatant was removed. The supernatant was incubated with anti-HA agarose (50 µl of a 50% slurry) for 16 hours at 4°C. The suspension was centrifuged and the supernatant was removed. The agarose was washed 3 times with wash buffer (1 ml) and sample buffer was added to a final volume of 50 µl and boiled. The samples were analyzed using Western blot and revealed with anti-HA, anti-TbPDI-1 and anti-TbPDI-2 antibodies.

2.38 Digitonin subcellular fractionation

Cells expressing TbEro 1 HA were cultured and harvested ($1x10^8$ cells). The cells were washed in PSG buffer with sucrose and glucose twice and divided into 5 aliquots. The aliquots of cells were centrifuged at 900 x g for 5 minutes and the supernatants were removed. The cells were resuspended in STE buffer (sucrose 0.25 M, Tris 25 mM, EDTA 1 mM, NaCl 50 mM and glucose 10mM) with a cocktail of protease inhibitors (Aprotinin, Bestatin, E-64, Leupeptin and Pepstatin A) and with varying concentrations of digitonin (0, 0.1, 0.2, 0.4 and 0.8 mg/ml). The cells were

incubated at room temperature for 4 minutes. The cells were then centrifuged at 18,000 x g for 2 minutes and the supernatants were separated. The pellet and supernatant fractions were resuspended in sample buffer (to a total volume of 200 µl) and analyzed using Western blot. Each lane represented $5x10^6$ cells (10 µl of samples).

2.39 Digestion of TbEro 1 HA with glycosidases

Cells expressing TbEro 1 HA were grown, harvested (1×10^8 cells) then washed using PBS with glucose and sucrose 4°C. The cells were resuspended in cold lysis buffer (200 µl of PBS, 1% CHAPS and a cocktail of protease inhibitors) and incubated for 1 hour at 4°C. The cells were centrifuged at 18,000 x g for 15 minutes at 4°C. The supernatant was removed and added to anti-HA agarose (50 µl of a 50% slurry). The mixture was incubated for 1 hour at 4°C.

The agarose (40 µl) was then washed with *N*-glycosidase F digestion buffer (Na_2HPO_4 50 mM, EDTA 10 mM, CHAPS 1%, β-mercaptoethanol 5 mM). Half of the washed agarose and bound TbEro 1 HA (20 µl) was then digested with *N*-glycosidase F overnight at 37°C with additional protease inhibitors. The resulting digested and undigested samples were analyzed by Western blot.

An additional alliqout of agarose (40 µl) was washed with denaturation buffer (PBS, SDS 2% and protease inhibitors). The agarose was then washed 4 times with wash buffer (PBS, Triton X100 2.5%). Half of the washed agarose and bound TbEro 1 HA (20 µl) was then digested with Endo H (5µl) overnight at 37°C. The resulting digested and undigested samples were analyzed by Western blot.

2.40 Purification of bacterial recombinant TbEro1

The BL21 TbEro 1 GST cell line was grown for 16 hours in LB broth (5 ml) containing ampicillin (50 µg/ml) at 37°C. The culture volume was brought to 1 l with new LB broth and induced by the addition of IPTG (35 µM). The cells were incubated at 30°C for 3 hours, then centrifuged at 5000 x g for 20 minutes and frozen (-20°C). A pellet

from 0.5 l was resuspended in 30 mls of NETN (Tris 20 mM, NaCl 100 mM, EDTA 1 mM, NP-40 0.5% pH 8.0), PMSF (1mM), lysozyme (1 mg/ml) and a cocktail of protease inhibitors. The cells were incubated on ice for 10 minutes then sonicated. The solution was centrifuged at 16,000 x g for 10 minutes at 4°C. The supernatant was incubated with anti-GST agarose (400 µl of a 50% slurry) at 4°C overnight. The anti-GST agarose was washed 3 times with NETN (500 µl) then washed with PBS a further 3 times. The agarose was then resuspended in 5 mls of elution buffer (Tris 50 mM and glutathione 10 mM, pH 8.0) and incubated at 4°C for 60 minutes. The agarose was centrifuged (4000 x g for 10 minutes) and the supernatant was recovered and the protein concentrated to 0.5 mls using Millipore ultracentrifugation kit.

3 Bioinformatics analysis of oxidative folding pathway of Trypanosoma brucei

3.1 Protein disulphide isomerases of *Trypanosoma brucei*

The protein disulphide isomerase family members have been suggested to be essential for the proper folding of secretory proteins in numerous multicellular species including humans (Bulleid and Ellgaard, 2011; Jessop et al., 2007; Soldà et al., 2006; Tokuhiro et al., 2012). A search for PDIs in *Homo sapiens* has revealed 19 possible PDIs with multiple members of the same family being present in every cell (Kozlov et al., 2010). Many PDIs have been shown to have specific substrates, which suggest PDIs play subtle but specific roles within the cell. Some of these roles may be attributed to domain structure, active site motifs, endoplasmic reticulum retention signals and transmembrane regions. All of these features are easily predicted using bioinformatics techniques and reveal the possible function of the PDIs.

Firstly, this bioinformatics analysis set out to identify all possible PDIs within the genome of *T. brucei*. A previous search for possible PDIs was carried out in conjunction with the characterization of two PDIs identified (Rubotham et al., 2005). While this search was comprehensive, the criteria used to identify possible PDIs does not reflect the current criteria for the identification of PDI family members (Kozlov et al., 2010). What defines PDI family members has been a topic of debate over the past decade but it has become widely accepted that PDIs must meet the following criteria: Firstly, the PDI must enter the endoplasmic reticulum to be involved in the folding of proteins of the secretory pathway and so the PDI must contain an *N*-terminal leader sequence. Secondly, the protein must contain at least one thioredoxin like domain. However, this domain does not have to possess any catalytic activity (Kozlov et al., 2010).

Secondly, one of the goals of this bioinformatics analysis was to speculate as to the possible functions of the PDIs identified. Many of the human PDIs have shown substrate specificity (Jessop et al., 2007, 2009; Tokuhiro et al., 2012). This finding suggested that the PDIs of *T. brucei* may also show substrate specificity; specifically increasing the rate of folding of VSG. These functions were to be predicted based upon

the features of the trypanosome PDIs and assessment of the known function of PDIs which showed sequence similarity.

The goals of this bioinformatics analysis of the PDIs of *T. brucei* were as follows: Firstly, identify all possible PDIs in the genome of *T. brucei* by carrying out a genome wide search for all thioredoxin like proteins. Secondly, assess the other attributes of these candidate proteins and identify the PDIs base upon the features found; specifically *N*-terminal leader sequence and thioredoxin like domain. Thirdly, to predict the possible functions of the candidate PDIs by comparison to well known PDIs from other species. Lastly, to choose the PDI most likely to be essential, as a candidate for further analysis.

3.1.1 Identification of possible PDIs of *Trypanosoma brucei*

The Sanger sequencing project search service was employed to search the genome of *T. brucei* for the Pfam keyword "Thioredoxin". The results were individually reviewed for PDI specific conserved domain hits with Pfam and Panther which narrowed the group of thioredoxin containing proteins to six PDIs (

Table 3.1). Domain hits were considered significant base on their E (expectation) values. The E value score is the theoretically expected number of false sequences retrieved per sequence query. It follows that the lower the E value score the more likely the match retrieved was not by chance and the proteins matched share a true identity. An E value score of 1 represents a match found entirely by chance. In this study an E value score of $1e^{-8}$ or less was considered significant. A second search for thioredoxin like proteins was carried out using the BLASTp analysis of the genome of *T. brucei*. The protein sequence of trypanosome thioredoxin (UniProt accession: Q38FJ1) was used as the query sequence (Table 3.2). One additional protein, hypothetical protein (UniProt accession: Q381H0), was identified during this search but when a signature recognition search was employed using the EMBL-EBI InterProScan tool, no specific PDI hits were found in any of the available structural databases searched. It was concluded that there are only 6 possible PDI candidates.

Assigned name	NCBI Accession	Systematic gene id
TbPDI-1	XP_844395	Tb927.4.2450
TbPDI-2	XP_823057	Tb927.10.8230
TbPDI-3	XP_845754	Tb927.7.1300
TbPDI-4	XP_844776	Tb927.5.1020
TbPDI-5	XP_846202	Tb927.7.5790
TbPDI-6	XP_846201	Tb927.7.5780

Table 3.1 List of candidate PDIs putative proteins.

A two part search was undertaken using 1) the Sanger sequencing project search service which was searched for the Pfam keyword "Thioredoxin", 2) the protein sequence of trypanosome thioredoxin (NCBI accession number XP_803641) was used as the query sequence to search for additional proteins using the BLASTp search tool. Proteins found were subjected to a signature recognition search using the EMBL-EBI InterProScan tool to identify PDI specific conserved domains. The resulting proteins were narrowed down to 6 PDIs.

3.1.2 Primary sequence analysis of candidate PDIs

Various physical parameters of the candidate TbPDIs were predicted using different programs from the ExPASy website (http://ca.expasy.org/tools) including predicted mass, isoelectric point, *N*-glycosylation sites, the presence of a signal peptide, transmembrane span and an endoplasmic reticulum retention signal. The results are summarized in Table 3.2.

The predicted protein sequences of the candidate PDIs were analysed using NCBI conserved domain search (www.ncbi.nlm.nih.gov.elib). Domain hits were considered significant base on their E (expectation) values score. The search revealed the domain structure of the PDIs and these were classified as to the system published by Kanai (Kanai *et al.* 1998). TbPDI-1 to 5 showed a clear domain structure which was classified according number of catalytic domains and distance between multiple domains. In the case of TbPDI-6 it was clear that it was likely to be a PDI but it has lost the catalytic function of its a-like-domain (identified using Pfam database search) due to the substitution of one of the active site cysteine residues. This was still classified as a single domain PDI due to the domain sequence homology observed.

Resident proteins of the endoplasmic reticulum usually contain characteristic features. These include an *N*-terminal leader sequence, a *C*-terminal retrieval signal sequences or association with the membrane of the endoplasmic reticulum via transmembrane regions. To identify these hydrophobic regions, TbPDI-1 to 6 were subjected to analysis by a series of three programs: TMpred, Kyte-Doolittle and TopPred from the ExPASy website (http://ca.expasy.org/tools). Several programs were used to predict hydrophobic regions. The regions which were found to be hydrophobic by all three programs were the most likely to be correct and so only those regions were taken into account when constructing Table 3.2. The resulting hydrophobic graphs can be seen in section 7.2 of the Appendix.

	TbPDI-1 Tb927.4.2450	TbPDI-2 Tb927.10.8230	TbPDI-3 Tb927.7.1300	TbPDI-4 Tb927.5.1020	TbPDI-5 Tb927.7.5790	TbPDI-6 Tb927.7.5780
Open reading frame size (aa)	411	497	377	589	135	151
Predicted mass (kDa)	44.49	55.58	41.94	65.72	15.4	17.03
Predicted isoelectric point (pH)	9.14	5.3	8.66	6.58	8.52	5.73
N-glycosylation sites	2	12	0	1	0	0
Signal peptide	Yes	Yes	Yes	Yes	Yes	Yes
Number of Thioredoxin domains and E-value scores	PDI a domain $1.91e^{-39}$	PDI a domain $7.55e^{-24}$ PDIb domain $1.82e^{-09}$ PDIb' domain $1.82e^{-09}$ PDIa' domain $1.97e^{-20}$ PDIc domain $2.9e^{-09}$	PDI a domain $2.02e^{-39}$ PDI a domain $8.20e^{-31}$ PDI c domain $2.07e^{-15}$	PDI a domain $3.75e^{-19}$	PDI a domain $7.93e^{-33}$	PDI a domain $1.50e^{-12}$
Classification	Single domain	Class 1	Class 2	Single domain	Single domain	Single domain
Transmembrane span	No	No	No	Yes	No	No
C-terminal sequence	EQPK	KQDL	IKGE	EFED	KSNM	FQNS
Previously reported activity	Yes Shown to possess isomerase activity and reduction activity	Yes Shown to possess isomerase activity	None	None	None	None

Table 3.2 Various physical parameters of candidate PDIs

Various physical parameters of the candidate TbPDIs were predicted using different programs from the ExPASy website (http://ca.expasy.org/tools) including predicted mass, isoelectric point, *N*-glycosylation sites and the presence of a signal peptide.
The conserved domain search was employed (www.ncbi.nlm.nih.gov.elib) to locate any PDI conserved domains. Each candidate PDI contained at least one domain hit and were classified according to (Kanai *et al.* 1998).
A search for transmembrane spanning regions of the candidate proteins was carried out using three separate programs: TMpred, Kyte-Doolittle and TopPred. Results which agree were considered to be strong hits and are indicated.

3.1.3 Phylogenetic analysis of candidate PDIs

The candidate PDIs from *T. brucei* were used to construct a multiple sequence alignment of the putative protein sequences to assess if there were any conserved regions between the proteins (see section 7.1 of the Appendix for the alignment). Not surprisingly there was little sequence homology between the PDI protein sequences, due to their different domain structures, with the exception of the *N*-terminal a-domain which is present in all candidates. Interestingly there was no sequence homology observed in the *C*-terminal end of any of the PDIs, which was not expected. PDIs are considered resident proteins of the endoplasmic reticulum and as such, these PDIs would contain some form of endoplasmic reticulum retrieval signal. This was observed only in TbPDI-2 in the form KQDL.

A phylogenetic tree was constructed of putative PDIs first by carry out a BLASTp search using the candidate PDIs sequences to find related proteins. Proteins were chosen from a wide range of species and aligned using ClustalW (www.ebi.ac.uk/Tools/clustalw)(Eddy, 1995). A phylogenetic tree was constructed using the neighbour joining method and sections of the sequence which contained gaps were ignored (Figure 3.1). Although this method of phylogenetic tree construction is by no means the most accurate, the resulting phylogenetic tree does illustrate the difference between the various families of PDIs found in *T. brucei*.

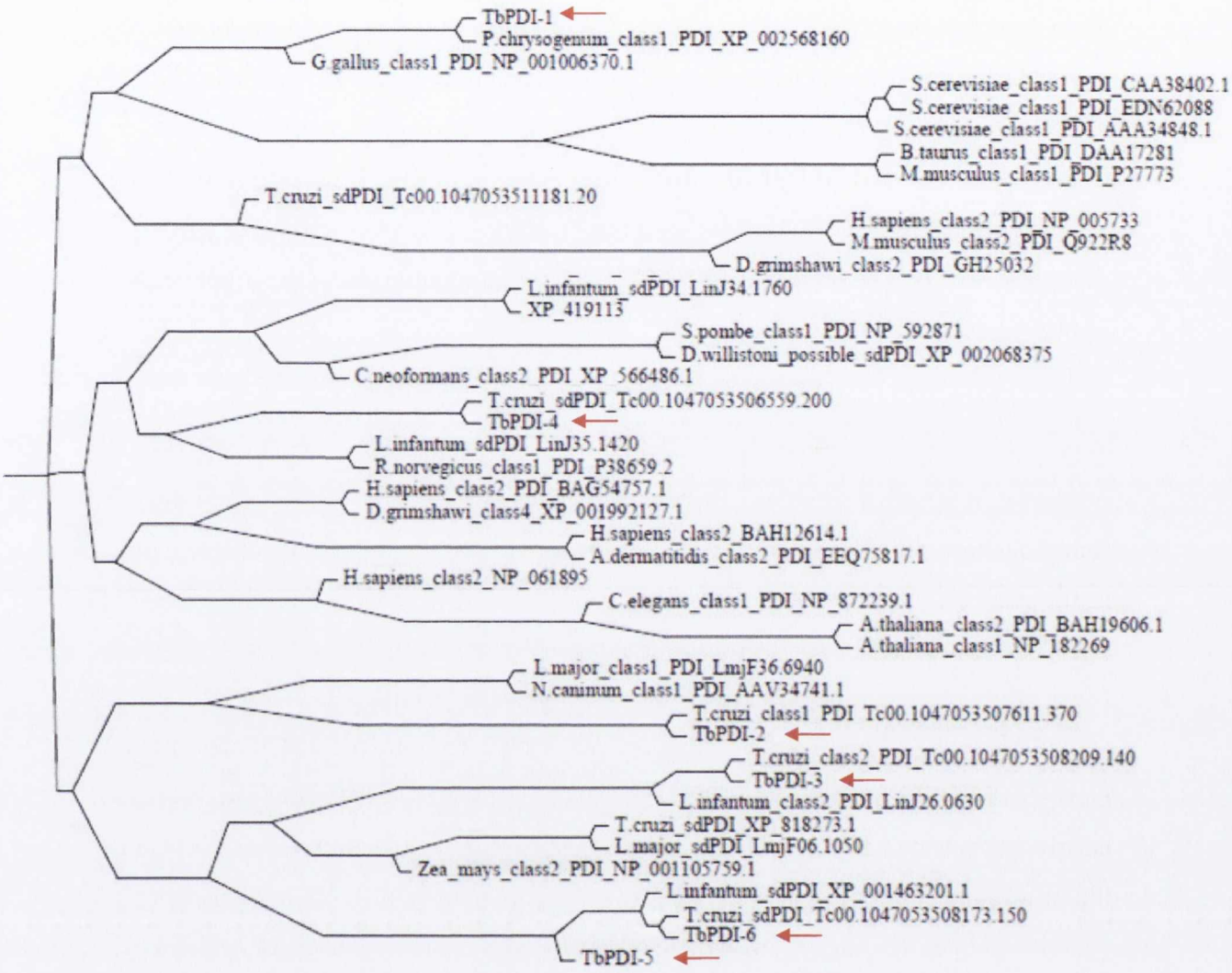


Figure 3.1 Phylogenetic tree of putative PDIs

A BLASTp search was carried out to identify the closest related proteins to the candidate PDIs. The chosen proteins were then aligned and a phylogenetic tree was constructed using ClustalW. The phylogenetic tree was constructed using the neighbour joining method and sections of the alignment which contained gaps were excluded. The phylogenetic tree was rendered using the program Phylodendron (iubio.bio.indiana.edu/treeapp). The species, classification of PDI and NCBI accession number are included in the names of the aligned proteins. The TbPDIs are marked with red arrows.

3.1.4 Multiple sequence alignment of TbPDI-1 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify TbPDI-1's closest related proteins in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The amino acid sequences of the proteins were reviewed for various attributes (Table 3.3) and were aligned using the program T-Coffee (Notredame et al., 2000)(Figure 3.2).

The protein found to be most similar to TbPDI-1 was the putative single domain PDI TcIL3000.4.2300 found in the protozoan parasite *Trypanosoma congolense,* the trypanosome species which causes nagana in cattle in east Africa. The alignment showed a high similarity with a 75% identity and an expectation value of $2.30e^{-166}$. Both sequences share a similar length; *T. congolense* PDI contained a predicted length of 409 amino acids and 411 amino acids in the case of TbPDI-1. The two proteins share other similar predicted attributes: *T. congolense* PDI has a pI of pH 9.06, a mass of 44.47 kDa and a single *N*-glycosylation site. TbPDI-1 possesses a pI of pH 9.14, a mass of 44.49 kDa and two possible *N*-glycosylation sites. Unfortunately no research has been published concerning the *T. congolense* PDI beyond sequence analysis and the same is true for the next most closely related sequence found in *Trypanosoma cruzi*. The alignment however reveals that TbPDI-1 has many conserved regions when compared to its closest relatives and so may serve an integral function as a PDI in trypanosomes. It appears that the *N*-glycosylation sites have not been highly conserved between these species but possible sites are present in all three protein sequences. It has been shown that *N*-glycosylation does occur in TbPDI-1 as a change in mass was observed when TbPDI-1 was digested by an *N*-glycopeptidase (Rubotham et al., 2005).

Other single domain PDIs were also found to be significantly similar to TbPDI-1 in the *Trypanosomatida Leishmania.* Three *Leishmania* single domain PDIs were chosen: a *Leishmania infantum* PDI (LinJ34_V3.1960) a *Leishmania mexicana* PDI (LmxM33.2200) and a *Leishmania major* PDI (LmjF34.2200). The sequences showed less similarity than the other trypanosome PDIs to TbPDI-1 but still remained significant (refer to Table 4-7 for values). Their sequence length, mass and domain

structure show similarity with TbPDI-1 although the *N*-glycosylation sites found in TbPDI-1 are not conserved in these Leishmania single domain PDIs.

The final putative single domain PDI included in the multiple sequence alignment was the PDI1 from *Giardia lamblia* (GL50803_29487 or gPDI-1). The protein gPDI-1 was included in the alignment due to extensive work previously carried out on single domain PDIs (see section 1.3.7). Although the sequence of gPDI-1 has an obvious difference in length, 234 residues versus 411 residues found in TbPDI-1, it does show some significant similarity when aligned. As expected the active site is highly conserved but what is more interesting is the presence of two *N*-glycosylation sites. Although the chosen PDIs of *Leishmania* appear not to be glycosylated, it seems to be a common feature of protozoan PDIs. The protein gPDI-1 has been shown to reside in the endoplasmic reticulum using antibody visualization techniques (Knodler et al., 1999).

Species	Name/accession	PDI class	Active site	pI (pH)	Mass (kDa)	E value score	% identity	size (aa)	*N*-glycosylation sites	signal peptide predicted cleavage site
T. brucei	TbPDI-1/ Tb927.4.2450	single domain	CGHC	9.14	44.49			411	2	23
T. congolense	TcIL3000.4.2300	single domain	CGHC	9.06	44.47	$2.30e^{-166}$	75	409	1	24
T. cruzi	Tc00.10470 53506559.200	single domain	CGHC	8.85	48.58	$6.40e^{-118}$	54	441	1	42
L. infantum	LinJ34_V3.1960	single domain	CGHC	6.98	47.22	$9.80e^{-86}$	65	433	0	28
L. mexicana	LmxM33.2200	single domain	CGHC	6.78	47.42	$1.70e^{-84}$	42	433	0	27
L. major	LmjF34.2200	single domain	CGHC	7.04	47.45	$9.30e^{-83}$	40	433	0	28
G. lamblia	gPDI-1/ GL50803_29487	single domain	CGHC	8.71	25.8	$3.00e^{-17}$	29	234	1	17

Table 3.3 Summary of various attributes of proteins similar to TbPDI-1

The PDI class was determined using the conserved domain search hosted by NCBI; active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated.

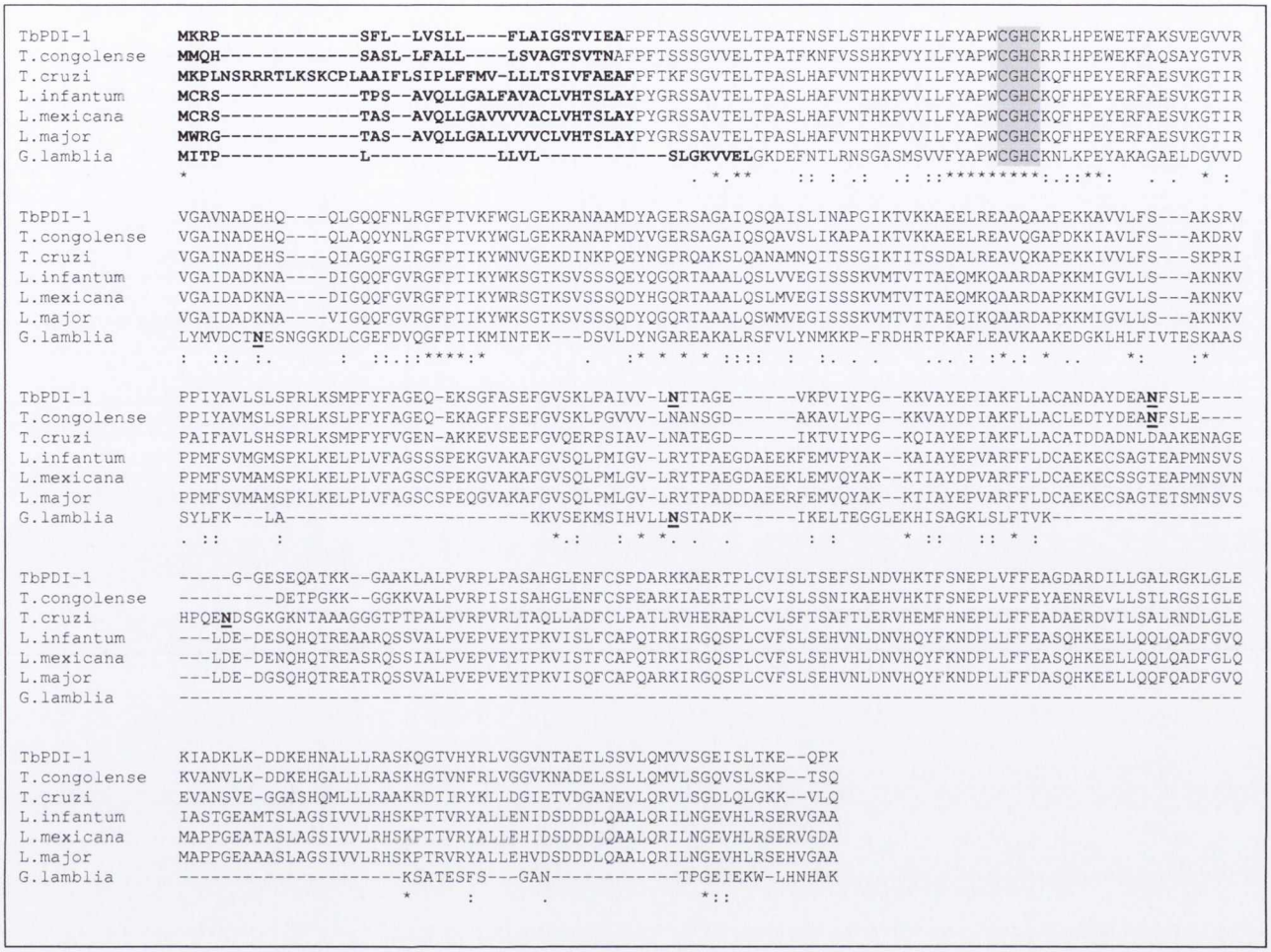


Figure 3.2 Multiple sequence alignment of closely related proteins of TbPDI-1

Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey and predicted *N*-glycosylation sites are underlined. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

3.1.5 Multiple sequence alignment of TbPDI-2 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify TbPDI-2's closest related proteins in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The sequences of the proteins were reviewed for various attributes (Table 3.4) and were aligned using the program T-Coffee (Figure 3.3).

The protein found to be most similar to TbPDI-2 was the Class 1 PDI TcIL3000_10_7120 found in *T. congolense*. The alignment showed a high similarity with a 65% identity and an expectation value of $5.5e^{-179}$. The two PDIs share many similar attributes: both have a similar length (TbPDI-2: 497aa and TcIL3000_10_7120: 496aa), contain many possible glycosylation sites (TbPDI-2: 12 sites and TcIL3000_10_7120: 8 sites) and both share identical active site motifs (CGYC/CGHC). To date there has been no published characterization of this *T. congolense* PDI as is the case with the other closely related PDI proteins found in the trypanosome species *T. cruzi* and *T. vivax*. The alignment of these trypanosome PDI sequences reveals some interesting differences: it appears the active site of the a-domain is not highly conserved throughout these trypanosomes (TbPDI-2: CGYC, *T. congolense* CGYC, *T. cruzi* CGHC and *T. vivax* CIFC) but still has the potential for catalytic properties due to possessing a CXXC motif. Also it appears that *N*-glycosylation is not a consistent feature as the PDI found in *T. cruzi* contains no *N*-glycosylation sites. Not all of these PDIs contain possible endoplasmic reticulum retention signals. *T. cruzi* has a very divergent signal (DDDL) which may have lost its ability to function as a classical endoplasmic reticulum retention signal.

The next most similar protein found in the BLAST search was the protein LmjF36.6940 found in *L. major* known as LmPDI-2. This protein was found to be conserved in other Leishmania species including *L. infantum*. It appears *N*-glycosylation does not play a role in this PDI as there were no possible sites for *N*-glycosylation to occur.

The most characterized PDI proteins to date are found in *H. sapiens* and *S. cerevisiae*. The most similar protein in the species *H. sapiens* to TbPDI-2 was the Class 1 PDI

known as ERp57 (UniProt accession: P30101; also known as Erp60, GPR58 and PDIA3). The alignment of TbPDI-2 and ERp57 showed a significant similarity with a % identity of 27% with an expectation value of $1e^{-42}$.

The PDI in *S. cerevisiae* found to give a significant alignment was the class 1 PDI known as Pdi1p (UniProt accession: A6ZTE9), which was discussed in section 1.3.2 of the introduction. The alignment of TbPDI-2 and Pdi1p showed a significant similarity with a % identity of 26% with an expectation value of $3e^{-37}$. As is the case with TbPDI-2, Pdi1p is *N*-glycosylated at four sites. The two active site motifs are the classical CGHC which are seen in ERp57 in contrast to the trypanosome CGYC active site motif.

The crystal structure of the full length protein has been solved which gives an insight into the possible structure of TbPDI-2. A secondary structure analysis was carried out to investigate if TbPDI-2 and Pdi1p share a similar secondary structure (Figure 3.4). The α helices and β regions of TbPDI-2 were predicted using PSIPRED (bioinf.cs.ucl.ac.uk). Although sequence similarity between the two proteins appears to be low, it is evident that the predicted secondary structure of TbPDI-2 shows clear similarities with the solved structure of Pdi1p.

Species	Name / accession	PDI class	Active site	pI (pH)	Mass (kDa)	E value score	% identity	size (aa)	*N*-glycosylation sites	signal peptide predicted cleavage site
T. brucei	TbPDI-2 / Tb927.10. 8230	Class 1	CGYC /CGHC	5.3	55.58			497	12	19
T. congolense	TcIL3000.10. 7120	Class 1	CGYC /CGHC	5.64	55.21	$5.50e^{-179}$	65	496	8	20
T. cruzi	Tc00.104705 3507611.370	Class 1	CGHC /CGHC	4.87	53.55	$5.70e^{-133}$	50	481	0	20
T. vivax	TvY486_100 8110	Class 1	CIFC /CGHC	4.99	55.64	$1.60e^{-124}$	45	504	15	18
L. major	LmPDI-2 / LmjF36.6940	Class 1	CGHC /CGHC	5.22	52.38	$6.00e^{-88}$	36	477	0	21
L. infantum	LinJ36_V3.72 80	Class 1	CGHC /CGHC	5.43	52.34	$1.30e^{-85}$	35	477	0	21
H. sapiens	PDIA3 / P30101.4	Class 1	CGHC /CGHC	5.98	56.8	$1.00e^{-42}$	27	505	0	25
S. cerevisiae	Pdi1p / EDN62088.1	Class 1	CGHC /CGHC	4.38	58.2	$3.00e^{-37}$	26	522	4	22

Table 3.4 Summary of various attributes of proteins similar to TbPDI-2

The PDI class was determined using the conserved domain search hosted by NCBI; active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated.

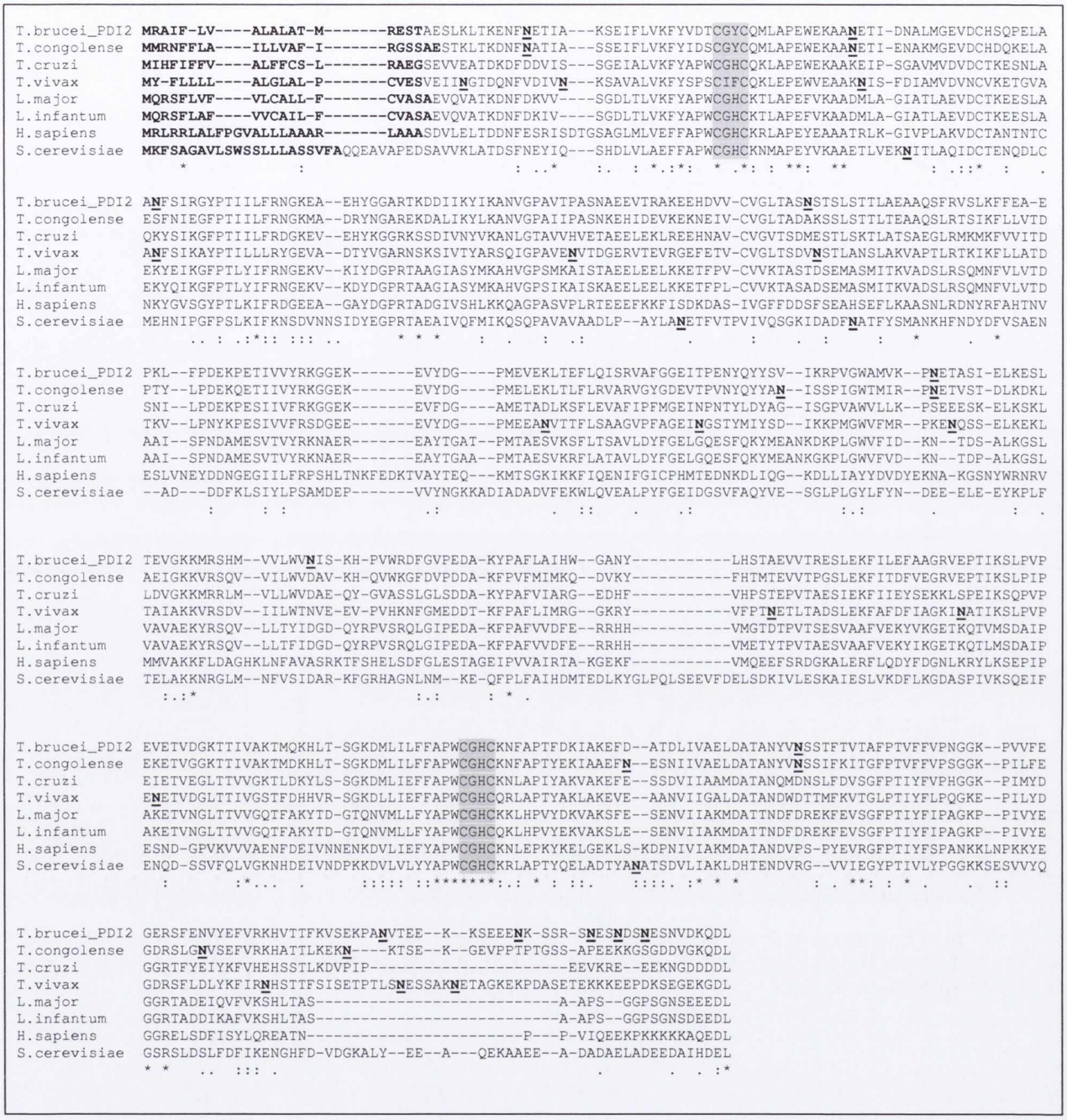


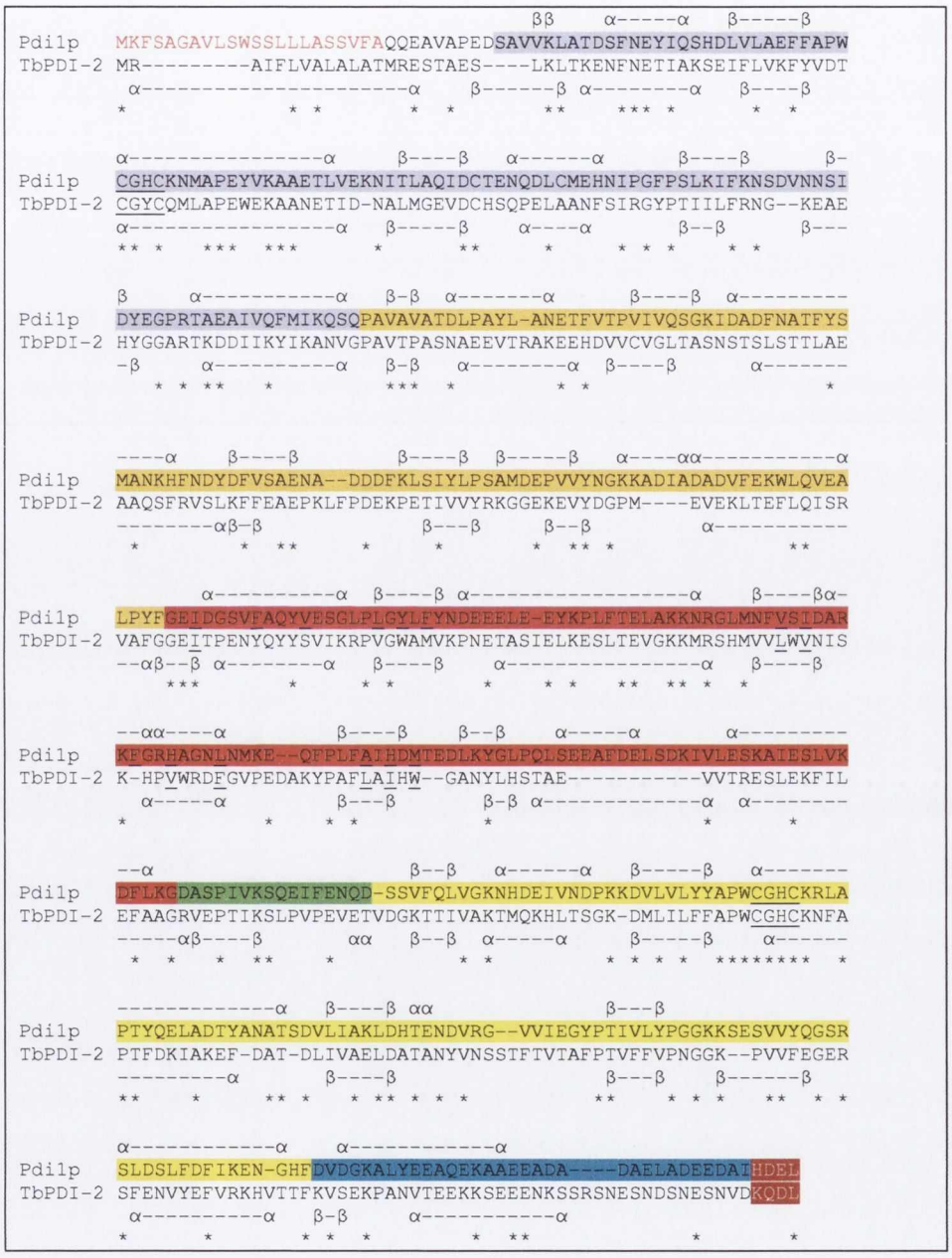
Figure 3.3 Multiple sequence alignment of closely related proteins of TbPDI-2

Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey and predicted *N*-glycosylation sites are underlined. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

Figure 3.4 Secondary structure comparison between TbPDI-2 and yeast Pdi1p.

Alignment produced using T-Coffee. The protein sequence of Pdi1p was labeled for known α helices, β regions and domain structure using the published 3d crystal structure (see section 0). The α helices and β regions of TbPDI-2 were predicted using PSIPRED (bioinf.cs.ucl.ac.uk). Although sequence similarity between the two sequences appears to be low between the two proteins, it is evident that the predicted secondary structure of TbPDI-2 shows clear similarities with the solved structure of baker's yeast Pdi1p. The domain structure of Pdi1p is represented in different colours as follows: N- a domain b domain b' domain Linker a'domain ER retention signal-C.

Figure 3.4 Secondary structure comparison between TbPDI-2 and yeast Pdi1p.



3.1.6 Multiple sequence alignment of TbPDI-3 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify the closest related proteins to TbPDI-3 in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The sequences of the proteins were reviewed for various attributes (Table 3.5) and were aligned using the program T-Coffee (Figure 3.5).

The protein found to be most similar to TbPDI-3 was the putative class 2 PDI TcIL3000_7_920 found in *T. congolense*. The alignment showed a strong sequence similarity with a % identity of 73% and an expectation number of $8.80e^{-149}$. The active site motif of these class 2 PDIs are identical as they were found to be in all the sequences aligned. This suggests class 1 PDIs may require both active sites to carry out their enzymatic function. Another feature of these PDIs, that can be observed, is their lack of *N*-glycosylation sites, except in the case of *T. vivax* which contains one predicted site for *N*-glycosylation. This PDI of *T .vivax* also has a high similarity with a % identity of 63% and an expectation value of 1.2e-127. The PDIs aligned from trypanosome species chosen (*T. brucei, T. congolense, T. vivax, T. cruzi, L. infantum* and *L. major*) all share similar predicted attributes including mass and pI suggesting they all undergo a similar conserved function.

The PDI from *Arabidopsis thaliana* known as PDIL2-1 (UniProt accession: O22263) gave a significant alignment with TbPDI-3. The alignment gave a very strong hit of 44% identity and an expectation value of $1e^{-78}$. The PDIL2-1 protein shares many of the attributes of TbPDI-3 including identical active site motif, PDI class and a lack of possible *N*-glycosylation sites. The PDI found to be most similar in *H. sapiens* to TbPDI-3 is the class 2 PDI known as PDIA6 (also known as PDI P5 UniProt accession: Q15084). The alignment gave a % identity of 34% with expectation value of $5e^{-35}$. This score is still significant although not as high as the other species included. This difference is due to the length of PDIA3 which is 63aa longer than TbPDI-3. The predicted pI of PDIA6 also shows a significant difference with TbPDI-3. One final difference observed is the presence of an endoplasmic reticulum retention signal

(KDEL) which is not present in any of the other PDIs aligned. Despite these differences PDIA3 shares a similar domain structure and so is comparable to TbPDI-3.

Species	Name/acession	PDI class	Active site	pI (pH)	Mass (kDa)	E value score	% identity	size (aa)	*N*-glycosylation sites	signal peptide predicted cleavage site
T. brucei	TbPDI-3/ Tb927.7.1300	Class 2	CGHC/CGHC	8.66	41.94			377	0	27
T. congolense	TcIL3000.7.920	Class 2	CGHC/CGHC	7.43	41.78	$8.80e^{-149}$	73	378	0	29
T. vivax	TvY486_0701000	Class 2	CGHC/CGHC	8.72	41.91	$1.20e^{-127}$	63	384	1	21
T. cruzi	Tc00.104705350 8209.140	Class 2	CGHC/CGHC	6.82	41.8	$3.60e^{-108}$	52	376	0	28
L. infantum	LinJ26_V3.0630	Class 2	CGHC/CGHC	6.67	40.79	$8.50e^{-87}$	43	377	0	25
L. major	LmjF26.0660	Class 2	CGHC/CGHC	7.37	40.7	$6.20e^{-86}$	43	377	0	25
A. thaliana	PDIL2-1/ NP_182269.1	Class 2	CGHC/CGHC	4.95	48.1	$1.00e^{-78}$	44	361	0	22
H. sapiens	NP_005733.1	Class 2	CGHC/CGHC	5.8	39.5	$5.00e^{-35}$	34	440	0	19

Table 3.5 Summary of various attributes of proteins similar to TbPDI-3

The PDI class was determined using the conserved domain search hosted by NCBI; active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated.

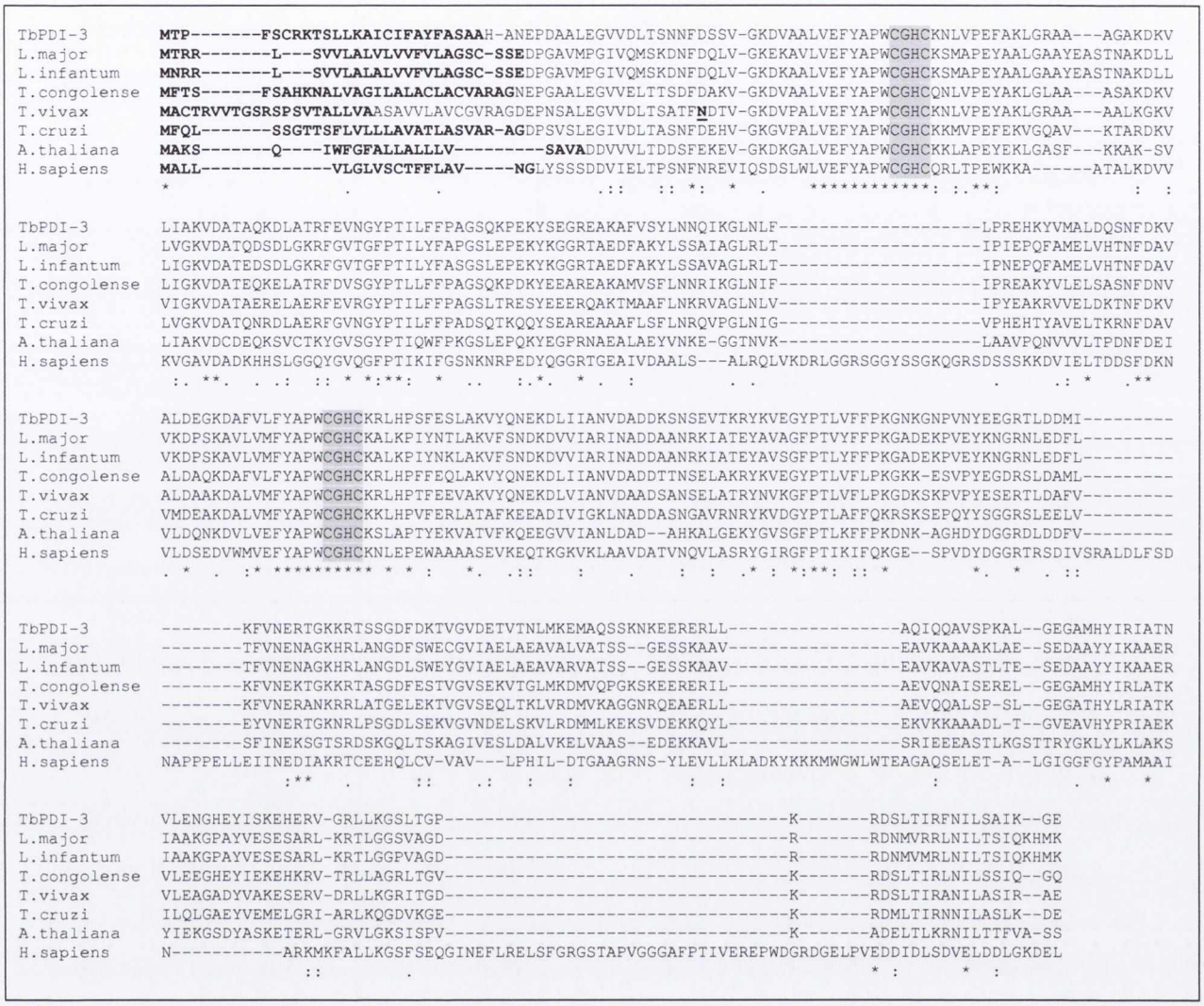


Figure 3.5 Multiple sequence alignment of closely related proteins of TbPDI-3

Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey and predicted *N*-glycosylation sites are underlined. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

3.1.7 Multiple sequence alignment of TbPDI-4 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify the proteins most closely related to TbPDI-4 in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The sequences of the proteins were reviewed for various attributes (Table 3.6) and were aligned using the program T-Coffee (Figure 3.6).

The single domain PDI known as TbPDI-4 was found to contain a very unusual feature, a transmembrane spanning region. While there are other examples of single domain PDIs with transmembrane spanning regions in other species, no significantly similar proteins were found using TbPDI-4 as a search query, besides the closely related trypanosome species which were included. The PDI which gave the most similar hit was the *T. congolense* PDI (TcIL3000_5_770). The PDI showed significant similarity with a % identity of 62% and an expectation value score of 2.5e-190. All possible TbPDI-4 homologues included in this alignment share similar attributes including size, pI, mass, signal peptide, predicted *N*-glycosylation sites and the very unique transmembrane spanning region. This finding suggested that this PDI is conserved and undertakes some function during proteins folding. One difference between the species *T .brucei* and *T .cruzi* or *L .major* is the substitution of a glycine for an alanine or a proline in the active site. This substitution would seem likely as all three amino acids are non-polar and neutral but structurally it would seem the active site would be altered by such a substitution. There have been no studies published on these single domain PDIs but their unusual structure would suggest a very specific and interesting function.

Species	Name/ accession	PDI class *	Active site	pI (pH)	Mass (kDa)	E value score	% identity	size (aa)	*N*-glycosylation sites	Signal peptide predicted cleavage site
T. brucei	TbPDI-4 / Tb927.5. 1020	sdTMPDI	CGHC	6.58	65.72			589	1	29
T congolense	TcIL3000. 5.770	sdTMPDI	CGHC	6.83	66.81	$2.50e^{-190}$	62	596	2	34
T. cruzi	Tc00.1047053511 181.20	sdTMPDI	CAHC	6.04	60.51	$3.00e^{-134}$	54	536	2	26
T. vivax	TvY486_0500620	sdTMPDI	CGHC	6.98	61.88	$1.10e^{-125}$	48	562	2	32
L. major	LmjF35.1330	sdTMPDI	CPHC	5.9	63.48	$2.10e^{-91}$	39	557	1	32

Table 3.6 Summary of various attributes of proteins similar to TbPDI-4

The PDI class was determined using the conserved domain search hosted by NCBI; active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated. * Single domain with transmembrane region (sdTMPDI).

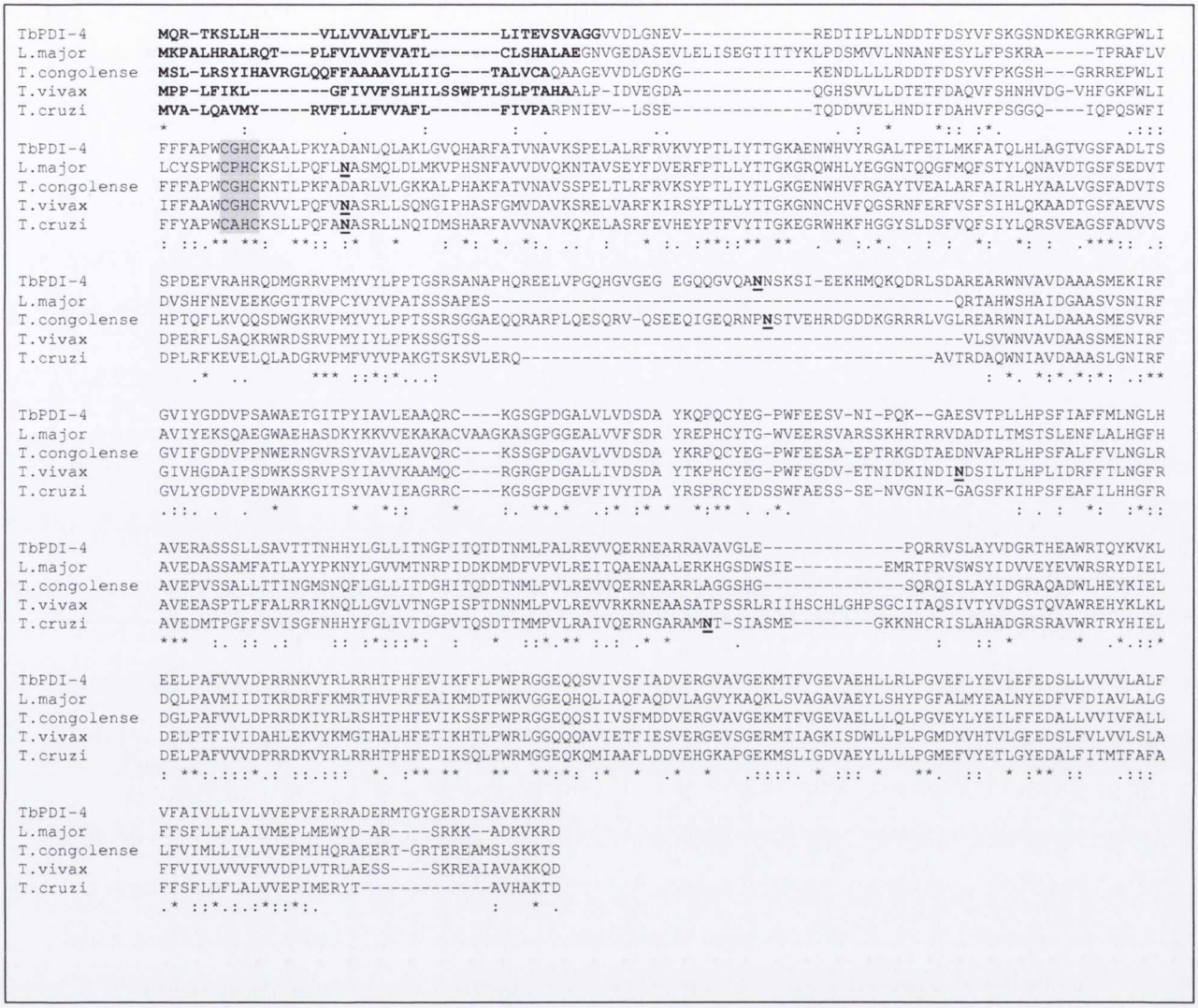


Figure 3.6 Multiple sequence alignment of closely related proteins of TbPDI-4

Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey and predicted *N*-glycosylation sites are underlined. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

3.1.8 Multiple sequence alignment of TbPDI-5 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify the proteins most closely related to TbPDI-5 in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The sequences of the proteins were reviewed for various attributes (Table 3.7) and were aligned using the program T-Coffee (Figure 3.7).

The protein which showed the highest similarity was the single domain PDI from *T. congolense*. The sequence showed a % identity of 68 and an expectation value of 2e-46. Other trypanosome species gave significant hits with very similar proteins. It appears TbPDI-5 is conserved across these species and share similar attributes including size, mass, signal peptide and no predicted sites for *N*-glycosylation. No work has been published on these possible homologous except for the *L. major* single domain PDI which was shown to be constitutively expressed at the mRNA level (Leifso et al., 2007). One interesting hit was found with a class 2 PDI found in the plant species *A. thaliana* (UniProt accession: B9DI05). This gave a significant score of 49% identity with an expectation score of $3.9e^{-29}$ despite having difference in sequence length of 114aa. Little is known about this class 2 PDI but it is interesting to see such a similarity and it is possible that these PDIs share a common function. Although TbPDI-5 only shares half the active sites of the class 2 proteins it may share similar substrates due to this active site similarity.

Species	Name/ accession	PDI class	Active site	pI (pH)	Mass (kDa)	E value score	% identity	Size (aa)	*N*-glycosylation sites	signal peptide predicted cleavage site
T. brucei	TbPDI-5/ Tb927.7.5790	Single domain	CGHC	8.52	15.4			135	0	23
T. congolense	TcIL3000. 7.4770	Single domain	CGHC	9.1	15.62	$2.00e^{-46}$	68	140	0	22
T. vivax	TvY486_07 05890	Single domain	CGHC	7.48	14.17	$3.60e^{-32}$	53	128	0	27
T. cruzi	Tc00.10470 53508173.150	Single domain	CVHC	9.74	16.05	$6.60e^{-29}$	48	141	0	30
L. major	LmjF06.1050	Single domain	CGHC	8.75	15.06	$3.90e^{-29}$	45	133	0	24
A. thaliana	BAH20372	Class 2	CGHC	5.49	36.56	$1.00e^{-27}$	49	249	0	22

Table 3.7 Summary of various attributes of proteins similar to TbPDI-5

The PDI class was determined using the conserved domain search hosted by NCBI; active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated.

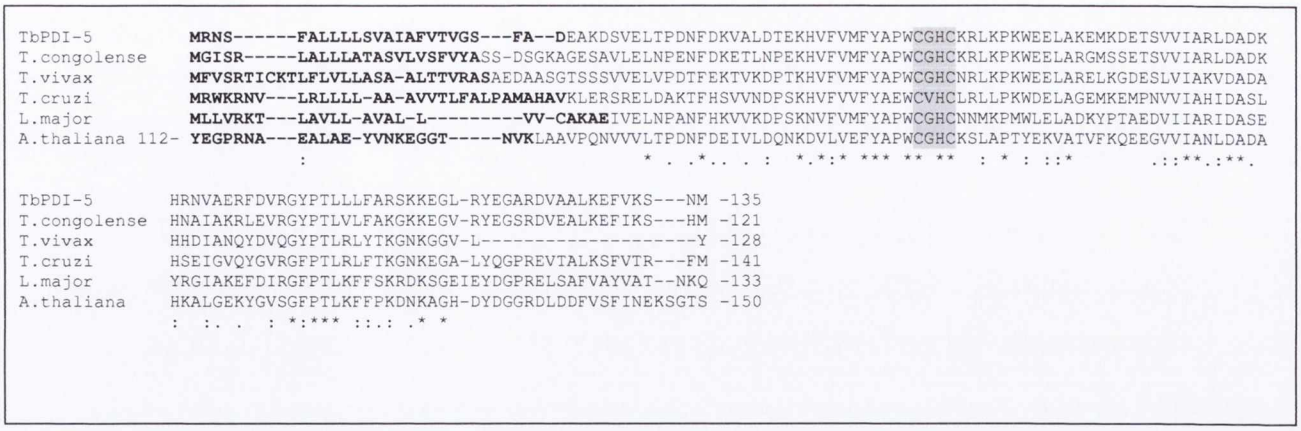


Figure 3.7 Multiple sequence alignment of closely related proteins of TbPDI-5

Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey and predicted *N*-glycosylation sites are underlined. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

3.1.9 Multiple sequence alignment of TbPDI-6 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify the proteins most closely related to TbPDI-6 in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The sequences of the proteins were reviewed for various attributes (Table 3.8) and were aligned using the program T-Coffee (Figure 3.8).

The protein found to be most similar to TbPDI-6 was the single domain PDI found in *T. congolense*. It gave a significant similarity with a % identity of 71% and an expectation value of $1.9e^{-50}$. The PDIs chosen shared common features including signal peptide, size and mass. An interesting difference from other single domain PDIs seen in *T. brucei* is the substitution of a serine for a cysteine in the active site. Included in the alignment is TbPDI-5 as it was located sequentially to TbPDI-6 on chromosome 7. However, the proteins were found to share little similarity with a % identity of 32% with an E value score of $9e^{-12}$.

Species	Name/ accession	PDI class	Active site	pI (pH)	Mass (kDa)	E value score	% identity	size (aa)	*N*-glycosylation sites	signal peptide predicted cleavage site
T. brucei	TbPDI-6 / Tb927.7. 5780	single domain	SRHC	5.73	17.03			151	0	23
T. congolense	TcIL3000. 7.4760	single domain	SRRC	5.89	16.56	$1.90e^{-50}$	71	148	0	25
T. vivax	TvY486_ 0705880	single domain	SRHC	5.44	16.3	$2.70e^{-50}$	67	144	0	21
T. cruzi	Tc00.1047 053508173. 160	single domain	SRHS	9.74	16.05	$6.00e^{-35}$	54	153	0	31
L. infantum	LinJ06_ V3.1240	single domain	SRQC	6.5	16.49	$7.90e^{-29}$	46	147	0	33
T. brucei	TbPDI-5/ Tb927.7. 5790	single domain	CGHC	8.52	15.4	$9.00e^{-12}$	32	135	0	23

Table 3.8 Summary of various attributes of proteins similar to TbPDI-6

The PDI class was determined using the conserved domain search hosted by NCBI; active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated.

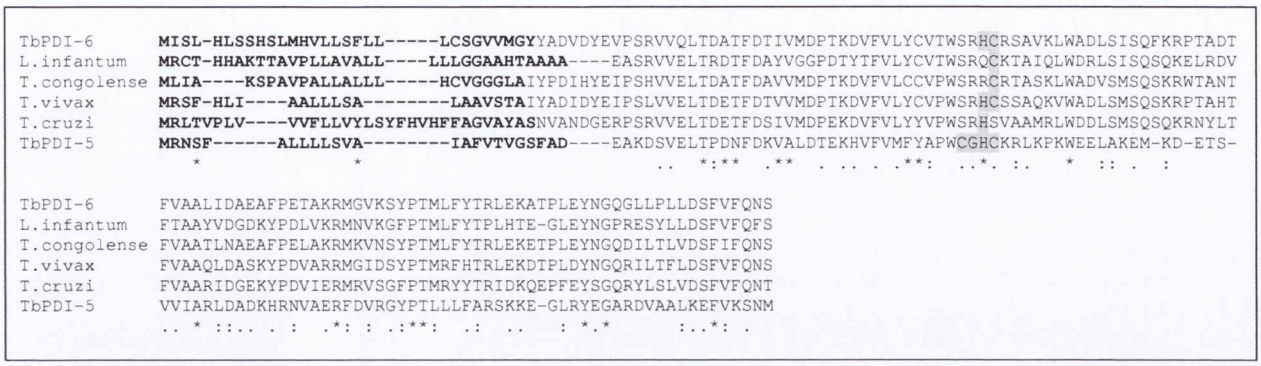


Figure 3.8 Multiple sequence alignment of closely related proteins of TbPDI-6

Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

3.2 Endoplasmic reticulum oxidoreductin of *Trypanosoma brucei*

The protein endoplasmic reticulum oxidoreductin 1 (Ero 1) is considered to be the principal provider of oxidizing equivalence to the endoplasmic reticulum. In many species it has been shown to interact with PDI and has been shown to have the ability to oxidize PDI *in vitro* (Frand and Kaiser, 1999, 2000; Tu et al., 2000). Currently, PDI and Ero 1 are considered to be the principal enzymes of the oxidative folding pathway of the endoplasmic reticulum and as part of this investigation into the oxidative folding pathway a bioinformatics search was carried out to identify all possible Ero 1 proteins in the genome of *T. brucei.*

Unlike the PDI enzymes, most species have very few Ero 1 enzymes. The unicellular yeast *S. cerevisiae* has a single well characterized Ero 1 enzyme where as *H. sapiens* possess two known Ero 1 enzymes (Ramming and Appenzeller-Herzog, 2012). As there were so few enzymes typically found in each species it was also expected that *T. brucei* may also contain few if only one Ero 1. The Ero 1 enzymes are defined by their single Ero 1 domain, which contain 5 conserved cysteine pairs.

While Ero1p, the Ero 1 of *S. cerevisiae,* was found to be essential for cell viability (Frand and Kaiser, 1998; Pollard et al., 1998) it is important to note that the human orthologs Ero1α and Ero1β were not (Zito et al., 2010). As such, one of the goals of this bioinformatics analysis of trypanosome Ero 1 was to determine if trypanosome Ero 1 proteins shared similar traits to the Ero1 proteins found in yeast or humans to predict if the trypanosome Ero 1 proteins are essential.

The goals of this bioinformatics analysis of the Ero 1s of *T. brucei* were as follows: Firstly, identify all possible Ero 1s in the genome of *T. brucei* by carrying out a genome wide search for all proteins possessing Ero 1 like domains. Secondly, assess the other attributes of these candidate proteins; specifically *N*-terminal leader sequence and catalytic domains. Lastly, to predict the possible functions of the candidate Ero 1s by comparison to well known Ero 1 orthologs from other species.

3.2.1 Identification of possible Ero 1s of *Trypanosoma brucei*

To identify possible candidate Ero 1 proteins, the Sanger sequencing project search service was employed to search the genome of *T. brucei* for the Pfam keyword "endoplasmic reticulum oxidoreduction". One protein was found with the systematic gene id: Tb927.8.4890. The sequence was reviewed for specific conserved domain hits with Pfam which returned a hit for ERO1 with an E value score of $9.5e^{-90}$. Panther also returned an ERO1-RELATED hit with an E value score of $3e^{-40}$. This protein was designated TbEro 1 (Tb927.8.4890).

3.2.2 Primary sequence analysis of candidate TbEro 1

Various physical parameters of the candidate TbEro 1 were predicted using different programs from the ExPASy website (http://ca.expasy.org/tools) including predicted mass, isoelectric point, *N*-glycosylation sites, the presence of a signal peptide, transmembrane span and endoplasmic reticulum retention signal. The results are summarized (Table 3.9). The putative protein TbEro 1 was found to have a predicted mass of 49.2 kDa which is typical of Ero 1 proteins. Unusually the number of possible *N*-glycosylation sites was quite high suggesting TbEro 1 may be glycosylated.

Resident proteins of the endoplasmic reticulum usually contain characteristic features. These include an *N*-terminal leader sequence, a *C*-terminal retrieval signal sequences or association with the membrane of the endoplasmic reticulum via transmembrane regions. To identify these hydrophobic regions, TbEro 1 was subjected to analysis by a series of three programs: TMpred, Kyte-Doolittle and TopPred from the ExPASy website (http://ca.expasy.org/tools). Several programs were used to predict hydrophobic regions. The regions which were found to be hydrophobic by all three programs were the most likely to be correct and so only those regions were taken into account when constructing Table 3.9. The resulting hydrophobic graphs can be seen in section 7.2 of the Appendix. The protein was found to contain a predicted *N*-terminal signal peptide and as with all Ero 1 proteins reported in the literature contained no predicted transmembrane span. Additionally no endoplasmic reticulum retention signal was identified.

Open reading frame size (aa)	Predicted mass (kDa)	Predicted isoelectric point (pH)	*N*-glycosylation sites	Signal peptide	Transmembrane span	Previously reported activity	ER retention signal
442	49.2	8.46	7	Yes	No	None	None

Table 3.9 Various physical parameters of candidate Ero 1

Various physical parameters of the candidate TbEro 1 were predicted using different programs from the ExPASy website (http://ca.expasy.org/tools) including predicted mass, isoelectric point, *N*-glycosylation sites and the presence of a signal peptide. The conserved domain search was employed (www.ncbi.nlm.nih.gov.elib) to locate any Ero 1 conserved domains. A search for transmembrane spanning regions of the candidate proteins was carried out using three separate programs: TMpred, Kyte-Doolittle and TopPred. Results which agree were considered to be strong hits and are indicated.

3.2.3 Multiple sequence alignment of TbEro 1 & closely related proteins

The program omniBLAST (hosted by the Sanger institute) and the program BLASTp (hosted by NCBI) were used to identify the proteins most closely related to TbEro 1 in the species contained within the Sanger Institute's database and the non-redundant protein data base hosted by NCBI. The proteins were selected based on their % identity and their E-value score. The sequences of the proteins were reviewed for various attributes (Table 3.10) and were aligned using the program T-Coffee (Figure 3.9).

The protein found to show the highest similarity to TbEro 1 was the uncharacterized Ero 1 found in *T. congolense* (TcIL3000_0_00060). The sequence showed a % identity of 56% and an E-value score of $1.3e^{-138}$. As with all species included in this multiple sequence alignment this protein contains a signal peptide, *N*-glycosylation sites and a highly conserved domain structure. The active site was found to be identical in TbEro 1 (CVTCEKC) as was the case for the Ero 1 homologues found in *T. cruzi* and *L. major*. The Ero 1 of *T. vivax* showed a high similarity with a % identity of 53% and an expectation value of $6.3e^{-115}$ but did not share this identical active site. All trypanosome species only contained one homologue of Ero 1 in their genome, a feature common to the yeast *S. cerevisiae*. This is interesting as it shows a distinct difference to metazoan species which usually contain two or more Ero 1 proteins in their genomes.

A homologue of TbEro 1 was found in the plant species *A. thaliana* known as AERO1 (UniProt accession: Q9C7S7). This protein has been characterized *in vitro* and has shown to be a glycoprotein although there is only one *N*-glycosylation site present in the sequence (Dixon et al., 2003). AERO1's subcellular location has not yet been shown experimentally and there have been no studies reported to date on this protein *in vivo*.

The genome of *H. sapiens* was searched for TbEro 1's closest match and the protein known as ERO1-β gave the most similar result with a % identity of 29% and an expectation value of $4e^{-34}$.The protein ERO1-β contained a slightly different active site motif (CVGCEKC), which was common to the other metazoan Ero 1 proteins included. The protein sequence contained 4 N- glycosylation sites which was similar to the trypanosome Ero 1 proteins which may suggests ERO1-β glycosylation may play an interesting role in the function of ERO1-β .

The sequence which gave a significant similarity in the yeast species *S. cerevisiae* to TbEro 1 was the protein Ero1p (UniProt accession: Q03103) which was extensively discussed in the introduction. The alignment gave a % identity of 25% with an E-value score of $5e^{-27}$. The difference in size between TbEro 1 and Ero1p is quite significant but these extra amino acids appear to reside in a non catalytic region. As can be seen in the alignment, two of the catalytic cysteines are well conserved in all species included, suggesting that domain structures are similar. Also a conserved *N*-glycosylation site can be seen preceding the active site. The similarities and differences were investigated further with a comparison of predicted secondary structure of TbEro1 and the solved secondary structure of Ero1p (Figure 3.10). As can be seen in the figure, TbEro 1 and Ero1p show a very high similarity between secondary structures. This suggests TbEro 1 may behave in a similar fashion to Ero1p.

Species	Name/ accession	Active site	pI (pH)	Mass (kDa)	E value score	% identity	Size (aa)	*N*-glycosylation sites	signal peptide predicted cleavage site
T. brucei	TbEro 1/ Tb927.8.4890	CVTCEKC	8.46	49.2			442	7	24
T. congolense	TcIL3000. 0.00060	CVTCEKC	6.31	49.68	$1.30e^{-138}$	56	444	6	19
T. vivax	TvY486_ 0804340	CVNCEKC	7.41	49.21	$6.30e^{-115}$	53	440	7	25
T. cruzi	Tc00.104705 3508837.189	CVTCEKC	5.81	50.18	$2.40e^{-111}$	48	443	5	18
L. major	LmjF16. 1530	CVTCEKC	5.19	48.18	$3.90e^{-91}$	46	432	3	33
A. thaliana	AERO1/NP_ 177372.1	CVGCEKC	6.17	53.83	$9.00e^{-53}$	31	469	1	36
H. sapiens	ERO1β/ AF252538_1	CVGCEKC	8.25	53.53	$4.00e^{-34}$	29	467	4	33
Drosophila. grimshawi	XP_0019 83494.1	CVGCEKC	5.18	56	$7.00e^{-35}$	28	486	1	33
S. cerevisiae	NP_013576.1	CVQCEKC	5.48	65	$5.00e^{-27}$	25	563	7	18

Table 3.10 Summary of various attributes of proteins similar to TbEro 1

The active site motifs were manually determined from the multiple sequence alignment of sequences; pI, mass, signal, *N*-glycosylation sites and peptide predicted cleavage sites were predicted using various predictive programs found at expasy.org; % identity and E value scores were determined from the BLASTp or omniBLAST search of the species indicated.

Figure 3.9 Multiple sequence alignment of proteins closely related to TbEro 1

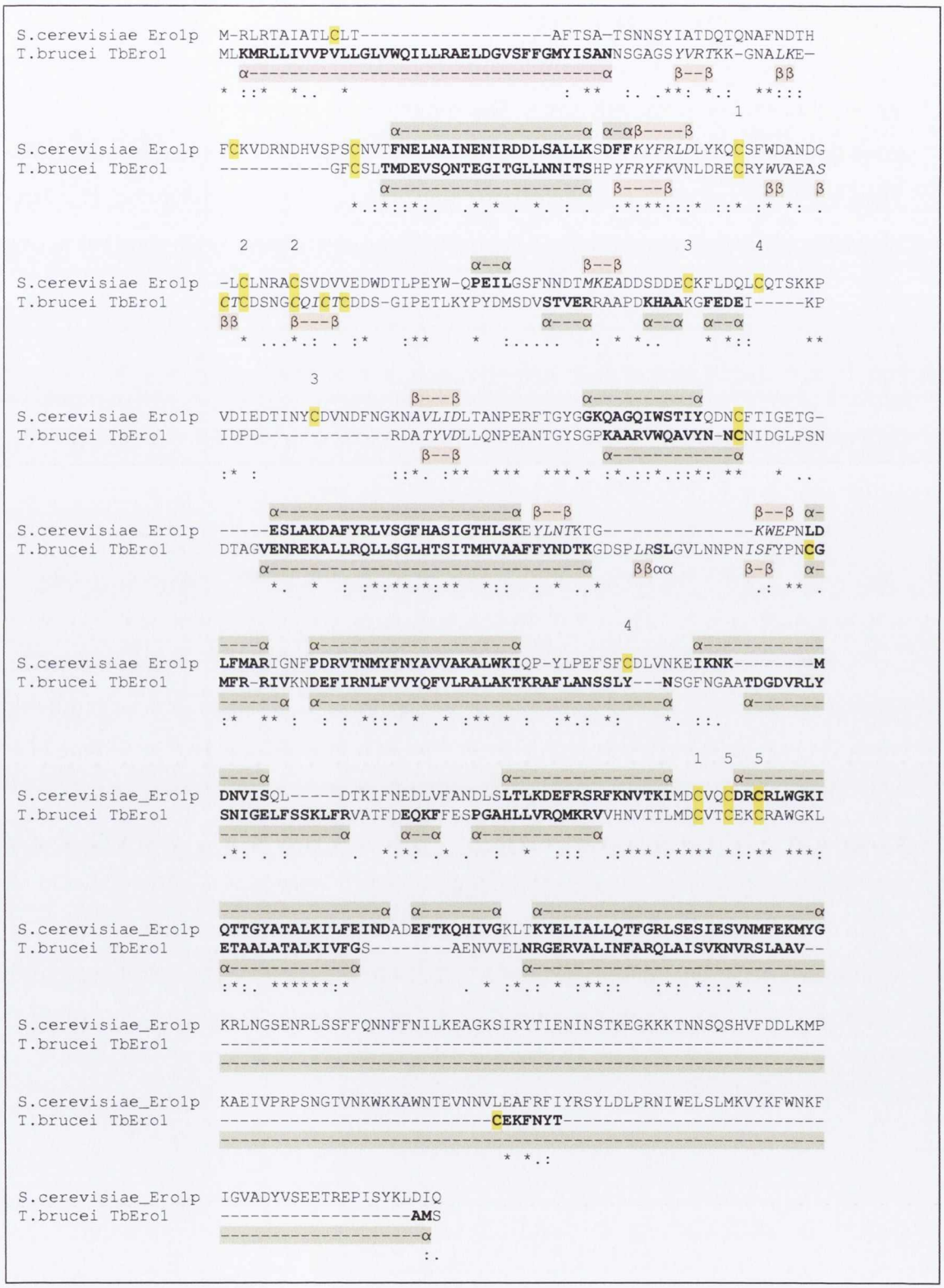
Predicted signal peptides are shown in bold, conserved thiol active sites are shaded in grey and predicted N-glycosylation sites are underlined. Conserved cysteine residues known to form disulphide bonds in the Ero 1 of *S.cerevisiae* are shown in yellow. Active site cysteines are shown in green. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (amino acids with similarly shape).

Figure 3.9 Multiple sequence alignment of proteins closely related to TbEro 1

Figure 3.10 Secondary structure comparison of TbEro 1 and yeast Ero1p

Both sequences were aligned using the program TCoffee. Cysteine residues are highlighted in yellow and those which are known to form disulphide bonds in Ero1p are marked by numbers showing their connectivity. Predicted α-helices and β-sheets are shown below the TbEro 1 sequence, which were predicted using PSIPRED (bioinf.cs.ucl.ac.uk) and known α-helices and β-sheets are shown above the Ero1p sequence for direct comparison. "*" represents identical residues, ":" indicates conserved substitutions and "." represents semi-conserved substitutions (similar shapes amino acids).

Figure 3.10 Secondary structure comparison of TbEro 1 and yeast Ero1p



3.3 Relative quantity of candidate PDIs in bloodstream versus procyclic cells

To determine if the candidate PDIs and TbEro 1 were upregulated in the bloodstream relative to procyclic forms of *T. brucei* the candidate genes were subjected to expression analysis using qRT-PCR. The aim of this experiment was to identify candidate PDIs most likely to be involved in the folding of VSG and other bloodstream stage specific surface proteins which are expressed in bloodstream forms. The PDI candidate, TbPDI-6 was excluded in this analysis due to the lack of a catalytic active site as previously discussed.

Procyclic and bloodstream forms of *T. brucei* were harvested during exponential growth phase. Total RNA was collected and primers were designed and optimized as described in the methods. The relative level of expression of the selected gene transcripts was investigated using qRT-PCR and relative expression ratios (bloodstream / procyclic form) were obtained using total RNA. Actin was chosen as the reference gene for normalization as it is a well established reference gene for normalization of the two *T. brucei* forms (Spitznagel *et al.* 2009). Reactions were carried out in triplicate and experiments were repeated at least once on different RNA preparations. The gene for TbPDI-6 was omitted from the analysis as it did not possess catalytic thiol domains and as such could not be involved in the folding of VSG. The candidate proteins TbPDI-2, TbPDI-5 and TbEro1 were ~three fold upregulated in the bloodstream forms relative to procyclic forms. This result suggests expression of these potential PDIs and the single Ero 1 proteins is upregulated in bloodstream forms. As TbPDI-5 was a single domain PDI it was less likely to play a significant role in the folding of VSG. As such, from the results obtained, it appeared TbPDI-2 and TbEro 1 were most likely to be the principal enzymes involved in the folding of VSG.

Gene ID	Designated name	BSF/PCF ratio
Tb927.8.4890	TbEro 1	2.7 ± 0.14
Tb927.4.2450	TbPDI-1	1.9 ± 0.12
Tb927.10.8230/Tb10.6K15.2290	TbPDI-2	3.0 ± 0.09
Tb927.7.1300	TbPDI-3	1.4 ± 0.10
Tb927.5.1020	TbPDI-4	0.9 ± 0.12
Tb927.7.570	TbPDI-5	3.8 ± 0.13

Table 3.11 Transcriptional analysis of possible catalytic proteins of the oxidative folding pathway in bloodstream and procyclic forms of *T. brucei*.

Total RNA was isolated from culture cells in the logarithmic phase of growth using the Absolutely RNA® kit from Stratagene. Relative quantification of the transcript was performed using the Brilliant® II Sybr green QRT-PCR one step kit and the MX3000P from Stratagene. Primers sets were designed by the Beacon Designer 7 software and were optimized for the RT and amplification steps. Each pair was also assessed to ensure > than 95% efficiency during the amplification stages. The transcript levels were determined relative to the actin transcript level and are expressed as the bloodstream/procyclic ratio (BSF/PCF) using the MX3000P relative quantification software.

3.4 Discussion

3.4.1 TbPDI candidate proteins

The possible function and location of the PDI-family members has grown over the past decade as more members are identified and other possible members have been classified as part of the PDI family of proteins. This project is concerned with the PDI enzymes considered to form part of the oxidative folding pathway found in the endoplasmic reticulum. As such the genome of *T. brucei* was searched for possible PDI family members and these candidate PDIs were assessed for their possible roles.

The criteria for proteins to be considered putative PDI family members is currently based only on sequence similarity with established PDI enzymes that are found in the endoplasmic reticulum and secretory pathway (Kozlov et al., 2010). Additional criteria were searched for, using various predictive programs, to identify and classify PDIs not only on sequence similarity. The PDI enzymes usually contain one or more thioredoxin like fold, a classical *N*-terminal signal peptide and may contain a *C*-terminal endoplasmic reticulum retention signal at the of the protein. The genome of *T. brucei* was searched for proteins, which had similarity to PDI conserved domains, within the Pfam and Panther databases and these proteins were searched for criteria found in PDIs. Six possible PDIs were identified, including two PDIs previously discussed in the introduction, TbPDI-1 and TbPDI-2. These PDIs were characterized in a previous study and shown to exhibit isomerase activity (Rubotham *et al.* 2005). Of the possible PDIs found, one was a class 1 PDI (TbPDI-2) considered to be a classical PDI, one previously uncharacterized class 2 PDI and four single domain PDIs, one of which was TbPDI-1. The number of PDIs found was considerably lower than the number found in metazoans (*H. sapiens* have 19) but similar to *S. cerevisiae* which have 5 (Ramming and Appenzeller-Herzog, 2012). These possible PDIs have been grouped into six categories: class 1 to 4 PDIs, single domain PDIs and transmembrane spanning PDIs. The lower number of possible PDIs found in *T. brucei* suggests these PDIs may play a similar role in the oxidative folding pathway to that found in yeast, one of which was shown to be essential in culture (Tachibana and Stevens, 1992; Tachikawa et al., 1995, 1997).

3.4.2 Single domain PDIs of *Trypanosoma brucei*

Of the four possible single domain PDIs (sdPDIs) found in *T. brucei*, only TbPDI-1 has been characterized and was shown to have the ability to reduce insulin and to refold scrambled bovine pancreatic RNase type III-A (Rubotham et al., 2005). Two additional sdPDIs were found during the search for possible PDIs carried out in this study; revising the number of sdPDIs found in *T. brucei* to three. The two possible PDIs found were designated TbPDI-5 (systematic gene ID: Tb927.7.5790) and TbPDI-6 (systematic gene ID: Tb927.7.5780). These sdPDIs were found to be located consecutively on chromosome seven, which suggested that a duplication event resulted in these two genes. This phenomenon of gene proximity was observed in the other sdPDIs found in trypanosome species included in the analysis and so it was concluded this duplication event occurred before the speciation of the trypanosomes investigated.

The active site of TbPDI-6 was found to be unusual (SRHC) containing only one cysteine residue. To date, the only PDIs reported to possess oxidation, reduction and isomerisation function also possessed a CXXC active site motif (Kozlov et al., 2010). This strongly suggests that TbPDI-6 may be an inactive PDI. Previous mutational studies of the a-domain active site of the yeast PDIs support this view. These studies were carried out by the substitution of a serine for an active site cysteine, which resulted in a reduction of activity of the PDIs studied (Nørgaard et al., 2001). The investigated PDIs possessed two active domains and as such the PDIs retained some activity through the second active domain. The resulting mutation caused a reduced ability of the yeast to fold protein carboxypeptidase Y (CPY) a substrate of the oxidative folding pathway in yeast (Nørgaard *et al.* 2001). In the case of an sdPDI, an active site mutation would be expected to compromise all activity as there would be no remaining second a-domain. The non-essential class 1 yeast PDI known as EUG1p (UniProt accession: P32474) possesses two catalytically inactive a-domain sites (a: CLHS a': CIHS). It was reported that substitution of the serine for a cysteine in either or both a-domain active site motifs caused EUG1p to become catalytically active to a level comparable to Pdi1p, the essential yeast class 1 PDI. This mutation allowed EUG1p to carry out isomerisation reactions on CPY (Nørgaard and Winther, 2001). It is still unclear what role EUG1p plays in the oxidative folding pathway within yeast, however, the presence of such catalytically inactive PDIs in many species suggests

they do play a role in oxidative folding of proteins within the endoplasmic reticulum, albeit a less critical role.

The mammalian class 1 PDI known as PDIA7 (UniProt accession: Q8N807) possesses two such alternate active site motifs within the two a and a'-domains of the protein (a: SKQS a':SKKC). This PDI has been shown also not to possess the ability to carry out oxidation, reduction or isomerization reactions *in vitro* (van Lith et al., 2007). It has been suggested that PDIA7 may be responsible for substrate retention within the endoplasmic reticulum via the binding of the substrates of PDIA7 through the use of the proteins catalytically inactive domains and the KEEL *C*-terminal endoplasmic reticulum retention motif. This may also be true for EUG1p as this yeast PDI also possesses a HDEL endoplasmic reticulum retrieval like signal. However, the trypanosome TbPDI-6 does not possess additional a and b-domains or any possible endoplasmic reticulum retention signal and as such it seems unlikely TbPDI-6 may share this ability to retrotranslocate proteins to the endoplasmic reticulum. The active site motif SRHC is of some interest however, as the substrate trapping experiments of Jessop et al. utilized the mutagenesis of the *C*-terminal active site cysteine to an alanine, in various PDIs, to identify possible PDI substrates (Jessop et al., 2007). This mutagenesis of the active sites of the a-domain allowed a longer interaction of substrates with the mutant PDIs to allow the purification and analysis of these possible substrates. Trypanosome TbPDI-6 may also possess this ability of substrate trapping and it is tempting to speculate that TbPDI-6 and other PDIs, which possess these CXXS active sites, may play a role in the oxidative folding pathway as PDIs which trap substrates. This seems likely as these PDIs are so numerous for example the following mammalian PDIs all possess only inactive a-domains: ERp44, TMX2, AGR2 and AGR3. The further investigation of this possibility however falls outside the scope of this study and so was not pursued any further.

Although TbPDI-5 is greatly reduced in size, when compare to other PDIs, it may still possess isomerase activity as other minimal single domain PDIs have been characterized in *G. lamblia* which have shown the ability to catalyze the refolding of "scrambled" bovine pancreatic RNase type III-A (Knodler *et al.* 1999). The proteins TbPDI-5 and to a lesser extent TbPDI-6 may represent the orthologs of these minimal PDIs found in unicellular organisms.

The sdPDI, TbPDI-4, was found to possess a possible transmembrane spanning region. This unusual subgroup of PDI has been observed in *Drosophila melanogaster*, and *Caenorhabditis elegans* (Ko and Chow, 2002) but not in plants, fungi or prokaryotes. It has also been observed in *Homo sapiens* where four putative transmembrane spanning PDIs have been identified and given the names TMX 1 to 4 (Kozlov et al., 2010). The thioredoxin-like domain of TMX-1 is considered to be present in the lumen of the endoplasmic reticulum and has been shown to exhibit reductase and isomerase activity *in vitro* (Matsuo et al., 2004). The physiological role of these TMX single domain PDIs is still unclear, but due to their location they may act on membrane anchored substrates. This presents an instance where *T. brucei* offers a simplified system of transmembrane PDIs as only one was found during a genome wide search of *T. brucei*. This offers an opportunity for the understanding of the possible role of such an enzyme *in vivo*. However, further investigation into this class of PDI fell outside of the scope of this investigation and so TbPDI-4 was not analyzed further using *in vivo* methods.

3.4.3 TbPDI-3 (Tb927.7.1300)

Of the putative PDIs identified during the search, a class 2 PDI was identified and was designated the name TbPDI-3. Class 2 PDIs possess two catalytic a-domains but, unlike class 1 PDIs, possess no b-domains between them. For example the human class 2 PDI PDIA6 (UniProt accession: Q15084) possesses the domain configuration a-a'-b (Jordan et al., 2005). The putative protein TbPDI-3 has not been characterized and it is not known if it possesses isomerase or reducing activity *in vitro*. The human PDIA6 has been shown to have both isomerase and chaperone activities (Kikuchi *et al.* 2002), but both were found to be lower than those found in the protein PDIA1 (class 1 PDI of *H. sapiens*). Mutational analysis of the active site cysteines of PDIA6 showed that the *N*-terminal active site cysteines were required for isomerase activity but the chaperone activities (as measured using citrate synthase as the substrate) were still present (Kikuchi et al., 2002). It has been shown that PDIA6 may be involved in a complex with the human proteins PDI, ERp72 and BiP which obstructs the cytosolic face of the translocation protein Sec61p the function of which has yet to be understood. However, it is unclear how a protein with an endoplasmic reticulum retrieval signal can exit the endoplasmic reticulum (Stockton *et al.* 2003). Due to the similarities in active site motif and domain structure of TbPDI-3 and PDIA6 it is likely

they may share a similar role, although according to current opinion it appears to be a less significant role in the oxidative folding pathway than that played by class 1 PDIs and so for the purposes of this study it was not the first choice for further investigation.

3.4.4 TbPDI-2 (Tb927.10.8230)

The class 1 PDI found in the genome of *T. brucei* was previously identified and characterized as a PDI , which was able to refold scrambled bovine pancreatic RNase type III-A but was unable to reduce the disulphide bonds between insulin chains (Rubotham *et al.* 2005). This experiment, however, was not attempted in the presence of an Ero 1 like protein, which may be required for such a reaction to occur. Other aspects were previously discussed in the introduction and so will not be repeated.

The human PDI found to be most similar to TbPDI-2 was the class 1 PDI known as PDIA3 (UniProt accession: P30101). This human PDI has been shown to catalyze oxidation, reduction and isomerisation reactions and is considered to be one of the principal proteins involved in these reactions within the lumen of the endoplasmic reticulum (Kozlov et al., 2010). The activity of PDIA3 was shown to be equal to other PDI-family member characterized but higher than the aforementioned class II PDIA6, which implies TbPDI-2 may be catalytically more active than TbPDI-3 (Alanen et al., 2006).

Substrate binding studies involving mutated PDIA3, which had a *C*-terminal active site mutation for the purposes of substrate trapping, was expressed in human fibroblast cells. The cells were treated with 25 mM *N*-ethylmaleimide to remove all proteins possessing free thiols and resulting PDIA3 & substrates were purified using anti-PDIA3 antibodies. The bound substrates were later identified using 2D gel electrophoresis and indicated PDIA3 interacted with an Ero 1 like protein (Ero1α) and glycoproteins destined to be secreted from the cell (Jessop *et al.* 2009). While this experiment indicated PDI may interact with Ero 1 and secretory substrates *in vivo*, artefacts have been known to occur during these types of mixed disulphide experiments leading to incorrect conclusions. This is unlikely in the case of Jessop et. al. as several approaches were taken to show this PDIA3 Ero1α interaction and all the results obtained were in agreement. The collective evidence suggests the class 1 PDI

of *H. sapiens* interact with Ero 1 and secretory proteins (Kozlov et al., 2010). Due to the sequence similarity found during the BLASTp search reported, it is likely that TbPDI-2 folds glycoproteins in the endoplasmic reticulum and is reoxidized by TbEro 1, which TbPDI-2 interacts with to form the major oxidative folding pathway in *T. brucei*.

Significantly PDIA3 has been shown to be essential *in vivo* in mice through gene knockout studies. However, cultured T-lymphocyte cells were able to survive without PDIA3 indicating the protein played an essential physiological role *in vivo* but was not required for cell proliferation in culture (Soldà et al., 2006). TbPDI-2 was also found to be similar to the yeast PDI known as PDIp not only in sequence and domain structure but also in secondary structure. This PDI was found to be essential in culture (Tachibana and Stevens, 1992). This finding suggests PDIs may be required in single celled organisms where the PDIs play an important role in cell growth; in contrast, metazoan species which use PDI to fold proteins destined for secretion from the cell (Jessop et al., 2007). From the data collected from these similar PDIs, it appears TbPDI-2 may also be essential in culture.

The leishmania ortholog of TbPDI-2, LmPDI-2 was found to share many similarities with TbPDI-2 and so may play a similar role in the endoplasmic reticulum. Interestingly, LmPDI-2 was found to be expressed at higher levels in the most virulent strains of wild type *L. major* collected directly from human lesions (Achour, Chenik et al. 2002). Western blots revealed expression levels were almost undetectable in less virulent strains and as such the protein was proposed as a possible drug target. Inhibitors of PDI were proposed to be administered in a topical cream to the site of infection. Unlike TbPDI-2, the mRNA level of LmPDI-2 was found to be the same for all life cycle stages but when measured at the protein expression level using Western blot, it was found to be highly expressed at the amastigote stage, which divides in the mammalian host. In this study it was shown by qRT-PCR that TbPDI-2 was upregulated three fold in the bloodstream form life cycle stage when compared to the procyclic form. This finding was previously shown at the protein level also (Rubotham et al., 2005). The similarities between TbPDI-2 and LmPDI-2 suggest TbPDI-2 may also play a role in the virulence of *T. brucei* by efficiently processing the VSG.

3.4.5 TbEro 1 (Tb927.8.4890)

As previously described in the introduction, Ero 1 proteins have been shown in yeast and mammals to form part of the oxidative folding pathway. Their role is to reoxidize PDIs within the endoplasmic reticulum and to impart this electron pair to molecular oxygen, producing hydrogen peroxide (Ramming and Appenzeller-Herzog, 2012). The genome of *T. brucei* was searched for Ero 1 like proteins using a Pfam keyword search as described previously. Only one Ero 1 like protein was found and was previously uncharacterized. This putative protein was reviewed for conserved domain hits with various databases. The protein was termed TbEro 1. A BLASTp was carried out using the sequence of TbEro 1 and the results indicated that all the trypanosome species only contained one Ero 1-like gene in their genome; a feature common to the yeast *S. cerevisiae*. Interestingly metazoan species usually contain two or more genes for putative Ero 1 like proteins in their genomes (Ramming and Appenzeller-Herzog, 2012).

The protein found to have the highest similarity to TbEro 1 in humans was the second Ero1-like protein known as ERO1-β. Extensive work on ERO1-β and the other Ero 1 homolog found in *H. sapiens* known as ERO1-α have revealed some interesting features of these proteins. Both are resident proteins of the endoplasmic reticulum but do not contain an endoplasmic reticulum retention signal. It has been suggested that retention of these proteins in the endoplasmic reticulum was due to ERO1-α/ β interacting with PDIA1, which contains a KDEL retention signal, similar to TbPDI-2 (Jasuja et al., 2012). Over expression of ERO1-α/ β resulted in the secretion of the proteins from transfected cells (Otsu *et al.* 2006). Human PDIA1 was also shown to prevent ERO1 aggregation through this interaction. The mouse protein homologue ERO1-β was found to be expressed in pancreatic mice cells and was linked to the production of insulin (Zito *et al.* 2010). Disruption of the production of ERO1-β through gene mutation caused homozygous mutant mice to show a stable diabetic phenotype.

The sequence of TbEro 1 was found to be similar to the essential yeast Ero 1, Ero1p, which was discussed in the introduction. The yeast protein Ero1p was found to be essential in culture and so it seemed likely that this would also be the case for TbEro 1 (Frand and Kaiser, 1999; Pollard et al., 1998). While there were differences

observed between TbEro 1 and Ero1p it has been shown that the Ero 1 enzymes of yeast and mammals are interchangeable. The loss of Ero1p could be complemented by the expression of human Ero1α, indicating that the Ero 1 proteins are functionally interchangeable between species, despite differences in primary sequence (Cabibbo et al., 2000). As Ero 1 proteins are interchangeable it was thought that they must share similar roles within the endoplasmic reticulum.

Interestingly both Ero1p and TbEro 1 were found to contain a high number of possible *N*-glycosylation sites, and in the case of Ero1p it has been shown that the protein is *N*-glycosylated (Frand and Kaiser 1998). The presence of many possible *N*-glycosylation sites, as was discussed in the introduction, appears to be a feature of TbPDI-2 and the TbPDI-2-like proteins of the African trypanosomes. This may be true for TbEro 1 as it contains many possible *N*-glycosylation sites and so this was taken into consideration when proceeding with the characterization of TbEro 1.

A comparison with the predicted secondary structure of TbPDI-2 and the known secondary structure of Pdi1p, the yeast PDI, indicated both share a similar structure and as such may share similar functions. This also holds true for the comparison of the secondary structure of Ero1p, the yeast Ero 1, and TbEro 1 with one exception: the yeast protein Ero1p appears to contain an additional *C*-terminal region. This region has been shown experimentally to be required for the peripheral association of Ero1p with the membrane of the endoplasmic reticulum (Pagani *et. al.* 2001). This finding suggests TbEro 1 does not share this same property, and like other Ero 1-like proteins reported in other organisms, is a luminal membrane associated endoplasmic reticulum protein which attaches via some means other than an additional transmembrane domain. However, the secondary structure observed outside of this *C*-terminal region suggests TbEro 1 will behave in the same manner as the yeast protein. As TbPDI-2 and TbEro 1 share similar secondary structures with their essential yeast counterparts it was expected that they would prove to be required components of the oxidative folding pathway in *T. brucei*. As such these two putative proteins, TbPDI-2 and TbEro 1 were chosen for further investigation.

4 Analysis of TbPDI-2 *in vivo*

Currently the role of TbPDI-2 (Tb927.10.8230/10.6k15.2290) in *T. brucei* is unclear. There is a discrepancy between published data concerning the requirements for the activity of the protein. One study functionally characterised TbPDI-2 while screening proteins which bound to tomato lectin (Rubotham et al., 2005). This study showed that the knockdown of TbPDI-2 by RNAi had no effect on cell growth in culture. The TbPDI-2 transcript and protein levels were shown to no longer be detectable in induced RNAi cells by Northern and Western blot respectively. Another study investigated TbPDI-2 as part of an evaluation of chaperone requirements for the biosynthesis of VSG (Field et al., 2010). This study revealed the knockdown of TbPDI-2 lead to a profound effect on growth. However both studies did not quantify the knockdown of TbPDI-2 at the transcript level. Significantly the knockdown in both studies was achieved using the same RNAi vectors and cell lines. However, Rubotham et al. specifically avoided the highly conserved thioredox motifs within TbPDI-2 when choosing an RNAi target region; whereas the region selected by Field et al. contained the *N*-terminal thioredoxin domain. The RNAi targeting of a conserved domain by Field et al. may have lead to the knockdown of other proteins including all PDIs. A nonspecific knockdown of thioredoxin containing proteins would lead to a significantly greater effect on growth. Unfortunately Field et al. did not determine if other PDIs were affected, nor did they quantify the knockdown of TbPDI-2 at the RNA or protein level.

In addition, there has been no attempt to knockout TbPDI-2 and so an unequivocal requirement for the activity of TbPDI-2 has yet to be established. It has been previously shown that the yeast PDI, which is similar in protein sequence to TbPDI-2 (see chapter 3), was essential in culture through the generation of a conditional knockout (Tachibana and Stevens, 1992; Tachikawa et al., 1995, 1997) and so it was thought TbPDI-2 may also prove to be essential. The aims of this study were as follows: 1) Assess if TbPDI-2 was essential for basic cellular function by attempting to knockout *TbPDI-2*. 2) Assess what level of knockdown was adequate to result in a growth phenotype and for the knockdown levels to be analyzed at both the transcript and protein levels. 3) Assess the phenotype of a *TbPDI-2* knockdown, principally by

assaying directly the rate of surface protein production. For a further discussion of PDIs see section 1.3.1.

4.1 Generation of constructs

4.1.1 Generation of knockout constructs

An attempt was made to knockout the two *TbPDI-2* alleles by replacing them with the resistance genes required for the inactivation of the selection drugs puromycin and blasticidin. The generation of the knockout constructs is described in detail in the Methods (Section 2.27). The strategy was to knockout the target *TbPDI-2* gene (Tb927.10.8230) by the insertion of a resistance gene (blasticidin / puromycin) flanked by nucleotide sequences corresponding to regions of 300 - 500 nucleotides immediately upstream and downstream of the *TbPDI-2* ORF. This strategy allowed for the replacement of the *TbPDI-2* gene, by the appropriate resistance gene, via homologous recombination with the flanking regions. Since trypanosomes are diploid, the process involved two rounds of replacement / selection to generate true TbPDI-2 null mutants.

The following steps were taken to generate the knockout constructs: A fragment of 2.3 kb containing the entire open reading frame of *TbPDI-2* (1494 bp) was amplified from genomic DNA using primers based on a *BamHI* site 526 bp upstream of the TbPDI-2 initiation codon and an *XbaI* site located ~300 bp downstream of the stop codon. This fragment was subcloned into the *BamHI* / *XbaI* sites of the Puc19 expression vector. This construct was used as a template to amplify a fragment, by PCR, containing the pUC19 plasmid and the TbPDI-2 flanking regions. This PCR involved the use of a forward and reverse primer that allowed the outward amplification of the entire plasmid and flanking regions but excluded the TbPDI-2 open reading frame (Figure 4.1). Included in these forward and reverse primers were sites for the restriction enzymes *MluI* and *SpeI* respectively. The resistance genes blasticidin-S deaminase and puromycin acetyltransferase were amplified by PCR. Forward and reverse primers contained sites for the restriction enzymes *MluI* and *SpeI* respectively. The two resistance gene PCR products were ligated to the pGEM-T Vector for amplification in *E. coli*, as was standard practice for PCR products. Once amplified the vectors were digested with *MluI* and *SpeI* and the regions to be ligated

were gel purified and ligated together. The critical steps of the cloning process were confirmed using agarose gel electrophoresis and are shown in Figure 4.1.

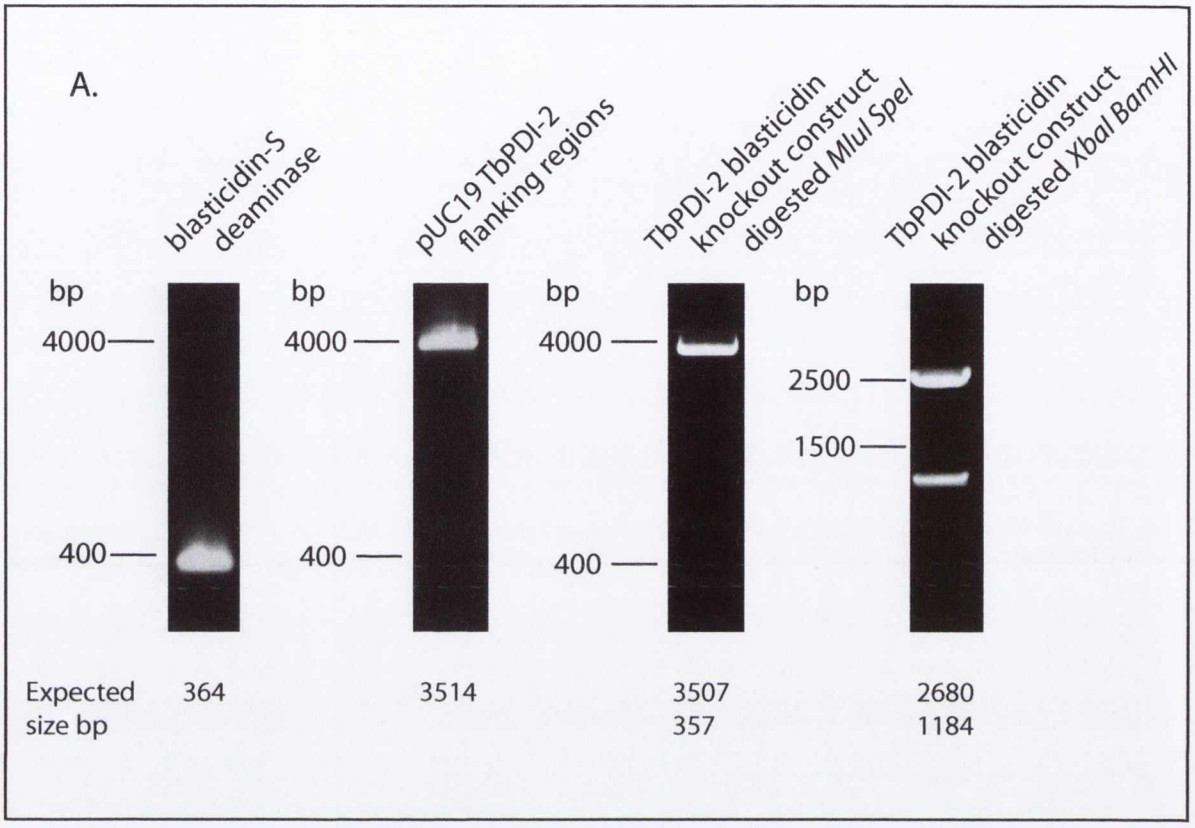
As was standard practice throughout this study, all amplified DNA fragments to be subcloned into expression/RNAi vectors were sequenced first in the cloning vectors, for example *TbPDI-2* in pUC19, and sequenced again once subcloned into their corresponding target vectors. The sequences observed were compared to those reported in the *T. brucei* genome database. The sequences for the *TbPDI-2* knockout constructs are presented (Figure 4.2 and 4.3). The completed knockout constructs were designated TbPDI-2 ko blast and TbPDI-2 ko puro. For a list of all constructs and their functions see the Appendix.

Figure 4.1 Analysis of amplification products of various cloning steps of TbPDI-2 knockout constructs

Panel A: The TbPDI-2 single allele knockout construct, containing blasticidin-S deaminase gene, was generated in three steps. The first step required the amplification of blasticidin-S deaminase gene (364 bp); using primers which contained *MluI* and *SpeI* restriction enzyme sites. The second step required the amplification of pUC19 expression vector, upstream *TbPDI-2* flanking region (520 bp) and downstream *TbPDI-2* flanking region (295 bp; pUC19 TbPDI-2 flanking regions). The primers used to amplify the pUC19 and *TbPDI-2* DNA (3514 bp) contained *MluI* and *SpeI* restriction enzyme sites, which were used to digest the resulting product. The final step required the ligation of the blasticidin-S deaminase gene to the pUC19 and TbPDI-2 flanking regions; the resulting ligation was confirmed by digestion with *MluI* and *SpeI* (3507 bp and 357 bp; TbPDI-2 blasticidin knockout construct digested *MluI SpeI*). Before transfection into the target cell line, TbPDI-2 ko blast was digested with *XbaI* and *BamHI* to release the blasticidin resistance gene flanked upstream and downstream by *TbPDI-2* upstream and downstream regions (2680 bp and 1184 bp; TbPDI-2 blasticidin knockout construct digested *XbaI BamHI*).

Panel B: The same TbPDI-2 single allele knockout construct was made using the puromycin acetyltransferase gene (608 bp) in the place of blasticidin-S deaminase gene. The amplification of puromycin resistance gene yielded a 608 bp PCR product (puromycin acetyltransferase) and was ligated as before to the pUC19 and TbPDI-2 flanking regions (pUC19 TbPDI-2 flanking regions). The resulting plasmid was confirmed by digested with *MluI* and *SpeI* restriction enzymes (3507 bp and 606 bp; TbPDI-2 puromycin knockout construct digested *MluI SpeI*). Before transfection into the target cell line, TbPDI-2 ko puro was digested with *XbaI* and *BamHI* to release the puromycin resistance gene flanked upstream and downstream by *TbPDI-2* upstream and downstream regions (2680 bp and 1433 bp; TbPDI-2 puromycin knockout construct digested *XbaI BamHI*).

Figure 4.1 Analysis of amplification products of various cloning steps of TbPDI-2 knockout constructs



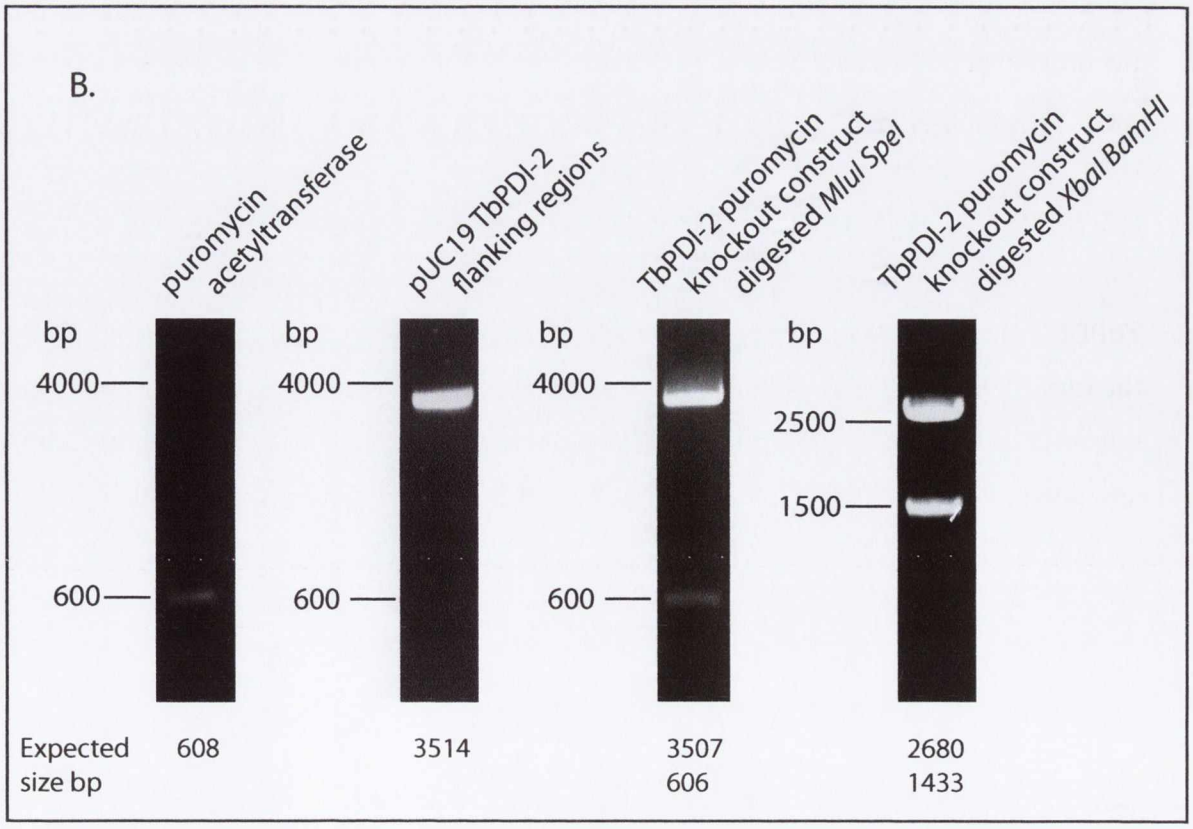


Figure 4.2 Analysis of TbPDI-2 blasticidin knockout construct sequence

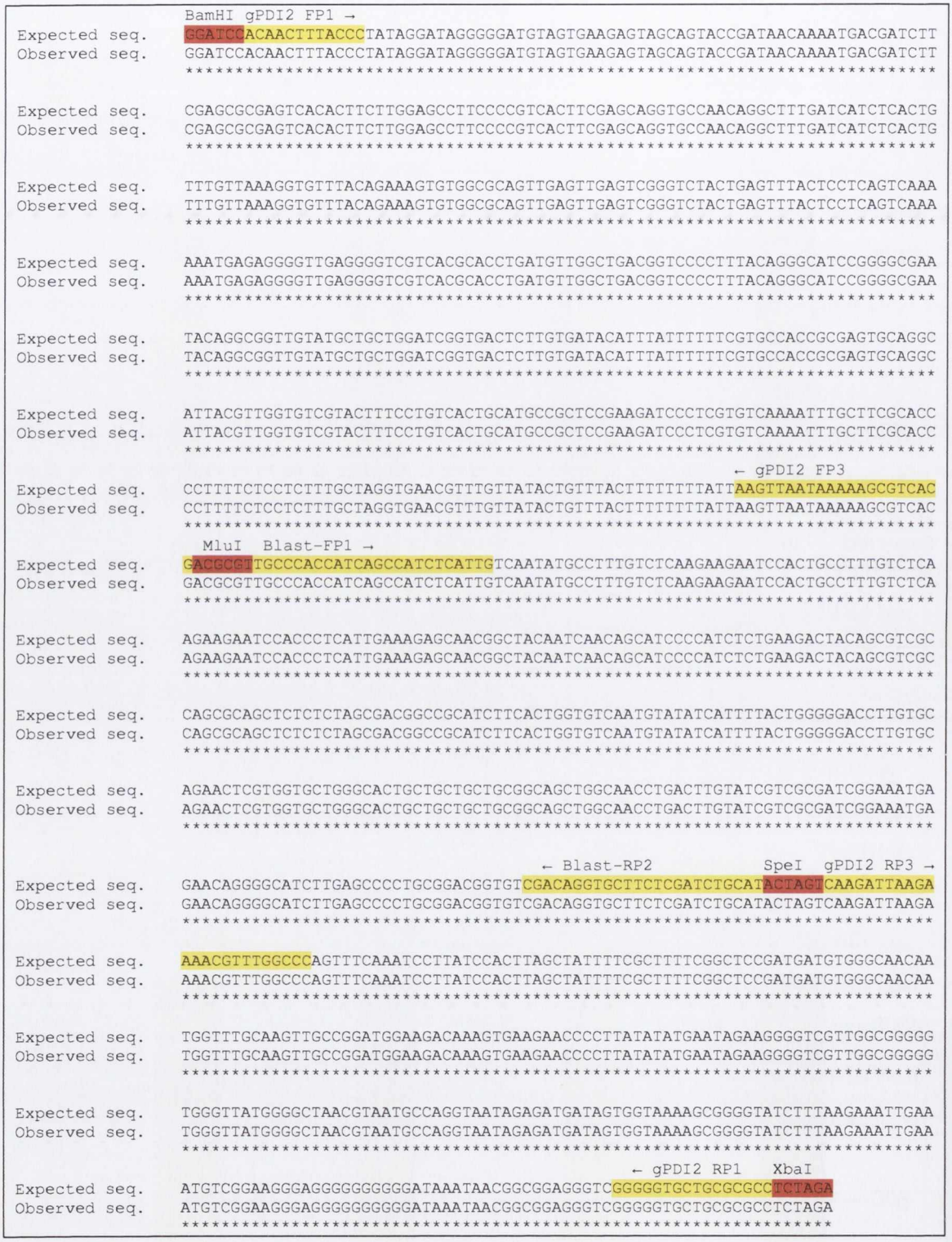
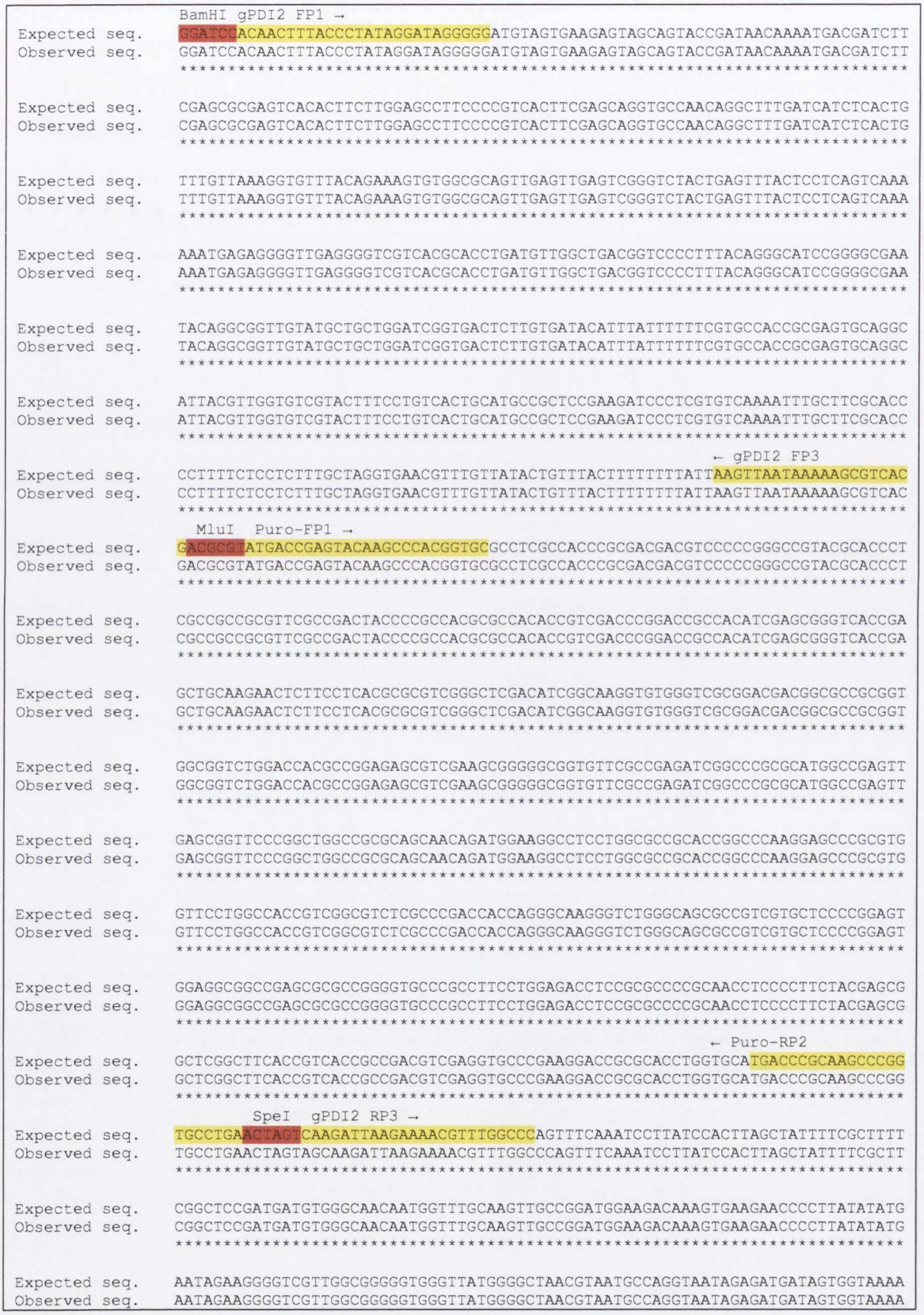


Figure 4.2 Analysis of TbPDI-2 blasticidin knockout construct sequence

The TbPDI-2 ko blast construct was sequenced to ensure that the resistance gene open reading frames and flanking regions were correct. The expected sequence and observed sequence obtained were compared. The location where primers bound to the sequence are shown in yellow and the primer name and direction are indicated above. Restriction sites are shown in red and the restriction enzymes at which these sites cut are indicated above. Agreement is indicated by a * below the sequence.

Figure 4.3 Analysis of TbPDI-2 puromycin knockout construct sequence



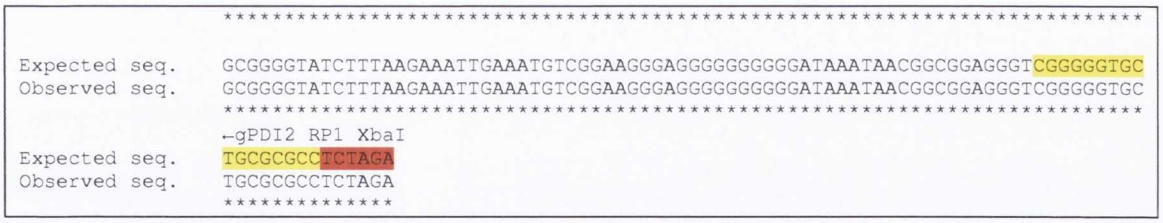


Figure 4.3 Analysis of TbPDI-2 puromycin knockout construct sequence

The TbPDI-2 ko puro construct was sequenced to ensure that the resistance gene open reading frames and flanking regions were correct. The expected sequence and observed sequence obtained were compared. The location where primers bound to the sequence are shown in yellow and the primer name and direction are indicated above. Restriction sites are shown in red and the restriction enzymes at which these sites cut are indicated above. Agreement is indicated by a * below the sequence.

4.1.2 Generation of ectopic expression construct

An ectopic expression vector for TbPDI-2 was also generated to allow the generation of conditional TbPDI-2 null mutants. The p1791 vector was employed for the generation of this construct, which was provided by the Mark Carrington lab, Department of Biochemistry, Cambridge University. The p1791 is an expression construct which is based on the pLEW100 (Wirtz et al., 1999). Once transfected, the p1791 vector incorporates into the trypanosome genome at the ribosomal locus and is under the control of the tetracycline operator.

A detailed description of how the vector was generated can be seen in the Methods (section 2.25). The strategy was to subclone the ORF of *TbPDI-2* into the p1791 using the restriction sites *BamHI* and *HindIII* to replace the existing *GFP* gene (see section 2.25 for a detailed diagram of the p1791 vector). An internal *HindIII* restriction site was observed in the sequence of TbPDI-2 found on the genome database. However, this restriction site was not found once the *TbPDI-2* ORF was sequenced.

The following steps were taken to generate the *TbPDI-2* ectopic expression construct: A fragment containing the *TbPDI-2* ORF (1500 bp) was amplified by PCR using the previously generated pUC19 *TbPDI-2* plasmid as a template. The primers employed during this amplification were designed to incorporate a *HindIII* site upstream of the ORF and a *BamHI* downstream of the ORF. This PCR product was ligated to the pGEM-T Vector, amplified in *E. coli*, digested *BamHI* and *HindIII* and gel purified (product may be seen in Figure 4.4). The p1791 plasmid was digested *BamHI* and *HindIII* and the vector was gel purified (product may be seen in Figure 4.4). The two gel purified products were then ligated together to produce the completed expression construct. This construct was amplified in *E. coli* and a digestion with *BamHI* and *HindIII* to confirm the ligation was successful. The construct was sequenced and the sequences observed were compared to the expected sequence (Figure 4.5). The completed ectopic expression construct was designated TbPDI-2 p1791. For a diagram of the completed construct see Section 2.25.

4.2 Generation and analysis of bloodstream form (TbPDI-2 +/-) cell lines

The TbPDI-2 gene knockout constructs TbPDI-2 ko blast and TbPDI-2 ko puro were digested with *BamHI* and *XbaI* to release the resistance genes flanked by *TbPDI-2* upstream and downstream regions (Figure 4.1). The DNA from the individual knockout constructs was transfected into two separate populations of bloodstream form (328.114) cells as described in the Methods (section 2.14). Positive wells containing resistant cells were diluted into fresh medium and cloned. The knockout of one of the *TbPDI-2* gene was confirmed in each case by PCR using genomic DNA extracted from positive clones as the template, which showed the amplification of the typical *TbPDI-2* gene and the altered *TbPDI-2* locus containing the resistance gene (Figure 4.6). This result demonstrated that the selected cells were single knockout (TbPDI-2 +/-) cells. The cell lines were designated (TbPDI-2 +/- blast) and (TbPDI-2 +/- puro). The agarose gel revealed a rather striking preference for the amplification of the shorter of the two alleles. This propensity for PCR to amplify the shorter of two products has been observed previously during the knockout of single alleles in *T. brucei* before (Spitznagel et al., 2009). One possibility for this preference could have been that a mixed population of (TbPDI-2 +/-) and (TbPDI-2 -/-) cells were growing in culture. To address this possibility the transcript level of *TbPDI-2* was measure by qRT-PCR analysis of the total RNA collected from the clones (Figure 4.7). The results of the qRT-PCR showed the transcript in both clones was ~50% of the normal level confirming that a mixed population did not exist. The level of TbPDI-2 protein was also measured using Western blot (Figure 4.8). The result also confirmed what appeared to be a knockdown of approximately 50% relative to the actin loading control, however the use of Western blot as a quantitative analytical technique is limited in comparison to qRT-PCR (VanGuilder et al., 2008). As expected the growth of (TbPDI-2 +/-) cell lines were unaffected at a 50% knockdown (Figure 4.9). The doubling times for the cell lines were calculated and were found to be 7.5 hours (parental) and 7.8 hours (TbPDI-2 +/- blast) 7.9 hours (TbPDI-2 puro) which suggested that the growth of single knockout cells was unaffected.

Figure 4.4 Analysis of amplification products of various steps involved in generating the p1791 TbPDI-2 expression construct

The p1791 TbPDI-2 expression construct was generated in three steps. The first step required the amplification of TbPDI-2 ORF (1512 bp); using primers which contained *BamHI* and *HindIII* restriction enzyme sites. The second step required the digestion of p1791 expression vector with *BamHI* and *HindIII* and gel purification of the 5540 bp fragment. The final step required the ligation of the TbPDI-2 ORF to the p1791 vector; the resulting ligation was confirmed by digestion with *BamHI* and *HindIII* as shown.

Figure 4.4 Analysis of amplification products of various steps involved in generating the p1791 TbPDI-2 expression construct

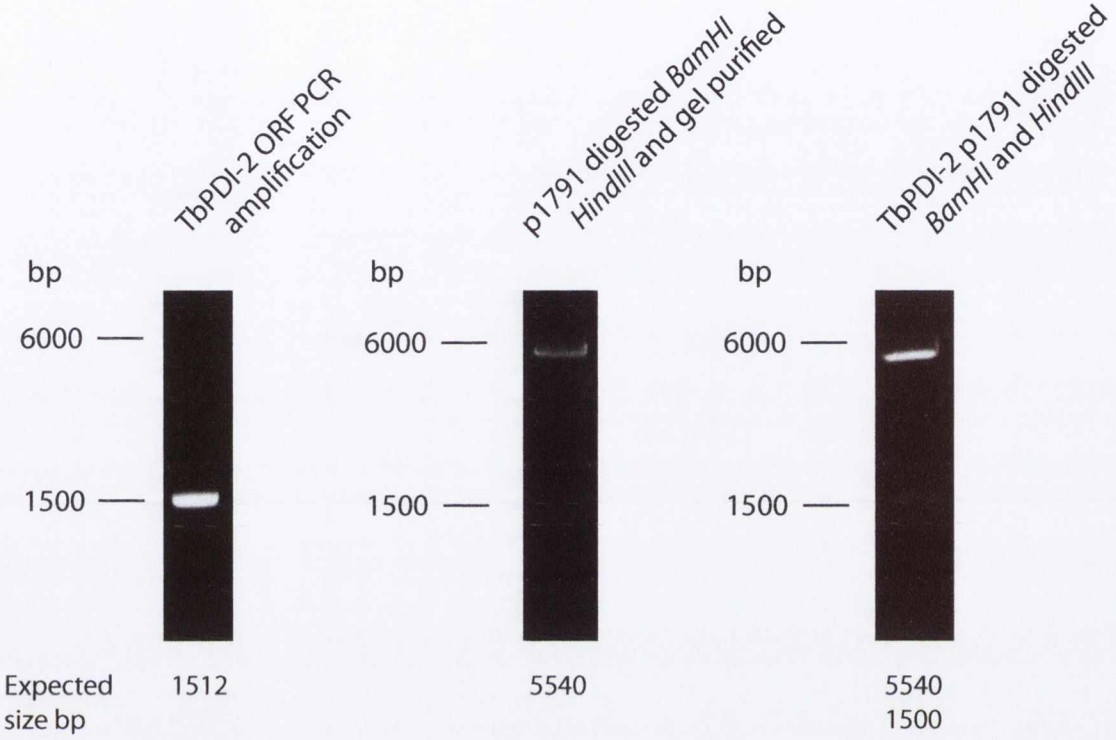
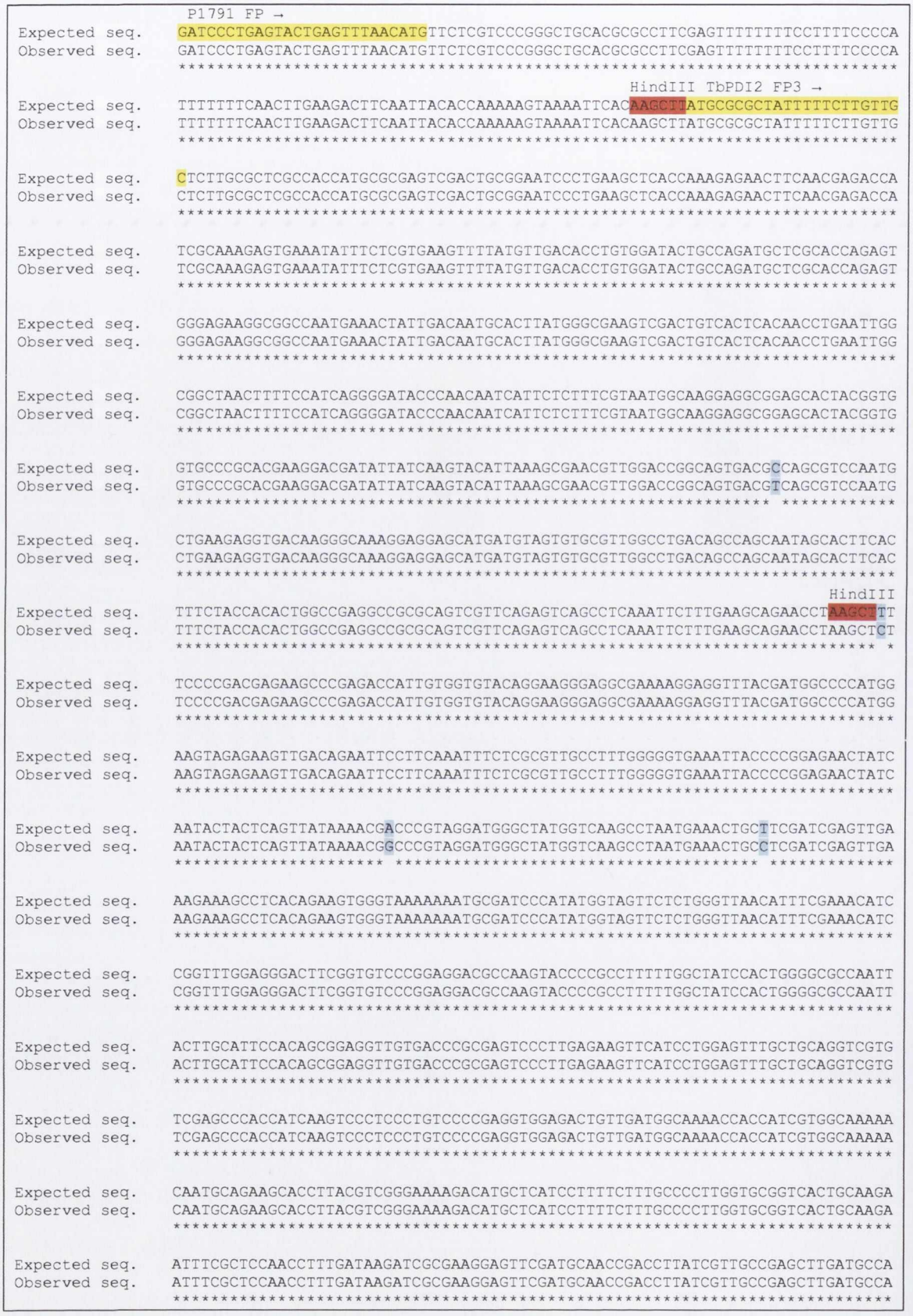
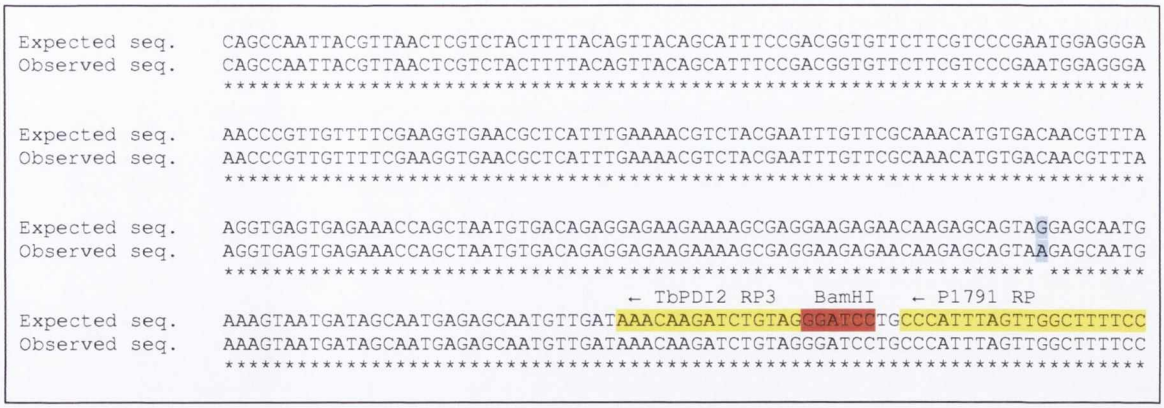


Figure 4.5 Analysis of TbPDI-2 p1791 expression construct sequence

A.




Expected seq. CAGCCAATTACGTTAACTCGTCTACTTTTACAGTTACAGCATTTCCGACGGTGTTCTTCGTCCCGAATGGAGGGA
Observed seq. CAGCCAATTACGTTAACTCGTCTACTTTTACAGTTACAGCATTTCCGACGGTGTTCTTCGTCCCGAATGGAGGGA
Expected seq. AACCCGTTGTTTTCGAAGGTGAACGCTCATTTGAAAACGTCTACGAATTTGTTCGCAAACATGTGACAACGTTTA
Observed seq. AACCCGTTGTTTTCGAAGGTGAACGCTCATTTGAAAACGTCTACGAATTTGTTCGCAAACATGTGACAACGTTTA
Expected seq. AGGTGAGTGAGAAACCAGCTAATGTGACAGAGGAGAAGAAAAGCGAGGAAGAGAACAAGAGCAGTAGGAGCAATG
Observed seq. AGGTGAGTGAGAAACCAGCTAATGTGACAGAGGAGAAGAAAAGCGAGGAAGAGAACAAGAGCAGTAAGAGCAATG
← TbPDI2 RP3 BamHI ← P1791 RP
Expected seq. AAAGTAATGATAGCAATGAGAGCAATGTTGATAAACAAGATCTGTAGGGATCCTGCCCATTTAGTTGGCTTTTCC
Observed seq. AAAGTAATGATAGCAATGAGAGCAATGTTGATAAACAAGATCTGTAGGGATCCTGCCCATTTAGTTGGCTTTTCC

B.

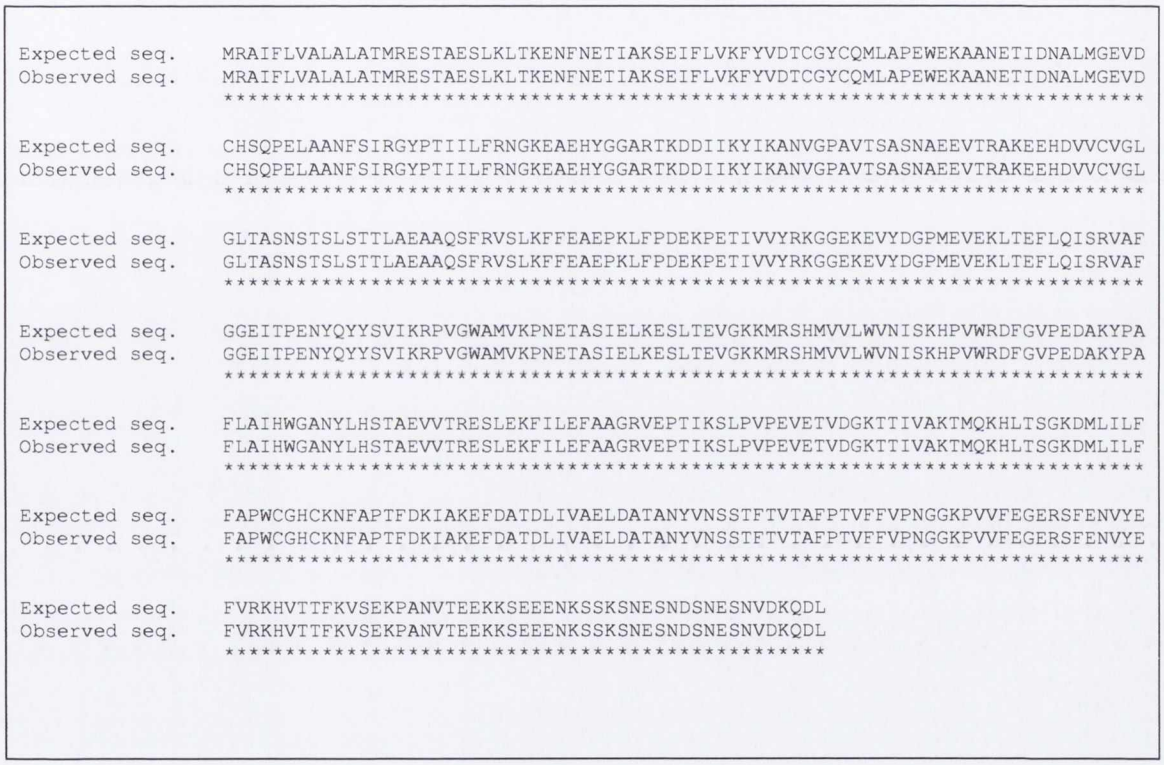

Expected seq. MRAIFLVALALATMRESTAESLKLTKENFNETIAKSEIFLVKFYVDTCGYCQMLAPEWEKAANETIDNALMGEVD
Observed seq. MRAIFLVALALATMRESTAESLKLTKENFNETIAKSEIFLVKFYVDTCGYCQMLAPEWEKAANETIDNALMGEVD
Expected seq. CHSQPELAANFSIRGYPTIILFRNGKEAEHYGGARTKDDIIKYIKANVGPAVTSASNAEEVTRAKEEHDVVCVGL
Observed seq. CHSQPELAANFSIRGYPTIILFRNGKEAEHYGGARTKDDIIKYIKANVGPAVTSASNAEEVTRAKEEHDVVCVGL
Expected seq. GLTASNSTSLSTTLAEAAQSFRVSLKFFEAEPKLFPDEKPETIVVYRKGGEKEVYDGPMEVEKLTEFLQISRVAF
Observed seq. GLTASNSTSLSTTLAEAAQSFRVSLKFFEAEPKLFPDEKPETIVVYRKGGEKEVYDGPMEVEKLTEFLQISRVAF
Expected seq. GGEITPENYQYYSVIKRPVGWAMVKPNETASIELKESLTEVGKKMRSHMVVLWVNISKHPVWRDFGVPEDAKYPA
Observed seq. GGEITPENYQYYSVIKRPVGWAMVKPNETASIELKESLTEVGKKMRSHMVVLWVNISKHPVWRDFGVPEDAKYPA
Expected seq. FLAIHWGANYLHSTAEVVTRESLEKFILEFAAGRVEPTIKSLPVPEVETVDGKTTIVAKTMQKHLTSGKDMLILF
Observed seq. FLAIHWGANYLHSTAEVVTRESLEKFILEFAAGRVEPTIKSLPVPEVETVDGKTTIVAKTMQKHLTSGKDMLILF
Expected seq. FAPWCGHCKNFAPTFDKIAKEFDATDLIVAELDATANYVNSSTFTVTAFPTVFFVPNGGKPVVFEGERSFENVYE
Observed seq. FAPWCGHCKNFAPTFDKIAKEFDATDLIVAELDATANYVNSSTFTVTAFPTVFFVPNGGKPVVFEGERSFENVYE
Expected seq. FVRKHVTTFKVSEKPANVTEEKKSEEENKSSKSNESNDSNESNVDKQDL
Observed seq. FVRKHVTTFKVSEKPANVTEEKKSEEENKSSKSNESNDSNESNVDKQDL

Figure 4.5 Analysis of TbPDI-2 p1791 expression construct sequence

The conditional expression vector containing the TbPDI-2 open reading frame was sequenced to ensure the open reading frame was inserted correctly into the p1791 expression vector.

Panel A: The expected sequence and observed sequence were compared. The location where primers bound to the sequence are shown in yellow and the primer name and direction are indicated above. Restriction sites are shown in red and the restriction enzymes at which these sites cut are indicated above. Agreement is indicated by a * below the sequence and disagreements are highlighted in light blue.

Panel B: A translation of the expected sequence and observed sequence obtained written in single letter amino acid abbreviation. Agreement is indicated by a * below the sequence.

Figure 4.6 Analysis of gDNA of bloodstream form (TbPDI-2 +/-) cell lines using PCR

Agarose gel electrophoresis of PCR amplification of TbPDI-2 gene was carried out using gDNA collected from (TbPDI-2 +/-). Primers were designed to amplify the TbPDI-2 locus. The amplified region consisted of the 5' upstream flanking region (526 bp), the TbPDI-2 open reading frame (1494 bp) and the 3' downstream flanking region (300 bp). Three samples of equal concentrations of genomic DNA were used as the PCR template: Parental cell line genomic DNA, (TbPDI-2 +/- blast) cell line and (TbPDI-2 +/- puro) cell line. The expected size of the unaltered TbPDI-2 locus PCR product was 2050 bp, the altered TbPDI-2 locus containing blasticidin-S deaminase PCR product was 919 bp and the altered TbPDI-2 locus containing puromycin acetyltransferase PCR product was 1167 bp.

Figure 4.6 Analysis of gDNA of bloodstream form (TbPDI-2 +/-) cell lines using PCR

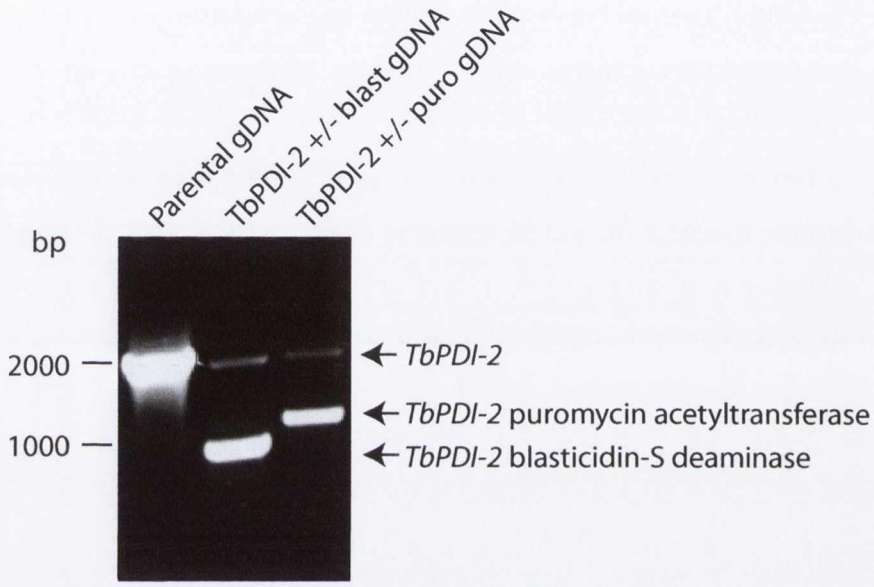


Figure 4.7 Analysis of TbPDI-2 RNA transcript of bloodstream form (TbPDI-2 +/-) cell lines using qRT-PCR

The relative level of expression of *TbPDI-2* RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of *TbPDI-2* RNA were normalized against actin RNA and the relative quantifications were calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the TbPDI-2 transcript in TbPDI-2 +/- blast was found to be 63 ± 2.6% and in TbPDI-2 +/- puro was found to be 52 ± 0.9% when compared to parental cell lines. In each case the results are expressed as the mean ± standard deviation of three replicates.

Figure 4.7 Analysis of TbPDI-2 RNA transcript of bloodstream form (TbPDI-2 +/-) cell lines using qRT-PCR

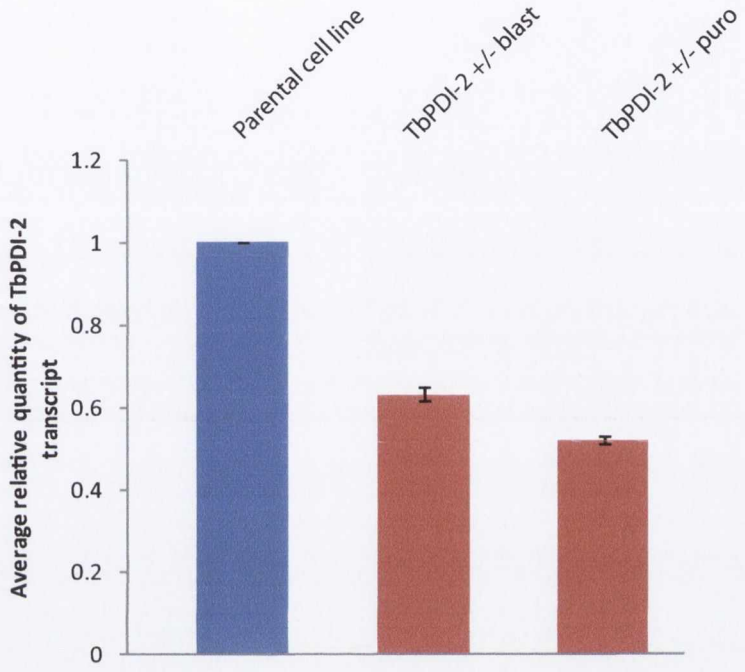
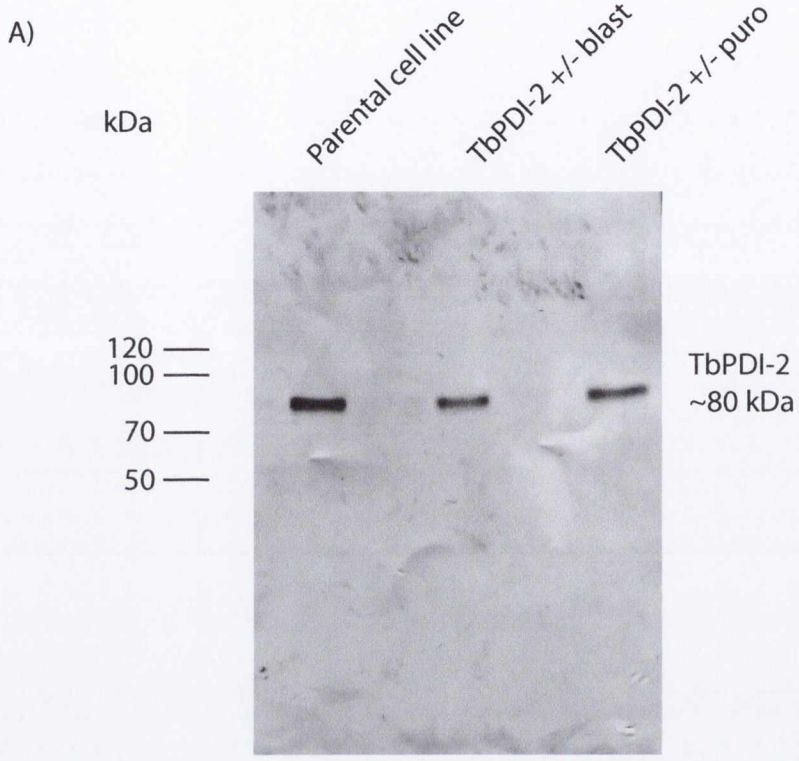


Figure 4.8 Analysis of TbPDI-2 expression in bloodstream form (TbPDI-2 +/-) cell lines and parental cells using Western blot

Panel A: Western blot of whole cell extracts from bloodstream form (TbPDI-2 +/-) cell lines and parental cell line probed with anti-TbPDI-2 antibody. Each lane represents $5x10^6$ cells.

Panel B: Western blot of whole cell extracts from bloodstream form (TbPDI-2 +/-) cell lines and parental cell line probed with anti-Actin antibody as a loading control. Each lane represents $5x10^6$ cells.

Figure 4.8 Analysis of TbPDI-2 expression in bloodstream form (TbPDI-2 +/-) cell lines and parental cells using Western blot



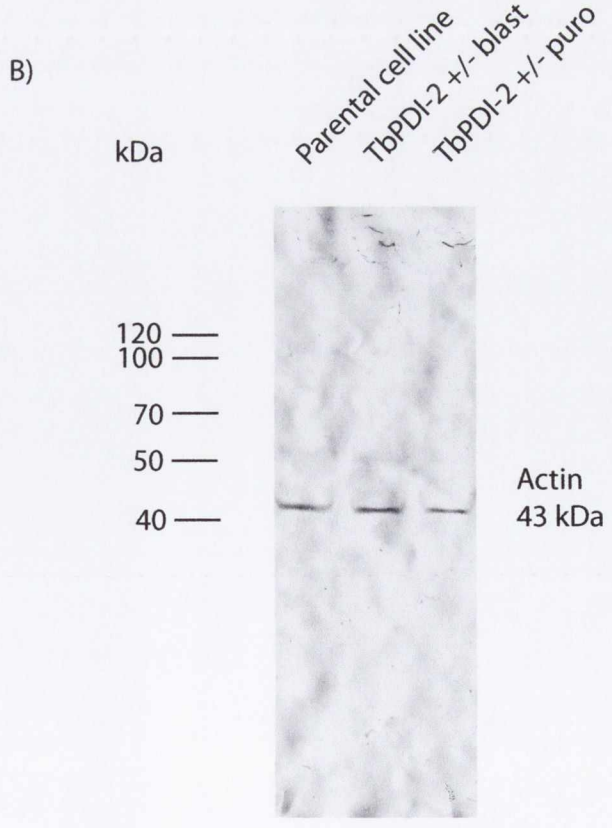


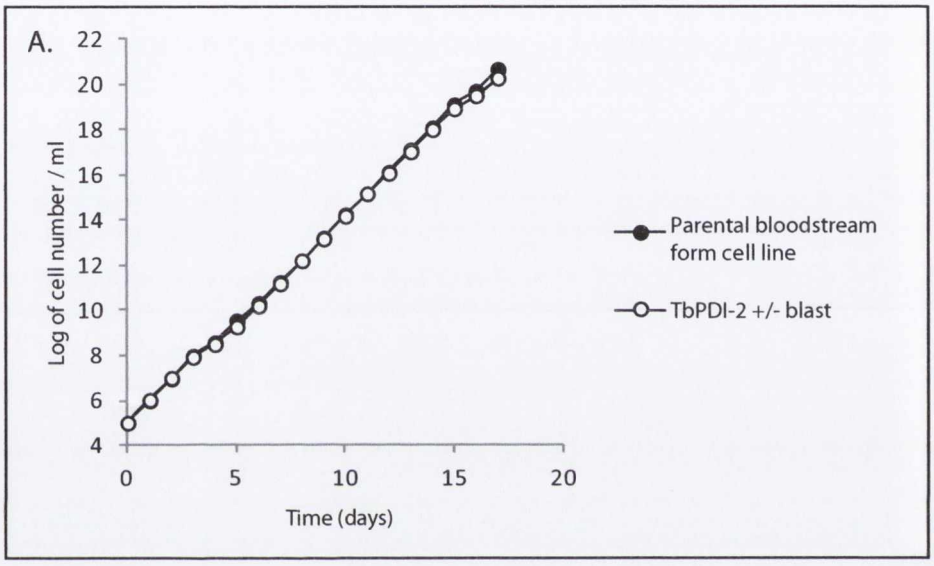
Figure 4.9 Analysis of growth rate of bloodstream form (TbPDI-2 +/-) cell lines

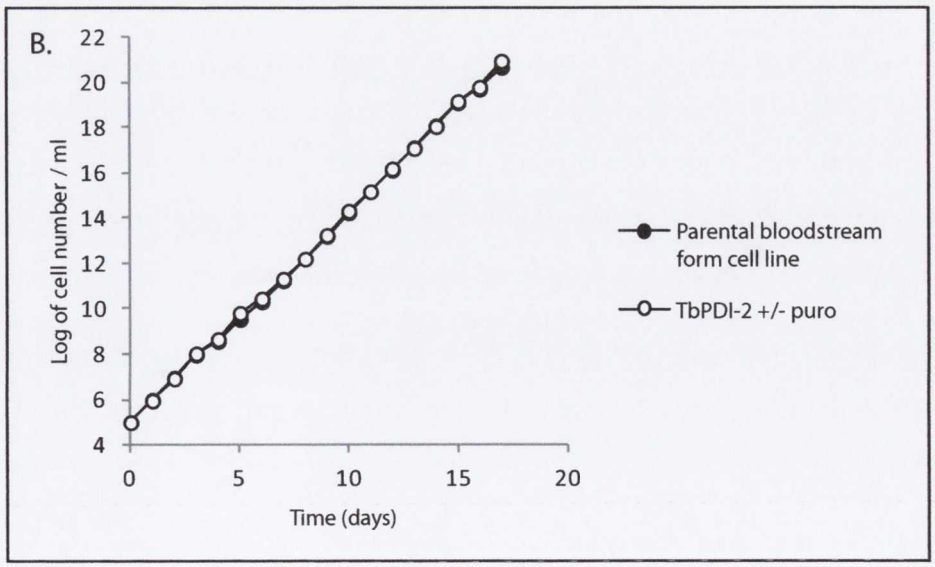
Bloodstream form (TbPDI-2 +/- blast) cell line, (TbPDI-2 +/- puro) cell line and parental bloodstream form cell line were grown in HMI9, as described in the Methods, over a period of 17 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 7.5 hours for the parental cell line, 7.9 hours for TbPDI-2 +/- puro and 7.8 hours for TbPDI-2 +/- blast.

Panel A: (TbPDI-2 +/- blast) cell line versus parental cell line.

Panel B: (TbPDI-2 +/- puro) cell line versus parental cell line.

Figure 4.9 Analysis of growth rate of bloodstream form (TbPDI-2 +/-) cell lines





4.3 Attempted generation of bloodstream form (TbPDI-2 -/-) null mutant cell line

As previous studies did not include the knockout of *TbPDI-2* it was thought that it may be possible to knockout the second *TbPDI-2* gene without a lethal effect on bloodstream form cells. If a (TbPDI-2 -/-) cell line could not be generated, as it was an essential gene, an ectopic copy of *TbPDI-2* could be introduced allowing the knockout of the second allele. This ectopic copy *TbPDI-2* would be under the control of the tetracycline operator and could be switched off allowing the study of the effects resulting from the absence of TbPDI-2.

The first attempt was to generate (TbPDI-2 -/-) null mutant cells by increasing the concentration of selection antibiotic, thus increasing the selection pressure for a recombination event to occur, leading to the displacement of *TbPDI-2* at the second locus with a second copy of the resistance gene. This strategy has been used to generate null mutant cell lines in *Leishmania* (Gueiros-Filho and Beverley, 1996). However, when this approach was employed with either the (TbPDI-2 +/- blast) and (TbPDI-2 +/- puro) cell lines a true TbPDI-2 null mutant could not be generated. When the concentration of blasticidin was increased to 10 times the original concentration (50 μg/ml) there was no loss of the second allele as determined by PCR (Figure 4.10). Puromycin resistant cells were able to tolerate 5 times (12.5 μg/ml) the original concentration of antibiotic before growth ceased but again no double allele knockout cells were obtained. If a recombinant event occurred, an alteration of the size of the TbPDI-2 locus would be expected to occur but as an upper band representing the expected size of the unaltered TbPDI-2 locus (2050 bp) was observed in both (TbPDI-2 +/- puro) and (TbPDI-2 +/- blast) cell lines it was concluded that the cells were not true TbPDI-2 null mutants.

As this method failed to produce a TbPDI-2 null mutant, an attempt to knockout the second allele with the second resistance gene marker was undertaken. The (TbPDI-2 +/- blast) and (TbPDI-2 +/- puro) cell lines were transfected as before with the alternate resistance marker knockout construct. Five separate attempts were made to achieve null mutant cells but no true null mutant cells survived selection. As a positive control, (TbPDI-2 +/+) cells were simultaneously transfected with the knockout constructs. Out of 240 wells, 61 positive wells were observed (~25%

success rate) when wild type cells were transfected with the puromycin TbPDI-2 construct and 53 out of 240 (~22% success rate) positive wells were observed when the blasticidin TbPDI-2 construct was transfected. The resulting control single knockout cell lines were confirmed, as before, by PCR assay of genomic DNA (Figure 4.11). This experiment suggested that TbPDI-2 was essential in culture as single allele knockout cells were easily generated but true null mutant cell lines could not be generated; but as this experiment yielded a negative result it could not be concluded that TbPDI-2 was essential in culture.

Figure 4.10 Analysis of gDNA of bloodstream form (TbPDI-2 +/-) cell lines incubated under additional selection pressure

Agarose gel electrophoresis of PCR amplification of TbPDI-2 gene was carried out using gDNA collected from (TbPDI-2 +/-). Primers were designed to amplify the TbPDI-2 locus. The amplified region consisted of the 5' upstream flanking region (526 bp), the TbPDI-2 open reading frame (1494 bp) and the 3' downstream flanking region (300 bp). Three samples of equal concentrations of genomic DNA were used as the PCR template: Parental cell line genomic DNA, (TbPDI-2 +/- blast) cell line incubated with blasticidin (50 μg/ml) and (TbPDI-2 +/- puro) cell line incubated with puromycin (12 μg/ml). The expected size of the unaltered TbPDI-2 locus PCR product was 2050 bp, the altered TbPDI-2 locus containing blasticidin-S deaminase PCR product was 919 bp and the altered TbPDI-2 locus containing puromycin acetyltransferase PCR product was 1167 bp.

Figure 4.10 Analysis of gDNA of bloodstream form (TbPDI-2 +/-) cell lines incubated under additional selection pressure

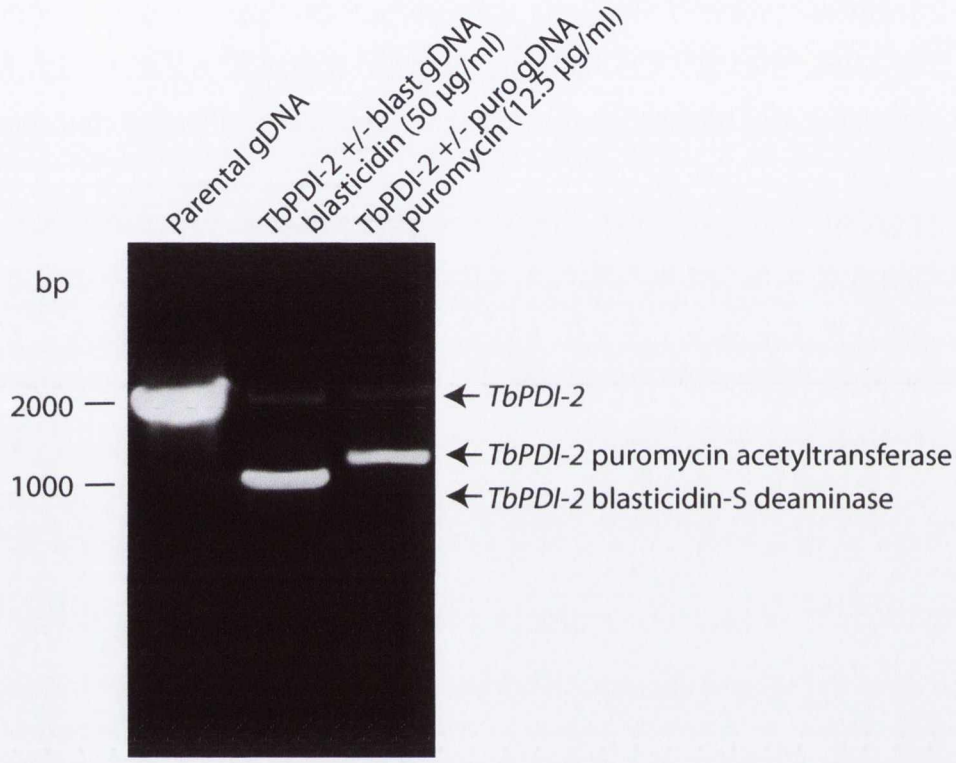


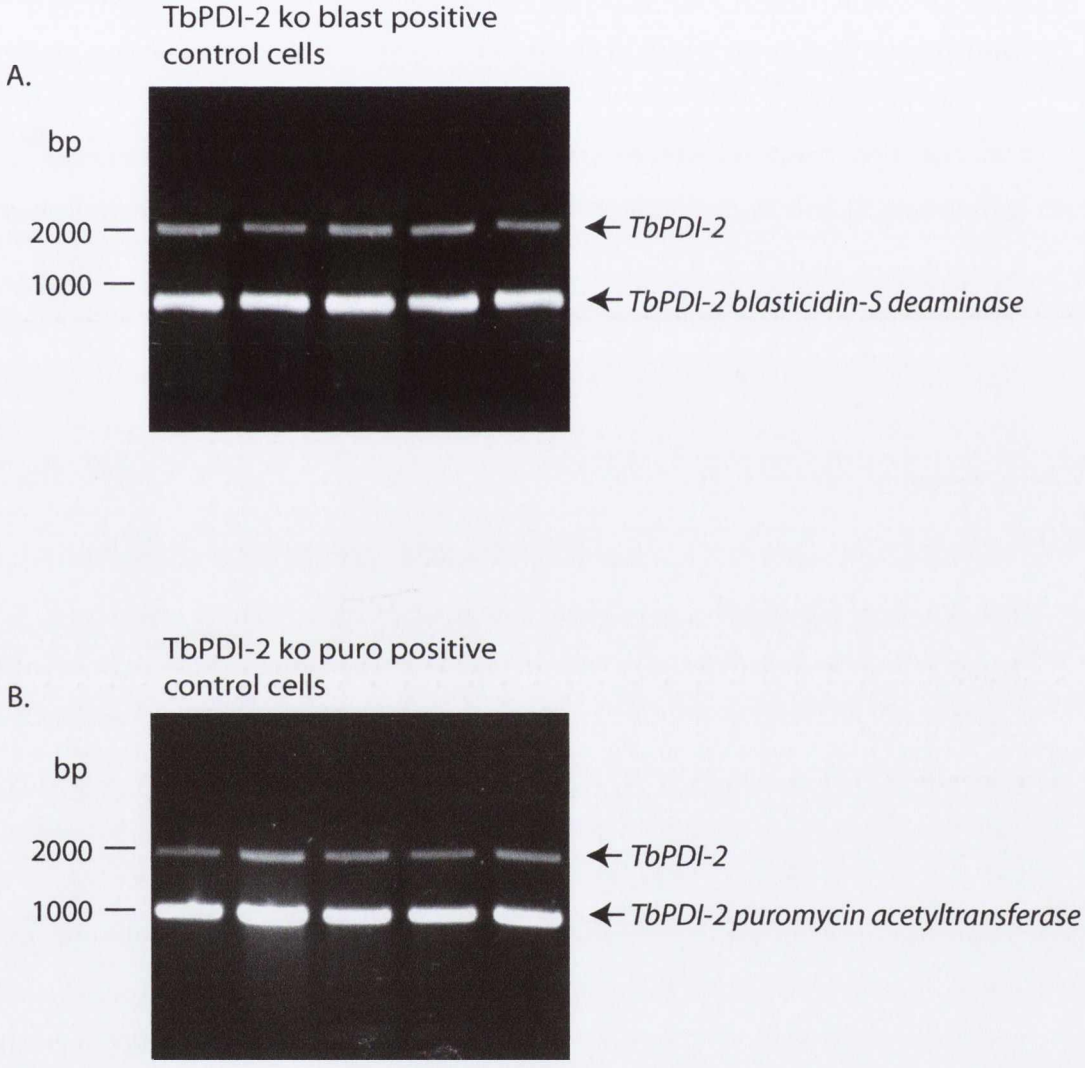
Figure 4.11 Analysis of gDNA of bloodstream form positive control (TbPDI-2 +/-) cells generated during attempted generation of (TbPDI-2 -/-) cells using PCR

An attempt to knockout the second *TbPDI-2* gene, by transfection of the (TbPDI-2 +/- puro) and (TbPDI-2 +/- blast) cell lines with the alternate resistance gene TbPDI-2 knockout constructs, resulted in no cells surviving the transfection (see section 4.3 for context). As a positive control, wild type cells were transfected simultaneously with TbPDI-2 puro or blast knockout constructs. 5 wells containing cells were grown from the two control transfections, and gDNA was collected. A PCR was run, using the gDNA as the template, with primers which were designed to amplify the TbPDI-2 locus. The amplified region consisted of the 5' upstream flanking region (526 bp), the TbPDI-2 open reading frame (1494 bp) and the 3' downstream flanking region (300 bp). The expected size of the unaltered TbPDI-2 locus PCR product was 2050 bp, the altered TbPDI-2 locus containing blasticidin-S deaminase PCR product was 919 bp and the altered TbPDI-2 locus containing puromycin acetyltransferase PCR product was 1167 bp.

Panel A. PCR amplification of the positive control cells collected from TbPDI-2 blast knockout construct transfection. The 5 wells chosen contained positive (TbPDI-2 +/- blast) cells, showing the knockout of *TbPDI-2* was successfully occurring during this experiment.

Panel B. PCR amplification of the positive control cells collected from TbPDI-2 puro knockout construct transfection. The 5 wells chosen contained positive (TbPDI-2 +/- puro) cells, showing the knockout of *TbPDI-2* was successfully occurring during this experiment.

Figure 4.11 Analysis of gDNA of bloodstream form positive control (TbPDI-2 +/-) cells generated during attempted generation of (TbPDI-2 -/-) cells using PCR



4.4 Attempted generation of bloodstream form (TbPDI-2 -/- TbPDI-2 p1791) cell line

To obtain bloodstream form TbPDI-2 null mutants, an ectopic copy of TbPDI-2 was introduced into the genome of bloodstream form (TbPDI-2 +/- puro) cells under the control of the tetracycline inducible promoter using the p1791 expression vector. The (TbPDI-2 +/- puro) cell line was transfected with linearized TbPDI-2 p1791 and cells were selected as described in the Methods (section 2.14). A positive clone was chosen and analysed using qRT-PCR to ensure the ectopic copy was expressing TbPDI-2 RNA. When induced the (TbPDI-2 +/- puro TbPDI-2 p1791) cell line was shown to produce TbPDI-2 RNA from the ectopic copy vector 119% when compared wild type levels (Figure 4.12). Interestingly the (TbPDI-2 +/- puro) cell line and the noninduced (TbPDI-2 +/- puro TbPDI-2 p1791) cell line were shown to express 52% and 79% respectively of wild type levels. This result indicated there was transcription of TbPDI-2 p1791 in the absence of tetracycline; showing that if a (TbPDI-2 -/- TbPDI-2 p1791) cell line was successfully cloned, the expression of TbPDI-2 could not be prevented.

The next step taken was to transfect the induced (TbPDI-2 +/- puro TbPDI-2 p1791) cell line with the TbPDI-2 blasticidin knockout construct. Three separate transfections were carried out, with the goal of generating a (TbPDI-2 -/- TbPDI-2 p1791) cell line, with one successful well obtained. The cells were grown to a concentration of $1x10^5$ cells/ml with appropriate selection drugs and subcultured into two flasks. The tetracycline was removed from one of the flasks by washing the cells in fresh media. Growth was monitored over a period of 10 days (Figure 4.13). The cells appeared not to grow during this period and declined in numbers until they could no longer be counted. The noninduced (TbPDI-2 -/- TbPDI-2 p1791) cell line only survived 3 days where as the induced (TbPDI-2 -/- TbPDI-2 p1791) cell line survived without growing for 7 days. The apparent growth arrest of the cell line may be explained by 3 possibilities: 1. The (TbPDI-2 -/- TbPDI-2 p1791) cell was a false positive and the blasticidin selection drug failed to kill the cells. 2. The cells had lost one of the selection drug resistance genes at the beginning of the growth curve.3. The cell line had lost the ability to express the TbPDI-2 protein at the level required for cell growth. This last possibility is supported by the earlier death of the noninduced cells before the induced cells. However, these possibilities could not be proven as

there were not enough cells to collect an RNA sample to carry out an analysis of the TbPDI-2 transcript.

4.5 Generation and analysis of bloodstream form (TbPDI-2 +/- TbPDI-2 RNAi) cell line

As a true null mutant of TbPDI-2 or a robust conditional null mutant could not be generated, an attempt was made to knockdown the expression of TbPDI-2 as low as possible. A fourth approach was employed to knock down the level of TbPDI-2 which combined the use of standard RNAi with a single allele null mutant. This technique has been employed successfully in previous knockdown studies of *T. brucei* (Spitznagel et al., 2009). The TbPDI-2 RNAi p2T7-Ver3 construct, which is described in detail in the Methods (section 2.24), was linearized and transfected into the (TbPDI-2 +/- puro) cell line. Selection of the cells was carried out using hygromycin (2.5 μg/ml) and cells collected from a positive well were grown, subcultured into two flasks and one of the flasks was induce using tetracycline (1 μg/ml).

The growth of cells was monitored for a period of 17 days (Figure 4.14). It was clear the growth of the induced (TbPDI-2 +/- puro) TbPDI-2 RNAi cell line was slower than the noninduced cell line. Typical growth of bloodstream form cell lines in HMI9 result in a doubling time of 7.5 to 8 hours (Field et al., 2010; Spitznagel et al., 2009; Webb et al., 1997). The doubling times of the two cell lines were calculated and found to be 7.8 hours for noninduced and 13 hours for induced cell line; indicating an increase of 67% in the doubling time due to the knockdown of TbPDI-2. During this time course samples of cells were removed to prepare total cell lysates and total RNA samples.

The transcript of TbPDI-2 was measured using qRT-PCR (Figure 4.15). Interestingly, the level of TbPDI-2 transcript in the induced cells decreased by day 6 suggesting RNAi required a minimum of 6 days to reach optimum levels of knockdown. The qRT-PCR results revealed the single knockout and RNAi combination achieved a knockdown of 99% of the total transcript when compared to the wild type (TbPDI-2 +/+) cell line. The levels of transcript of the other TbPDIs were also measured using qRT-PCR (Figure 4.16). The level of transcript of the (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line was measured in the other PDIs identified in the genome of *T. brucei.* This experiment was carried out firstly to ensure no knockdown was achieved in

Figure 4.12 Analysis of TbPDI-2 RNA transcript of bloodstream form (TbPDI-2 +/- puro TbPDI-2 p1791) and parental cell lines using qRT-PCR

The relative level of expression of TbPDI-2 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbPDI-2 RNA were normalized against actin RNA and the relative quantification was calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of 1) the TbPDI-2 transcript of the (TbPDI-2 +/-) cell line was found to be 52.2 ± 9.9%, 2) the (TbPDI-2 +/- TbPDI-2 p1791) cell line noninduced was found to be 79 ± 14.3% and 3) (TbPDI-2 +/- TbPDI-2 p1791)cell line induced was found to be 118.6 ± 20.2% when compared to parental cell line. In each case the results are expressed as the mean ± standard deviation of three replicates. Noninduced is indicated by –tet Induced is indicated by +tet.

Figure 4.12 Analysis of TbPDI-2 RNA transcript of bloodstream form (TbPDI-2 +/- puro TbPDI-2 p1791) and parental cell lines using qRT-PCR

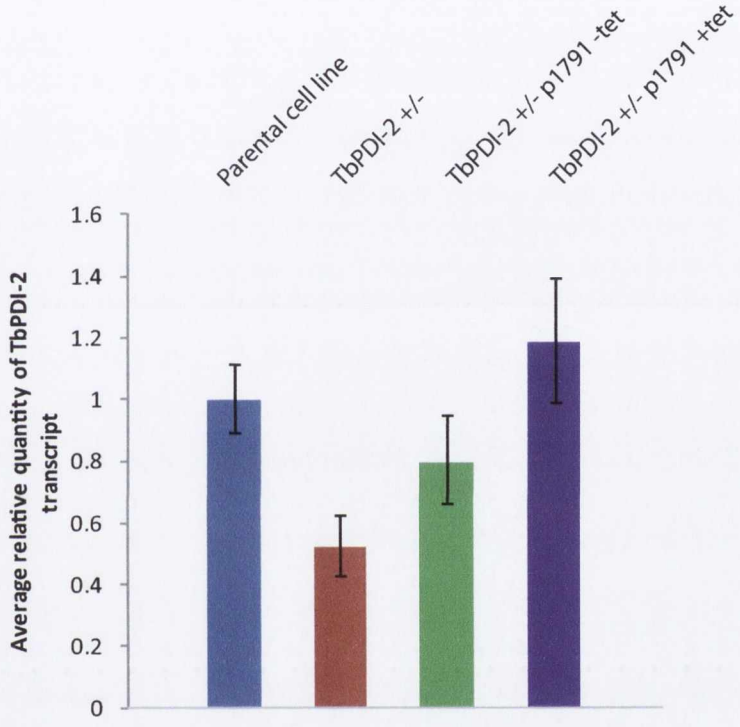


Figure 4.13 Analysis of growth rate of bloodstream form (TbPDI-2 -/- TbPDI-2 p1791) cell line and parental cell lines

Bloodstream form TbPDI-2 -/- p1791 cell line induced and noninduced; and parental bloodstream form cell line were grown in HMI9, as described in the Methods, over a period of 10 days. The data presented is of the log of cell number/ml versus time (days). Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.13 Analysis of growth rate of bloodstream form (TbPDI-2 -/- TbPDI-2 p1791) cell line and parental cell lines

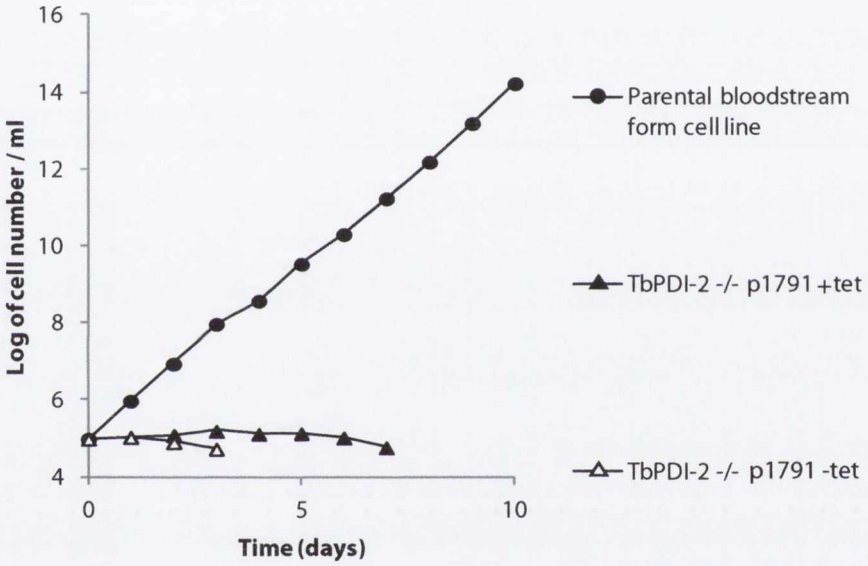


Figure 4.14 Analysis of growth rate of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

Bloodstream form TbPDI-2 +/- puro RNAi cell line was grown in HMI9 under noninduced and induced conditions as described in the Methods over a period of 17 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 7.8 hours for TbPDI-2 +/- puro RNAi noninduced and 13 hours induced. This result indicated a loss of TbPDI-2 caused an increase of 67% in doubling time. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.14 Analysis of growth rate of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

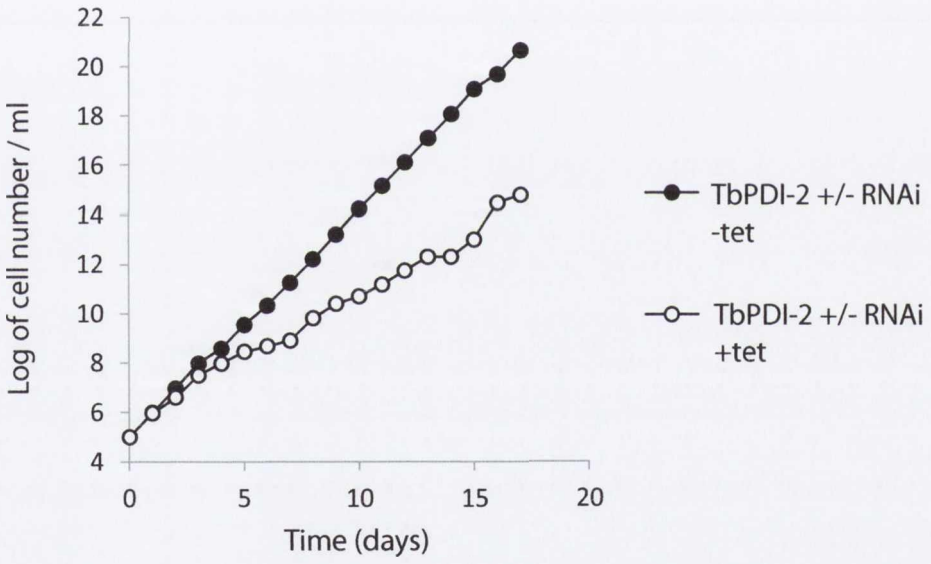


Figure 4.15 Analysis of TbPDI-2 RNA transcript of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) and parental cell lines using qRT-PCR

The relative level of expression of TbPDI-2 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbPDI-2 RNA were normalized against actin RNA and the relative quantification was calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the transcript found were as follows:

(TbPDI-2 +/- puro) 52 ± 1%;

(TbPDI-2 +/- puro TbPDI-2 RNAi) noninduced 35 ± 5.6%;

(TbPDI-2 +/- puro TbPDI-2 RNAi) induced day 1 22 ± 1%;

(TbPDI-2 +/- puro TbPDI-2 RNAi) induced day 6 1.7 ± 1%;

and (TbPDI-2 +/- puro TbPDI-2 RNAi) induced day 17 1 ± 0.1%

when compared to parental cell line. In each case the results are expressed as the mean ± standard deviation of three replicates. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.15 Analysis of TbPDI-2 RNA transcript of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) and parental cell lines using qRT-PCR

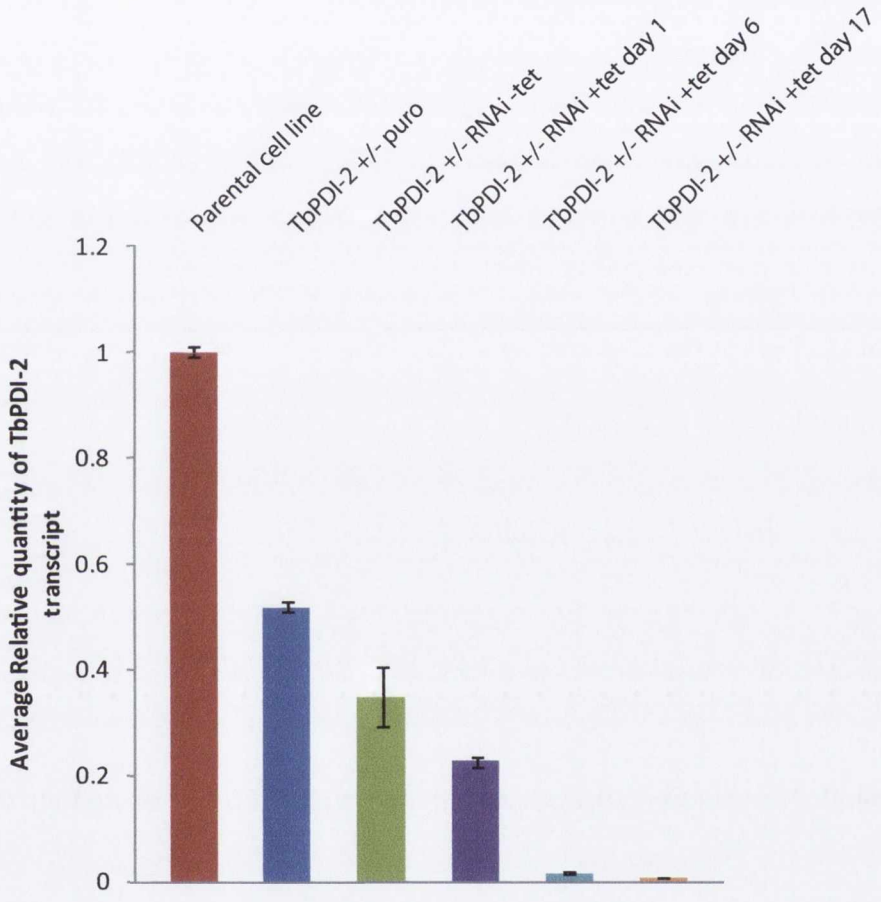


Figure 4.16 Analysis of other TbPDI RNA transcripts of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line using qRT-PCR

The relative level of expression of TbPDI-2 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbPDI-2 RNA were normalized against actin RNA and the relative quantifications were calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. In each case the results are expressed as the mean ± standard deviation of three replicates.

Panel A. The average relative quantity of the TbPDI-1 transcript found in 1) the TbPDI-2 +/- RNAi -tet was 88.3 ± 25% and 2) TbPDI-2 +/- RNAi +tet was 96.4 ± 26%.

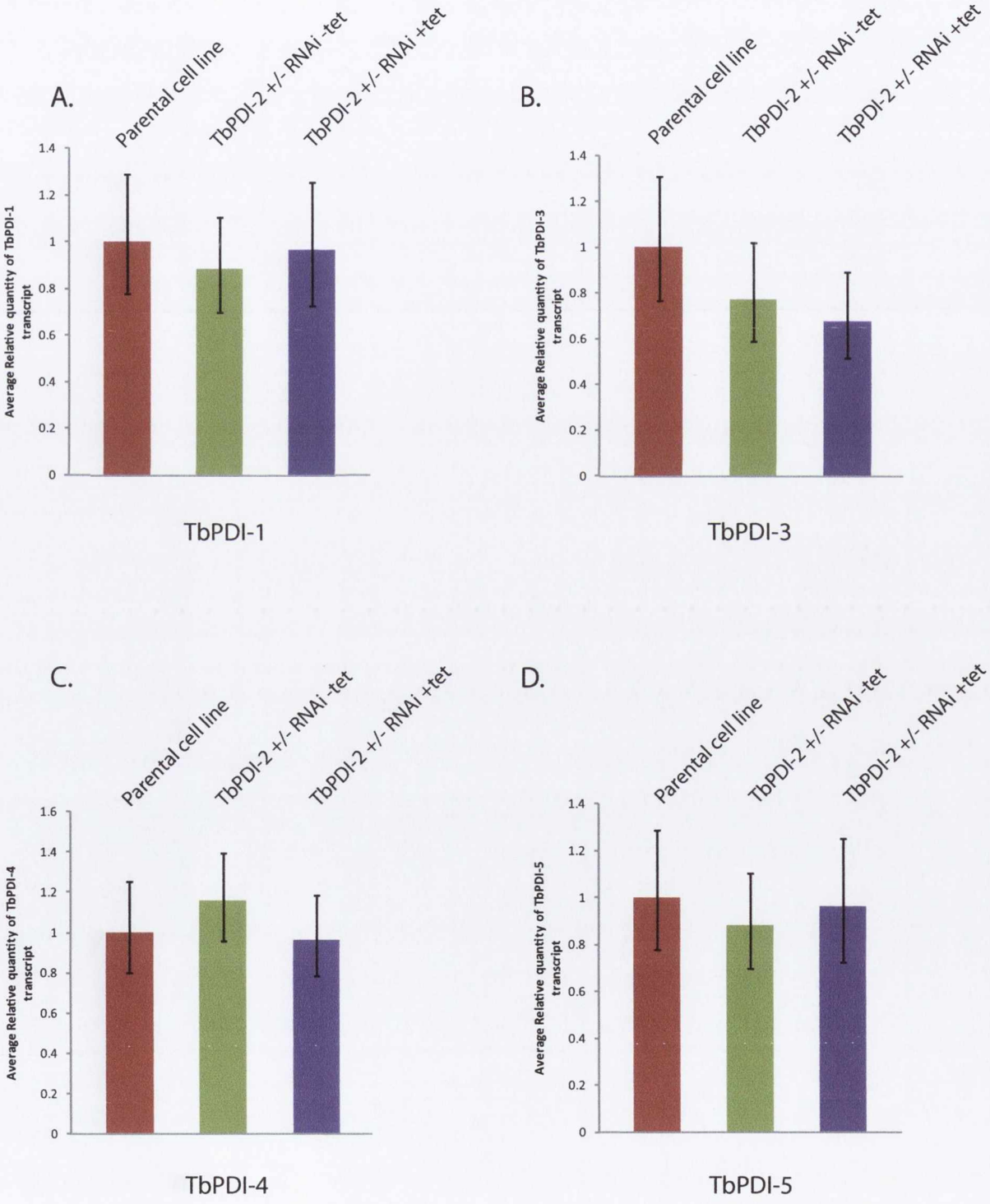
Panel B. The average relative quantity of the TbPDI-3 transcript found in 1) the TbPDI-2 +/- RNAi -tet was 77 ± 22% and 2) TbPDI-2 +/- RNAi +tet was 67.3 ± 19%.

Panel C. The average relative quantity of the TbPDI-4 transcript found in 1) the TbPDI-2 +/- RNAi -tet was found to be 115.4 ± 22% and 2) TbPDI-2 +/- RNAi +tet was found to be 96 ± 20%.

Panel D. The average relative quantity of the TbPDI-5 transcript found in 1) the TbPDI-2 +/- RNAi -tet was found to be 88.3 ± 20.2% and 2) TbPDI-2 +/- RNAi +tet was found to be 96.4 ± 26.3%.

Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.16 Analysis of other TbPDI RNA transcripts of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line using qRT-PCR



other PDIs due to sequence similarity and secondly to see if any of the PDIs were upregulated as a result of the loss of TbPDI-2. The resulting levels of transcript of the knockdown cell line versus the parental cell line were as follows: TbPDI-1 (Tb927.4.2450) 96.4 ± 26%, TbPDI-3 (Tb927.7.1300) 67.3 ± 19%, TbPDI-4 (Tb927.5.1020) 96 ± 20% and TbPDI-5 (Tb927.7.5780) 96.4 ± 26.3%. In each case the results are expressed as the mean ± standard deviation of three replicates. A fall in the level of TbPDI-3 transcript was observed to 67 % remaining when compared to the TbPDI-2 +/+ cell line however, while this appeared to be a significant result, when the sequence of *TbPDI-3* gene was used in a BLASTn of the sequence of the qRT-PCR target region of *TbPDI-2*, the program reported no significant similarity (refer to section 3.1 for parameters used). Additionally the error observed during this qRT-PCR run showed significant overlap which would be sufficient to explain the apparent fall in the TbPDI-3 transcript.

The protein level of TbPDI-2 was analyzed by Western blot analysis (Figure 4.17). The protein level of TbPDI-2 was significantly knocked down by day 6 and was not detectable by day 17 of induction although the expression of the loading control actin remained the same. From the Western blot, it was observed that the TbPDI-2 protein levels persisted at a relatively higher concentration when compared to the transcript level of TbPDI-2 indicating the rate of loss of the protein was slower than the transcript but this result was expected as it takes time for protein turnover to occur.

As there was an observable growth defect in the induced (TbPDI-2 +/- puro) TbPDI-2 RNAi cell line, a cell cycle analysis was carried out to assess if the cells were arresting during a particular stage of the cell cycle (Figure 4.18)(for a discussion of the cell cycle refer to section 1.1.3). The results showed that fewer cells were found in G_0/G_1 stage of the cell cycle which is characterized by the cell having 1 nucleus (N) and 1 kinetoplast (K) (McKean, 2003). The percentage of cells had fallen from ~80% 1N1K to ~60% 1N1K after 24 hours of induction and remained at this percentage till day 17. Interestingly, with longer induction times the occurrence of abnormal (>2N2K, 2N1K and Zoid) cell types increased from ~3% in the noninduced to ~25% in the induced cell line at day 17.

The protein TbPDI-2 was thought to be responsible for the folding of proteins in the endoplasmic reticulum; especially proteins destined for the secretory pathway such as VSG (see section 1.1.6). The VSG represents almost the entirety of the protein

produced by the secretory pathway and is rich in disulphide bonds. As such VSG was expected to be the only protein to reach the cell surface during a period of 90 minutes and so, an analysis of the rate of export of newly synthesized protein to the cell surface was carried out in bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cells. The export assay employed was designed to directly assay the rate at which newly synthesized proteins reached the surface of the cell (see section 2.34 of the Methods)(Nolan and Garcia-Salcedo, 2008). The (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line was induced for 10 days and subjected to a pulse chase analysis (Figure 4.19). The labelled proteins were analyzed by SDS PAGE autoradiography. A significant slowing of the rate of protein export to the cell surface was observed in the (TbPDI-2 +/- puro) TbPDI-2 RNAi cell line when compared to wild type cell line. The total surface protein produced by the (TbPDI-2 +/- puro) TbPDI-2 RNAi cell line during the chase period never amounted to the level observed in the control cell line after 20 minutes. In the (TbPDI-2 +/- puro) TbPDI-2 RNAi cell line two proteins of different sizes were observed. While this larger protein does not appear to be a VSG it however remains a surface protein which was folded at a reduced rate. This result indicates TbPDI-2 is directly involved in the folding of the majority of surface proteins and is required for normal cell function.

Figure 4.17 Analysis of TbPDI-2 expression in bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line and parental cell lines using Western blot

Panel A: Western blot of whole cell extracts from bloodstream form TbPDI-2 +/- RNAi cell line and parental cell lines probed with anti-TbPDI-2 antibody. Each lane represents $5x10^6$ cells.

Panel B: Western blot of whole cell extracts from bloodstream form TbPDI-2 +/- RNAi cell line and parental cell lines probed with anti-Actin antibody as a loading control. Each lane represents $5x10^6$ cells.

Figure 4.17 Analysis of TbPDI-2 expression in bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line and parental cell lines using Western blot

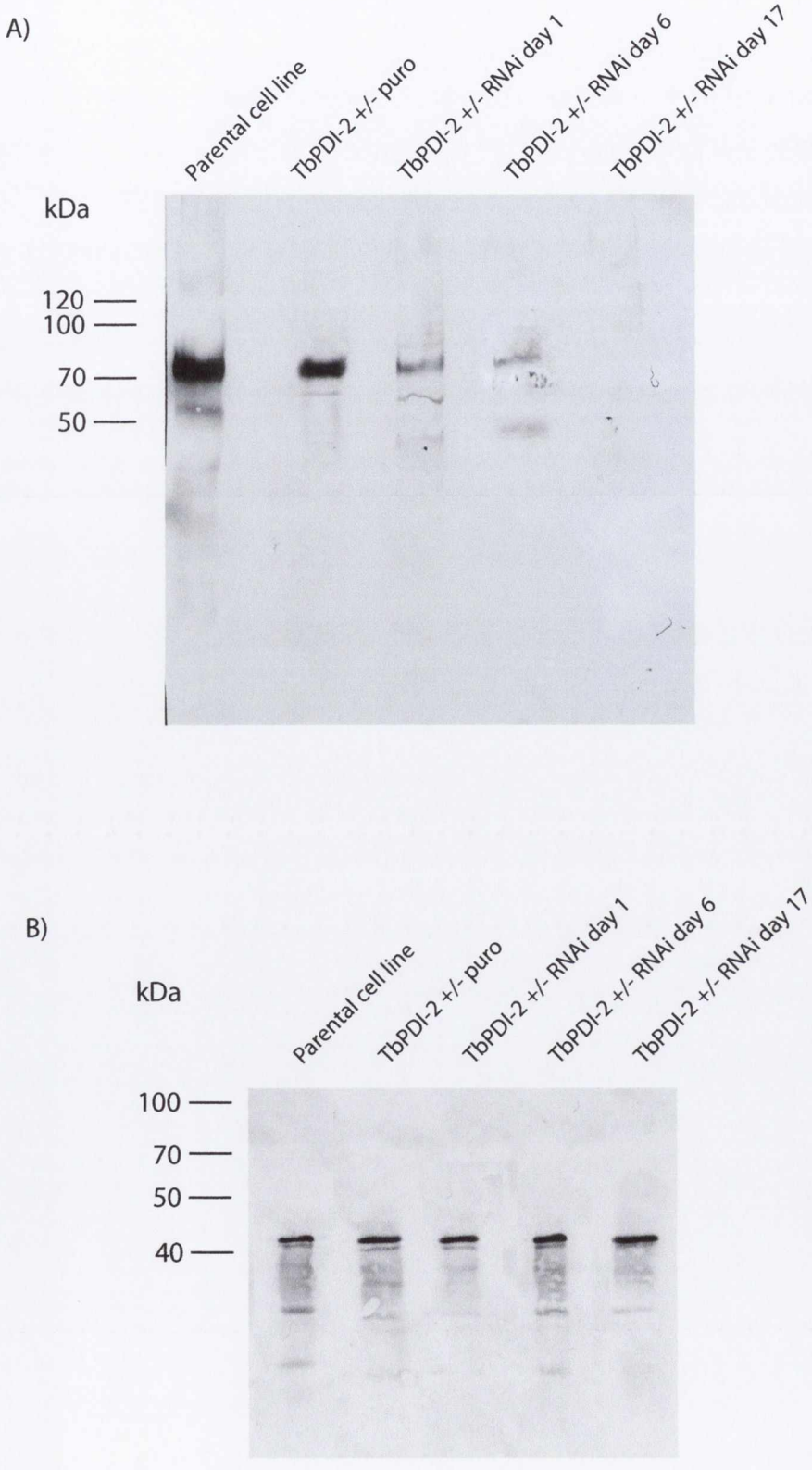


Figure 4.18 Analysis of cell cycle of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

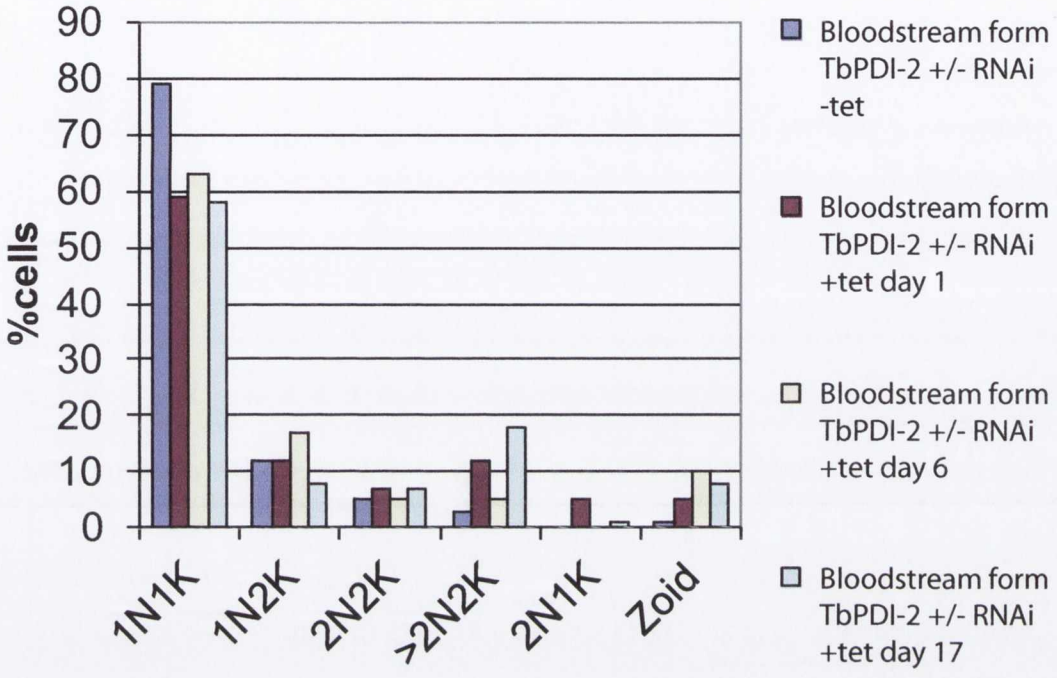
Bloodstream form TbPDI-2 +/- puro RNAi cell line was grown in HMI9, noninduced and induced over a period of 17 days. Cells were fixed at the time points indicated, stained using DAPI and viewed using fluorescence microscopy. The number of nuclei (N) and kinetoplast (K) per cell were observed for ~150 cells of each cell line. Noninduced is indicated by –tet and induced is indicated by +tet.

Panel A. The % of the total number of cells observed in various stages of the cell cycle.

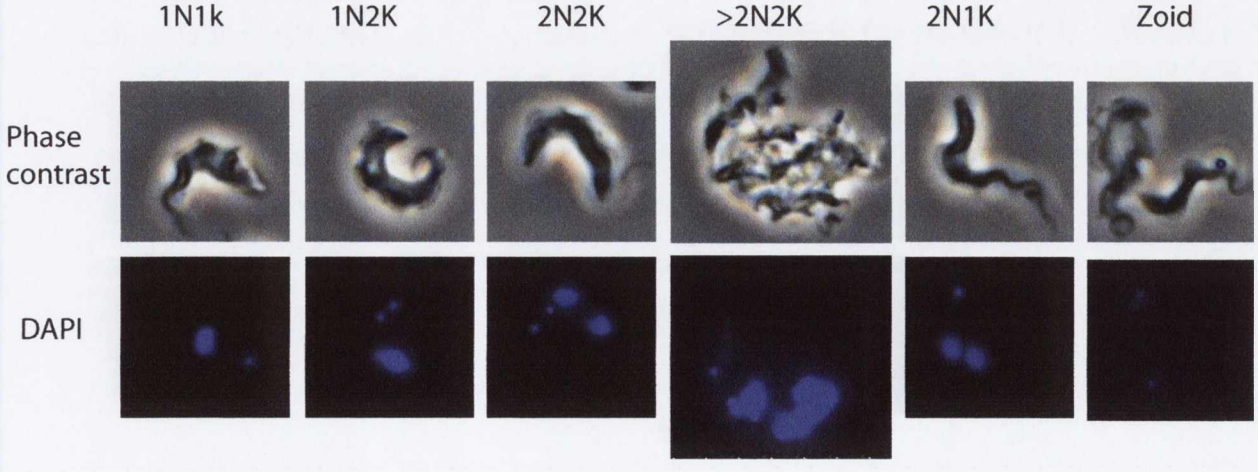
Panel B. Examples of the various stages of the cell cycle observed during fluorescence microscopy.

Figure 4.18 Analysis of cell cycle of bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

A.



B.



4.6 Generation and analysis of procyclic form TbPDI-2 knockout cell line

Previously it was shown by qRT-PCR analysis that TbPDI-2 transcription was 3 fold lower in procyclic form cells and the protein TbPDI-2 could not be detected by Western blot analysis (see section 3). This suggested that TbPDI-2 was not required in procyclic form cells to the extent it was in bloodstream form cells. However, in order to complete the study it was important to characterize procyclic TbPDI-2 in the same manner to determine if TbPDI-2 may be essential in procyclic form cells.

4.6.1 Generation and analysis of procyclic form TbPDI-2 single allele knockout cell lines

The knockout vectors TbPDI-2 KO blast and TbPDI-2 KO puro were employed in the same manner as before. Cells were transfected with the knockout vectors and the genomes of resistant clones were assessed using PCR (Figure 4.20). The approach showed the same preference for the amplification of the shorter of the two products. The transcript level of *TbPDI-2* was assessed in the single knockout cell lines using qRT-PCR and showed 60% knockdown of the parental cell line (Figure 4.21). As previously discussed in chapter 3, the level of TbPDI-2 transcript is 33% in procyclic form cells relative to bloodstream form cells. Therefore the level of TbPDI-2 transcript of (TbPDI-2 +/- puro) procyclic cells would be ~20% of the bloodstream form wild type cells.

The level of TbPDI-2 protein found in wild type procyclic form cells was below the level of detection by Western blot (Figure 4.22). As a control the same samples were blotted and probed for actin and indicated the levels of actin remained the same in both cell lines. As TbPDI-2 could not be detected in procyclic form cells, the quantity of TbPDI-2 could not be assessed in the procyclic (TbPDI-2 +/- puro) cell line. The (TbPDI-2 +/- puro) cell line showed no growth defect when assessed over a period of 15 days as was found in the bloodstream form (Figure 4.23).

4.6.2 Generation and analysis of procyclic form (TbPDI-2 -/-) cell line

As the requirement of TbPDI-2 was reduced in procyclic form cells, it was thought TbPDI-2 may not be required for normal cell growth in procyclic form cells. To address this possibility an attempt was made to generate a procyclic form TbPDI-2 null mutant. The knockout the second allele was attempted as was attempted in bloodstream forms. The procyclic form (TbPDI-2 +/- blast) and (TbPDI-2 +/- puro) cell lines were transfected with the alternate resistance marker knockout construct. Five attempts were made to achieve null mutant cells but no clones survived selection. As a positive control, parental cells were simultaneously transfected with the knockout constructs. When wild type cells were transfected with the puromycin TbPDI-2 construct, 92 positive wells were observed out of 240 wells (~38 % success rate) and 101 out of 240 positive wells were observed (~42 % success rate) when the blasticidin TbPDI-2 construct was transfected. As was found with bloodstream form cells, a procyclic form TbPDI-2 null mutant could not be generated. While this experiment suggested that TbPDI-2 was essential in culture, the experiment yielded a negative result so it could not be concluded that TbPDI-2 was essential for procyclic form cells in culture.

The expression vector TbPDI-2 p1791 was employed to allow expression of an ectopic copy of TbPDI-2 at another locus (rRNA spacer region) as was attempted with the bloodstream form cells. Both procyclic form (TbPDI-2 +/- blast) and (TbPDI-2 +/- puro) single allele knockout cells were transfected with TbPDI-2 p1791 and positive clones were selected. The (TbPDI-2 +/- puro TbPDI-2 p1791) cells were induced and the transcript level of TbPDI-2 was assessed using qRT-PCR (Figure 4.22). The induced cell line produced more than the basal amount of TbPDI-2 observed in the parental cell line. The induced (TbPDI-2 +/- puro TbPDI-2 p1791) cell line was transfected with the second TbPDI-2 knockout vector but no positive doubly resistant clones survived the selection during the 4 attempts made to generate a (TbPDI-2 -/- TbPDI-2 p1791) cell line. It was concluded that this approach could not be employed to investigate the essential nature of TbPDI-2 in procyclic form cells. Instead a combination of knockout and RNAi was employed. These experiments, although unsuccessful, indicated that a TbPDI-2 could not be knocked out without cell death occurring.

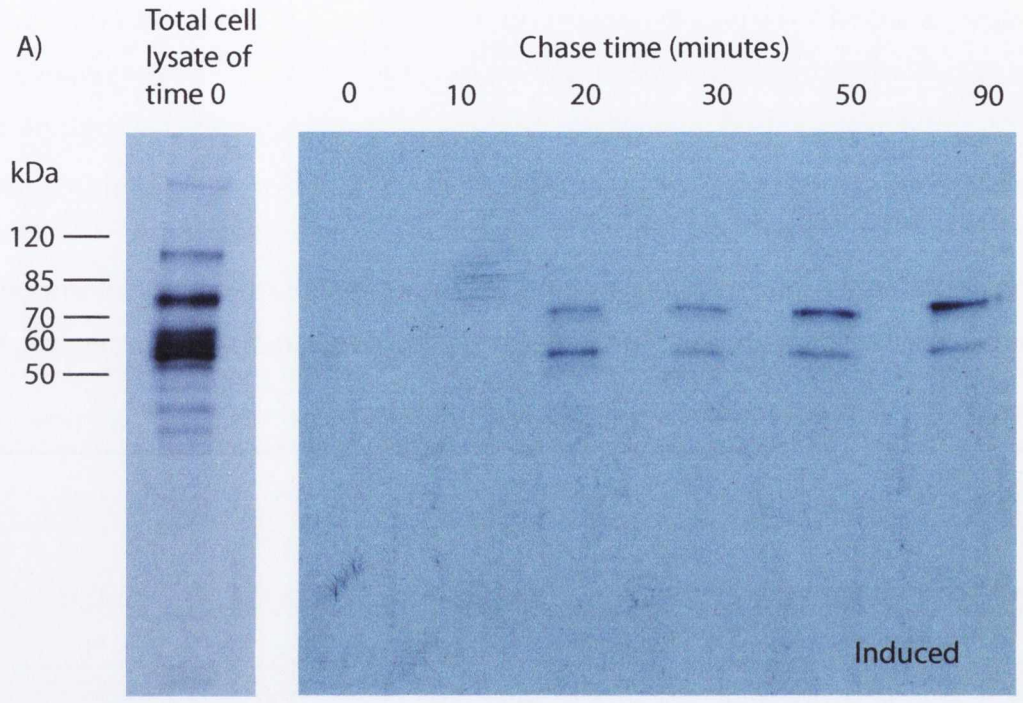
Figure 4.19 Analysis of the rate of surface protein production of bloodstream form (TbPDI-2 +/- TbPDI-2 RNAi) cell line and parental cell line

Bloodstream form cells were subjected to pulse labelling as described in the Methods. Briefly, after a 10 minute pulse of [^{35}S] methionine/cysteines medium the cells were diluted six fold in HMI9 medium and samples of cells were removed at various times (0, 10, 20, 30, 50 and 90 minutes). Cell fractions were washed at 4°C and their surface proteins were biotinylated. The cells were then lysed and the biotinylated proteins were precipitated using streptavidin agarose. The labelled proteins were analyzed by SDS PAGE autoradiography. Each lane represented 5x10^6 cells. All cell lines was analyzed at the same time.

Panel A: SDS PAGE autoradiography of fractions prepared from (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line. The total cell lysate was shown followed by biotinylated fractions (surface proteins) of the (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line.

Panel B: SDS PAGE autoradiography of fractions prepared from parental cell line. The total cell lysate was shown followed by biotinylated fractions (surface proteins) of parental cell line.

Figure 4.19 Analysis of the rate of surface protein production of bloodstream form (TbPDI-2 +/- TbPDI-2 RNAi) cell line and parental cell line



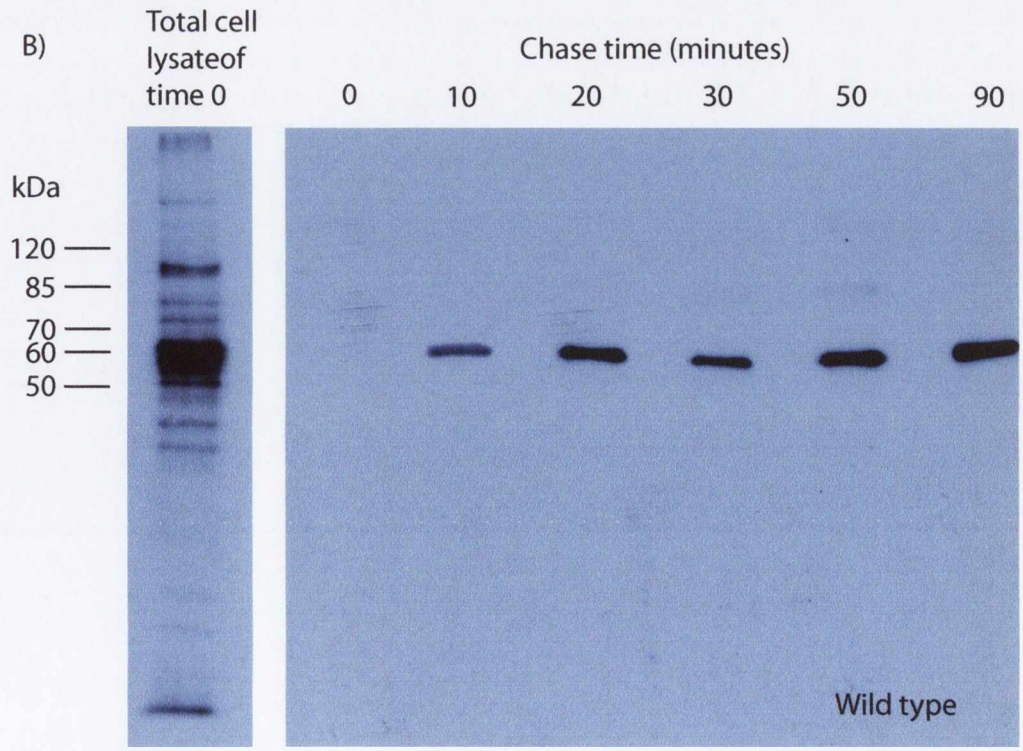


Figure 4.20 Analysis of gDNA of procyclic form TbPDI-2 single allele knock out using PCR

Agarose gel electrophoresis of PCR amplification of TbPDI-2 gene. Primers were designed to amplify the genomic region the TbPDI-2 gene. The amplified region consisted of the 5' upstream flanking region (526 bp), the TbPDI-2 open reading frame (1494 bp) and the 3' downstream flanking region (300 bp). Three samples of equal concentrations of genomic DNA were used as the PCR template: Parental cell line genomic DNA, TbPDI-2 +/- blast and TbPDI-2 +/- puro. The expected size of the unaltered TbPDI-2 PCR product was 2050 bp, the altered blasticidin-S deaminase PCR product was 919 bp and the altered puromycin acetyltransferase PCR product was 1167 bp.

Figure 4.20 Analysis of gDNA of procyclic form (TbPDI-2 +/-) cell lines using PCR

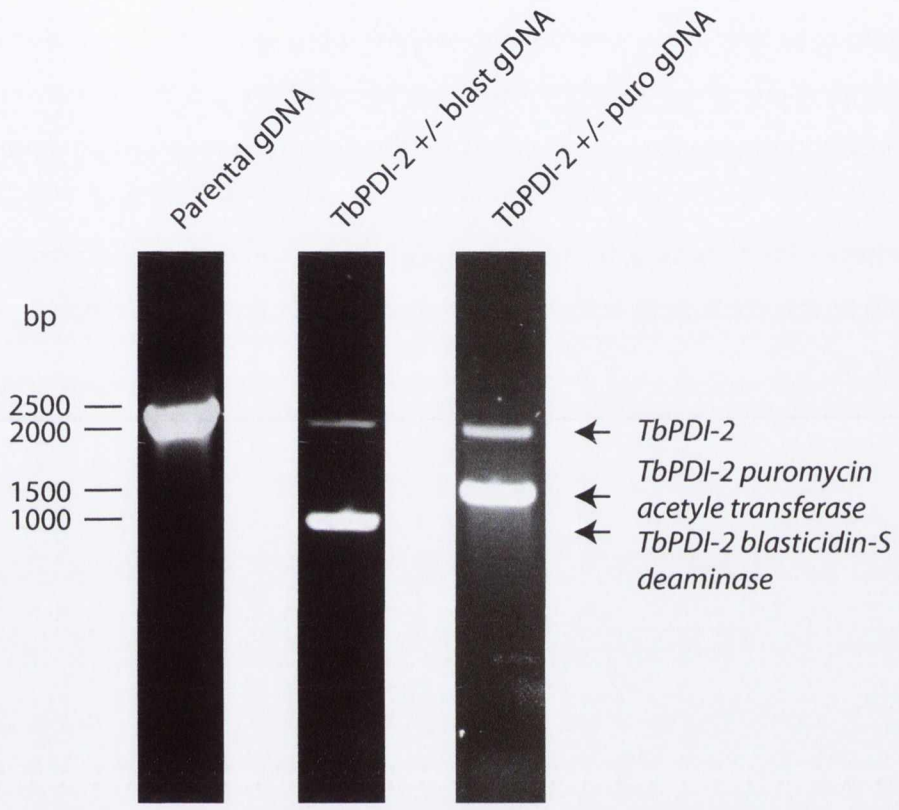


Figure 4.21 Analysis of TbPDI-2 RNA transcript of procyclic form (TbPDI-2 +/-) cell lines using qRT-PCR

The relative level of expression of TbPDI-2 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbPDI-2 RNA were normalized against actin RNA and the relative quantifications was calculated by the ΔΔCt method as described in the Methods and is expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the TbPDI-2 transcript in (TbPDI-2 +/- puro) was found to be 60 ± 8.1%. In each case the results are expressed as the mean ± standard deviation of three replicates.

Figure 4.21 Analysis of TbPDI-2 RNA transcript of procyclic form (TbPDI-2 +/-) cell lines using qRT-PCR

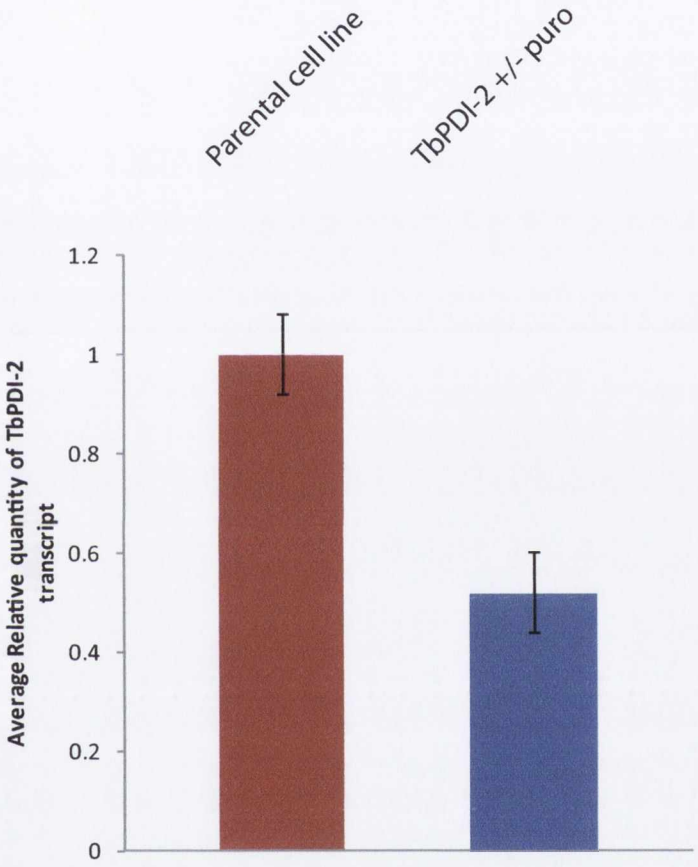


Figure 4.22 Analysis of TbPDI-2 expression in wild type bloodstream and procyclic form cells using Western blot

Panel A: Western blot of whole cell extracts from wild type bloodstream form and procyclic form cell lines probed with anti-TbPDI-2 antibody. Each lane represents $5x10^6$ cells.

Panel B: Western blot of whole cell extracts from wild type bloodstream form and procyclic form cell lines probed with anti-Actin antibody. Each lane represents $5x10^6$ cells.

Figure 4.22 Analysis of TbPDI-2 expression in wild type bloodstream and procyclic form cells using Western blot

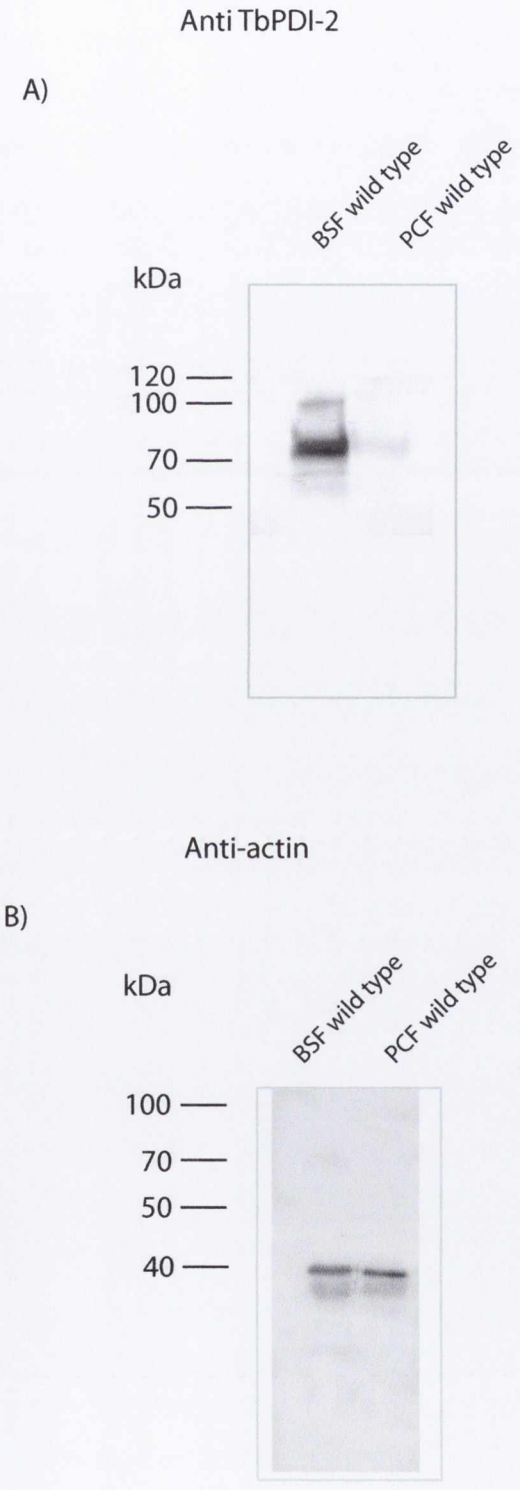
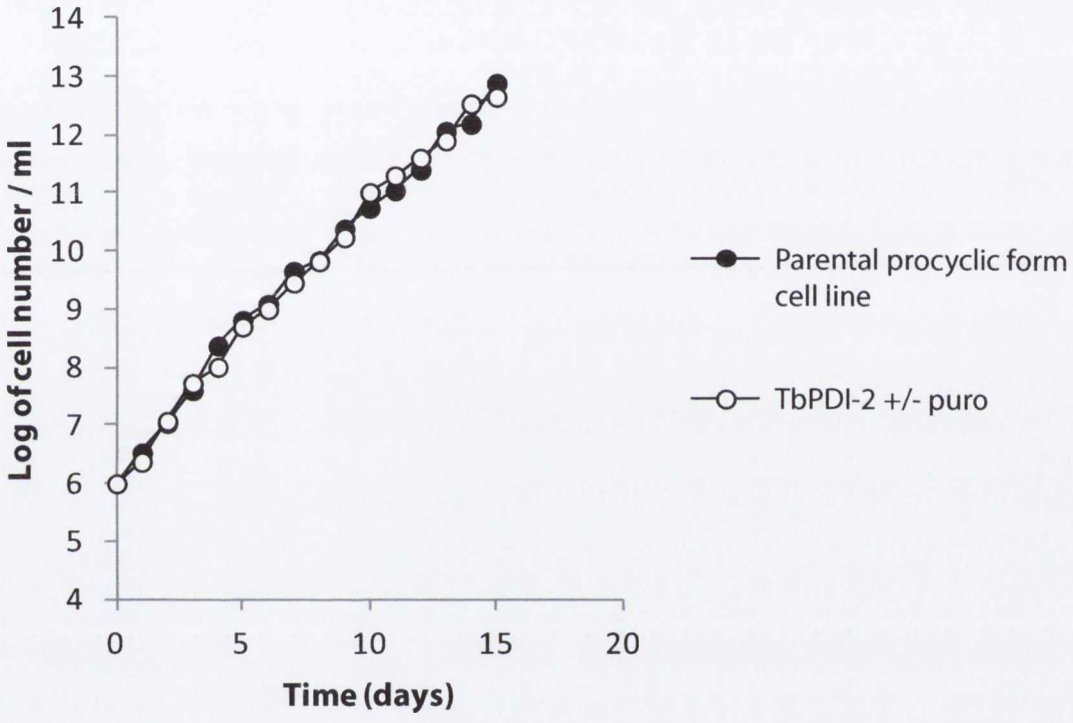


Figure 4.23Analysis of growth rate of procyclic form (TbPDI-2 +/-) cell line

Procyclic form (TbPDI-2 +/- puro) cell line and parental procyclic form cell line were grown in SDM79, as described in the Methods, over a period of 15 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 16.4 hours for the procyclic form parental cell line, 16 hours for (TbPDI-2 +/- puro) cells.

Figure 4.23 Analysis of growth rate of procyclic form (TbPDI-2 +/-) cell line



4.7 Generation and analysis of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

From the previous attempts to knockout TbPDI-2 in procyclic form cells, it appeared TbPDI-2 was essential for growth of procyclic cells in culture, as a true null mutant of TbPDI-2 or a conditional null mutant could not be generated. To further investigate the role of TbPDI-2 in procyclic form cells, a combination of (TbPDI-2 +/- puro) and RNAi knockdown was attempted. Procyclic form (TbPDI-2 +/- puro) cells were transfected with the TbPDI-2 p2T7-177 RNAi vector and cells were selected using phleomycin (5 μg/ml). The cells were grown in SDM79 with selection drugs, subcultured into two flasks and one of the flasks was induced with tetracycline (1 μg/ml). The growth of (TbPDI-2 +/- puro TbPDI-2 RNAi) induced/noninduced was monitored over a period of 15 days (Figure 4.25). Procyclic form cells typically have a doubling time of 16 hours in SDM79 (McKean, 2003). It appeared there was a growth defect in the induced cell line. The doubling time increased from 16.4 hours to 22 hours, an increase of 34% over the noninduced cell line. The growth rate of bloodstream form cells appeared to be affected to a greater degree, an increase in 67% in the rate of cell growth, than was found in the procyclic form when compare to the noninduced cell line. This result supported the hypothesis that TbPDI-2 was principally involved in the folding of VSG as procyclic form cells do not express VSG and so the loss of TbPDI-2 would have a lesser effect on procyclic form cell growth in culture.

The level of TbPDI-2 transcript was measured using qRT-PCR and was found to be knocked down by 96 % of the parental cell line transcript on day 6 of induction (Figure 4.26). Interestingly the results of the transcript level observed, due to this knockdown, are comparable to the bloodstream form knockdown. It was shown that the *TbPDI-2* transcript was 3 times greater in bloodstream form cells when compared to procyclic form cells. Thus a knockdown of 96 % in procyclic form cells would be equivalent to a knockdown of 99 % in bloodstream form cells as the starting concentration of TbPDI-2 was lower to begin with. This indicated that the two bloodstream and procyclic cell lines are comparable with regards to the level of knockdown achieved. This indicates that the greater effect on the rate of cell growth observed was not due to a lower level of *TbPDI-2* transcript.

The cell cycle was assessed to investigate if the procyclic form cells appeared to share a similar cell cycle arrest phenotype observed in the bloodstream form cell line (Figure 4.27). There appeared to be no discernible difference between induced and noninduced cell lines. Typically procyclic cells show a cell cycle profile similar to the one observed in Figure 4.27 (Castillo-Acosta et al., 2012). This result suggested that the mechanism responsible for slowing the bloodstream form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell cycle was not present in the procyclic form equivalent, although the absence of TbPDI-2 had an overall slowing effect on cell growth.

Figure 4.24 Analysis of TbPDI-2 RNA transcript of procyclic form (TbPDI-2 +/- puro TbPDI-2 1791) and parental cell line using qRT-PCR

The relative level of expression of TbPDI-2 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbPDI-2 RNA were normalized against actin RNA and the relative quantifications were calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the transcript found were as follows:

(TbPDI-2 +/- puro) 51.2 ± 8%;

(TbPDI-2 +/- puro TbPDI-2 p1791) noninduced 62.8 ± 2.7%;

and (TbPDI-2 +/- puro TbPDI-2 p1791) induced was found to be 111.4 ± 0.9% when compared to parental cell line. In each case the results are expressed as the mean ± standard deviation of three replicates. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.24 Analysis of TbPDI-2 RNA transcript of procyclic form (TbPDI-2 +/- puro TbPDI-2 1791) and parental cell line using qRT-PCR

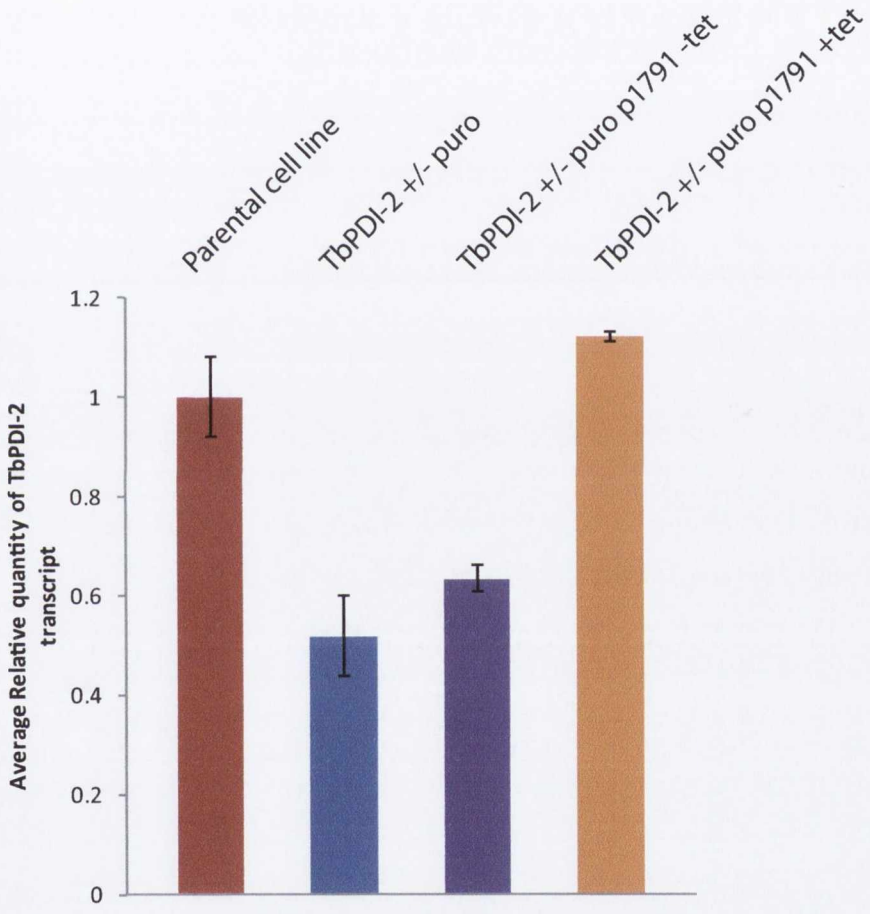


Figure 4.25 Analysis of growth rate of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

Procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line was grown in SDM79 induced; as described in the Methods over a period of 15 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 22 hours for TbPDI-2 +/- puro RNAi induced and 16 hours for parental cell line. This result represented a 34% increase in doubling time observed. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.25 Analysis of growth rate of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

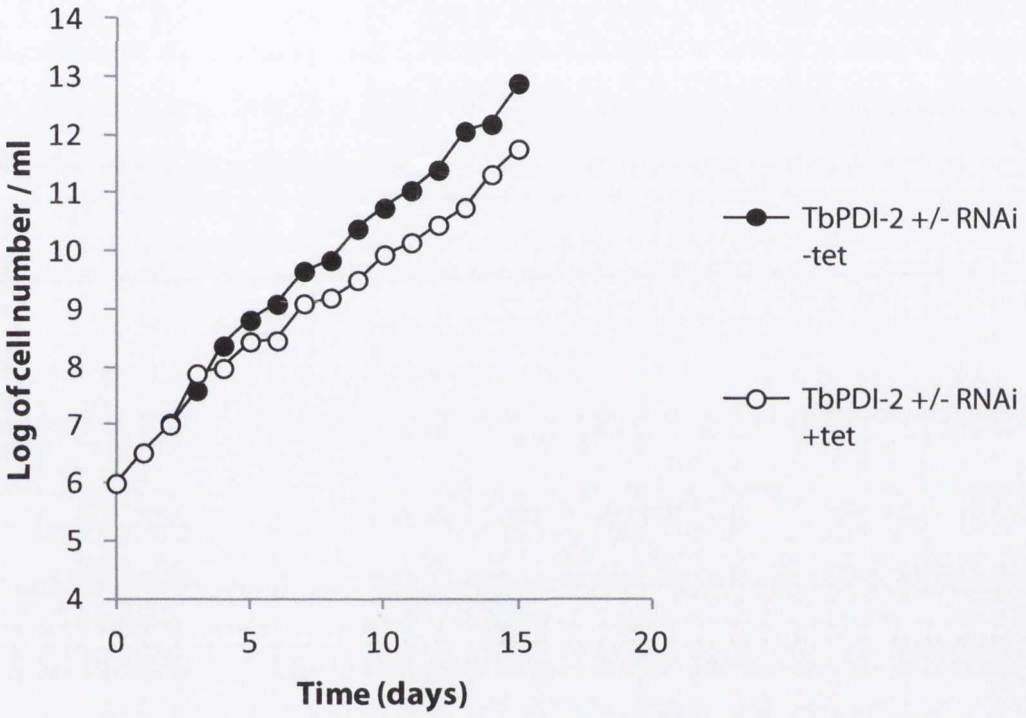


Figure 4.26 Analysis of TbPDI-2 RNA transcript of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) and parental cell lines using qRT-PCR

The relative level of expression of TbPDI-2 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbPDI-2 RNA were normalized against actin RNA and the relative quantifications were calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the transcript was as follows:

(TbPDI-2 +/- puro) was 52 ± 8 %;

(TbPDI-2 +/- puro TbPDI-2 RNAi) noninduced was found to be 29 ± 3 %;

and (TbPDI-2 +/- puro TbPDI-2 RNAi) induced day 6 was found to be 4 ± 0.5 %

when compared to parental cell line. In each case the results are expressed as the mean ± standard deviation of three replicates. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 4.26 Analysis of TbPDI-2 RNA transcript of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) and parental cell lines using qRT-PCR

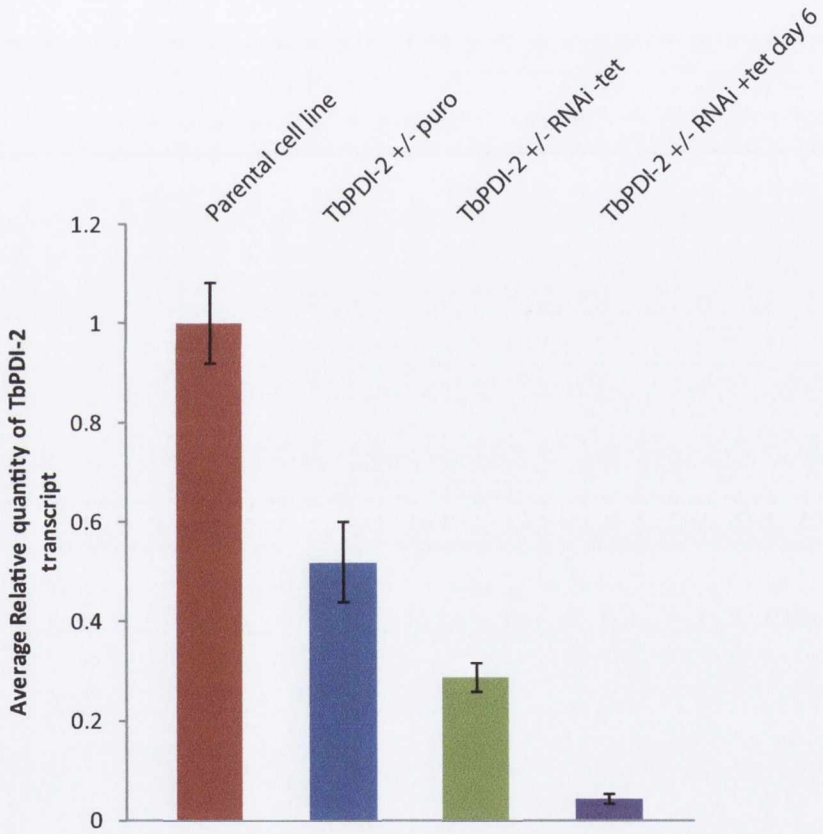


Figure 4.27 Analysis of cell cycle of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

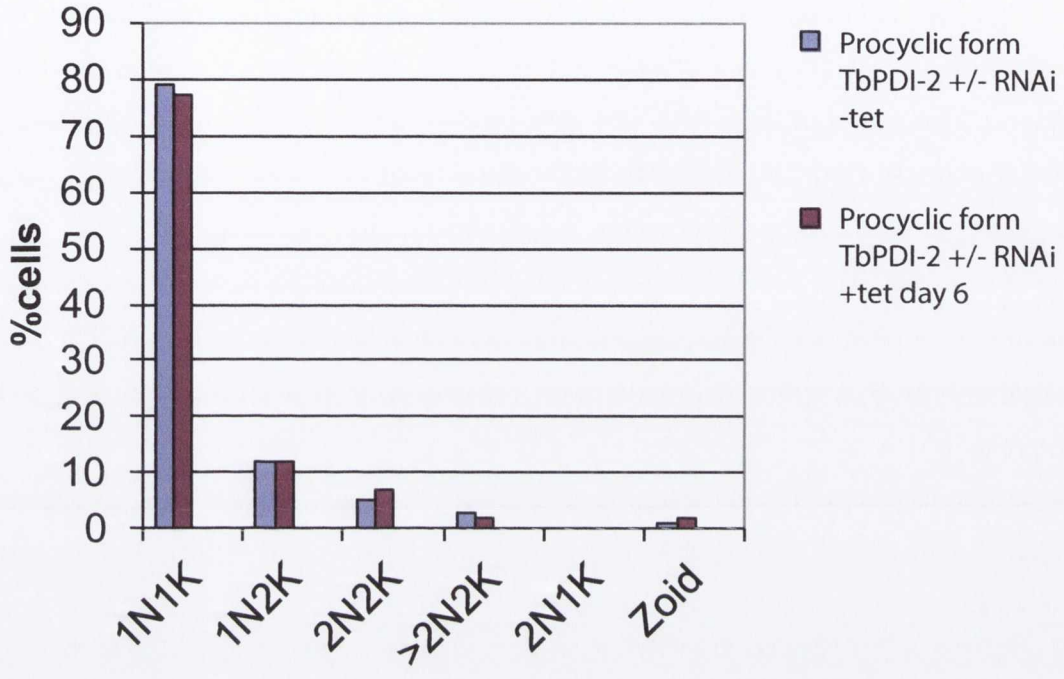
Procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line was grown in SDM79, noninduced and induced over a period of 6 days. Cells were fixed, stained using DAPI and viewed using fluorescence microscopy. The number of nuclei (N) and kinetoplast (K) per cell were observed for ~150 cells of each cell line. Noninduced is indicated by -tet and induced is indicated by +tet.

Panel A. The % of the total number of cells observed in various stages of the cell cycle.

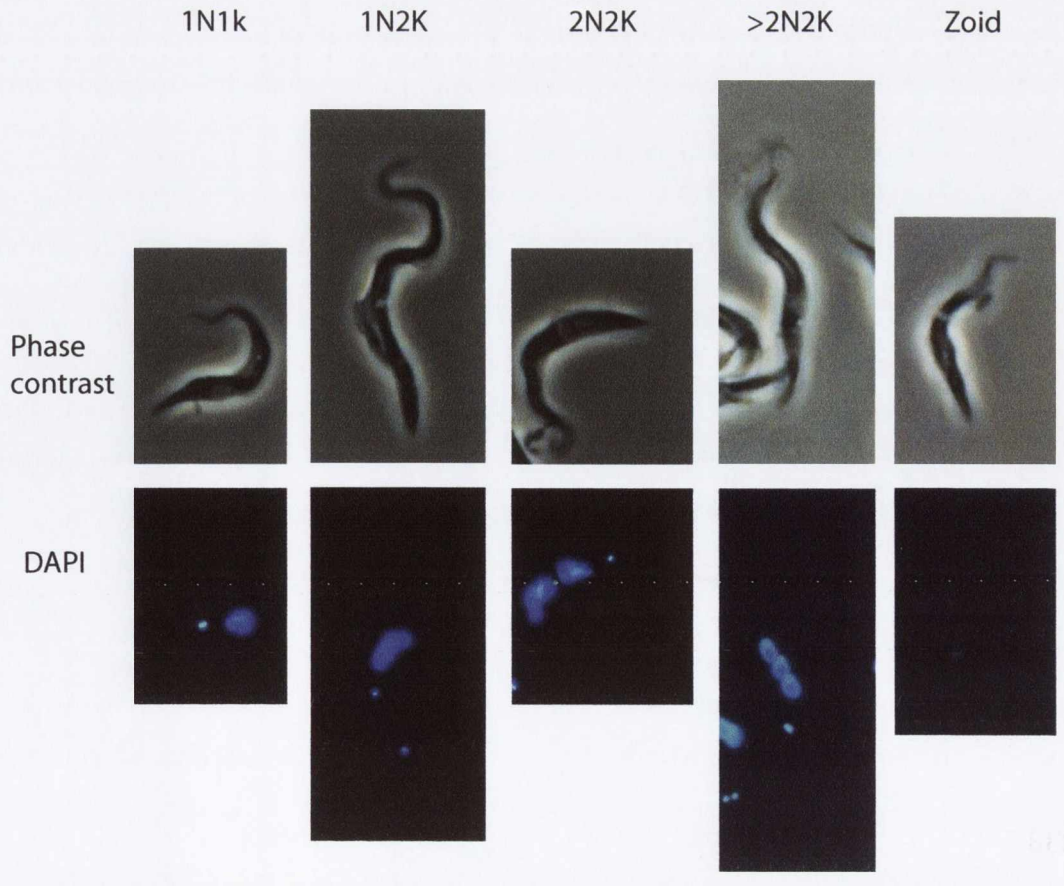
Panel B. Examples of the various stages of the cell cycle observed during fluorescence microscopy.

Figure 4.27 Analysis of cell cycle of procyclic form (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line

A.



B.



4.8 Discussion

The role of PDI has been controversial and unclear in many species, not least the case in *T. brucei*, where there have been conflicting reports as to the essential requirement of the enzyme. The results presented in this chapter support the view that TbPDI-2 is an essential activity in bloodstream and procyclic forms. Firstly, despite repeated efforts, a knockout of TbPDI-2 could not be generated or even a robust conditional null mutant. In contrast, single TbPDI-2 gene knockout cells were easily generated. Secondly, it was not possible to generate a robust conditional null mutant due to difficulties with the TbPDI-2 ectopic expression construct. Thirdly, in the case of a combination of single gene knockout and knockdown by RNAi, a slow growth phenotype was observed in bloodstream and procyclic forms. Interestingly, this effect on growth was more pronounced in the bloodstream than the procyclic form. Fourthly, at least in bloodstream forms, the effect on growth in TbPDI-2 knockdown may be due to a decrease in the rate of newly synthesized VSG exported to the cell surface.

Other class I PDI proteins, similar to TbPDI-2, have been shown to be essential for normal cell growth in other organisms. These studies reached similar conclusions as to the role of PDI in the cell and suggested that class I PDIs were essential *in vivo*. In the single celled yeast, *S. cerevisiae*, PDIp (UniProt accession: P17967) a class I PDI similar to TbPDI-2 was shown to be essential using a combination of gene disruption and tetrad analysis (Tachikawa et al., 1991). Additionally, a conditional knockout was generated by placing the PDIp gene under the control of the repressible GAL1 promoter (Günther et al., 1991; Tachibana and Stevens, 1992). As was found with the TbPDI-2 knockdown cell generated in this study, once production of PDIp in yeast was reduced to 5 to 10% of wild type levels the doubling time increased 4 fold (Tachibana and Stevens, 1992). Additionally, accumulation of procarboxypeptidase, the precursor of protein carboxypeptidase Y a protein of the yeast secretory pathway, was observed in the cells. This finding showed PDIp to be essential in yeast and suggested that PDIs may play an integral role in the growth of single celled organisms.

Conversely, in mammals, the class I PDIA3 (UniProt accession: P30101) that was found to be the most similar to TbPDI-2, was shown to not be essential for cell growth

in cultured T-lymphocyte cells or B cells (Garbi et al., 2006; Soldà et al., 2006). However, PDIA3 was shown to be essential for the folding of two secretory substrates in culture: clusterin (5 disulphide bonds) and β1 intergrin subunit (21 disulphide bonds)(Jessop et al., 2007). This finding was similar to that found in bloodstream form *T. brucei* as the export of the secretory protein, VSG, was slowed in TbPDI-2 knockdown cells. Although PDIA3 was shown to be dispensable in cultured cells, when the protein was knocked out in mice the deletion was found to be embrionically lethal, indicating PDIA3 was essential *in vivo* (Garbi et al., 2006).

From the transcript assay data observed in this study, it was shown that TbPDI-2 was knocked down ~99% of the transcript relative to the noninduced while in procyclic form cells a knockdown of 96% of the transcript relative to the noninduced cell was found. The knockdown of TbPDI-2 in procyclic form cells was comparable to the bloodstream form knockdown as the level of TbPDI-2 found in noninduced procyclic form cells was one third of that found in noninduced bloodstream form cells. As this level of knockdown was required to cause a growth defect in the cells, this indicated a significant amount of redundancy of TbPDI-2 in both lifecycle forms. Studies of the class I human PDI, known as PDIA1, have shown the export of MHC Class I proteins to the surface of the cell, in which PDI plays an essential role, was reduced only after a knockdown of the protein levels below 10%, similarly indicating a significant redundancy in the majority of the protein present (Park et al., 2006).

The dispensable activity of TbPDI-2 may be compensated for by the complementation of other PDIs present in *T. brucei*. In mammals 19 possible PDIs exist and as such some enzymes would be expected to be interchangeable in such a complex oxidative folding system. The evidence to date suggests that PDIs tend to interact with specific substrates but it has not been shown that other PDIs do not have the ability to fold the same substrates (Benham, 2012). It has been shown, using substrate trapping techniques, that the mammalian class I PDI, PDIA3, possesses 25 secreted substrate proteins (Jessop et al., 2007). Interestingly, three PDIs including PDIA3 were shown to share common substrates (Jessop et al., 2009). The PDIs PDIA3 and ERp46 (UniProt accession: Q8NB59 class 4) were observed in mixed disulphides with eight secretory proteins. Of these eight proteins, two were also found in mixed disulphides with the PDI ERp18 (UniProt accession: O95881 sdPDI). This finding supported the view that other PDIs are able to compensate for the loss of one PDI as their substrates

interacted indiscriminately with all three PDIs. However, the protein substrates shared by the PDIs were large heavily disulphide-bonded secreted proteins such as the laminins, which were likely to require more than one PDI to assist in the folding and quality control of the secretory protein, whereas other PDI substrates were shown to have restricted interactions (Jessop et al., 2009). Of the twenty five substrates shown to interact with PDIA3, sixteen were shown to solely interact with PDIA3. This finding suggested that the folding of proteins by the class I PDIs are indispensible. However, it has been reported that the substrate trapping experiments employed during this study of PDIA3 only revealed the substrate enzyme interactions which were abundant enough to be detected using this technique. It may have been possible that all nineteen human PDIs could have interacted with the substrates of PDIA3 but could not be detected. To date no knockout studies have been carried out on PDIA3 that demonstrate inhibition of these specific secretory proteins, with the exception of the aforementioned clusterin and β1 intergrin subunit. As such a specific substrate enzyme relationship between PDIA3 and these substrates has yet to be demonstrated in cultured cells or *in vivo*.

The oxidative folding pathway in *T. brucei* is less complex than the mammalian system. In this study only six TbPDIs were identified; one of which was a class I PDI. Yeast also has a simple oxidative folding pathway with only five PDIs, two of which are class I PDIs. The first class I PDI, PDIp has been shown to be essential in culture as shown by knockout of the PDIp gene (Tachikawa et al., 1991). Interestingly, the loss of PDIp gene could be rescued by the over expression of the second class I PDI, EUG1p. This finding supported the view that one class I PDI could compensate for the loss of another. However, a 10 fold increase of EUG1p protein would not occur normally *in vivo*. Additionally EUG1p was found to have a much lower activity than PDIp due to the CXXS CXXS active site motifs, suggesting that *in vivo* EUG1p would not be able to replace the loss of PDIp (Kimura et al., 2005). The trypanosome *T. brucei* only possesses one class I PDI, TbPDI-2, and as such there are no others to compensate for the loss of the enzyme. It does however possess the class II PDI TbPDI-3. This PDI has two active sites with the catalytic motif CXXC in both a-domains. As such, there was a strong possibility that TbPDI-3 could compensate for the loss of TbPDI-2. The possibility of compensation may explain how only a knockdown of TbPDI-2 transcript of 99% in bloodstream form cells and 96% in procyclic form cells resulted in a growth phenotype, which was not found to be

severe. However, if TbPDI-3 could fully replace TbPDI-2, a TbPDI-2 null mutant cell could be easily generated, which was not the case in either form. It is apparent that the majority of the activity of TbPDI-2 is dispensable but the activity still remains essential, which is typically a feature of other essential trypanosome enzymes (Chang et al., 2002; Roper et al., 2002; Turnock et al., 2007).

Single allele null mutants were easily obtainable in both bloodstream form and procyclic forms yet true null mutant cell lines could not be generated. Two approaches were employed to knockout TbPDI-2 but both methods failed to generate true TbPDI-2 null mutant cells.

The first approach was to replace the open reading frames of TbPDI-2 with resistance genes but this approach failed to produce a true null mutant, which suggested that the activity of TbPDI-2 was essential. There are few examples of true null mutant cell lines of proteins in *T. brucei* and all were shown to be non-essential in culture. The proteins GPI-PLC (Webb et al., 1994), glutamate dehydrogenase (Estévez et al., 1999), mitogen activated protein kinase 1 and 5 (Jensen et al., 2011) and aquaglyceroporin 2 (Baker et al., 2012) were all knocked out by replacement of the open reading frames of the target genes with two resistance genes using homologous recombination. In the case of Estévez et al., the glutamate dehydrogenase gene was knocked out using resistance genes flanked by different length homologous UTR regions (3' homologous region were 522 nt or 752 in length). However, GPI-PLC, mitogen activated protein kinase 1 and 5, and aquaglyceroporin were all knocked out using precisely the same strategy employed in this study.

The second approach employed was the generation of a conditional null mutant of TbPDI-2 using an inducible ectopic copy gene. This attempt however failed to produce a robust conditional null mutant in bloodstream or procyclic cells. The introduction of an ectopic copy of TbPDI-2 resulted in a twofold increase in expression of TbPDI-2 at the RNA level. The level of protein was not assayed by Western blot but as the sequence of the ectopic copy vector was found to be correct and it was shown that RNA levels of TbPDI-2 appeared to be equivalent to protein levels, it was reasonable to assume expression of the TbPDI-2 was occurring. To date there have been many examples of conditional null mutants of essential proteins generated in *T. brucei*. These studies employed the same strategy and vectors used during this study. However, all were able to assay the protein directly with ease and

demonstrated a direct increase of expression by the inducible ectopic gene at the protein level quantitatively (Chang et al., 2002; Krieger et al., 2000; Martin and Smith, 2006a, 2006b; Milne et al., 2001; Roper et al., 2002; Stokes et al., 2008; Turnock et al., 2007). Interestingly, a fourfold increase was reported from the induction of the ectopic system employed in these previous studies, which was similar to the twofold increase observed in this study. This finding suggested that the ectopic copy of TbPDI-2 was functioning correctly. However, attempts to knockout the second TbPDI-2 allele resulted in cell death or growth arrest. There were two possibilities for this result: Either the cells obtained were false positives, however this seemed unlikely as the selection drug used (blasticidin) was shown to be effective in the generation of single allele null mutants of TbPDI-2 and has been reported to be effective in the literature (Brooks et al., 2000). Alternatively, a problem had arisen with the ectopic expression system. Interestingly, this second possibility was reported to have occurred during the attempt to generate a conditional null mutant of alanine amino transferase (TbAAT) (Spitznagel et al., 2009). It was reported that the insertion of the ectopic copy gene and subsequent induction resulted in a fourfold increase in enzyme activity of TbAAT. However, when both alleles of TbAAT were knocked out, the activity fell below 5% of wild type levels, suggesting the responsiveness of the TbAAT rescue copy to tetracycline appeared to have been lost during the second round of selection. This result indicated that there was a problem with the T7 promoter causing the gene to be transcribed by endogenous read through of the ribosomal locus only. After the second round of transfection the cell growth ceased in both induced and noninduced cells. The cells were incubated further and subsequently, by the 8th day of incubation, the level of TbAAT enzyme activity rose in both induced and noninduced cells to 5% of wild type levels and growth of the cells resumed. TbPDI-2 could not be assayed directly with the same ease as TbAAT and so it was unclear if any TbPDI-2 was expressed after the second round of transfection. Furthermore, there were not enough cells to assay the levels of TbPDI-2 RNA by qRT-PCR or protein by Western blot. It is likely that the mechanism responsible for the loss of function of p1791 TbAAT was similarly responsible in this study.

As the knockout of TbPDI-2 did not allow the investigation of the role of TbPDI-2, beyond the conclusion that the enzyme was essential in cultured cells, a combination of single allele knockout and RNAi (TbPDI-2 +/- TbPDI-2 RNAi) was utilized in both bloodstream and procyclic form cells. This third approach was successfully employed

in the previous study of TbAAT (Spitznagel et al., 2009). The TbPDI-2 RNAi resulted in a knockdown of ~99% of the transcript of bloodstream forms and ~96% of the transcript in procyclic forms when compared to wild type levels. Previous attempts to knockdown TbPDI-2 did not directly measure the level of the TbPDI-2 transcript (Field et al., 2010). The absence of this measurement raised some questions as to the specificity of the RNAi employed during this previous study, not only because the knockdown was not quantified directly but also the target region chosen for RNAi was within the thioredoxin like domain of TbPDI-2. This domain is common to all thioredoxin like proteins, some of which have been shown to be essential in culture (Comini et al., 2007). The RNAi employed during this study targeted the regions outside of the conserved thioredoxin domains of TbPDI-2, to ensure only TbPDI-2 was knocked down. To further indicate the specific knockdown of TbPDI-2, other possible TbPDIs were assayed at the transcript level. The results indicated no other TbPDIs were affected by the RNAi. There was a possibility TbPDI-3 may have been knocked down. However, the error observed during this qRT-PCR analysis showed significant overlap which would be sufficient to suggest there may have been no significant fall in the level of TbPDI-3 transcript. The knockdown of TbPDI-2 was assayed at the protein level using Western blot in bloodstream form cells. Interestingly the protein level of TbPDI-2 could not be detected in procyclic form cells. This result was consistent with previous reports and may have been due to the transcript not being translated (Rubotham et al., 2005). The results of the protein level analysis showed TbPDI-2 was knocked down after 24 hours of induction and continued to fall until the protein was undetectable by day 17. The knockdown of TbPDI-2 resulted in a phenotype in both bloodstream and procyclic form cell. This phenotype and the possible function of TbPDI-2 will now be the focus of the remainder of this discussion.

From the results observed during this study, it was apparent that TbPDI-2 was essential in culture in both bloodstream and procyclic cells. However, the growth effect of the knockdown of TbPDI-2 was greater in bloodstream form cells. These results and other findings of this study support the view that the main function of TbPDI-2 is to fold proteins of the secretory pathway in both forms studied but plays a more important role in bloodstream forms. Additionally, in bloodstream forms, the principal role of TbPDI-2 is to fold VSG. A number of findings supported these views. Firstly, a growth phenotype was observed in both bloodstream and procyclic form

cells but was more pronounced in bloodstream forms. This observation suggested that TbPDI-2 played a more important role in bloodstream form cells. Secondly, the expression of TbPDI-2 was found to be three fold greater in bloodstream form cells when compared to procyclic form cells at the RNA level. This finding was also reflected at the protein level where TbPDI-2 could only be detected in bloodstream form cells. Thirdly, in bloodstream form cells, a cell cycle arrest was observed in TbPDI-2 knockdown cells, characteristic of the knockdown of VSG or proteins involved in the processing of VSG in the endoplasmic reticulum. Fourthly, in the absence of TbPDI-2, the surface proteins of bloodstream form cells required a longer time to reach the cell surface, including possibly VSG. Lastly, TbPDI-2 could not be knocked out in procyclic form cells suggesting that the protein played some essential role in the folding of procyclic proteins.

The improved knockdown of TbPDI-2 resulted in a slow growth phenotype in bloodstream and procyclic form cells but to a lesser extent in procyclic forms. The larger effect observed in bloodstream forms compared to procyclic forms was not surprising as it was likely TbPDI-2 played a more important role in bloodstream form cells; as was reflected in the transcript and protein levels in bloodstream compared to procyclic forms. The growth analysis of bloodstream and procyclic knockdown cell lines indicated the level of knockdown required for a growth defect was reached by day 5 in both cases. This resulting phenotype was most likely due to the persistence of TbPDI-2 after translation had ceased, which is a typical result of knockout/knockdown studies (Chang et al., 2002; Krieger et al., 2000; Milne et al., 2001; Spitznagel et al., 2009; Stokes et al., 2008).

To further characterise the growth phenotype observed, a cell cycle analysis was carried out. A specific cell cycle arrest was observed only in the bloodstream form cells indicating a bloodstream specific mechanism was halting the cell cycle. Previous studies have demonstrated that when the VSG is knocked down in bloodstream forms the cells enter a cell cycle arrest in the M phase of the cell cycle (2 kinetoplast and 2 nucleus are observed)(Sheader et al., 2005). This precytokinesis cell cycle arrest was caused by a yet uncharacterized mechanism which senses the amount of VSG on the cell surface and triggers a general arrest in translation initiation (Smith et al., 2009). A precytokinesis arrest was also observed when the cell cycle of the bloodstream form TbPDI-2 knockdown cell line was analyzed and this arrest may be caused by the

slowing of VSG production. However, while the bloodstream form cells could not undergo cell cytokinesis, the nucleus and kinetoplast duplication cycles continued, as was indicated by the increased proportion of multinucleated multi-kinetoplast cells observed (Das et al., 1994; LaCount et al., 2002). Interestingly, a previous study which showed the knockdown of the protein GPI8 resulted in a very similar cell cycle phenotype of an increased proportion of multinucleated muti-kinetoplast cells (Lillico et al., 2003). The protein GPI8 was shown to be responsible for the addition of GPI anchors onto nascent VSG. As GPI8 has been shown to be directly involved in the synthesis of VSG and both cell cycle phenotypes observed were similar, it was concluded that TbPDI-2 may also similarly be responsible for the direct folding of VSG.

The cell cycle arrest observed in bloodstream form cells was not observed in procyclic form cells. This finding indicated that the mechanism responsible for the phenotype observed in the bloodstream forms was not present in procyclic forms. This result again supported the view that TbPDI-2 is responsible for the folding of VSG, as VSG is not expressed in procyclic form cells. In the place of VSG, the protein procyclin is produced and is the major surface protein found in procyclic form cells. Procyclin however contains no disulphide bonds suggesting it does not require the oxidative folding pathway (Güther et al., 2006). However, the absence of VSG in procyclic forms does not rule out the ability of TbPDI-2 to fold other substrates within the endoplasmic reticulum of procyclic form cells. In support of this view it was found that a procyclic form null mutant of TbPDI-2 could not be generated. The results of the cell cycle analysis of bloodstream form and procyclic form RNAi cells suggested that the principal role of TbPDI-2 is to fold VSG in bloodstream form cells.

To further investigate if TbPDI-2 was responsible for the folding of VSG in bloodstream forms, the rate of protein export of surface proteins was analyzed using a pulse chase assay. The results clearly indicated that the (TbPDI-2 +/- puro TbPDI-2 RNAi) cell line showed a significant slowing of the rate of surface protein production when compared to wild type cells. Unexpectedly, two proteins of different sizes were exported to the surface of the RNAi cells. The first was the typical VSG (~60 kDa) and the second was a larger protein (~80 kDa). There was a possibility that the TbPDI-2 RNAi cell line had switched VSG, as switching may occur due to the stress of repeated transfection and selection of cells (Aitcheson et al., 2005). The two protein sizes could

be explained by two cell lines, each expressing a different VSG, producing the two observed proteins. However, this explanation seems unlikely as the two proteins appear to have equal concentrations across all samples taken. This finding would suggest the two proteins were equally present within a population derived from one clone as a mixed population would most likely exhibit unequal intensities due to one cell line being more abundant. Typically, cell lines expressing two different VSG exhibit different growth rates. For example, Mitat 1.2 grows at a faster rate than Mitat 1.5 (Smith et al., 2009). Additionally the second protein appears to be larger than all known VSG proteins (Smith et al., 2009). This additional size may be due to misfolding or incorrect processing of VSG within the endoplasmic reticulum or may represent some other surface protein such as invariant surface protein 75 (ISG75). However, it is unclear at present if the larger protein is a VSG. The typical $T_{1/2}$ of Mitat 1.2 (VSG of the cell line used) is ~14 minutes (Bangs et al., 1986). Other VSG have been known to fold at a slower rate compared to Mitat 1.2, for example Mitat 1.5, but generally it is considered that VSG have evolved to fold efficiently (Personal communication with Mark Carrington; Field et al., 2010). While the second protein observed during this experiment may not be a VSG, it is still evident that both proteins reached the cell surface at a slower rate when compared to wild type cells. The ~60 kDa protein was very likely to be a VSG as it is the correct size. It is generally accepted that VSG is the only protein that could be synthesized *de novo* and secreted to the cell surface in such a short period of time (Bangs et al., 1986; Field et al., 2010; Nolan and Garcia-Salcedo, 2008). Additionally the total amount of protein which reached the cell surface was less when compared to the wild type cell line. It was therefore clear that TbPDI-2 played a direct role as a catalyst in the folding of bloodstream surface proteins, most likely VSG *in vivo*.

It was evident that TbPDI-2 must have played a role in procyclic form cells as a true null mutant could not be generated and a growth phenotype was observed in the TbPDI-2 knockdown cell line. The major surface protein of procyclic form cells is procyclin. Of the procyclins found, none contain cysteine residues, as such do not contain any disulphide bonds and so would not require the oxidative folding pathway. However, it has been shown that at least twelve separate proteins are transported to the surface of the cell, which contain cysteine residues, two of which appear to be relatively abundant (Nolan and Garcia-Salcedo, 2008). Typically surface proteins contain disulphide bonds, as discussed in chapter 1, and consequently some of these

proteins may be folded by TbPDI-2. A surface protein of procyclic form cells, PSSA-2 (UniProt accession: Q389L7), has been shown to contain cysteine residues and it has been postulated that this protein may contain disulphide bonds (Fragoso et al., 2009). However a deletion of the PSSA-2 gene had no discernible effect on growth or morphology of procyclic forms in culture and so could not explain the growth defect observed in the procyclic form TbPDI-2 knockdown cells. It seems likely that the reduced folding of one of these twelve possible cysteine containing surface proteins was responsible for this growth effect. However, at present it is unclear which of these proteins are responsible.

From the results of the experiments carried out on TbPDI-2 in this study, a picture of the function of TbPDI-2 has emerged. The protein TbPDI-2 is essential in both bloodstream form and procyclic form cells indicating TbPDI-2 plays a general housekeeping role in protein folding. However, TbPDI-2 also plays a bloodstream stage specific role in the folding of VSG as indicated by the cell cycle arrest phenotype and protein export assays. This finding is very much in line with current thinking in the field of oxidative folding where it has been shown in numerous studies that PDIs are involved in folding extracellular proteins (Grubb et al., 2012; Jessop et al., 2009; Narayan, 2012; Sciandra et al., 2012; Walker et al., 2012). In the case of TbPDI-2 it is apparent that if disrupted, export of VSG is slowed which leads to a distinctive cell cycle arrest indicating a link between VSG and TbPDI-2.

5 Analysis of TbEro 1

The protein Ero 1 has been the topic of much research over the past four years during which the structure of mammalian Ero1α was solved. However, controversy over the essential nature of the protein still remains due to discrepancies presented in the literature. The protein was shown to be essential in yeast, which only possess one Ero 1 protein and is thought to be the sole provider of oxidizing equivalence to the endoplasmic reticulum. However, yeast are the only single celled organism in which this finding has been demonstrated. To date, there has been no characterization of TbEro 1 reported in the literature. As part of the investigation into the oxidative folding pathway of an African trypanosome the protein was chosen for further investigation.

A bioinformatics analysis of the oxidative folding pathway of *T. brucei* revealed that only one Ero 1 protein was present in the genome. Additionally, through expression analysis of the transcript of TbEro 1, it was found that expression of the protein was upregulated in the bloodstream stage of the life cycle. This finding suggested that the demand of TbEro 1 was increased in bloodstream form cells, which was most likely caused by the additional oxidative folding, which takes place as a result of the production of VSG. From these findings and from a review of the literature it was hypothesized that TbEro 1 must play an important role in the folding of VSG. For TbEro 1 to be involved in the oxidative folding of the VSG it must be a resident of the endoplasmic reticulum that interacts with a PDI and is most likely to play an essential role in this pathway. To test this theory the following questions were asked: Firstly, is TbEro 1 essential for the normal cell growth of bloodstream form cells and procyclic form cells? Secondly, is TbEro 1 involved in the folding of VSG? Thirdly, is TbEro 1 an endoplasmic reticulum resident protein? And finally, does TbEro 1 interact with the two known PDIs of *T. brucei* (TbPDI-1 and TbPDI-2)?

5.1 Generation and analysis of bloodstream form TbEro 1 knockdown cell line

The protein TbEro 1 was thought to be the main electron accepting enzyme of the endoplasmic reticulum. To test this hypothesis the function of TbEro 1 was to be assessed using reverse genetics. The TbEro 1 p2T7-177 RNAi vector was cloned as described in the Methods and the various stages of the process were assessed using agarose gel electrophoresis (Figure 5.1). The completed vector was sequenced to ensure the correct orientation of the TbEro 1 fragments in the RNAi vector (Figure 5.2). The TbEro 1 RNAi vector was then linearized and transfected into bloodstream form cells. A positive clone was selected using phleomycin (5 μg/ml), grown, divided into two flasks and one flask was induced. The growth of the cells was monitored over a period of 10 days and revealed a slow growth phenotype in the induced cell line (Figure 5.3). The doubling times were calculated and found to be 7.5 hours for the parental cell line and 13 hours for the induced (TbEro 1 RNAi) cell line; an increase of 73% in doubling time. The transcript level of *TbEro 1* was measured using qRT-PCR (Figure 5.4) and the results showed a knockdown to 15.2% remaining of the *TbEro 1* transcript was achieved relative to the parental cell line. A small amount of leak was observed in the noninduced cell line.

As TbPDI-2 and TbEro 1 were thought to be part of a the electron transfer chain from unfolded proteins in the endoplasmic reticulum to molecular oxygen, it would follow that the same cell cycle arrest phenotype should be observed by the knockdown of TbEro 1, as was seen in the TbPDI-2 deficient bloodstream form cells. To test this hypothesis a cell cycle analysis was carried out (Figure 5.5). Cells in the 1N1K variant rapidly fell from ~80% in the noninduced form to ~60% after 1 day and continued to fall to 50% on day 5 then to below 40% on day 10. The occurrence of abnormal cell lines rose rapidly from 5% of the noninduced to 13% after day 1 of induction and continued to rise to 35% at day 5 and finally to 41% by day 10 of induction. This phenotype seemed to be a more severe growth arrest, when compared to the growth arrest observed in the TbPDI-2 knockdown cell line but both cell lines seemed to display deficiencies undergo cytokinesis. The cell cycle was instead continuing resulting in multiple kinetoplast and nucleus. The severity of the phenotype observed in the induced (TbEro 1 RNAi) cells was however not a surprise, as it is thought TbEro 1 is the sole provider of oxidizing equivalence to the endoplasmic reticulum.

To observe how surface protein production rates were affected by the knockdown of TbEro 1, a pulse chase analysis was carried out (Figure 5.6). The results showed the folding of surface protein was slowed by 20 minutes when compared to the parental cell line. The total amount of surface protein folded in the (TbEro 1 RNAi) cell line does not appear to reach the same levels as those observed in the parental cell line even after 90 minutes.

5.2 Generation and analysis of a bloodstream form cell line that expresses an epitope tagged TbEro 1

The protein Ero 1 is thought to be resident of the endoplasmic reticulum which interacts directly with the PDIs to form disulphide bonds within substrate proteins of the oxidative folding pathway in the endoplasmic reticulum. To establish if TbEro 1 was an endoplasmic reticulum resident and to find if the PDIs TbPDI-1 and TbPDI-2 interact with TbEro 1, an expression system for an epitope tagged TbEro 1 was cloned as described in the Methods. Amplification products of various steps were analyzed using agarose gel electrophoreses (Figure 5.7). The resulting construct TbEro 1 p2620 was sequenced to ensure the TbEro 1 ORF contained no errors (Figure 5.8). Bloodstream form cells were transfected with the TbEro 1 p2620 vector and positive clones were selected using phleomycin (5μg/ml). The cells were cloned as described in the methods to ensure the population of cells was derived from a single clone. These cells were then grown, divided into two flasks and one flask was induced using tetracycline (1 μg/ml). The growth was monitored over a period of 6 days to ensure the expression system did not affect normal growth of the bloodstream form cell line (Figure 5.9). The protein expression of epitope tagged TbEro 1 was analyzed using Western blot (Figure 5.10) and found a epitope tagged protein of ~120 kDa was expressed in induced cells. The predicted mass of epitope tagged TbEro 1 was found to be 68 kDa and so the increased mass of the epitope tagged TbEro 1 expressed in the bloodstream form cells was surprising. The extra mass was thought to be caused by extensive *N*-glycosylation of the protein due to the seven possible *N*-glycosylation sites identified during the bioinformatics analysis. In order to investigate *N*-glycosylation, epitope tagged TbEro 1 was immunopurified and digested with *N*-glycosidase F, which cleaves most *N*-glycans or endoglycosidase H, which cleaves high mannose or hybrid but not complex glycans (Figure 5.11). The

results showed a fall in molecular weight from ~120 kDa to ~100 kDa when TbEro 1 HA was digested with *N*-glycosidase F and an identical reduction in mass when digested with endoglycosidase H. These results taken together suggest TbEro1 is heavily *N*-glycosylated and that these glycosylations are not complex in structure. It is likely the additional 32 kDa in apparent mass of the protein was due to glycans which could not be cleaved by the glycosidases.

To assess if TbEro 1 was a membrane associated/bound protein or a soluble protein, bloodstream form TbEro 1 cells were subjected to detergent lysis or mechanical lysis (Figure 5.12). The detergent lysis fractions once analyzed by Western blot revealed all of the TbEro 1 protein had been detected in the supernatant whereas after mechanical lysis all the TbEro 1 protein was found in the pellet fraction. This result suggested that TbEro 1 was a membrane associated protein.

To investigate if TbEro 1 did interact with TbPDI-1 or TbPDI-2 *in vivo*, a co-immunoprecipetation analysis was carried out. TbEro 1 HA was immunoprecipetated using anti-HA agarose and washed as described in the Methods. The resulting protein samples were analyzed using Western blot and probed for TbPDI-1 and TbPDI-2 (Figure 5.13). The results clearly indicated TbPDI-1 and TbPDI-2 were co-immunoprecipetated with TbEro1 HA showing they interacted *in vivo*.

To establish what subcellular organelle TbEro 1 resided in, a subcellular fractionation of bloodstream form cells expressing TbEro1 HA was carried out (Figure 5.14). TbEro 1 was compared to other established proteins and was found to behave as the endoplasmic reticulum resident protein BiP and shared a somewhat similar profile to TbPDI-2 but not TbPDI-1. This result suggests TbEro 1 is a resident of the endoplasmic reticulum.

To further investigate the localization of TbEro 1, bloodstream form cells expressing TbEro 1 HA were subjected to immunofluorescence microscopy (Figure 5.15). The distribution of TbEro 1 appeared to be similar to the distribution captured during the immunnofluorescence of BiP suggesting the TbEro 1 was in the endoplasmic reticulum (Bangs et al., 1993). Co-localization of TbEro 1 and TbPDI-2 revealed the two proteins co-localized but TbEro 1 showed a wider distribution throughout the endoplasmic reticulum where as TbPDI-2 appeared to have a more discrete distribution near the flagellar pocket (Figure 5.16).

Figure 5.1 Analysis of amplification products of various cloning steps of TbEro 1 RNAi construct

To produce the p2T7-177 TbEro 1 RNAi vector two cloning steps were required to subclone two opposing complementary *TbEro 1* fragments. Once transfected into trypanosome cells, the vector was transcribed and produced a RNA hairpin.

Panel A: A fragment of the TbEro 1 ORF was amplified by PCR with primers which included *XhoI* and *HindIII* restriction sites (expected size 607 bp). This PCR was subcloned into the p2T7-177 RNAi vector and the product was digested with *XhoI* and *HindIII* for confirmation (expected size 6028 and 596 bp).

Panel B: The same fragment of the TbEro 1 ORF was amplified by PCR with primers which included *BamHI* and *SpeI* (expected size 607 bp). This PCR was subcloned into the p2T7-177 & TbEro 1 vector and the plasmid DNA obtained from positive clones was subjected to digestion with *BamHI* and *SpeI* (expected size 6603 and 599 bp).

Figure 5.1 Analysis of amplification products of various cloning steps of TbEro 1 RNAi construct

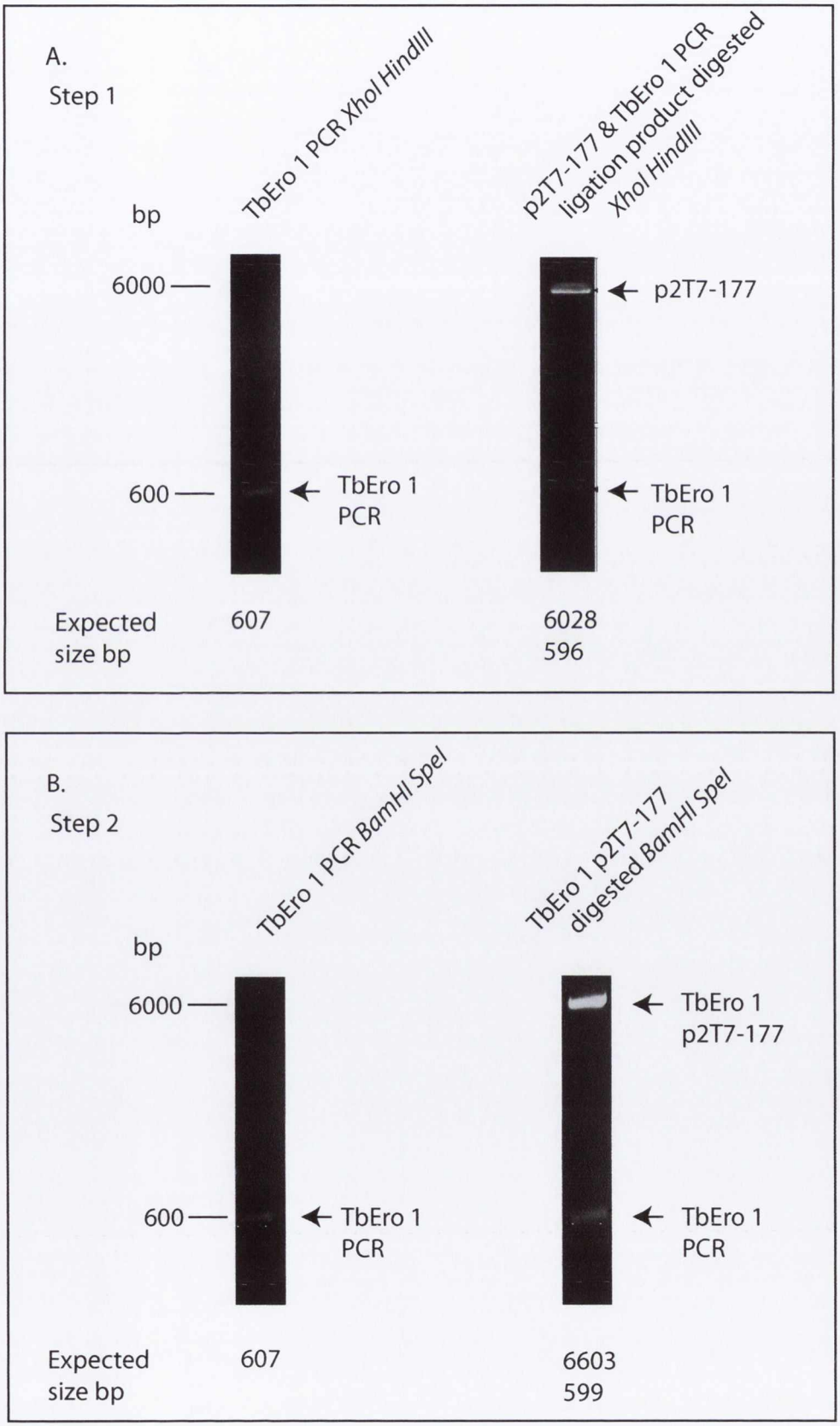
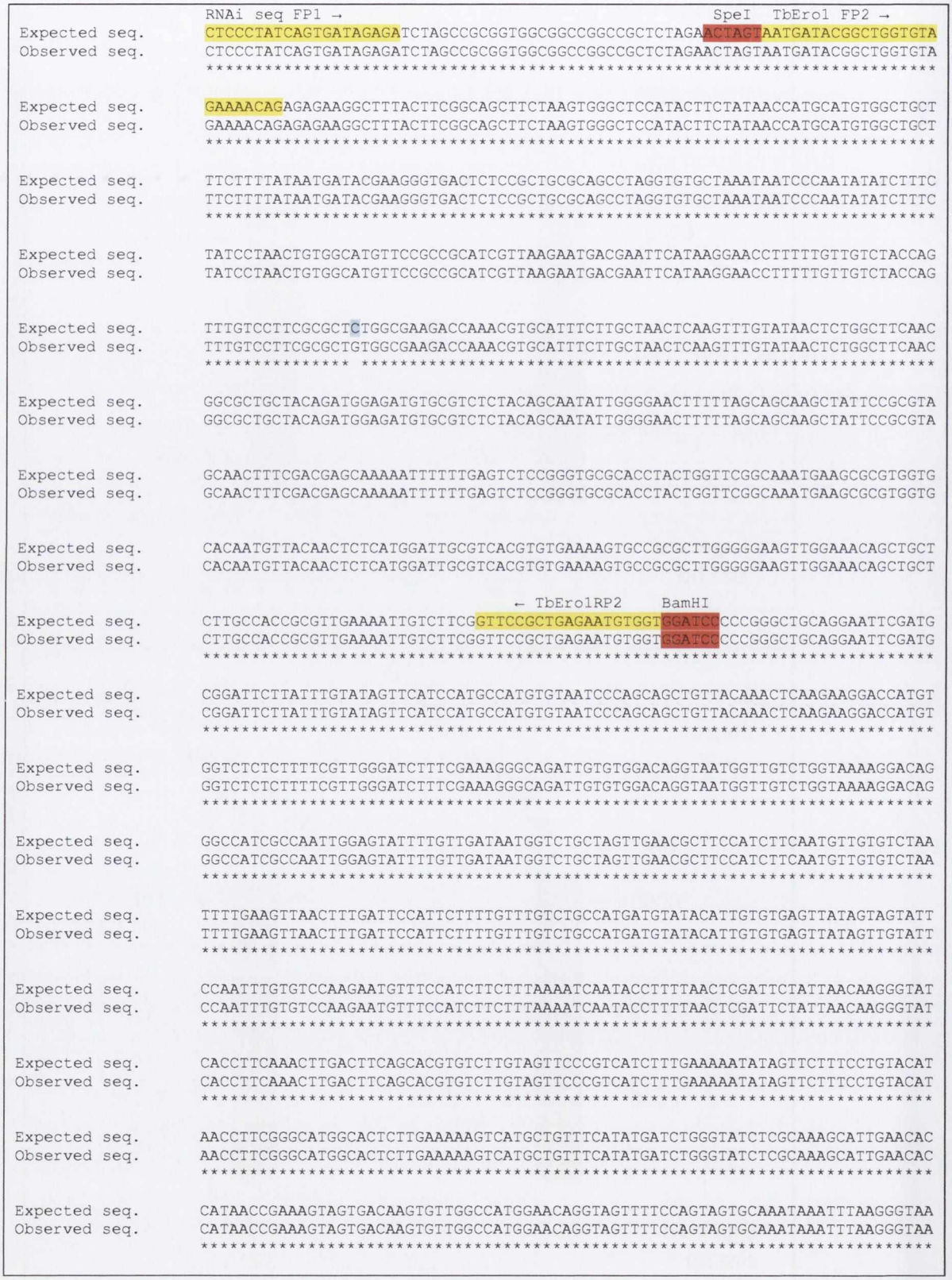


Figure 5.2 Analysis of TbEro 1 p2T7-177 construct sequence

A.



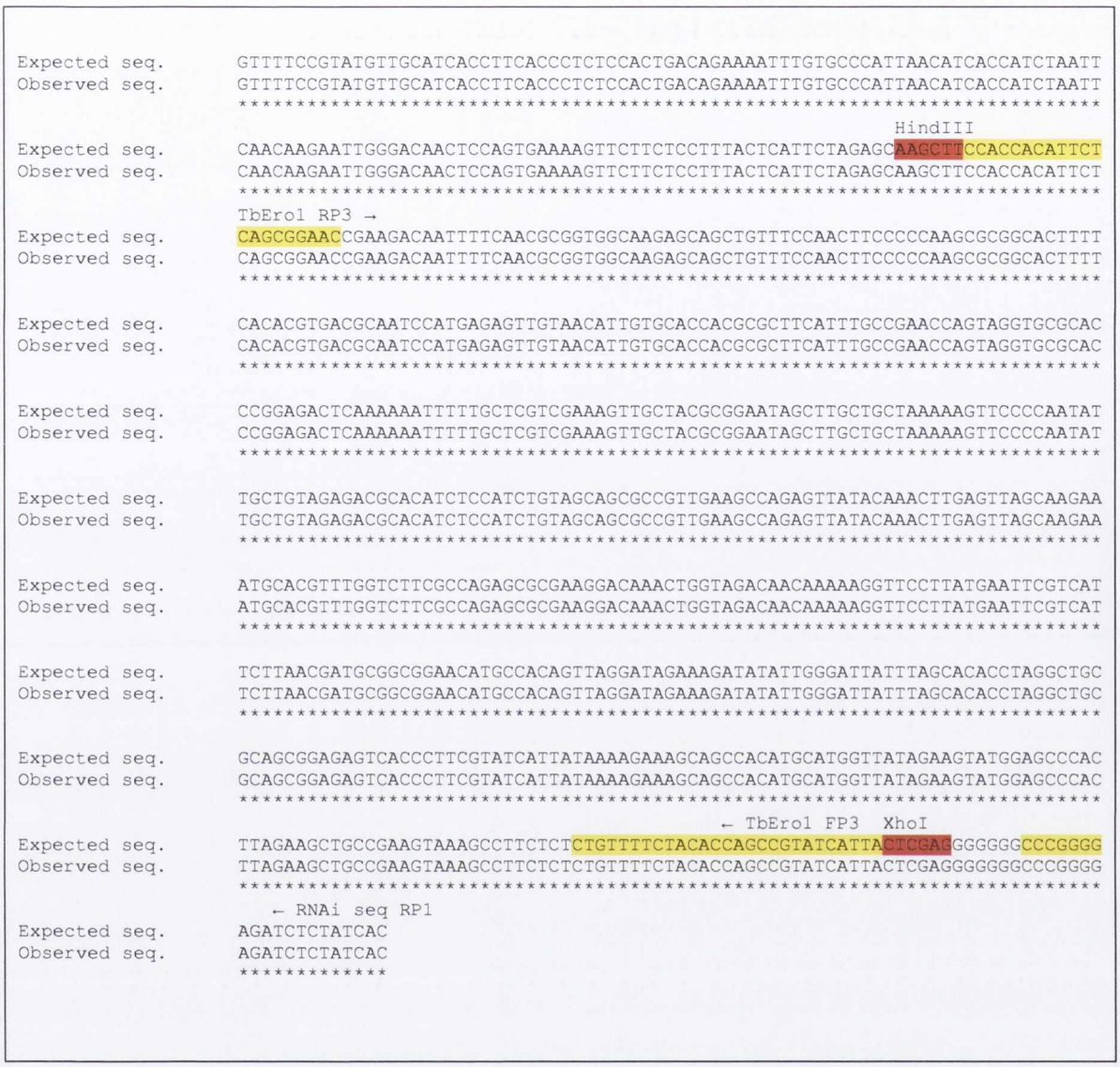

Expected seq. GTTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTGTGCCCATTAACATCACCATCTAATT
Observed seq. GTTTTCCGTATGTTGCATCACCTTCACCCTCTCCACTGACAGAAAATTTGTGCCCATTAACATCACCATCTAATT
HindIII
Expected seq. CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTACTCATTCTAGAGCAAGCTTCCACCACATTCT
Observed seq. CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTACTCATTCTAGAGCAAGCTTCCACCACATTCT
TbEro1 RP3 →
Expected seq. CAGCGGAACCGAAGACAATTTTCAACGCGGTGGCAAGAGCAGCTGTTTCCAACTTCCCCCAAGCGCGGCACTTTT
Observed seq. CAGCGGAACCGAAGACAATTTTCAACGCGGTGGCAAGAGCAGCTGTTTCCAACTTCCCCCAAGCGCGGCACTTTT
Expected seq. CACACGTGACGCAATCCATGAGAGTTGTAACATTGTGCACCACGCGCTTCATTTGCCGAACCAGTAGGTGCGCAC
Observed seq. CACACGTGACGCAATCCATGAGAGTTGTAACATTGTGCACCACGCGCTTCATTTGCCGAACCAGTAGGTGCGCAC
Expected seq. CCGGAGACTCAAAAAATTTTTGCTCGTCGAAAGTTGCTACGCGGAATAGCTTGCTGCTAAAAAGTTCCCCAATAT
Observed seq. CCGGAGACTCAAAAAATTTTTGCTCGTCGAAAGTTGCTACGCGGAATAGCTTGCTGCTAAAAAGTTCCCCAATAT
Expected seq. TGCTGTAGAGACGCACATCTCCATCTGTAGCAGCGCCGTTGAAGCCAGAGTTATACAAACTTGAGTTAGCAAGAA
Observed seq. TGCTGTAGAGACGCACATCTCCATCTGTAGCAGCGCCGTTGAAGCCAGAGTTATACAAACTTGAGTTAGCAAGAA
Expected seq. ATGCACGTTTGGTCTTCGCCAGAGCGCGAAGGACAAACTGGTAGACAACAAAAAGGTTCCTTATGAATTCGTCAT
Observed seq. ATGCACGTTTGGTCTTCGCCAGAGCGCGAAGGACAAACTGGTAGACAACAAAAAGGTTCCTTATGAATTCGTCAT
Expected seq. TCTTAACGATGCGGCGGAACATGCCACAGTTAGGATAGAAAGATATATTGGGATTATTTAGCACACCTAGGCTGC
Observed seq. TCTTAACGATGCGGCGGAACATGCCACAGTTAGGATAGAAAGATATATTGGGATTATTTAGCACACCTAGGCTGC
Expected seq. GCAGCGGAGAGTCACCCTTCGTATCATTATAAAAGAAAGCAGCCACATGCATGGTTATAGAAGTATGGAGCCCAC
Observed seq. GCAGCGGAGAGTCACCCTTCGTATCATTATAAAAGAAAGCAGCCACATGCATGGTTATAGAAGTATGGAGCCCAC
← TbEro1 FP3 XhoI
Expected seq. TTAGAAGCTGCCGAAGTAAAGCCTTCTCTCTGTTTTCTACACCAGCCGTATCATTACTCGAGGGGGGGCCCGGGG
Observed seq. TTAGAAGCTGCCGAAGTAAAGCCTTCTCTCTGTTTTCTACACCAGCCGTATCATTACTCGAGGGGGGGCCCGGGG
← RNAi seq RP1
Expected seq. AGATCTCTATCAC
Observed seq. AGATCTCTATCAC

Figure 5.2 Analysis of TbEro 1 p2T7-177 construct sequence

The TbEro 1 p2T7-177 construct was sequenced to ensure the TbEro 1 fragments were correctly ligated into the p2T7-177 RNAi vector. The expected sequence and observed sequence obtained were compared. The location where primers bound to the sequence are shown in yellow and the primer names and their direction are indicated above. Restriction sites are shown in red and the restriction enzymes at which these sites cut are indicated above. Agreement is indicated by a * below the sequence and disagreements are highlighted in light blue.

Figure 5.3 Analysis of growth rate of bloodstream form TbEro 1 RNAi cell line

Bloodstream form TbEro 1 RNAi cell line was grown in HMI9 and induced; as described in the Methods over a period of 10 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 7.5 hours for parental cell line and 13 hours for induced TbEro 1 RNAi. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.3 Analysis of growth rate of bloodstream form TbEro 1 RNAi cell line

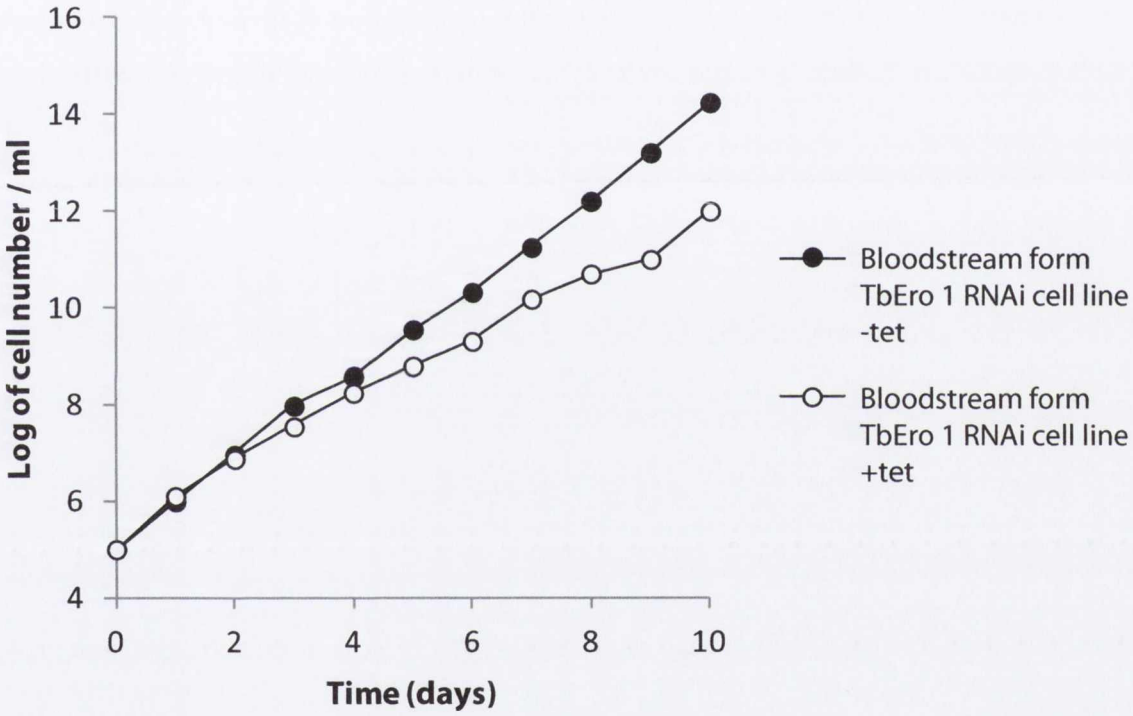


Figure 5.4 Analysis of TbEro 1 RNA transcript of bloodstream form TbEro 1 RNAi and parental cell lines using qRT-PCR

The relative level of expression of TbEro 1 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbEro 1 RNA were normalized against actin RNA and the relative quantifications were calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the TbEro 1 transcript in noninduced TbEro 1 RNAi was found to be 84% ± 18.5% and induced TbEro 1 RNAi was found to be 15% ± 6% when compared to parental cell lines. In each case the results are expressed as the mean ± standard deviation of three replicates. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.4 Analysis of TbEro 1 RNA transcript of bloodstream form TbEro1 RNAi and parental cell lines using qRT-PCR

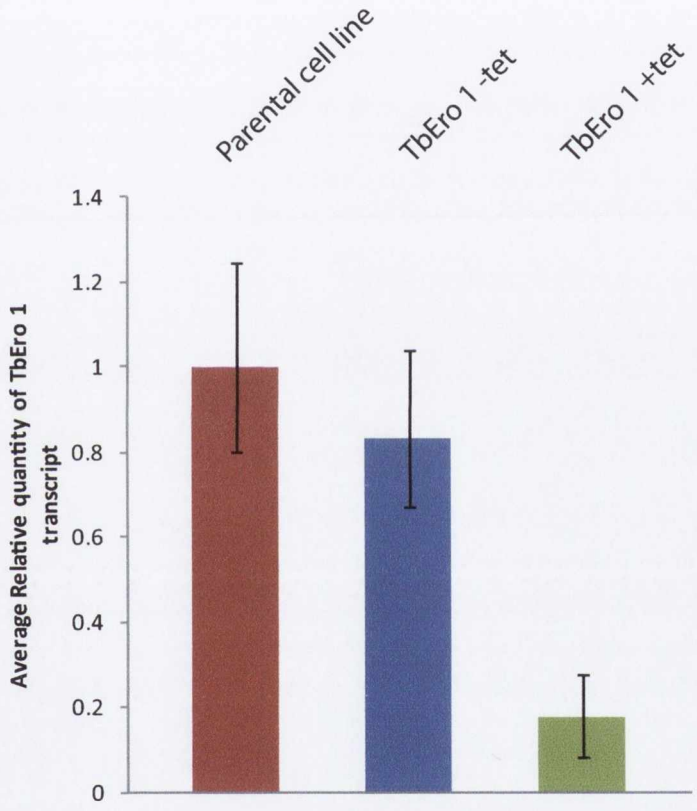


Figure 5.5 Analysis of cell cycle of bloodstream form TbEro 1 RNAi cell line

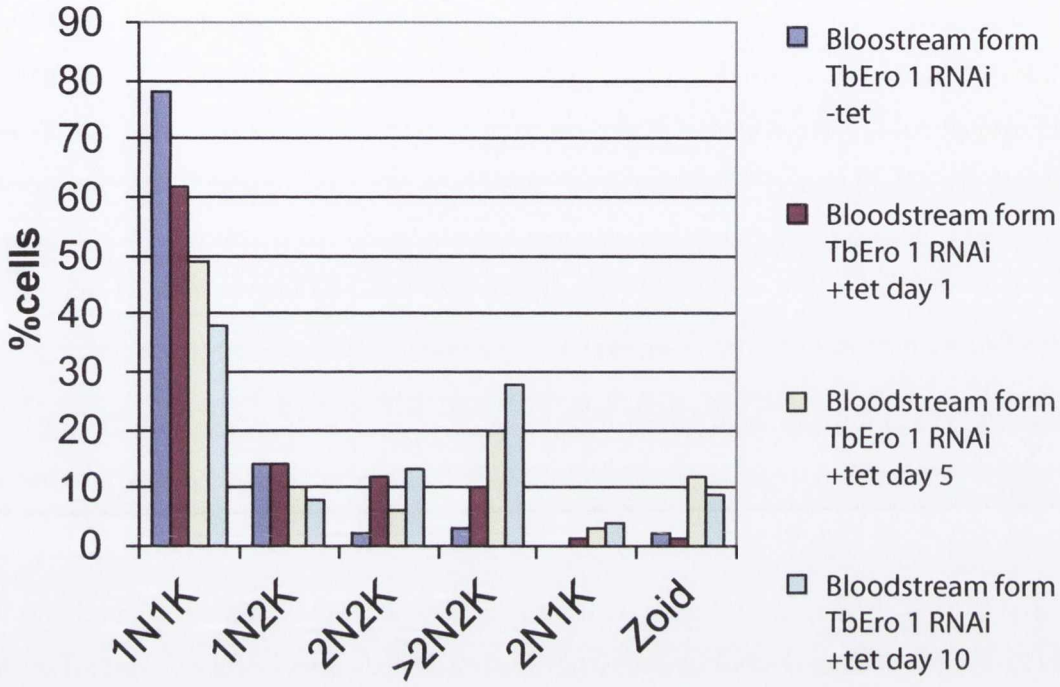
Bloodstream form TbEro 1 RNAi cell line was grown in HMI9, noninduced and induced over a period of 10 days. Cells were fixed at the time points indicated, stained using DAPI and viewed using fluorescence microscopy. The number of nuclei (N) and kinetoplast (K) per cell were observed for ~150 cells of each cell line. Noninduced is indicated by –tet and induced is indicated by +tet.

Panel A. The % of the total number of cells observed in various stages of the cell cycle.

Panel B. Examples of the various stages of the cell cycle observed during fluorescence microscopy.

Figure 5.5 Analysis of cell cycle of bloodstream form TbEro 1 RNAi cell line

A.



B.

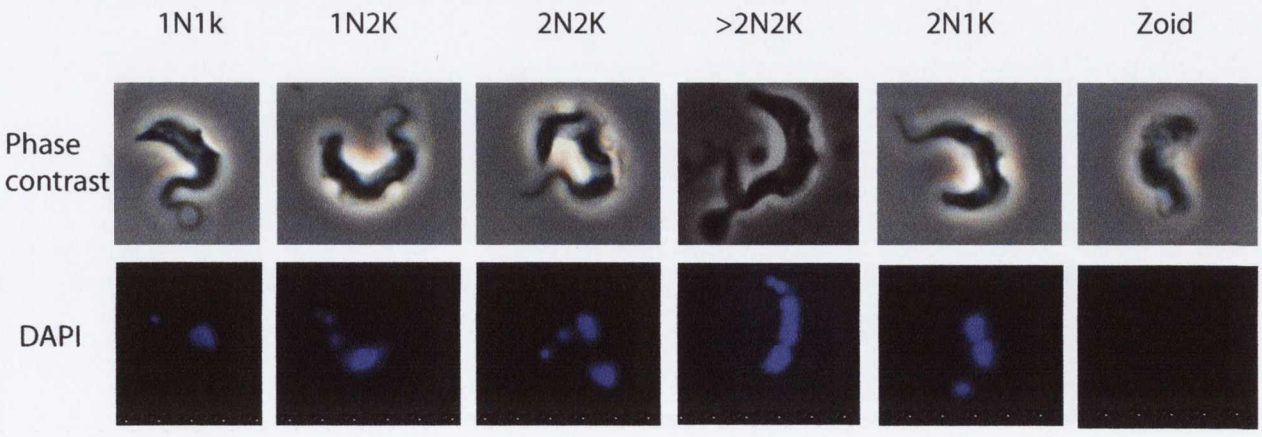


Figure 5.6 Analysis of the rate of surface protein production of bloodstream form TbEro 1 RNAi versus parental cell line

Bloodstream form cells were subjected to pulse labelling as described in the Methods. Briefly, after a 10 minute pulse of [^{35}S] methionine/cysteines medium the cells were diluted six fold in HMI9 medium and samples of cells were removed at various times (0, 10, 20, 30, 50 and 90 minutes). Cell fractions were washed at 4°C and their surface proteins were biotinylated. The cells were then lysed and the biotinylated proteins were precipitated using streptavidin agarose. The labelled proteins were analyzed by SDS PAGE autoradiography. Each lane represented 5x10^{6} cells.

Panel A: SDS PAGE autoradiography of fractions prepared from induced TbEro 1 RNAi cell line. The total cell lysate is shown followed by biotinylated fractions (surface proteins) of the cell line.

Panel B: SDS PAGE autoradiography of fractions prepared from parental cell line. The total cell lysate is shown followed by biotinylated fractions (surface proteins) of the cell line.

Figure 5.6 Analysis of the rate of surface protein production of bloodstream form TbEro 1 RNAi versus parental cell line

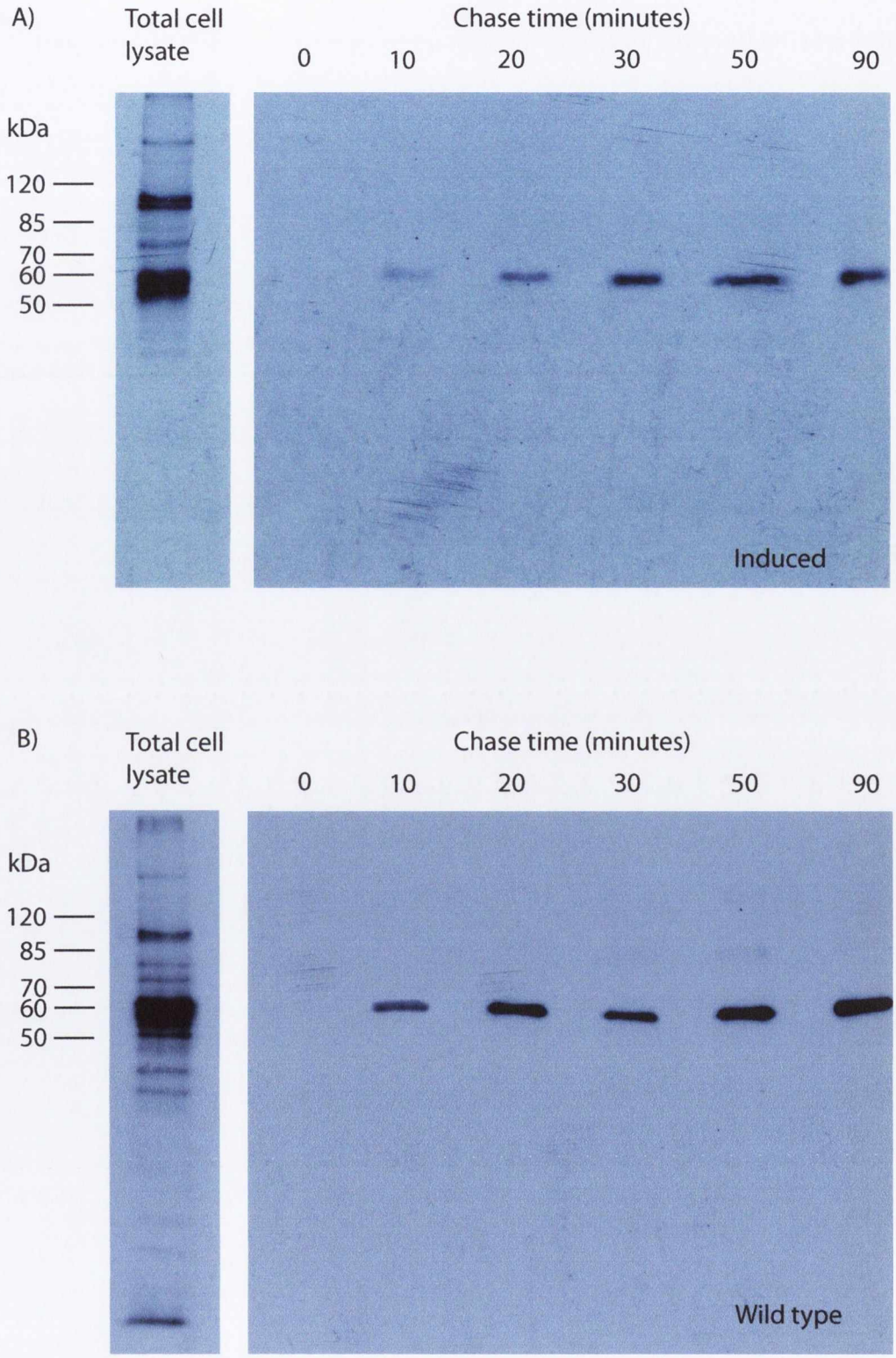


Figure 5.7 Analysis of amplification products of various cloning steps of p2620 TbEro 1 expression construct

The TbEro 1 ORF was amplified by PCR with primers which included *BamHI* and *HindIII* (expected size 1339 bp). This PCR was subcloned into the p2620 expression vector and the product was digested with *BamHI* and *HindIII* for confirmation (expected size 6044 and 1332 bp).

Figure 5.7 Analysis of amplification products of various cloning steps of p2620 TbEro 1 expression construct

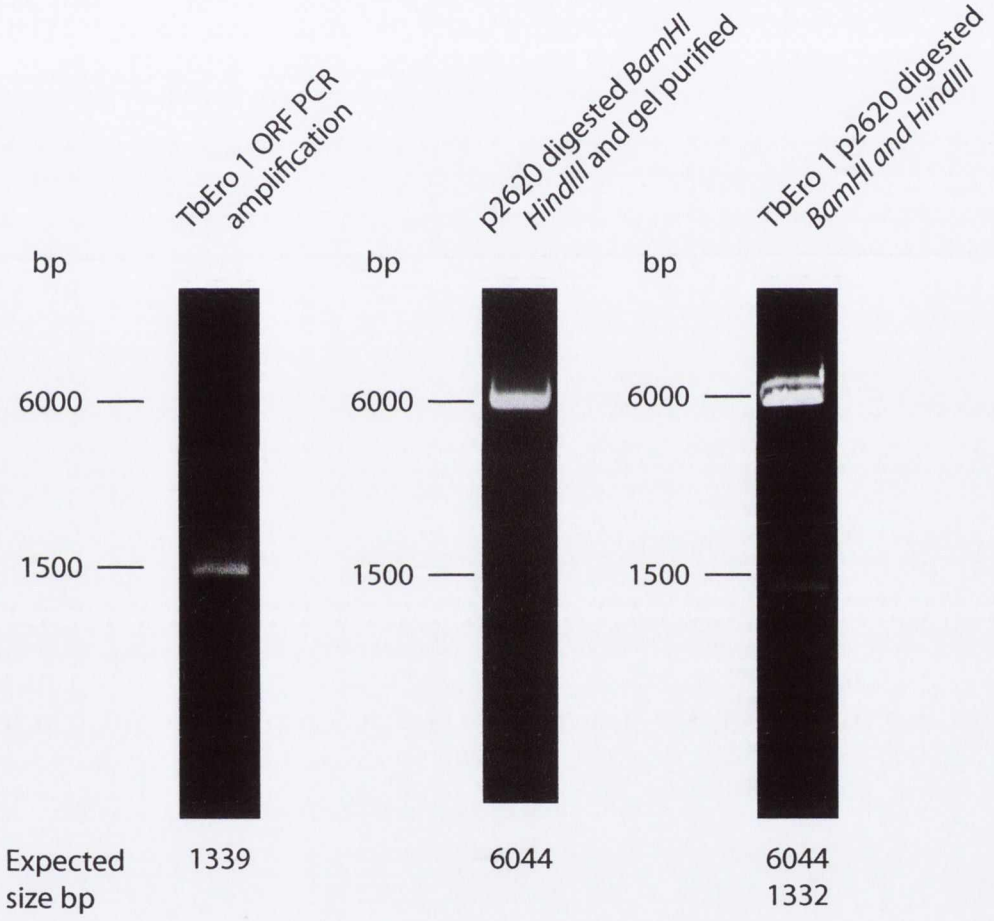
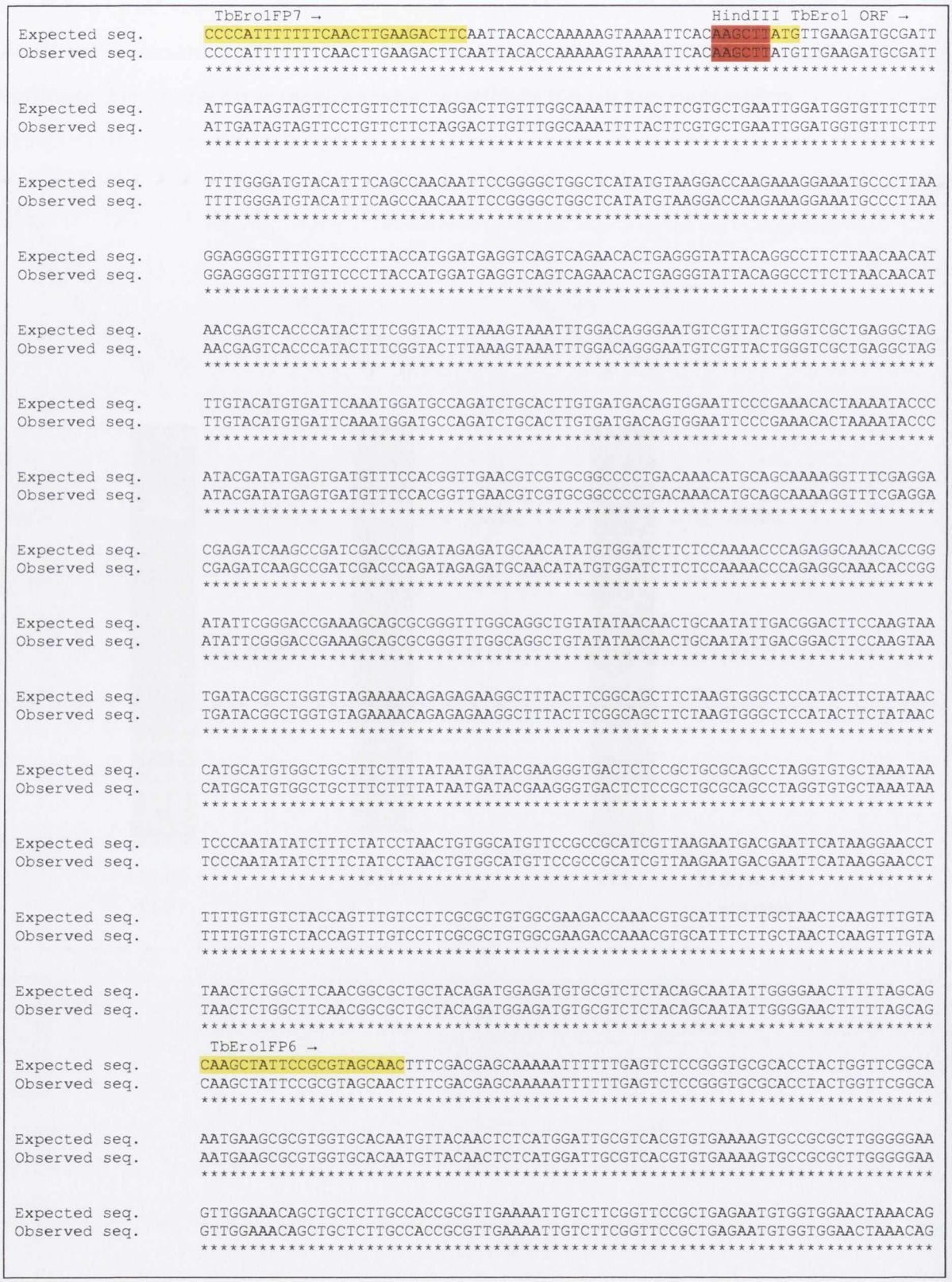
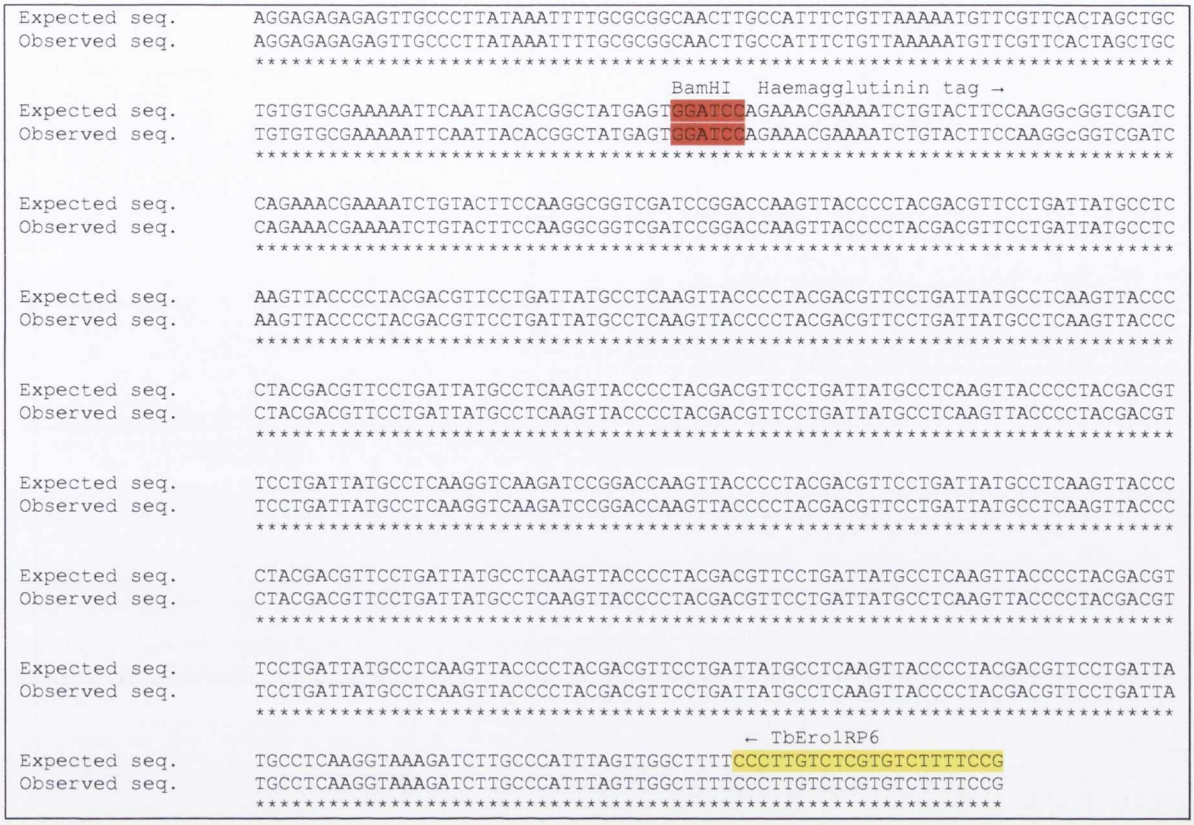


Figure 5.8 Analysis of TbEro 1 p2620 construct sequence

A.




Expected seq. AGGAGAGAGAGTTGCCCTTATAAATTTTGCGCGGCAACTTGCCATTTCTGTTAAAAATGTTCGTTCACTAGCTGC
Observed seq. AGGAGAGAGAGTTGCCCTTATAAATTTTGCGCGGCAACTTGCCATTTCTGTTAAAAATGTTCGTTCACTAGCTGC
BamHI Haemagglutinin tag →
Expected seq. TGTGTGCGAAAAATTCAATTACACGGCTATGAGTGGATCCAGAAACGAAAATCTGTACTTCCAAGGcGGTCGATC
Observed seq. TGTGTGCGAAAAATTCAATTACACGGCTATGAGTGGATCCAGAAACGAAAATCTGTACTTCCAAGGcGGTCGATC
Expected seq. CAGAAACGAAAATCTGTACTTCCAAGGCGGTCGATCCGGACCAAGTTACCCCTACGACGTTCCTGATTATGCCTC
Observed seq. CAGAAACGAAAATCTGTACTTCCAAGGCGGTCGATCCGGACCAAGTTACCCCTACGACGTTCCTGATTATGCCTC
Expected seq. AAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCC
Observed seq. AAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCC
Expected seq. CTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGT
Observed seq. CTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGT
Expected seq. TCCTGATTATGCCTCAAGGTCAAGATCCGGACCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCC
Observed seq. TCCTGATTATGCCTCAAGGTCAAGATCCGGACCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCC
Expected seq. CTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGT
Observed seq. CTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGT
Expected seq. TCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTA
Observed seq. TCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTATGCCTCAAGTTACCCCTACGACGTTCCTGATTA
← TbEro1RP6
Expected seq. TGCCTCAAGGTAAAGATCTTGCCCATTTAGTTGGCTTTTCCCTTGTCTCGTGTCTTTTCCG
Observed seq. TGCCTCAAGGTAAAGATCTTGCCCATTTAGTTGGCTTTTCCCTTGTCTCGTGTCTTTTCCG

B.

```
Expected seq.    MLKMRLLIVVPVLLGLVWQILLRAELDGVSFFGMYISANNSGAGSYVRTKKGNALKEGFCSLTMDEVSQNTEGIT
Observed seq.    MLKMRLLIVVPVLLGLVWQILLRAELDGVSFFGMYISANNSGAGSYVRTKKGNALKEGFCSLTMDEVSQNTEGIT
                 **********************************************************************************

Expected seq.    GLLNNITSHPYFRYFKVNLDRECRYWVAEASCTCDSNGCQICTCDDSGIPETLKYPYDMSDVSTVERRAAPDKHA
Observed seq.    GLLNNITSHPYFRYFKVNLDRECRYWVAEASCTCDSNGCQICTCDDSGIPETLKYPYDMSDVSTVERRAAPDKHA
                 **********************************************************************************

Expected seq.    AKGFEDEIKPIDPDRDATYVDLLQNPEANTGYSGPKAARVWQAVYNNCNIDGLPSNDTAGVENREKALLRQLLSG
Observed seq.    AKGFEDEIKPIDPDRDATYVDLLQNPEANTGYSGPKAARVWQAVYNNCNIDGLPSNDTAGVENREKALLRQLLSG
                 **********************************************************************************

Expected seq.    LHTSITMHVAAFFYNDTKGDSPLRSLGVLNNPNISFYPNCGMFRRIVKNDEFIRNLFVVYQFVLRAVAKTKRAFL
Observed seq.    LHTSITMHVAAFFYNDTKGDSPLRSLGVLNNPNISFYPNCGMFRRIVKNDEFIRNLFVVYQFVLRAVAKTKRAFL
                 **********************************************************************************

Expected seq.    ANSSLYNSGFNGAATDGDVRLYSNIGELFSSKLFRVATFDEQKFFESPGAHLLVRQMKRVVHNVTTLMDCVTCEK
Observed seq.    ANSSLYNSGFNGAATDGDVRLYSNIGELFSSKLFRVATFDEQKFFESPGAHLLVRQMKRVVHNVTTLMDCVTCEK
                 **********************************************************************************

Expected seq.    CRAWGKLETAALATALKIVFGSAENVVELNRGERVALINFARQLAISVKNVRSLAAVCEKFNYTAMSGSRNENLY
Observed seq.    CRAWGKLETAALATALKIVFGSAENVVELNRGERVALINFARQLAISVKNVRSLAAVCEKFNYTAMSGSRNENLY
                 **********************************************************************************

Expected seq.    FQGGRSRNENLYFQGGRSGPSYPYDVPDYASSYPYDVPDYASSYYDVPDYASSYPYDVPDYASSYPYDVPDYASS
Observed seq.    FQGGRSRNENLYFQGGRSGPSYPYDVPDYASSYPYDVPDYASSYYDVPDYASSYPYDVPDYASSYPYDVPDYASS
                 **********************************************************************************

Expected seq.    YPYDVPDYASRSRSGPSYPYDVPDYASSYPYDVPDYASSYPYDVPDYASSYPYDVPDYASSYPYDVPDYASSYPY
Observed seq.    YPYDVPDYASRSRSGPSYPYDVPDYASSYPYDVPDYASSYPYDVPDYASSYPYDVPDYASSYPYDVPDYASSYPY
                 **********************************************************************************

Expected seq.    DVPDYASRstop
Observed seq.    DVPDYASRstop
                 ********
```

Figure 5.8 Analysis of TbEro 1 p2620 construct sequence

The TbEro 1 haemagglutinin construct was sequenced to ensure the open reading frame was correctly ligated into the p2620 expression vector.

Panel A: The expected sequence and observed sequence obtained were compared. The location where primers bound to the sequence are shown in yellow and the primer name and direction are indicated above. Restriction sites are shown in red and the restriction enzymes at which these sites cut are indicated above. Agreement is indicated by a * below the sequence and disagreements are highlighted in light blue. The TbEro1 open reading frame and the haemagglutinin sequence are indicated above with the direction of the transcription indicated with an arrow (→).

Panel B: A translation of the expected sequence and observed sequence obtained written in single letter amino acid abbreviation. Agreement is indicated by a * below the sequence. The HA tag is highlighted yellow.

Figure 5.9 Analysis of growth rate of bloodstream form epitope tagged TbEro 1 cell line

Bloodstream form TbEro 1 p2620 cell line was grown in HMI9 and induced; as described in the Methods over a period of 6 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 7 hours for noninduced TbEro 1 p2620 cell line and 7 hours for induced TbEro 1 p2620 cell line. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.9 Analysis of growth rate of bloodstream form epitope tagged TbEro 1 cell line

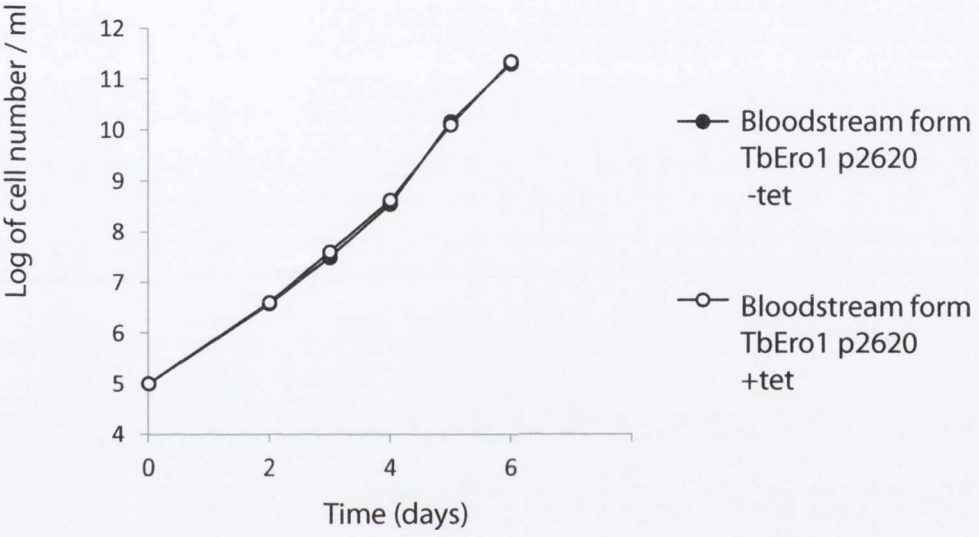


Figure 5.10 Analysis of epitope tagged TbEro 1 expression in bloodstream form cells using Western blot

Panel A: Western blot of whole cell extracts from bloodstream form TbEro 1 p2620 cell line noninduced and induced. Each lane represents $5x10^6$ cells. The membrane was first stained with ponceau S red reversible stain then washed and probed with anti-HA antibody as described in the methods.

Panel B: Western blot of varying concentrations of whole cell extracts from induced bloodstream form TbEro 1 p2620 cell line probed with anti-HA antibody to determine antibody sensativity.

Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.10 Analysis of epitope tagged TbEro 1 expression in bloodstream form cells using Western blot

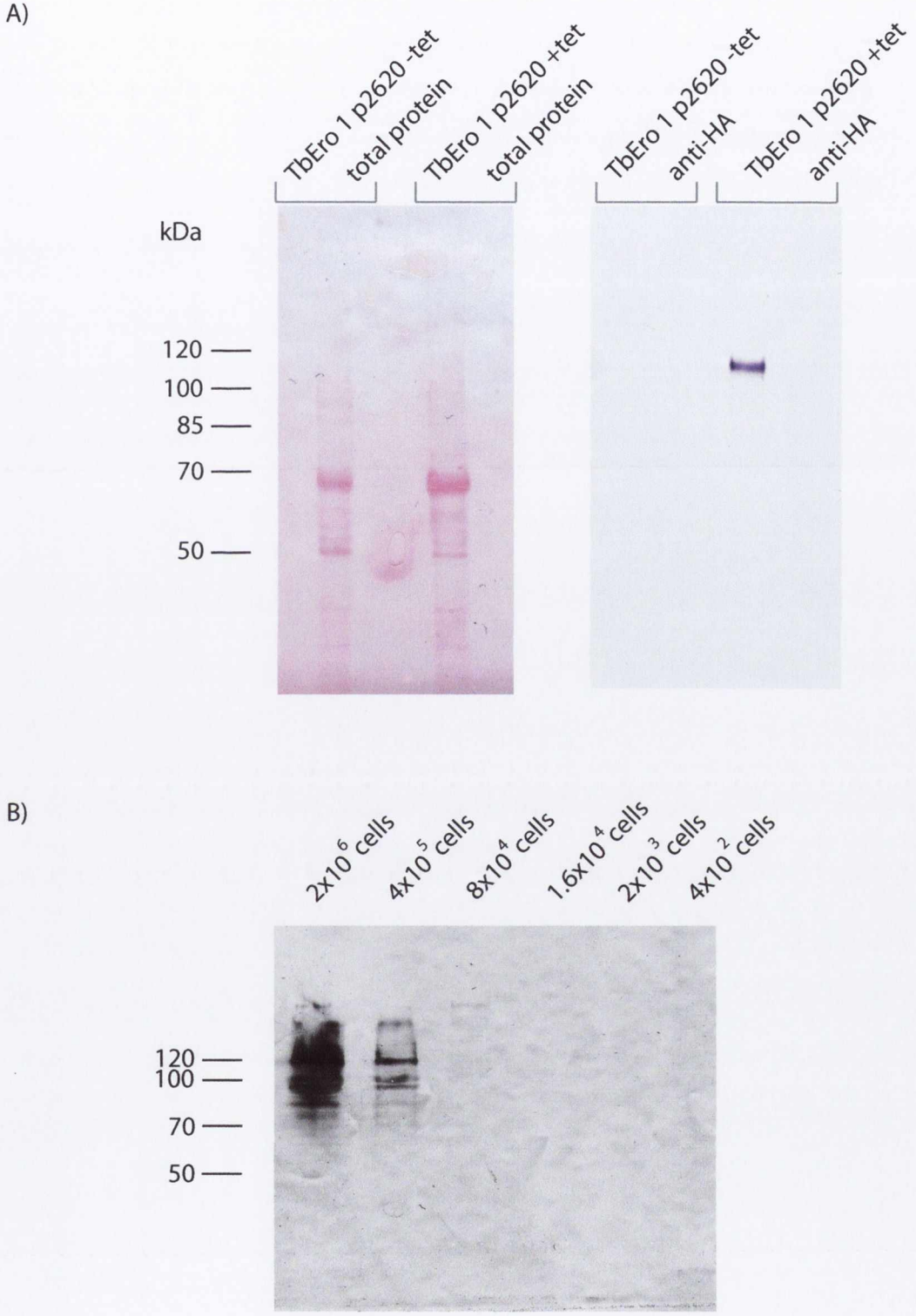


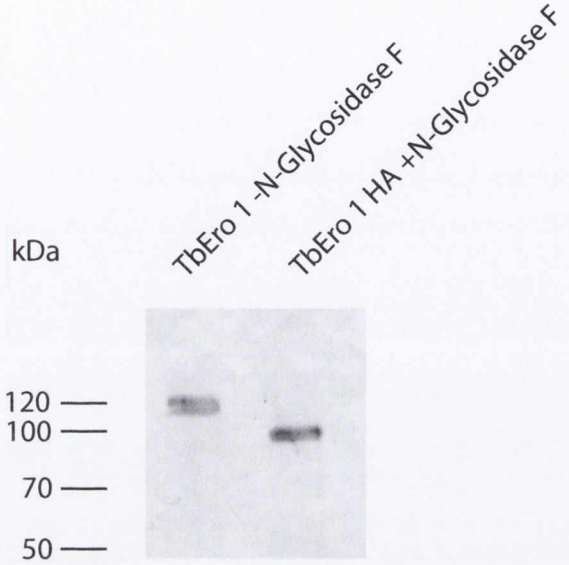
Figure 5.11 Analysis of epitope tagged TbEro 1 glycosylation using glycosidases and Western blot

Panel A: Western blot of epitope tagged TbEro 1 digested with *N*-Glycosidase F and immunoprecipitated probed with anti-HA antibody as described in the Methods. Each well represents the equivalent of $5x10^6$ cells.

Panel B: Western blot of epitope tagged TbEro 1 digested with Endo H and immunoprecipitated probed with anti-HA antibody as described in the Methods. Each well represents the equivalent of $5x10^6$ cells.

Figure 5.11 Analysis of epitope tagged TbEro 1 glycosylation using glycosidases and Western blot

A.



B.

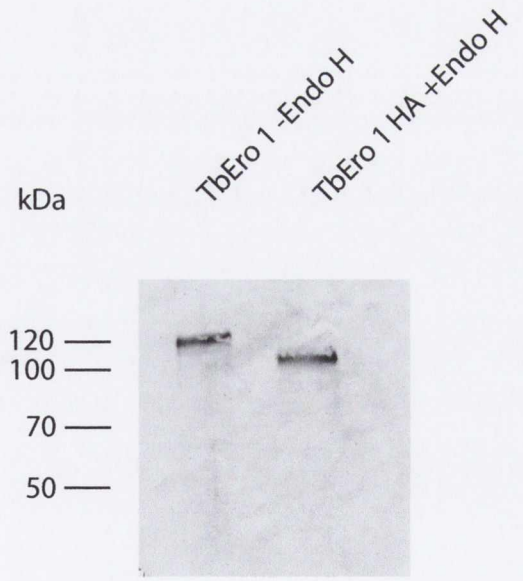


Figure 5.12 Analysis of TbEro 1 solubility using Western blot

Cells expressing epitope tagged TbEro 1 were mechanically lysed using sonication. The lysed cells were then divided into two batches, one of which contained 1% CHAPS to allow the release of any membrane bound proteins. The samples were centrifuged and the supernatant and pellet fractions were separated. A Western blot of the fractions was run and probed with anti-HA antibodies (epitope attached to TbEro 1) as described in the Methods. Pellet fraction, supernatant fraction and presence or absence of CHAPS is indicated. Each lane represents the equivalent of $5x10^6$ cells per well.

Figure 5.12 Analysis of TbEro 1 HA solubility using Western blot

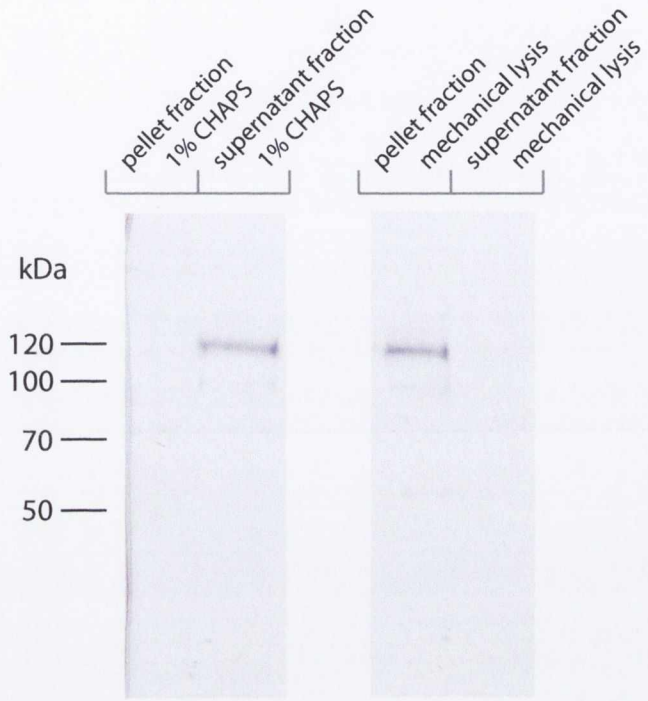


Figure 5.13 Analysis of TbPDIs co-immunoprecipetated with epitope tagged form of TbEro 1 using Western blot

Western blot of total cell lysates of bloodstream form TbEro 1 p2620 cell line induced versus immunoprecipetated TbEro 1 HA fractions; probed with various antibodies.

Figure 5.13 Analysis of TbPDIs co-immunoprecipetated with epitope tagged form of TbEro 1 using Western blot

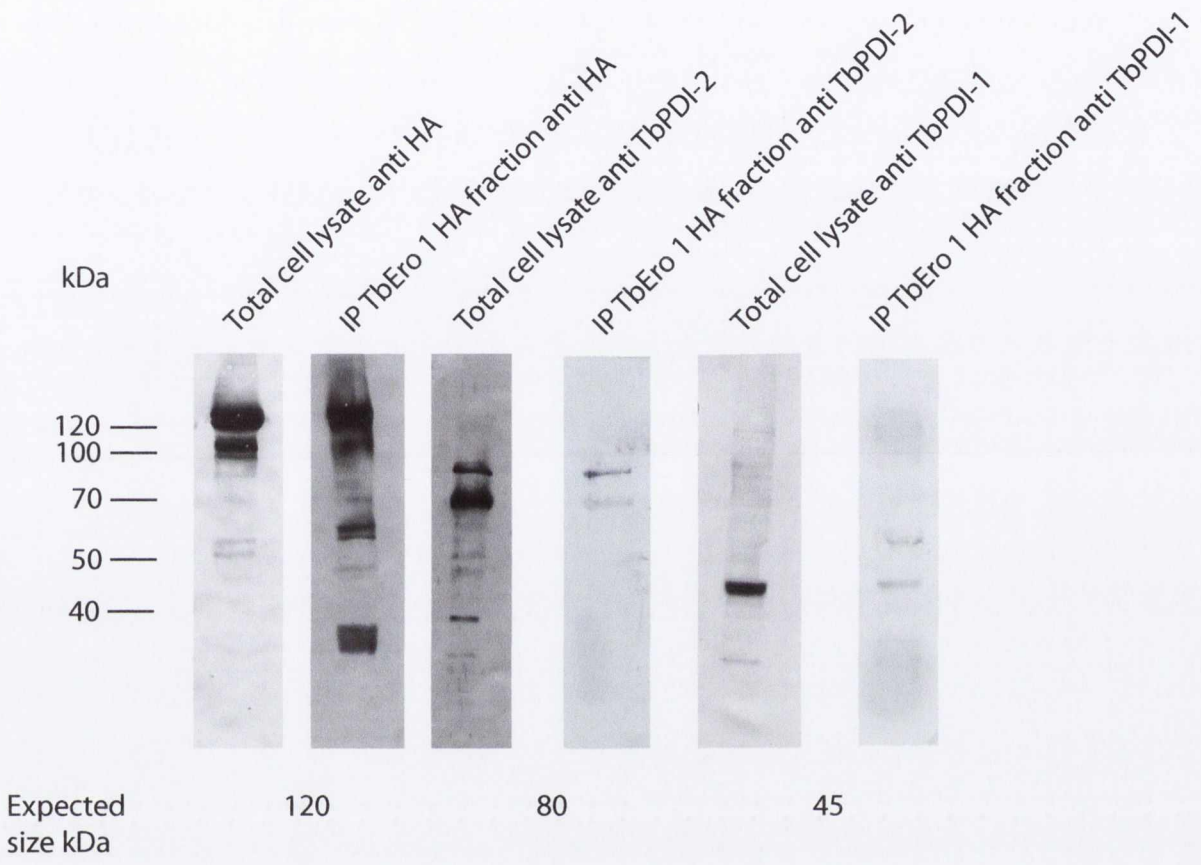


Figure 5.14 Analysis of subcellular fractions of bloodstream form cells using digitonin-based subcellular fractionation and Western blot

Bloodstream form cells expressing epitope tagged TbEro1 were induced for 3 days and then harvested. The cells were incubated for 4 minutes at increasing concentrations of digitonin and were then centrifuged to yield a pellet (P) and supernatant fraction (S), as described in the Methods. Several Western blots of the samples were run and probed with various antibodies which are indicated.

Note: Lanes 9 and 10 of the anti-ectophosphatase blot were swapped. Lane 9: Supernatant and lane 10: Pellet fraction.

Figure 5.14 Analysis of subcellular fractions of bloodstream form cells using digitonin-based subcellular fractionation and Western blot

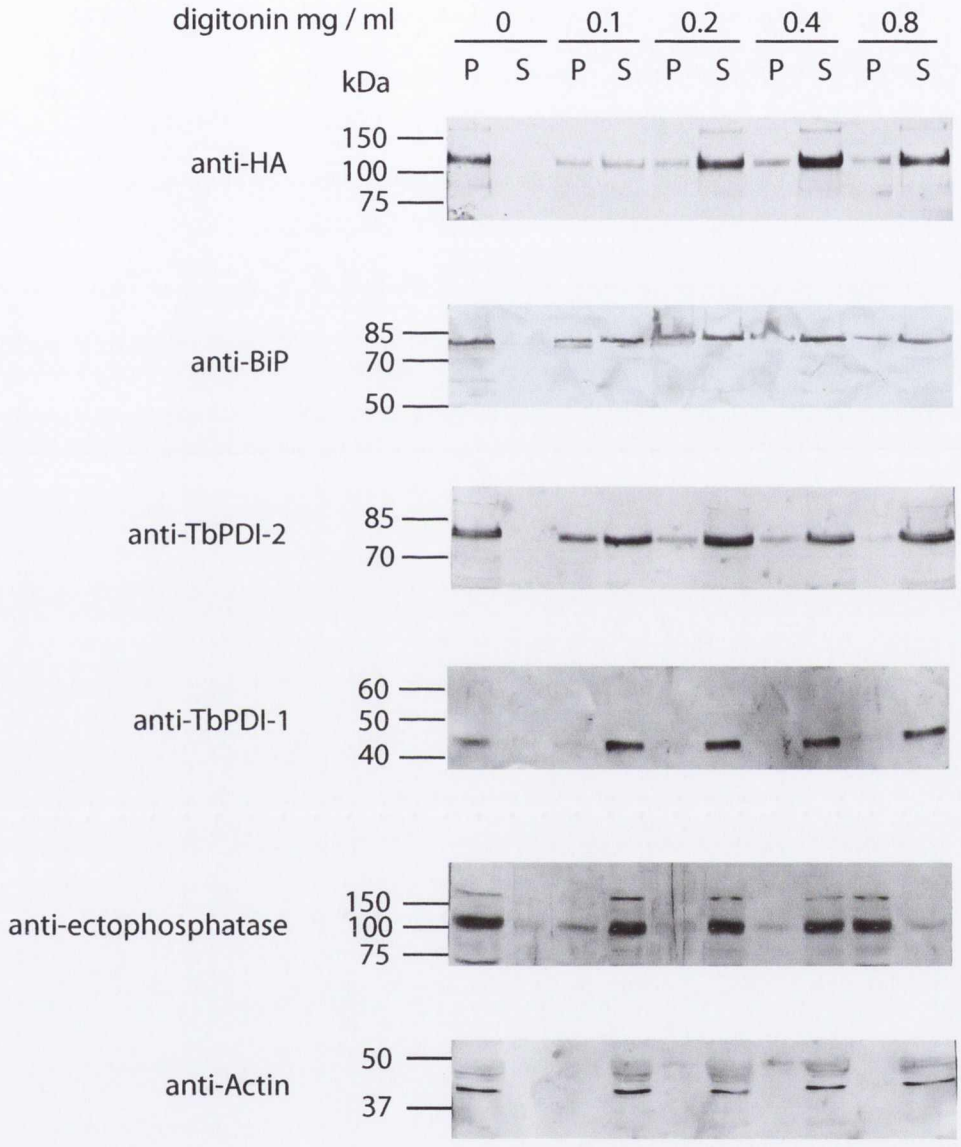


Figure 5.15 Analysis of TbEro 1 location in bloodstream form cells using immunnofluorescence microscopy

Bloodstream form cells, expressing epitope tagged TbEro 1, were induced and culture for 5 days. The cells were prepared, probed with anti-HA antibodies, which binds to the chosen epitope and visualized as described in the Methods. Alexa fluor dye 488 (green) secondary antibody revealed the location of TbEro 1 HA and DAPI stain (blue) indicated the location of nucleus and kinetoplast DNA. Both induced and noninduced cells (do not express epitope tagged TbEro 1) are shown.

Figure 5.15 Analysis of TbEro 1 location in bloodstream form cells using immuno fluorescence microscopy

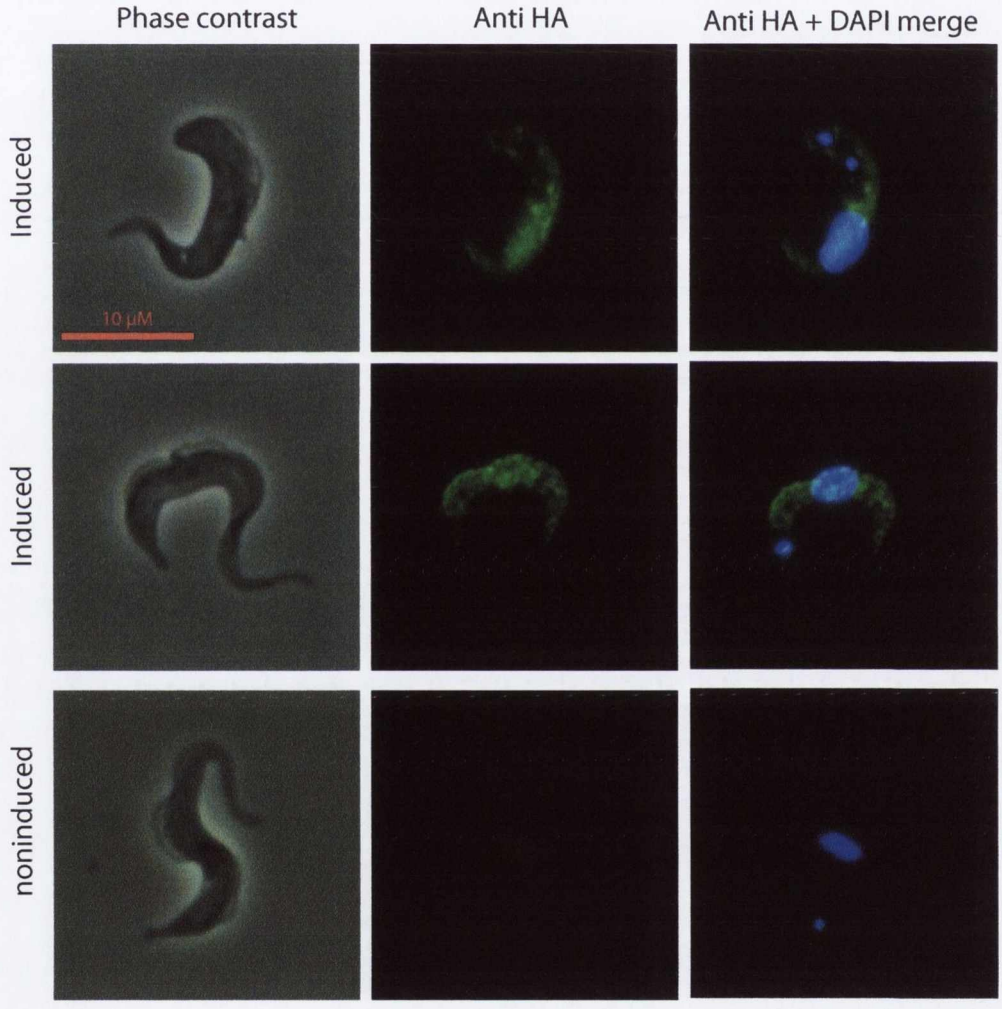
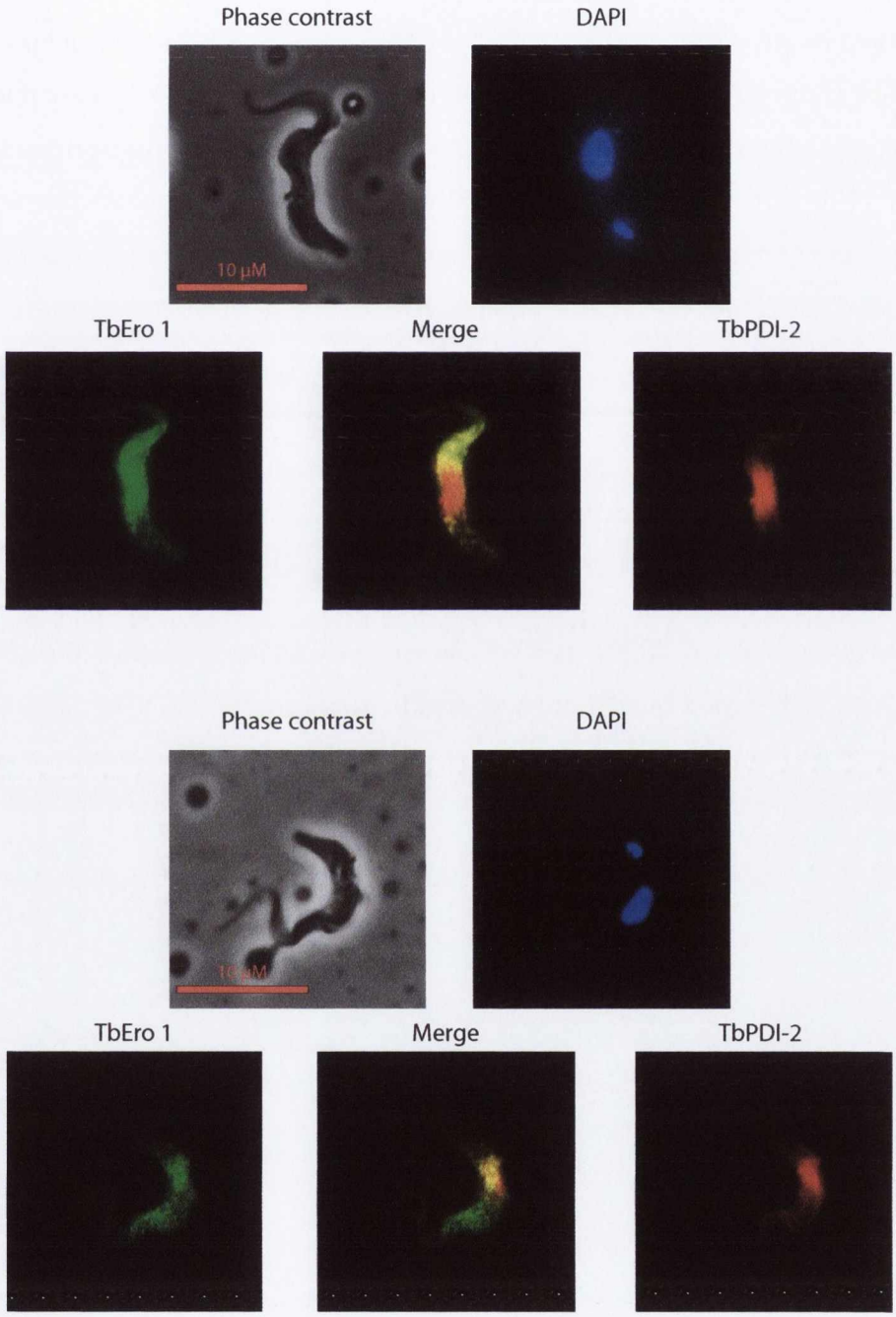


Figure 5.16 Analysis of TbEro 1 and TbPDI-2 co-localization in bloodstream form cells using immunofluorescence microscopy

Bloodstream form cells expressing epitope tagged TbEro 1 were induced and culture for 5 days. The cells were prepared, probed with anti-HA, which interacts with the chosen epitope and TbPDI-2 antibodies and visualized as described in the Methods. Alexa fluor dye 488 (green) secondary antibody revealed the location of TbEro 1 HA, alexa fluor dye 568 (red) secondary antibody revealed the location of TbPDI-2 and DAPI stain (blue) indicated the location of kinetoplast and genomic DNA.

Figure 5.16 Analysis of TbEro 1 and TbPDI-2 co-localization in bloodstream form cells using immunofluorescence microscopy



5.3 Generation and analysis of procyclic form (TbEro 1 RNAi) cell line

The knockdown experiments of TbPDI-2 indicated that procyclic form cells showed a slightly different phenotype to bloodstream form cells. To investigate if this was true for a TbEro 1 knockdown in procyclic form cells the TbEro 1 RNAi vector was transfected into procyclic form cells. The positive clones were selected using phleomycin (5μg/ml), cells were grown, divided into two flasks and one of the flasks was induced with tetracycline (1 μg/ml). The cell growth was monitored over a period of 17 days (Figure 5.17) and showed only a slight growth defect. Doubling times were calculated and found to be 16 hours for the parental cell line and 19 hours for the induced cell line, an increase of only 16% doubling time. This result is significantly less dramatic than the slow growth phenotype observed in the bloodstream TbEro 1 knockdown cell line. The transcript level was assessed using qRT-PCR (Figure 5.18) which showed the transcript level of *TbEro 1* was knocked down to 18% remaining which was equivalent to the knockdown observed in the bloodstream form cells.

As the knockdown of TbPDI-2 in procyclic forms showed little effect on the cell cycle it was expected that the same would be found in the procyclic form (TbEro 1 RNAi) cells. A cell cycle analysis was carried out (Figure 5.19) and revealed there was no observable change in the cell cycle.

5.4 Generation and analysis of a procyclic form cell line that expresses an epitope tagged TbEro 1

The procyclic form of *T. brucei* shows a different morphology to bloodstream form cells and expresses TbPDI-2 and TbEro 1 to different extents. As such, it could not be assumed TbEro 1 behaved in an identical manner in procyclic form cells. To address the characterization of TbEro 1 in procyclic form cells the TbEro 1 p2620 expression vector was employed. Procyclic form cells were transfected and selected with phleomycin (5 μg/ml). The cells were grown then divided into two flasks and one flask was induced using tetracycline (1 μg/ml). The growth of the cells was monitored over a period of 17 days to ensure that the expression of an epitope tagged TbEro 1 did not inhibit growth (Figure 5.20). The protein expression of TbEro 1 HA was analyzed using Western blot (Figure 5.21) and, as was found in bloodstream form cells, a ~120 kDa protein was detected in the induced procyclic form cells. This increase in observed mass over the expected mass of TbEro 1 was thought to be caused by heavy glycosylation of the protein. To test this theory epitope tagged TbEro 1 was subjected to digestion using *N*-glycosidase F and endoglycosidase H as previously undertaken with the bloodstream form TbEro 1 HA (Figure 5.22). The results of these digestions were identical to those found with the bloodstream form epitope tagged TbEro 1 suggesting TbEro 1 is heavily *N*-glycosylated and these glycosylations are not complex in structure.

To establish what subcellular organelle TbEro 1 was found in, in procyclic form cells, a subcellular fractionation experiment was carried out (Figure 5.23). The results indicated TbEro 1 shared the same profile as BiP a resident of the endoplasmic reticulum indicating TbEro 1 was a resident of the endoplasmic reticulum, as was found to be the case with bloodstream form TbEro 1.

To confirm the findings of the subcellular fractionation experiment, immunnofluorescence was carried out on epitope tagged TbEro 1 expressing cells using anti-HA antibodies, which bound to the HA epitope tag (Figure 5.24). As was found in the bloodstream form cells TbEro 1 showed a wide distribution throughout the cell in a pattern similar to BiP, indicating TbEro 1 was within the endoplasmic reticulum. As a negative control, it was shown that the noninduced cells did not interact at all with the anti-HA antibody.

Figure 5.17 Analysis of growth rate of procyclic form TbEro 1 RNAi cell line

Procyclic form TbEro 1 RNAi cell line was grown in SDM79 and induced; as described in the Methods over a period of 10 days. The data was presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the methods and were found to be 16 hours for parental cell line and 19 hours for induced TbEro 1 RNAi. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.17 Analysis of growth rate of procyclic form TbEro 1 RNAi cell line

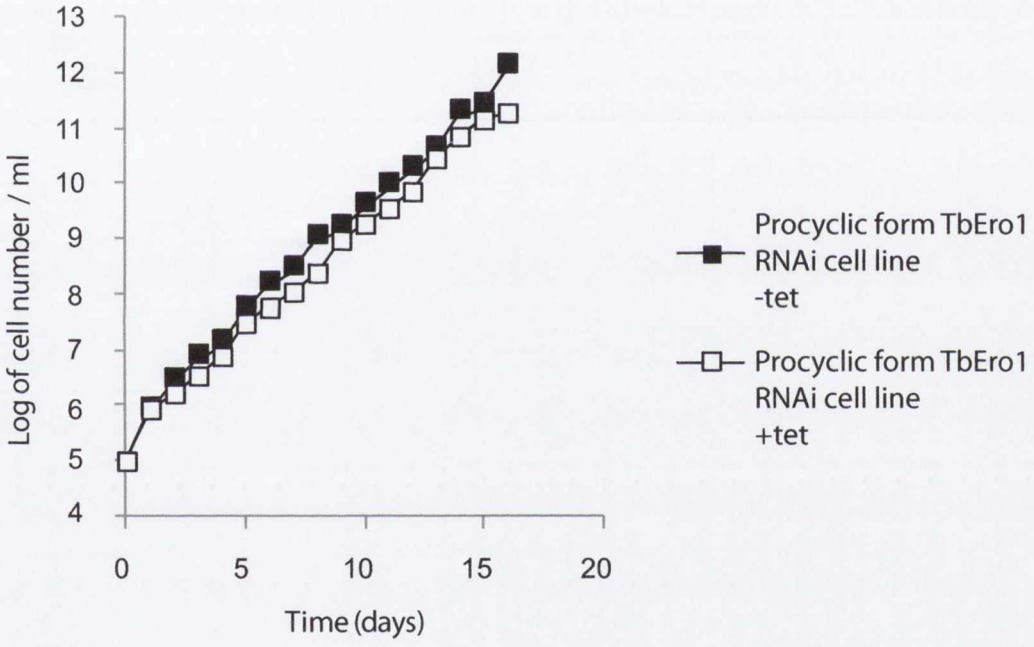


Figure 5.18 Analysis of TbEro 1 RNA transcript of procyclic form TbEro 1 RNAi cell line using qRT-PCR

The relative level of expression of TbEro 1 RNA was estimated by quantitative reverse transcriptase PCR (qRT-PCR). The analysis was conducted using total RNA fractions isolated using the Stratagene Absolutely RNA kit. The levels of TbEro 1 RNA were normalized against actin RNA and the relative quantifications were calculated by the ΔΔCt method as described in the Methods and are expressed as the mean of 3 replicates of the ratio of knockdown cell line to parental cell line. The average relative quantity of the TbEro 1 transcript in the noninduced TbEro 1 RNAi cell line was found to be 85% ± 3.5% and induced 18% ± 3.3%. In each case the results are expressed as the mean ± standard deviation of three replicates. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.18 Analysis of TbEro 1 RNA transcript of procyclic form TbEro 1 RNAi cell line using qRT-PCR

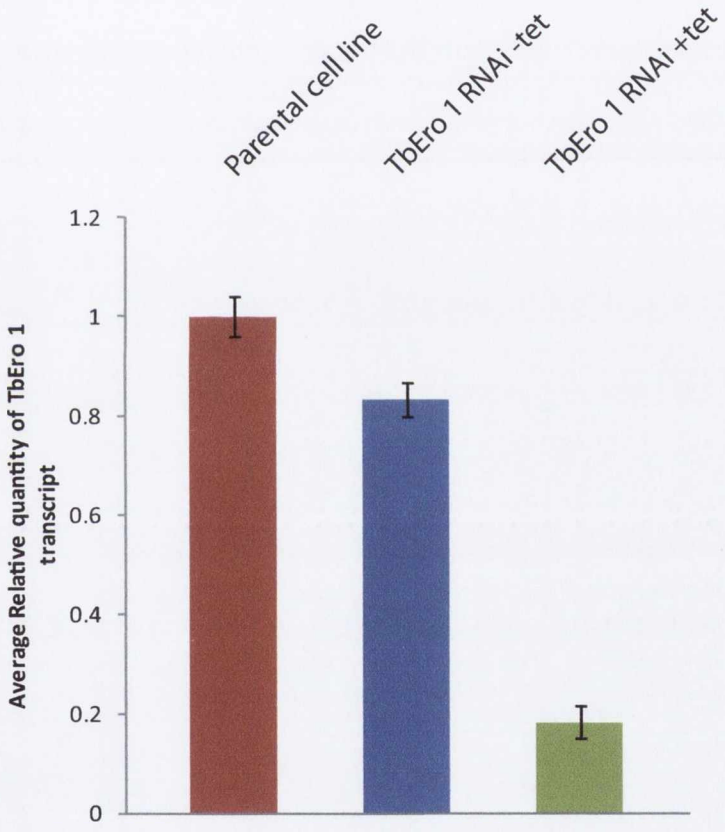


Figure 5.19 Analysis of cell cycle of procyclic form TbEro 1 RNAi cell line

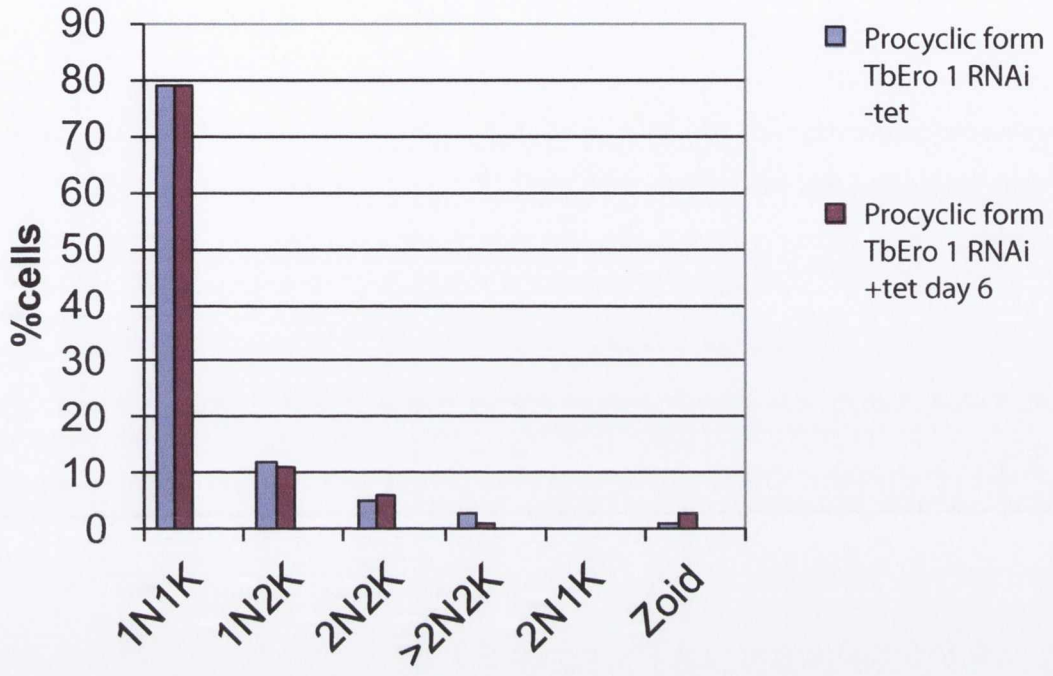
Procyclic form TbEro 1 RNAi cell line was grown in SDM79, noninduced and induced over a period of 6 days. Cells were fixed, stained using DAPI and viewed using fluorescence microscopy. The number of nuclei (N) and kinetoplast (K) per cell were observed for ~150 cells of each cell line. Noninduced is indicated by –tet and induced is indicated by +tet.

Panel A. The % of the total number of cells observed in various stages of the cell cycle.

Panel B. Examples of the various stages of the cell cycle observed during fluorescence microscopy.

Figure 5.19 Analysis of cell cycle of procyclic form TbEro 1 RNAi cell line

A.



B.

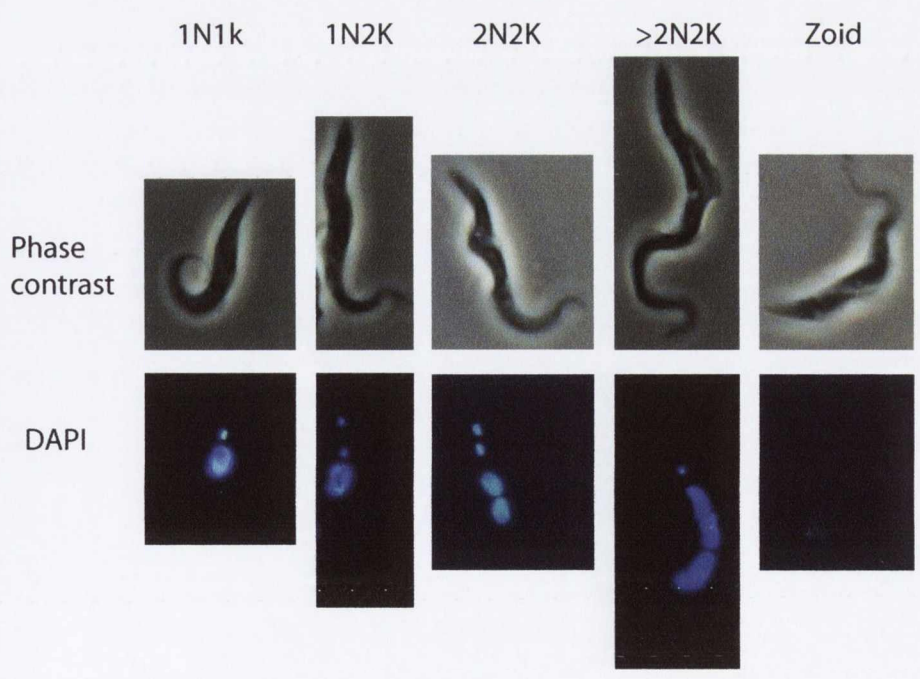


Figure 5.20 Analysis of growth rate of procyclic form epitope tagged TbEro 1 cell line 2

Procyclic form TbEro 1 p2620 cell line was grown in SDM79 and induced; as described in the Methods over a period of 17 days. The data is presented as the log of cell number/ml versus time (days). The doubling times were calculated as described in the Methods and were found to be 20 hours for noninduced TbEro 1 p2620 cell line and 20 hours for induced TbEro 1 p2620 cell line. Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.20 Analysis of growth rate of procyclic form epitope tagged TbEro 1 cell line 2

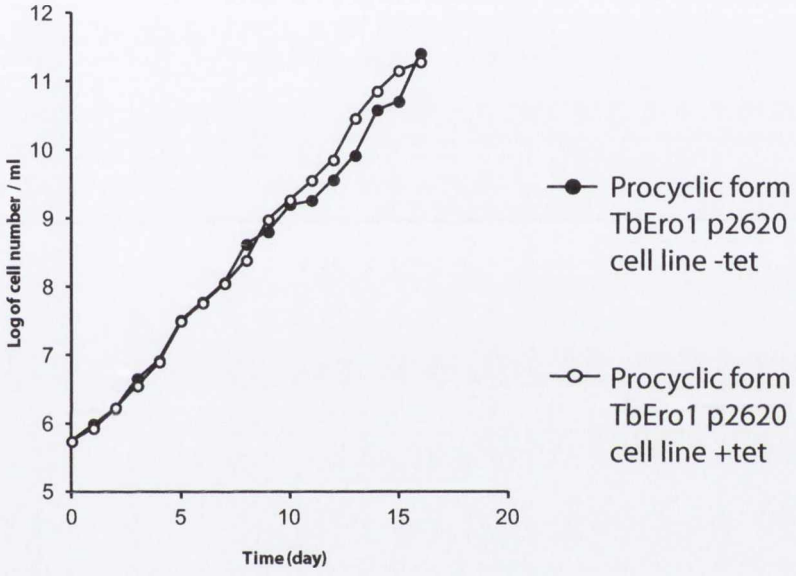


Figure 5.21 Analysis of epitope tagged TbEro 1 expression in procyclic form cells using Western blot

Panel A: Western blot of whole cell extracts from procyclic form TbEro 1 p2620 cell line noninduced and induced. Each lane represents $5x10^6$ cells. The membrane was first stained with ponceau S red reversible stain then washed and probed with anti-HA antibody as described in the methods.

Panel B: Western blot of varying concentrations of whole cell extracts from induced procyclic form TbEro 1 p2620 cell line probed with anti-HA antibody to determine antibody sensitivity.

Noninduced is indicated by –tet and induced is indicated by +tet.

Figure 5.21 Analysis of epitope tagged TbEro 1 expression in procyclic form cells using Western blot

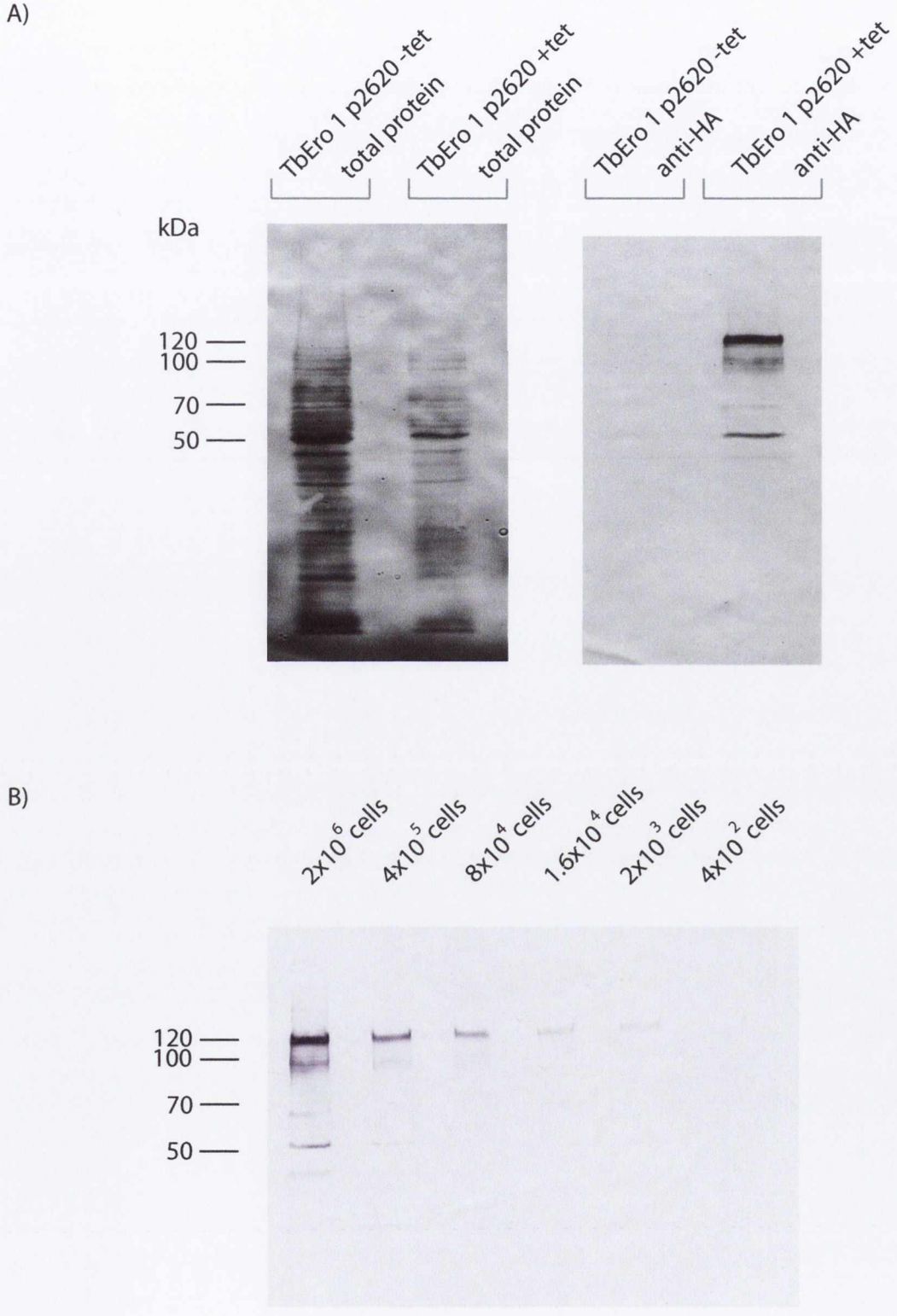


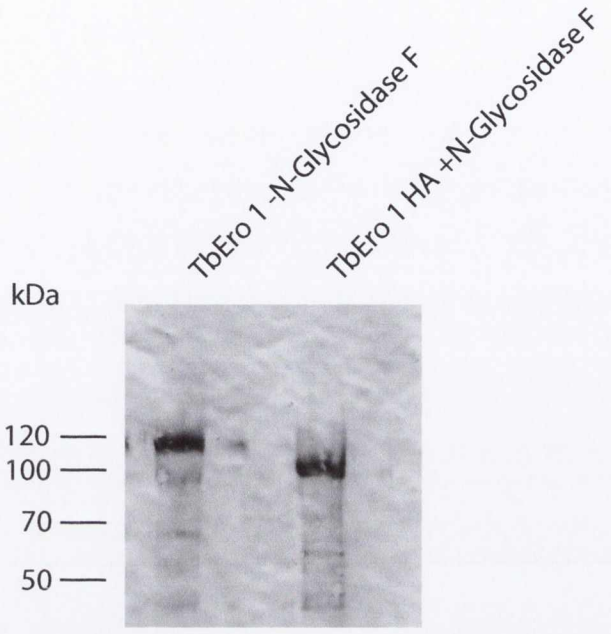
Figure 5.22 Analysis of procyclic form epitope tagged TbEro 1 glycosylation using glycosidases and Western blot

Panel A: Western blot of TbEro 1 HA digested with *N*-Glycosidase F and immunoprecipitated probed with anti-HA antibody as described in the Methods. Each well represents the equivalent of $5x10^6$ cells.

Panel B: Western blot of TbEro 1 HA digested with Endo H and immunoprecipitated probed with anti-HA antibody as described in the Methods. Each well represents the equivalent of $5x10^6$ cells.

Figure 5.22 Analysis of procyclic form epitope tagged TbEro 1 glycosylation using glycosidases and Western blot

A.



B.

Figure 5.23 Analysis of subcellular fractions of procyclic form cells using digitonin-based subcellular fractionation and Western blot

Procyclic form cells expressing epitope tagged TbEro1 were induced for 3 days and then harvested. The cells were incubated for 4 minutes at increasing concentrations of digitonin and were then centrifuged to yield a pellet (P) and supernatant fraction (S), as described in the Methods. Several Western blots of the samples were run and probed with various antibodies, which are indicated.

Figure 5.23 Analysis of subcellular fractions of procyclic form cells using digitonin-based subcellular fractionation and Western blot

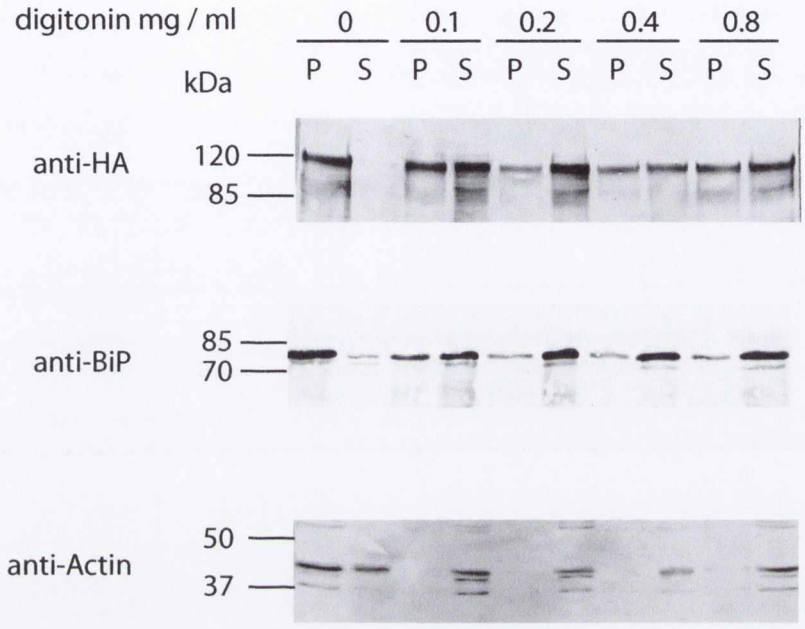
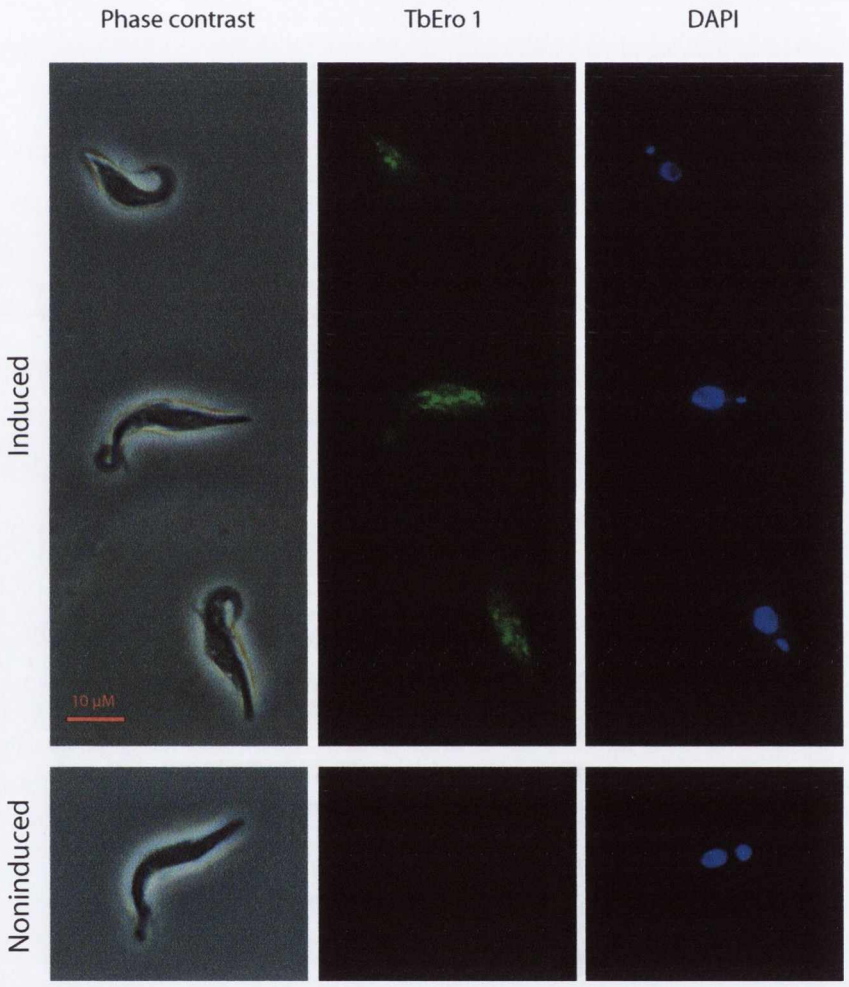


Figure 5.24 Analysis of TbEro 1 localization in procyclic form cells using immunofluorescence microscopy

Procyclic form cells, expressing epitope tagged TbEro 1, were induced and culture for 5 days. The cells were prepared, probed with anti-HA antibodies, which binds to the chosen epitope and visualized as described in the Methods. Alexa fluor dye 488 (green) secondary antibody revealed the location of TbEro 1 HA and DAPI stain (blue) indicated the location of nucleus and kinetoplast DNA. Both induced and noninduced cells (do not express epitope tagged TbEro 1) are shown.

Figure 5.24 Analysis of TbEro 1 localization in procyclic form cells using immuno fluorescencemicroscopy



5.5 Generation and purification of recombinant TbEro 1 GST fusion protein

In order to carry out *in vitro* assays of TbEro 1 a recombinant expression vector was designed and cloned to express GST tagged TbEro 1 in BL21 cells. The cloning and description of the expression vector are described in the Methods section and various steps of the cloning process were confirmed using agarose gel electrophoresis (Figure 5.25). The TbEro 1 pGST-Parallel 3 vector was sequenced to ensure the TbEro 1 ORF was in frame (Figure 5.26). The expression vector was transfected into BL21 cells and viable cells were grown, induced and the TbEro 1 GST protein was purified as using immunopurification as described in the Methods (Figure 5.27). The expected size of the epitope tagged TbEro 1 was 76 kDa and a protein of that size was expressed in the induced BL21 cells. The same protein was purified using anti-GST agarose to a concentration of ~150µg/ml in 5mls. Interestingly a peptide generated antibody originally intended for use with TbEro 1, which failed to react with trypanosome TbEro 1, readily bound to recombinant TbEro 1. The reason anti-TbEro 1 did not interact with trypanosome TbEro 1 is likely due to the heavy *N*-gylcosylation shielding the protein. In contrast, recombinant TbEro 1 was not glycosylated and so the antibody could be used to detect the protein on a Western blot. While the recombinant TbEro 1 was not used to carry out *in vitro* assays during this study, due to time constraints, it has been shown that the protein is readily purified from BL21 *E. coli* protein expression cell line in spite of the protein being a membrane associated protein.

Figure 5.25 Analysis of amplification products of various cloning steps of pGST-Parallel 3 expression construct

The TbEro 1 ORF was amplified by PCR with primers which included *BamHI* and *SpeI* (expected size 1363 bp). This PCR was subcloned into the pGST-Parallel 3 expression vector and the product was digested with *BamHI* and *SpeI* to confirm the ligation product (expected size 4982 and 1363 bp).

Figure 5.25 Analysis of amplification products of various cloning steps of TbEro 1 pGST-Parallel 3 expression construct

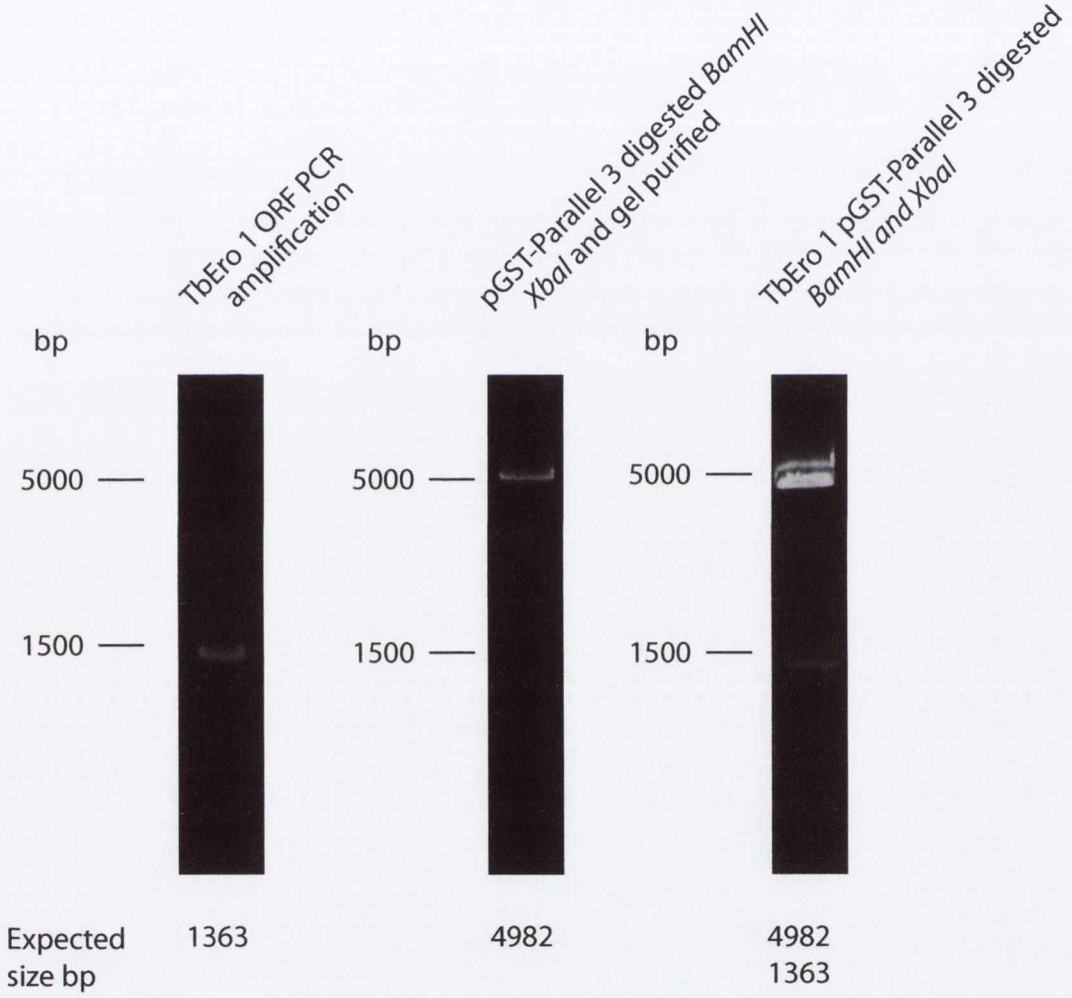


Figure 5.26 Analysis of TbEro 1 pGST-Parallel 3 expression construct sequence

A.

```
                  Ero1 p3 seq FP →        BamHI  FP4Ero1 →
Predicted seq.    CCTGTATTTTCAGGGCGCCATGGGGATCCATGTTGAAGATGCGATTATTGATAGTAGTTCCTGTTCTTCTAGGA
Resulting seq.    CCTGTATTTTCAGGGCGCCATGGGGATCCATGTTGAAGATGCGATTATTGATAGTAGTTCCTGTTCTTCTAGGA
                  ********************************************************************************

Predicted seq.    CTTGTTTGGCAAATTTTACTTCGTGCTGAATTGGATGGTGTTTCTTTTTTTGGGATGTACATTTCAGCCAACAAT
Resulting seq.    CTTGTTTGGCAAATTTTACTTCGTGCTGAATTGGATGGTGTTTCTTTTTTTGGGATGTACATTTCAGCCAACAAT
                  ********************************************************************************

Predicted seq.    TCCGGGGCTGGCTCATATGTAAGGACCAAGAAAGGAAATGCCCTTAAGGAGGGGTTTTGTTCCCTTACCATGGAT
Resulting seq.    TCCGGGGCTGGCTCATATGTAAGGACCAAGAAAGGAAATGCCCTTAAGGAGGGGTTTTGTTCCCTTACCATGGAT
                  ********************************************************************************

Predicted seq.    GAGGTCAGTCAGAACACTGAGGGTATTACAGGCCTTCTTAACAACATAACGAGTCACCCATACTTTCGGTACTTT
Resulting seq.    GAGGTCAGTCAGAACACTGAGGGTATTACAGGCCTTCTTAACAACATAACGAGTCACCCATACTTTCGGTACTTT
                  ********************************************************************************

Predicted seq.    AAAGTAAATCTGGACAGGGAATGTCGTTACTGGGTCGCTGAGGCTAGTTGTACATGTGATTCAAATGGATGCCAG
Resulting seq.    AAAGTAAATCTGGACAGGGAATGTCGTTACTGGGTCGCTGAGGCTAGTTGTACATGTGATTCAAATGGATGCCAG
                  ********************************************************************************

Predicted seq.    ATCTGCACTTGTGATGACAGTGGAATTCCCGAAACACTAAAATACCCATACGATATGAGTGATGTTTCCACGGTT
Resulting seq.    ATCTGCACTTGTGATGACAGTGGAATTCCCGAAACACTAAAATACCCATACGATATGAGTGATGTTTCCACGGTT
                  ********************************************************************************

Predicted seq.    GAACGTCGTGCGGCCCCTGACAAACATGCAGCAAAAGGTTTCGAGGACGAGATCAAGCCGATCGACCCAGATAGA
Resulting seq.    GAACGTCGTGCGGCCCCTGACAAACATGCAGCAAAAGGTTTCGAGGACGAGATCAAGCCGATCGACCCAGATAGA
                  ********************************************************************************

Predicted seq.    GATGCAACATATGTGGATCTTCTCCAAAACCCAGAGGCAAACACCGGATATTCGGGACCGAAAGCAGCGCGGGTT
Resulting seq.    GATGCAACATATGTGGATCTTCTCCAAAACCCAGAGGCAAACACCGGATATTCGGGACCGAAAGCAGCGCGGGTT
                  ********************************************************************************

Predicted seq.    TGGCAGGCTGTATATAACAACTGCAATATTGACGGACTTCCAAGTAATGATACGGCTGGTGTAGAAAACAGAGAG
Resulting seq.    TGGCAGGCTGTATATAACAACTGCAATATTGACGGACTTCCAAGTAATGATACGGCTGGTGTAGAAAACAGAGAG
                  ********************************************************************************

Predicted seq.    AAGGCTTTACTTCGGCAGCTTCTAAGTGGGCTCCATACTTCTATAACCATGCATGTGGCTGCTTTCTTTTATAAT
Resulting seq.    AAGGCTTTACTTCGGCAGCTTCTAAGTGGGCTCCATACTTCTATAACCATGCATGTGGCTGCTTTCTTTTATAAT
                  ********************************************************************************

Predicted seq.    GATACGAAGGGTGACTCTCCGCTGCGCAGCCTAGGTGTGCTAAATAATCCCAATATATCTTTCTATCCTAACTGT
Resulting seq.    GATACGAAGGGTGACTCTCCGCTGCGCAGCCTAGGTGTGCTAAATAATCCCAATATATCTTTCTATCCTAACTGT
                  ********************************************************************************

Predicted seq.    GGCATGTTCCGCCGCATCGTTAAGAATGACGAATTCATAAGGAACCTTTTTGTTGTCTACCAGTTTGTCCTTCGC
Resulting seq.    GGCATGTTCCGCCGCATCGTTAAGAATGACGAATTCATAAGGAACCTTTTTGTTGTCTACCAGTTTGTCCTTCGC
                  ********************************************************************************

Predicted seq.    GCTCTGGCGAAGACCAAACGTGCATTTCTTGCTAACTCAAGTTTGTATAACTCTGGCTTCAACGGCGCTGCTACA
Resulting seq.    GCTCTGGCGAAGACCAAACGTGCATTTCTTGCTAACTCAAGTTTGTATAACTCTGGCTTCAACGGCGCTGCTACA
                  ********************************************************************************

Predicted seq.    GATGGAGATGTGCGTCTCTACAGCAATATTGGGGAACTTTTTAGCAGCAAGCTATTCCGCGTAGCAACTTTCGAC
Resulting seq.    GATGGAGATGTGCGTCTCTACAGCAATATTGGGGAACTTTTTAGCAGCAAGCTATTCCGCGTAGCAACTTTCGAC
                  ********************************************************************************

Predicted seq.    GAGCAAAAATTTTTTGAGTCTCCGGGTGCGCACCTACTGGTTCGGCAAATGAAGCGCGTGGTGCACAATGTTACA
Resulting seq.    GAGCAAAAATTTTTTGAGTCTCCGGGTGCGCACCTACTGGTTCGGCAAATGAAGCGCGTGGTGCACAATGTTACA
                  ********************************************************************************

Predicted seq.    ACTCTCATGGATTGCGTCACGTGTGAAAAGTGCCGCGCTTGGGGGAAGTTGGAAACAGCTGCTCTTGCCACCGCG
Resulting seq.    ACTCTCATGGATTGCGTCACGTGTGAAAAGTGCCGCGCTTGGGGGAAGTTGGAAACAGCTGCTCTTGCCACCGCG
                  ********************************************************************************

Predicted seq.    TTGAAAATTGTCTTCGGTTCCGCTGAGAATGTGGTGGAACTAAACAGAGGAGAGAGAGTTGCCCTTATAAATTTT
Resulting seq.    TTGAAAATTGTCTTCGGTTCCGCTGAGAATGTGGTGGAACTAAACAGAGGAGAGAGAGTTGCCCTTATAAATTTT
                  ********************************************************************************
```

```
                                                                  ← RP4Ero1
Predicted seq.      GCGCGGCAACTTGCCATTTCTGTTAAAAATGTTCGTTCACTAGCTGCTGTGTGCGAAAAATTCAATTACACGGCT
Resulting seq.      GCGCGGCAACTTGCCATTTCTGTTAAAAATGTTCGTTCACTAGCTGCTGTGTGCGAAAAATTCAATTACACGGCT
                    ********************************************************************************

                             BamHI      ← Ero1 p3 seq RP
Predicted seq.      ATGAGTACTAGTGCGGCCGCTTTCGAATCTAGAG
Resulting seq.      ATGAGTACTAGTGCGGCCGCTTTCGAATCTAGAG
                    ***********************************
```

B.

```
Expected seq.    MLKMRLLIVVPVLLGLVWQILLRAELDGVSFFGMYISANNSGAGSYVRTKKGNALKEGFCSLTMDEVSQNTEGIT
Observed seq.    MLKMRLLIVVPVLLGLVWQILLRAELDGVSFFGMYISANNSGAGSYVRTKKGNALKEGFCSLTMDEVSQNTEGIT
                 ********************************************************************************

Expected seq.    GLLNNITSHPYFRYFKVNLDRECRYWVAEASCTCDSNGCQICTCDDSGIPETLKYPYDMSDVSTVERRAAPDKHA
Observed seq.    GLLNNITSHPYFRYFKVNLDRECRYWVAEASCTCDSNGCQICTCDDSGIPETLKYPYDMSDVSTVERRAAPDKHA
                 ********************************************************************************

Expected seq.    AKGFEDEIKPIDPDRDATYVDLLQNPEANTGYSGPKAARVWQAVYNNCNIDGLPSNDTAGVENREKALLRQLLSG
Observed seq.    AKGFEDEIKPIDPDRDATYVDLLQNPEANTGYSGPKAARVWQAVYNNCNIDGLPSNDTAGVENREKALLRQLLSG
                 ********************************************************************************

Expected seq.    LHTSITMHVAAFFYNDTKGDSPLRSLGVLNNPNISFYPNCGMFRRIVKNDEFIRNLFVVYQFVLRALAKTKRAFL
Observed seq.    LHTSITMHVAAFFYNDTKGDSPLRSLGVLNNPNISFYPNCGMFRRIVKNDEFIRNLFVVYQFVLRALAKTKRAFL
                 ********************************************************************************

Expected seq.    ANSSLYNSGFNGAATDGDVRLYSNIGELFSSKLFRVATFDEQKFFESPGAHLLVRQMKRVVHNVTTLMDCVTCEK
Observed seq.    ANSSLYNSGFNGAATDGDVRLYSNIGELFSSKLFRVATFDEQKFFESPGAHLLVRQMKRVVHNVTTLMDCVTCEK
                 ********************************************************************************

Expected seq.    CRAWGKLETAALATALKIVFGSAENVVELNRGERVALINFARQLAISVKNVRSLAAVCEKFNYTAMSTSAAAFES
Observed seq.    CRAWGKLETAALATALKIVFGSAENVVELNRGERVALINFARQLAISVKNVRSLAAVCEKFNYTAMSTSAAAFES
                 ********************************************************************************

Expected seq.    R
Observed seq.    R
                 *
```

Figure 5.26 Analysis of TbEro 1 pGST-Parallel 3 expression construct sequence

The TbEro 1 parallel 3 construct was sequenced to ensure the open reading frame was correctly ligated into the parallel 3 expression vector.

A. The predicted sequence and resulting sequence obtained were compared. The location where primers bound to the sequence are shown in yellow and the primer name and direction are indicated above. Restriction sites are shown in red and the restriction enzymes at which these sites cut are indicated above. Agreement is indicated by a * below the sequence and disagreements are highlighted in light blue.

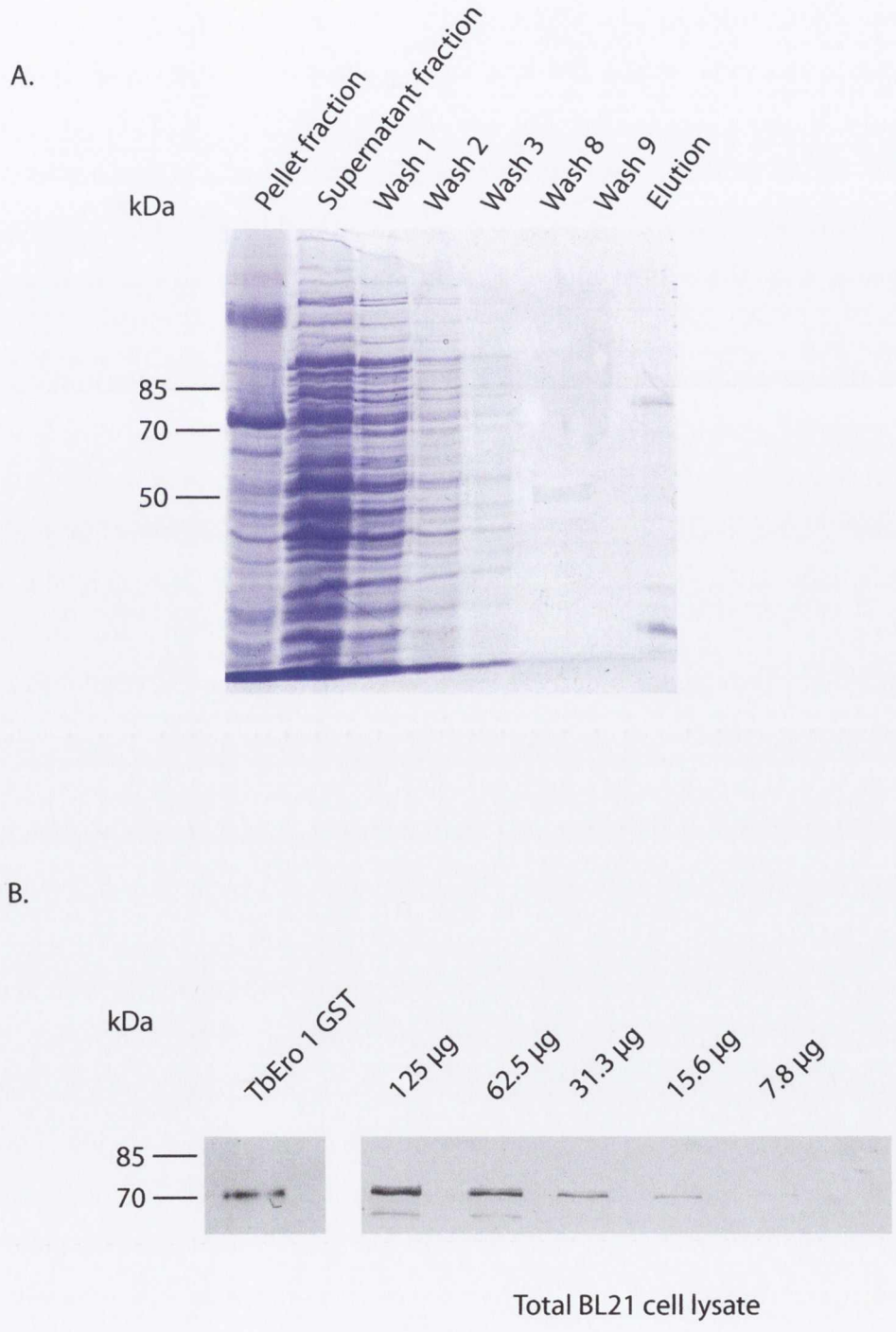
Panel B: A translation of the expected sequence and observed sequence obtained written in single letter amino acid abbreviation. Agreement is indicated by a * below the sequence. The beginning of the GST tag is highlighted yellow.

Figure 5.27 Purification of recombinant TbEro 1 GST

Panel A: SDS-PAGE of various fractions collected during the affinity purification of TbEro 1 GST using anti-GST agarose. All lanes represent 20µl of fractions collected.

Panel B: Western blot of affinity purified TbEro 1 GST versus total BL21 cell lysate; probed with anti-TbEro 1 antibody as described in the Methods.

Figure 5.27 Purification of recombinant TbEro 1 GST



5.6 Discussion

The oxidative folding pathway within the endoplasmic reticulum of mammals and yeast has been a topic of much research over the past decade and new light has been shed on the subject. Presently, it is widely accepted that Ero 1 and PDIs form the major enzymes of the oxidative folding pathway of the endoplasmic reticulum. This pathway allows the transport of electrons from unfolded proteins of the secretory pathway to PDI, then to Ero 1 and finally molecular oxygen to generate hydrogen peroxide; thus allowing the formation of disulphide bonds within the substrate proteins through the net loss of an electron pair (Ramming and Appenzeller-Herzog, 2012). While Ero 1 has been a topic of much research in species such as mammals and yeast, this is the first instance where an Ero 1 has been characterized in trypanosomes. The results presented in this study indicate that TbEro 1 is an essential membrane associated protein of the endoplasmic reticulum and is heavily *N*-glycosylated. Additionally, the data suggests that TbEro 1, with TbPDI-2, forms the major oxidative folding pathway in the endoplasmic reticulum; of which VSG is a substrate. These conclusions are consistent with the following important findings: Firstly, the knockdown of TbEro 1 resulted in a slow growth defect in bloodstream and procyclic forms indicating that TbEro 1 is essential for normal growth in culture. This effect on growth appeared to be more pronounced in bloodstream forms than procyclic form cells. This result suggested that bloodstream forms required higher levels of TbEro1 activity than procyclic forms. Secondly, the knockdown of TbEro 1 resulted in a failure of cell cytokinesis. This phenotype was only observed in the bloodstream form cell line. However, the cells then re-entered mitosis resulting in multinucleated and multiflagellate cells. Interestingly, this cell cycle arrest was similar to the phenotype observed as a result of the disruption of the protein GPI8, which was shown to be involved in the direct processing of VSG (Lillico et al., 2003). Thirdly, TbEro 1 was found to interact with TbPDI-1 and TbPDI-2 *in vivo* by analysis of proteins that co-immunoprecipetated with TbEro 1 using Western blot. Fourthly, TbPDI-2 and TbEro 1 were found to be endoplasmic reticulum resident proteins as the two proteins were found to co-localize within this cellular organelle. Finally, the knockdown of TbEro 1 and TbPDI-2 resulted in a slowing of the rate of newly synthesized proteins reaching the cell surface, proteins which were most likely to be

VSG. These findings, their consequences and the current views expressed in the field of oxidative folding will be the topic of this discussion.

With the use of detergent lysis and mechanical lysis, TbEro 1 was shown to be a membrane associated protein. Similar methods have also demonstrated that the same is true for the two mammalian Ero 1 proteins (Ero 1α/β) and the single yeast Ero 1 (Cabibbo et al., 2000; Frand and Kaiser, 1998; Pagani et al., 2000). To date no Ero 1 proteins have been shown to be soluble *in vivo* (Ramming and Appenzeller-Herzog, 2012). While it has been suggested that the human proteins Ero 1α and Ero 1β are membrane associated proteins, the molecular mechanism for this association has yet to be explained (Ramming and Appenzeller-Herzog, 2012). Additionally, the problem has been made more complex by virtue of the fact that Ero1p, the yeast Ero 1, possesses an additional 18 kDa *C*-terminal domain, while both human Ero 1 proteins do not (Pagani et al., 2001). A truncated Ero1p protein, expressed without the *C*-terminal domain, was found to be soluble in cultured cells and the absence of this domain resulted in a loss of function (Pagani et al., 2001). In contrast, the two mammalian and the single trypanosome Ero 1 proteins possessed no such *C*-terminal domain but were still found to be membrane associated proteins. The results of the bioinformatics analysis presented in this thesis demonstrated that TbEro 1 possesses no obvious *C*-terminal hydrophobic region, which would allow association with the membrane, with the exception of the *C*-terminal active site. This is similarly the case with the mammalian Ero 1 proteins. As such, membrane association of TbEro 1 cannot be explained by hydrophobic interactions with the membrane. Interestingly, recombinant TbEro 1 expressed in *E. coli* was shown to be soluble under physiological conditions. This finding suggests that the post translational modifications, which are typically carried out in eukaryotic cells, were required for membrane association. Recombinant forms of the mammalian Ero 1 proteins were also shown to be soluble, supporting this view. This solubility did not affect the activity of the mammalian Ero 1 proteins *in vitro* (Baker et al., 2008). This finding indicates that membrane association of mammalian Ero 1 proteins is not required for the function of the protein *in vitro.* However, it appears that membrane association may play an important role *in vivo,* as was shown in the case of the yeast Ero1p.

It was found that TbEro 1 is glycosylated in both bloodstream form and procyclic form cells. This was shown by a preliminary characterization of the glycans of the HA

tagged protein. When analyzed via Western blot, the mass observed of the HA epitope tagged TbEro 1 protein expressed in *T. brucei* was found to be ~52 kDa larger than expected. It was thought this increase in size may be due to post translational modifications. This increase in observed size is characteristic of glycoproteins; not only because of the added mass of the glycans but also because the binding of SDS is lower due to the interference caused by the carbohydrate chains (Fukuda and Kobata, 1993). Additionally, TbEro 1 migrated with a characteristic 'batwing' band, which is typical of highly *N*-glycosylated proteins, which is said to be caused by microheterogeneity of the *N*-glycans (Fukuda and Kobata, 1993). It was found that TbEro 1 had seven possible *N*-glycosylation sites and so it appeared likely that TbEro 1 was heavily glycosylated. Interestingly, the recombinant TbEro 1 expressed in *E. coli* migrated at the expected protein size unlike the TbEro 1, which was expressed in *T. brucei*. This result suggested that the modification that occurred in *T. brucei* was not carried out in *E. coli* as *N*-glycosylation is not a feature of bacteria. This finding suggested that TbEro 1 may possess these *N*-glycans. Furthermore, this same bacterially expressed TbEro 1 was detected by Western blot using a polyclonal rabbit antibody generated during this study. This result was surprising as the TbEro1 expressed in *T. brucei* failed to react with the same antibody. This finding suggested that the epitope chosen to raise the antibodies was blocked by a post translational modification. The region of TbEro 1 sequence, from which the target peptide was engineered, was 12 amino acids immediately downstream of a highly conserved *N*-glycosylation site, which was present in all the Ero 1 proteins from the various species reviewed in the bioinformatics analysis (see methods). This site may have been the location of the *N*-glycosylation that blocked the antibody from binding to TbEro 1.

These findings taken together suggested that some of the additional mass may have been due to glycosylation of TbEro 1, which was confirmed by digestion with two glycosidases. The size of TbEro 1 fell by ~20 kDa when digested with *N*-glycosidase F and endoglycosidase H (Endo H), suggesting the protein was *N*-glycosylated. While this difference in mass was smaller than expected it was likely that all of the glycans were not removed. This is however a common occurrence when digesting glycans either because the enzyme was unable to reach all of the *N*-glycans or because other carbohydrates which could not be digested by the two enzymes were present; for example *O*-linked glycans however this is unlikely as *O*-glycosylations are added in

the Golgi apparatus (Rubotham et al., 2005; Spitznagel et al., 2009; Woods et al., 2013). The enzymes *N*-glycosidase F digests all *N*-glycans where as Endo H only digests high mannose-type *N*-linked chains, which become resistant to digestion by Endo H after modification in the Golgi apparatus. The results showed that TbEro 1 contained at least some high mannose-type *N*-glycosylation which indicated that the protein does not progress through the Golgi apparatus. However, some mature proteins have been known to contain high mannose N-glycosylations beyond the Golgi, although this is uncommon (Fukuda and Kobata, 1993). This result therefore supported the view that TbEro 1 was a glycosylated resident of the endoplasmic reticulum.

The yeast Ero 1 protein (Ero1p) was shown also to be glycosylated and interestingly these glycans were also susceptible to digestion with Endo H (Frand and Kaiser, 1998). The yeast Ero 1 also possesses seven potential *N*-glycosylation sites. There was an ~30 kDa difference in the mass of Ero1p between the expected and the observed mass; as determined by Western blot. Once digested with Endo H the apparent size of the proteins fell by 15 kDa. Similarly in mammals, Ero1 α and Ero1 β were found to be glycosylated (Benham et al., 2000; Cabibbo et al., 2000; Lewandrowski et al., 2006). The glycans of Ero1α and Ero1β were also found to be sensitive to digestion with Endo H. Similarly, the conclusion that Ero1α and Ero1β did not progress to the Golgi as a result of this finding was also reached (Cabibbo et al., 2000; Pagani et al., 2000). Interestingly, Ero1β possesses 3 unique *N*-glycosylation sites surrounding an active site cysteine residue. While it has not been experimentally proven if any of these unique sites are indeed glycosylated, it has been suggested that the presence of these bulky oligosaccharides may provide a mechanism to modulate substrate recruitment (Ramming and Appenzeller-Herzog, 2012). This may explain how Ero1β is able to interact preferentially with substrates different to Ero1α. However, to date, such a mechanism has yet to be investigated (Ramming and Appenzeller-Herzog, 2012).

To further investigate the possibility that TbEro 1 was a resident of the endoplasmic reticulum, as was indicated by the ER specific glycans of the protein, a subcellular fractionation was carried out of both bloodstream and procyclic form cells. The results indicated that TbEro 1 shared a similar profile and as such a similar location to BiP, an endoplasmic reticulum resident protein, and a somewhat similar location to

TbPDI-2, also considered a resident of the endoplasmic reticulum. Additionally, immunofluorescence of TbEro 1 indicated a pattern similar to the distribution of BiP (Bangs et al., 1993). These results taken together indicate TbEro 1 is a resident of the endoplasmic reticulum which does not progress into the Golgi. The results of this study mirrored those found in the mammalian and yeast Ero 1 proteins. Immunofluorescence microscopy of myc-tagged Ero1α and Ero1β revealed the proteins co-localized with two endoplasmic reticulum resident proteins in cultured mammalian cells (Cabibbo et al., 2000; Pagani et al., 2000). Additionally, yeast Ero1p was shown to reside only in the endoplasmic reticulum similarly by immunofluorescence microscopy and co-localization with BiP (Frand and Kaiser, 1998; Pagani et al., 2001; Pollard et al., 1998).

The mechanism by which Ero 1 proteins remain in the endoplasmic reticulum is still under question. It has been speculated that association of Ero 1 proteins with PDIs that contain KDEL like retrieval signals may ensure Ero 1 is returned from the Golgi in the proposed 'thiol retention' model (Anelli et al., 2003). Data has indicated that ERp44, a human PDI, which possesses a RDEL retrieval signal and a single non-catalytic active site CXXS may play a role in Ero1α and Ero1β retrieval from the Golgi (Anelli et al., 2003). This PDI has been shown to interact with many secretory substrates. They include IgM (Anelli et al., 2003), adipose-derived hormone adiponectin (Wang et al., 2007) and suphatase modifying factor 1 (Fraldi et al., 2008; Mariappan et al., 2008). These substrates all required cysteine-dependent dimerization/oligomerization to be secreted. However, it is important to note that ERp44 does not have the required catalytic site to achieve such reactions and it was shown that knockout of ERp44 resulted in no change in the secretion of these proteins (Wang et al., 2007). Interestingly, the location of ERp44 was found to be primarily in the cis-Golgi using immunofluorescence microscopy. Due to the location of ERp44, the inactive CXXS a-domain, which was shown to bind to proteins of the secretory pathway, and the RDEL retention signal, it was thought ERp44 was responsible for the retro-translocation of these substrate proteins back to the endoplasmic reticulum. These substrates were said to be sequestered in the secretory pathway, by ERp44 until needed (Mariappan et al., 2008; Wang et al., 2007). Crucially, it is thought that ERp44 interacts preferentially with Ero1α over these sequestered proteins. This finding was indicated by the overexpression of Ero1α, which resulted in the increased secretion of substrates of ERp44 (Anelli et al., 2003; Wang et al.,

2007). It was thought this additional secretion was caused by the occupation of ERp44 in returning the excess Ero1α proteins to the endoplasmic reticulum. This 'thiol retention' model may provide a mechanism for the retention of TbEro 1 within the endoplasmic reticulum. However, the only TbPDI found to possess a possible endoplasmic reticulum retention signal is TbPDI-2. TbPDI-2 possesses two CXXC catalytic domains and so does not share the CXXS active site of ERp44 which has been proposed to form the mix disulphide trap required for thiol mediated retention. It remains to be seen if the thiol retention model will hold true; however, as of yet there is not enough direct evidence to prove the mechanism is responsible for the retention of Ero1 with the endoplasmic reticulum (Ramming and Appenzeller-Herzog, 2012). In particular, further co-localization studies of ERp44 and Ero 1 *in vivo* is needed to provide more conclusive evidence to support this possible mechanism.

The knockdown of TbEro 1 by the use of RNAi resulted in a growth phenotype in both bloodstream and procyclic form cells. This indicated TbEro 1 was essential for normal growth of both lifecycle forms in culture. While it was likely TbEro 1 was essential in culture, it was surprising to find that a knockdown of 85% of TbEro 1 at the transcript level resulted in a growth phenotype. In other species a measurable phenotype did not occur, even when the Ero 1 proteins were knockdown below 10% of wild type levels. The mammalian Ero 1 proteins were said to be nonessential (Zito et al., 2010). However, the results obtained during the study of Ero1α and Ero1β are misleading. The gene trapping techniques employed to knockdown the Ero1α and Ero1β allowed a residual amount of enzyme to remain, as was confirmed at the RNA and protein levels (Chin et al., 2011; Zito et al., 2010). As such, it was concluded that only a fraction of the total amount of Ero1 was required to maintain cell viability (Ramming and Appenzeller-Herzog, 2012). Mutation of the Ero1β gene in mice *in vivo,* however, resulted in a delay in the maturation of pro-insulin in pancreatic islets, which manifests in a diabetic phenotype (Zito et al., 2010). This indicated the Ero1 β gene is essential *in vivo* for the proper folding of pro-insulin. Additionally Ero1 α could not compensate for the loss of Ero1 β, which indicated the two enzymes possess different substrates (Zito et al., 2010a).

Similarly in yeast, a temperature sensitive mutant of Ero1p failed to grow indicating the protein was essential (Frand and Kaiser, 1998; Pollard et al., 1998). The introduction of an ectopic copy of the Ero1p gene, under the control of the GAL1

promoter, restored the growth of the temperature sensitive mutant cells; further supporting this view. Interestingly, the cells were able to grow when the promoter of the ectopic gene was repressed. This result indicated that, like mammalian Ero1α and Ero1β knockdown cells, the yeast cells were able to survive with 10% of the wild type amount of Ero 1 protein (~90% knockdown)(Ramming and Appenzeller-Herzog, 2012). The knockdown studies in yeast and mammals highlight a significant difference in the oxidative folding pathway between *T. brucei* and these species. This difference may be explained by the possible alternative pathways for *de novo* disulphide-bond generation in the mammalian endoplasmic reticulum, which are not present in *T. brucei* (Bulleid and Ellgaard, 2011).

Many different pathways have been proposed for the production of disulphide bonds in mammalian cells, which may compensate for the loss of Ero 1. These pathways involve the production and transfer of disulphide bonds within the endoplasmic reticulum through the following enzymes: peroxiredoxin 4, glutathione peroxidase 7, glutathione peroxidase 8, quiescin sulfhydryl oxidase and vitamin K epoxide reductase. These enzymes have been proposed to provide disulphide bonds to PDIs or their substrates through the same thiol disulphide reactions described in the introduction *in vivo* (Bulleid and Ellgaard, 2011). However, the majority of the proteins listed have only been shown to provide disulphide bonds *in vitro* and little evidence exists to support these roles *in vivo*. Additionally, of these proteins proposed to interact with PDIs within the endoplasmic reticulum, *T. brucei* only possesses one protein which may carry out similar reactions: TbQSOX.

The protein TbQSOX, as mentioned in the introduction, is another sulfhydryl oxidase family member and has the same ability to generate disulphide bonds *de novo* by imparting an electron pair to molecular oxygen. The difference between QSOX and Ero 1 proteins is the ability of QSOX to directly reduce proteins within the endoplasmic reticulum without the use of PDIs. However, QSOX is unable to carry out isomerization reactions (Rancy and Thorpe, 2008). Additionally, in mammalian cells this protein is thought to reside in the Golgi apparatus not the endoplasmic reticulum. This has yet to be established in *T. brucei* (Mairet-Coello et al., 2004). While QSOX seems to be a critical member of the oxidative folding pathway of secretory proteins, it appears the location of the protein indicates TbQSOX is likely to have a separate role to TbEro 1; a role found later in the secretory pathway. Therefore compensation

of one by the other seems unlikely. Currently no *in vivo* investigation of TbQSOX has been carried out. However, in yeast it has been shown that overexpression of yeast QSOX was able to compensate for the knockout of Ero1p (Chakravarthi et al., 2007). In the fruit fly *D. melanogaster* knockdown of QSOX resulted in no phenotype. However, an Ero 1 and QSOX knockdown lead to a greater phenotype than a knockdown of Ero 1 alone (Tien et al., 2008). These findings indicate that QSOX and Ero 1 work together for the contribution of disulphide bonds to the secretory pathway but Ero 1 plays a more crucial role. In trypanosomes, TbQSOX may compensate for the loss of TbEro 1 however this has yet to be established using knockout of TbEro 1 *in vivo*. What is clear from the results of this study is that TbEro 1 is required for normal cell growth in culture in spite of the possibility of compensation by TbQSOX.

While mammalian Ero 1 proteins may not be the sole provider of oxidizing equivalence in the endoplasmic reticulum as previously thought, the knockdown of Ero 1 proteins directly resulted in the slowing of the proliferation of cells in culture. The growth of cultured mice cells, which usually divide rapidly, was slowed by the mutation of Ero1α/β (Ester Zito, David Ron, and Christan Appenzeller-Herzog, unpublished observations 2012 ;Ramming and Appenzeller-Herzog, 2012). As a result of the RNAi of TbEro 1, a growth phenotype was observed in bloodstream and procyclic form cells. However, the effect on growth was greater in bloodstream forms than procyclic form cells. This indicated TbEro 1 played a more crucial role in bloodstream form cells over procyclic forms, as had been similarly shown with TbPDI-2. In added support of this view, when the transcript level of TbEro 1 in wild type cells was analyzed it was found to be three fold greater in bloodstream forms when compared to procyclic forms.

Mammalian Ero1β was found to be a tissue specific protein, which is similar to the life cycle stage specific upregulation of TbEro 1, observed in bloodstream form cells compared to procyclic form cells. The protein Ero1β was found to be expressed in the cells of the islets of Langerhans. The demand for oxidative folding pathway of this tissue, like that of the bloodstream form trypanosomes, is greater than other cells. This increase in demand is caused by the production and correct folding of proinsulin and as such an additional Ero 1 enzyme is required, which has shown to be involved

in the folding of proinsulin through knockdown studies (Ramming and Appenzeller-Herzog, 2012).

It is evident that the demand on the oxidative folding pathway of bloodstream form cells is greater than that of the procyclic form cells, as was previously discussed in chapter 4. This demand is likely to be greater due to the folding of VSG in bloodstream form cell, which must be folded correctly to maintain cell growth. Procyclic form cells do not have this same demand as the major surface protein, which is expressed in place of VSG known as procyclin, does not contain any disulphide bonds. As this demand is greater in bloodstream form cells versus procyclic form cells it seems likely that an equivalent loss of activity of TbEro 1 in both bloodstream and procyclic forms would lead to a greater effect in bloodstream form cells. This view was supported by the results observed during the knockdown of TbEro 1.

The knockdown of TbEro 1 in the procyclic form cells did however cause a growth phenotype. As was discussed in chapter 4, it seems likely that this was caused by the reduced rate of folding of other important proteins, which contained disulphide bonds. It is apparent that at least twelve separate proteins are transported to the surface of the cell, which contain cysteine residues. Two of these proteins appear to be relatively abundant and so it is likely that the loss one of these proteins was responsible for the slow growth phenotype observed in procyclic form cells (Nolan and Garcia-Salcedo, 2008).

From the results of the knockdown study of TbEro 1, it appeared that the level of TbEro 1 was insufficient to support normal growth in either bloodstream or procyclic form cells, regardless of the life cycle investigated. This result indicated that there are no other enzymes with the ability to compensate for the loss of TbEro 1.

From the findings of this study, which indicated TbEro 1 played a greater role in bloodstream form cells, it was thought TbEro 1 may play a role in the folding of the bloodstream stage specific protein VSG. To further investigate this possibility a cell cycle analysis was carried out and revealed an almost identical cell cytokinesis arrest phenotype to the knockdown of TbPDI-2. As previously discussed this phenotype was only seen in bloodstream form cells and was similar to that seen in the knockdown of GPI8, a protein involved in the processing of VSG (Lillico et al., 2003). To assess the rate of surface protein production the same pulse chase assay was employed to assess

the rate of newly synthesized proteins reaching the cell surface. Interestingly, the surface proteins were slowed by the absence of TbEro 1 as they were in the absence of TbPDI-2. The size of the protein produced was ~60 kDa and was most likely to be VSG supporting the view that TbEro 1 was required for the efficient *de novo* folding of VSG within the endoplasmic reticulum.

The slowed rate of protein folding and export from the endoplasmic reticulum caused by a loss of activity of Ero 1 has been observed in other species. It was shown in the yeast temperature sensitive null mutant of Ero1p that two proteins failed to mature and progress from the endoplasmic reticulum, as was indicated by pulse chase analysis. The first was carboxypeptidase Y (CPY), which was shown to be slowed by the knockdown of Ero1p and the second, and more significantly with respect to comparison with VSG, was the GPI-linked plasma-membrane protein Gas1p (Frand and Kaiser, 1999; Pollard et al., 1998). Interestingly, the loss of Ero1p could be complemented by the expression of human Ero1α, indicating that the Ero 1 proteins are functionally interchangeable between species despite differences in primary sequence (Cabibbo et al., 2000).

From the results taken together it is evident that a link between VSG, TbPDI-2 and TbEro 1 exists as each are intrinsically linked. It is well established that Ero 1 proteins interact with class 1 PDIs in yeast and mammals and that these enzymes are required for the efficient folding of proteins of the secretory pathway within the endoplasmic reticulum *in vivo*. From the data collected from this study, this too appears to be the case in *T. brucei*. While many questions remain as to the flow of electron within the endoplasmic reticulum, TbEro 1, TbPDI-2 and VSG appear to provide the first of such pathways investigated in *T. brucei in vivo*.

6 Conclusions and Future Work

The primary aim of this thesis was to provide the first, overall description of the oxidative folding pathway in the endoplasmic reticulum of an African trypanosome. Taken together, the results presented here support the view that variant surface glycoprotein (VSG) serves as a major protein substrate for a pathway involving a specific protein disulphide isomerase (TbPDI-2) that is reoxidized by a single trypanosomal oxidase (TbEro 1) in the archetypical trypanosome, *T. brucei*. In broad outline this trypanosome pathway concurs with current ideas presented in the literature about the oxidative folding pathway in higher eukaryotes. However, a number of novel features have emerged from this study, which probably reflect the adaptation of African trypanosomes to life as an exclusively extracellular parasite in mammals as well as the fact that these unicellular eukaryotes possess a less complex pathway, in terms of possible constituent.

A bioinformatics analysis of the complete genome of *T. brucei* indicated that these parasites were likely to possess only six candidate genes for PDIs, which is almost a third less than the number present in the human genome. These six putative PDIs were classified as class 1 PDI (TbPDI-2), one class 2 PDI (TbPDI-3), three single domain PDIs (TbPDI-1, TbPDI-5 and TbPDI-6) and finally a single transmembrane PDI (TbPDI-4). Additionally, only a single possible TbEro 1 was found during the analysis. Therefore, *T. brucei* possesses a much less complex oxidative folding pathway in comparison to higher eukaryotes, for example mammals, which possess 19 possible PDIs and two Ero 1 proteins (Kozlov et al., 2010; Ramming and Appenzeller-Herzog, 2012). A comparison of the genomes of other trypanosomes indicated the presence of these predicted enzymes in *T. congolense*, *T. vivax*, *T. cruzi* and *L. major*. This finding demonstrates that the enzymes of the oxidative folding pathway are conserved across the different trypanosome species, which differ greatly in lifecycle and morphologies.

Interestingly, *T. brucei* was found to have a similar number of PDIs to the yeast, *S. cerevisiae*. However it was shown that yeast only possess one catalytically active class 1 PDI (Pdi1p) and have no class 2 PDIs. In contrast *T. brucei* possesses a single class 2 PDI (TbPDI-3). This indicates that African trypanosomes possess a more complex

oxidative folding pathway than yeast and as such offers more opportunity to understand the oxidative folding pathway.

The bioinformatics analysis determined that *T. brucei* possessed only one Ero 1 (TbEro 1). It was also established that between the trypanosome, yeast, drosophila and mammalian species included in the analysis, the domain structure, active site sequence and predicted size of the putative Ero 1 proteins were found to be similar. Additionally the percentage identity, in the least similar Ero 1 included (yeast Ero1p), was found to have a 25% identity with TbEro 1. This result suggested Ero 1, as a protein, is highly conserved across the evolutionary phylogenetic scale and therefore plays a critical role in the eukaryotes.

Interestingly, the size of the trypanosome Ero 1 was found to be similar to that of the mammalian Ero 1 proteins. This came as a surprise as it was expected that TbEro 1 would be most similar to the yeast Ero 1, which possessed an additional *C*-terminal domain required for membrane association (Pagani et al., 2001). Additionally, no transmembrane spanning region was identified in TbEro 1, which was similarly found in the mammalian Ero 1.

It was found that the expression of two of the elements of the oxidative folding pathway, TbPDI-2 and TbEro 1, are clearly upregulated in bloodstream forms when compared to procyclic forms from an analysis of the transcript of the putative enzymes of the oxidative folding pathway. This requirement for TbPDI-2 and TbEro 1 reflects the short generation time of these bloodstream forms of *T. brucei* and their need to cover their surface with ~10^7 copies of a monolayer of a single protein: the variant surface glycoprotein, whose structure is underpinned by key disulphide bonds.

It was demonstrated in a previous study that a knockdown of TbPDI-2 resulted in no measurable phenotype in *T. brucei* (Rubotham et al., 2005). This result was thought to be caused by one of two possibilities: Either TbPDI-2 was not required for growth in culture or the knockdown of TbPDI-2 was insufficient to cause a phenotype. As TbPDI-2 was found to be the only class 1 PDI with an endoplasmic reticulum retention signal, the first possibility seemed unlikely and so it was decided that further analysis of TbPDI-2 was required *in vivo*.

From the analysis of TbPDI-2 presented in this thesis it was shown that TbPDI-2 is essential for normal growth in culture in both bloodstream and procyclic form cells. The loss of the protein resulted in a slow growth phenotype in both bloodstream and procyclic form cells, which suggested that the enzyme was responsible for the folding of required proteins in both lifecycles. Interestingly the phenotype was more pronounced bloodstream than in procyclic form cells. This finding was also reflected in the results of a cell cycle analysis, which showed a failure of cytokinesis in bloodstream forms but not in procyclic forms. This result suggests that the demand on the oxidative folding of the bloodstream form cells was greater than that of procyclic form cells. This outcome is most likely due to the folding of VSG, which puts additional demand on the oxidative folding pathway, causing a more severe phenotype in this lifecycle.

An *in vivo* analysis of TbEro 1 also revealed the protein was essential for normal growth of both bloodstream and procyclic form trypanosomes in culture. The results of the knockdown of TbEro 1 mirrored those observed for the loss of TbPDI-2. The knockdown caused a slow growth phenotype, which was more pronounced in the bloodstream forms compared to the procyclic forms. The phenotype, however, was found to be more severe with the knockdown of TbEro 1 when compared to that of a knockdown of TbPDI-2. Additionally, a similar, yet more severe failure of cell cytokinesis was also observed. These findings were unsurprising as it is thought that TbEro 1 is the main provider of oxidizing equivalence to the endoplasmic reticulum and a knockdown of TbEro 1 would be equivalent to the knockdown of all PDIs at once.

In this study it was clear that both TbPDI-2 and TbEro 1 are directly involved in the correct and efficient folding of VSG as a loss of the activity of either enzyme resulted in a slowing of newly synthesized VSG reaching the cell surface. This is the first reported study to show that a surface protein is a substrate of the oxidative folding pathway. However, there was an additional unexpected ~70 kDa protein produced in the TbPDI-2 knockdown cell line, which reached the surface of the cell. It is not yet known what this protein is and further investigation is required to determine the proteins identity.

Prior to this study, it was shown that when VSG production is disrupted at the transcript level, the cell cycle is halted at the precytokinesis stage and cells do not

undergo reinitiation of S phase of the cell cycle (Sheader et al., 2005). It is thought this halting of the cell cycle is due to a yet uncharacterized mechanism which "senses" the lack of mature VSG on the cell surface and triggers a cell wide arrest of protein production (Smith et al., 2009). Surprisingly, the loss of TbPDI-2 and TbEro 1 did not result in this characteristic cell cycle arrest. Instead the knockdown resulted in a failure of cytokinesis, similar to that reported as a result of the loss of GPI8, an enzyme involved in the processing of GPI-anchored proteins (Lillico et al., 2003). This finding suggested that while TbPDI-2 and TbEro 1 were directly involved in VSG production, both enzymes were not exclusively involved in this task. This outcome was not entirely unexpected in the case of TbEro 1 as it is thought to be the sole provider of oxidizing equivalence to the endoplasmic reticulum. However, it was surprising that the same was also true for TbPDI-2, suggesting that TbPDI-2 is involved in the oxidative folding of the majority of the proteins of the endoplasmic reticulum. This finding also suggests that both proteins provide effective drug targets for the treatment of trypanosomiasis.

The demand for the oxidative folding pathway is comparatively lower in procyclic forms, as the major surface protein, procyclin, does not contain any disulphide bonds. This was reflected in the less dramatic phenotype observed as a result of the knockdown of TbPDI-2 or TbEro 1. However, it is clear that the oxidative folding pathway is responsible for the folding of other essential proteins in procyclic form cells, which contain disulphide bonds. In a previous study it was demonstrated that at least twelve separate proteins are transported to the surface of the cell, which contain cysteine and methionine residues (Nolan and Garcia-Salcedo, 2008). These proteins are likely also to contain disulphide bonds. Two of the proteins appear to be relatively abundant and so it is likely that the loss of one was responsible for the slow growth phenotype observed in procyclic form cells. The genes ESAG 6 and 7 both encode the essential transferrin receptor, which possesses disulphide bonds. This protein is an example of an essential receptor that could possibly be a substrate of the oxidative protein folding pathway.

The characterization of TbEro 1 demonstrated that the protein is a resident of the endoplasmic reticulum in both bloodstream and procyclic form cells, as indicated by localization using immunofluorescence microscopy, subcellular fractionation studies and by analysis of endoplasmic reticulum specific *N*-glycans of the protein.

Additionally TbEro 1 was co-localized with TbPDI-2 using immunofluorescence microscopy and subcellular fractionation studies, and both proteins were shown to interact, using co-immunoprecipitation and Western blot analysis. These findings have established that TbPDI-2 and TbEro 1 form the principal oxidative folding pathway within the endoplasmic reticulum with VSG as the major substrate of this pathway. Not only is this pathway important in bloodstream form cells, it is also interesting to note that the same can be said in procyclic form cells.

It has been well documented that Ero 1 is a membrane associated protein, which was also found to be true for TbEro 1 during this study. This has been attributed to a *C*-terminal domain in the yeast Ero 1 (Pagani et al., 2001). However, TbEro 1 does not possess this *C*-terminal domain, which was also found to be a feature of mammalian Ero 1. Significantly, it was demonstrated that TbEro 1 is soluble when expressed as a recombinant protein in *E. coli*. This suggested that TbEro 1 associates with the membrane through post translational modifications, which are not carried out in *E. coli*. This was also found to be the case in mammalian Ero 1 (Ramming and Appenzeller-Herzog, 2012). Additionally it was found that TbEro 1 is heavily *N*-glycosylated and that this post translational modification may allow association with the membrane. Mammalian Ero 1 proteins are similarly *N*-glycosylated. This membrane attachment may possibly occur through a lectin present in the endoplasmic reticulum. While the way in which Ero 1 associates with the cell membrane remains unknown, *T. brucei* may provide a useful model for further investigation into this mechanism in future.

While investigating the effects of knocking out TbPDI-2 on the rate of VSG folding using a pulse chase assay, it was revealed that a loss of TbPDI-2 activity caused a slowing of the rate of newly synthesized VSG reaching the cell surface. There was, however, an additional unexpected ~70 kDa protein produced, which reached the surface of the cell. It is not yet known what this protein is and further investigation is required to determine the identity of this protein. This can be achieved using mass spectrometry of the excised protein. However, due to time constraints this was not carried out during this study. Additionally, it would be interesting to see what effect the loss of TbPDI-2 may have on the rate of folding of different VSG as it has been shown that different VSG are more efficient at folding than others (Bangs et al., 1986; Field et al., 2010; Smith et al., 2009). This suggests that other *T. brucei* cell lines

expressing different VSG, for example Mitat 1.2 or Mitat 1.5, may have a more or less severe phenotype.

The other PDIs identified possess intriguing features and in future it would be interesting to see what roles these enzymes play in oxidative folding pathway. With the exception of TbPDI-1, none of the other PDIs have been investigated. The class 2 PDI (TbPDI-3) is the only PDI of this classification in *T. brucei* and so this PDI presents an opportunity to study the loss of such an enzyme in a single celled organism. The study of a class 2 PDI has not been reported in the literature in any single celled organism; as *S. cerevisiae* does not possess any class 2 PDIs. Additionally no knockout studies have been carried out on the only human class 2 PDI known as PDIA6, most likely due to the difficulties associated with knocking out essential PDIs (Ramming and Appenzeller-Herzog, 2012). As previously discussed the protein PDIA6 is considered to be important in platelet function in humans and so additional understanding of class 2 PDIs may offer new insights into their function (Jordan et al., 2005).

While an RNAi study has been carried out on TbPDI-1, a single domain PDI, the two remaining trypanosome single domain PDIs have yet to be investigated (Rubotham et al., 2005). The two additional single domain PDIs, identified during this investigation, offer a unique opportunity to investigate the role of both active and inactive small single domain PDIs as TbPDI-5 was found to be possesses the CXXC active site motif whereas TbPDI-6 possesses a catalytically inactive SXXC motif. Additionally it is not known whether the single domain PDIs are essential as no knockout studies have been previously carried out.

Of the PDIs identified only TbPDI-4 was found to possess a transmembrane domain. Yeast also possess a single domain PDI with a transmembrane spanning region, but this PDI was shown to be catalytically inactive (Kimura et al., 2005). Therefore *T. brucei* offers a unique opportunity to study a catalytically active transmembrane PDI. To date no knockout studies have been carried out on this class of PDI and currently little is known about transmembrane PDIs. Additionally it would be interesting to see if any of the yet unstudied PDIs interact with TbEro 1 or if they interact with TbQSOX later in the secretory pathway.

This study has conclusively proven that TbPDI-2 and TbEro 1 form the principal oxidative folding pathway within the endoplasmic reticulum of *T. brucei* and this pathway is indispensible for cell growth. Therefore, these two proteins offer new drug targets in the treatment of trypanosomiasis. Additionally this study and the work of Joyce Rubotham have provided the tools for the further *in vitro* analysis of this pathway by the expression and purification of recombinant forms of TbPDI-2 and TbEro 1. While the road to a treatment of trypanosomiasis is doubtlessly long, this study has begun the journey and has conclusively shown that *T. brucei* has a part to play in the wider understanding of the oxidative folding pathway.

7 Appendix

7.1 Genomic sequences of enzymes of the oxidative folding pathway in *Trypanosoma brucei* and the multiple sequence alignment of them using T. Coffee

TbPDI-1 (Tb927.4.2450)

```
ATGAAGCGCC CTTCTTTTCT TCTTGTGAGC CTTTTGTTTT TGGCAATTGG CAGCACTGTC
ATAGAAGCCT TCCCCTTCAC AGCCTCCTCT GGTGTTGTGG AGCTAACCCC TGCGACTTTC
AACAGCTTCC TCGGCTCTCA CAAACCCGTT TTTATCCTTT TCTACGCACC GTGGTGTGGG
CACTGCAAGC GCCTTCACCC GGAGTGGGAA AAGTTTGCCA AATCTGTCGA GGGAATTGTT
CGTGTGGGTG CTGTGAATGC CGATGAGCAC CAGCAACTTG GCCAACAATT CAACCTGCGC
GGTTTCCCTA CGGTTAAGTT TTGGGGTTTG GGAGAAAAGC GTGCCAATGC AGCTATGGAC
TACGCCGGTG AGCGCAGCGC GGGAGCCATT CAATCGCAAG CCATCAGTCT CATCAATGCA
CCAGGAATCA AGACGGTGAA AAAAGCTGAG GAGCTACGCG AGGCGGCTCA GGCGGCACCA
GAGAAGAAGG CTGTTGTACT ATTCAGTGCT AAAAGCCGCG TGCCGCCAAT TTATGCCGTT
CTTTCGCTCT CACCCCGACT CAAGTCGATG CCGTTTTACT TTGCAGGGGA GCAAGAAAAG
TCTGGATTCG CTTCAGAATT TGGCGTAAGT AAACTTCCAG CCATTGTTGT CCTCAACACC
ACCGCTGGCG AAGTGAAGCC GGTGATATAC CCGGGGAAGA AAGTTGCTTA TGAGCCCATT
GCAAAGTTTT TACTGGCCTG TGCAAACGAT GCTTACGATG AGGCTAACTT CTCCTTGGAG
GGTGGTGAGA GTGAGCAAGC TACAAAGAAG GGTGCTGCAA AGTTGGCGCT GCCAGTCCGG
CCGCTCCCCG CGTCAGCTCA TGGTCTCGAG AATTTTTGCT CCCCTGACGC ACGGAAGAAG
GCTGAGCGAA CGCCACTGTG TGTGATTTCA CTCACTTCTG AGTTTTCGCT TAACGACGTG
CATAAAACCT TCAGCAATGA GCCACTTGTT TTCTTCGAGG CTGGGGATGC TCGTGATATC
CTTCTTGGGG CCCTTCGGGG AAAGCTGGGA CTAGAGAAGA TTGCAGACAA ACTGAAGGAC
GACAAGGAAC ACAACGCACT TCTGCTTCGC GCCTCCAAGC AGGGCACCGT ACACTACAGA
CTCGTTGGAG GCGTAAACAC AGCTGAAACT CTCAGCTCAG TCTTACAAAT GGTTGTATCA
GGCGAGATAT CCCTCACGAA GGAGCAGCCG AAGTAA
```

TbPDI-2 (Tb10.6k15.2290/Tb927.10.8230)

ATGCGCGCTA TTTTTCTTGT TGCTCTTGCG CTCGCCACCA TGCGCGAGTC GACTGCGGAA
TCCCTGAAGC TCACCAAAGA GAACTTCAAC GAGACCATCG CAAAGAGTGA AATATTTCTC
GTGAAGTTTT ATGTTGACAC CTGTGGATAC TGCCAGATGC TCGCACCAGA GTGGGAGAAG
GCGGCCAATG AAACTATTGA CAATGCACTT ATGGGCGAAG TCGACTGTCA CTCACAACCT
GAATTGGCGG CTAACTTTTC CATCAGGGGA TACCCAACAA TCATTCTCTT TCGTAATGGC
AAGGAGGCGG AGCACTACGG TGGTGCCCGC ACGAAGGACG ATATTATCAA GTACATTAAA
GCGAACGTTG GACCGGCAGT GACGCCAGCG TCCAATGCTG AAGAGGTGAC AAGGGCAAAG
GAGGAGCATG ATGTAGTGTG CGTTGGCCTG ACAGCCAGCA ATAGCACTTC ACTTTCTACC
ACACTGGCCG AGGCCGCGCA GTCGTTCAGA GTCAGCCTCA AATTCTTTGA AGCAGAACCT
AAGCTTTTCC CCGACGAGAA GCCCGAGACC ATTGTGGTGT ACAGGAAGGG AGGCGAAAAG
GAGGTTTACG ATGGCCCCAT GGAAGTAGAG AAGTTGACAG AATTCCTTCA AATTTCTCGC
GTTGCCTTTG GGGGTGAAAT TACCCCGGAG AACTATCAAT ACTACTCAGT TATAAAACGA
CCCGTAGGAT GGGCTATGGT CAAGCCTAAT GAAACTGCTT CGATCGAGTT GAAAGAAAGC
CTCACAGAAG TGGGTAAAAA AATGCGATCC CATATGGTAG TTCTCTGGGT TAACATTTCG
AAACATCCGG TTTGGAGGGA CTTCGGTGTC CCGGAGGACG CCAAGTACCC CGCCTTTTTG
GCTATCCACT GGGGCGCCAA TTACTTGCAT TCCACAGCGG AGGTTGTGAC CCGCGAGTCC
CTTGAGAAGT TCATCCTGGA GTTTGCTGCA GGTCGTGTCG AGCCCACCAT CAAGTCCCTC
CCTGTCCCCG AGGTGGAGAC TGTTGATGGC AAAACCACCA TCGTGGCAAA AACAATGCAG
AAGCACCTTA CGTCGGGAAA AGACATGCTC ATCCTTTTCT TTGCCCCTTG GTGCGGTCAC
TGCAAGAATT TCGCTCCAAC CTTTGATAAG ATCGCGAAGG AGTTCGATGC AACCGACCTT
ATCGTTGCCG AGCTTGATGC CACAGCCAAT TACGTTAACT CGTCTACTTT TACAGTTACA
GCATTTCCGA CGGTGTTCTT CGTCCCGAAT GGAGGGAAAC CCGTTGTTTT CGAAGGTGAA
CGCTCATTTG AAAACGTCTA CGAATTTGTT CGCAAACATG TGACAACGTT TAAGGTGAGT
GAGAAACCAG CTAATGTGAC AGAGGAGAAG AAAAGCGAGG AAGAGAACAA GAGCAGTAGG
AGCAATGAAA GTAATGATAG CAATGAGAGC AATGTTGATA AACAAGATCT GTAG

TbPDI-3 (Tb927.7.1300)

ATGACCCCGT TTTCATGTCG CAAAACCAGC CTGTTGAAAG CCATTTGCAT CTTTGCTTAT
TTTGCATCTG CAGCGCATGC TAACGAACCG GATGCGGCGC TGGAGGGTGT TGTAGACCTC
ACCAGCAATA ACTTTGACAG TAGTGTGGGA AAGGACGTGG CAGCGCTGGT GGAGTTTTAC
GCGCCATGGT GTGGACACTG CAAAAACCTC GTACCGGAGT TTGCCAAACT CGGCCGCGCC
GCCGCTGGGG CCAAGGATAA AGTTCTCATT GCTAAAGTGG ATGCCACAGC CCAAAAGGAC
CTCGCAACTC GCTTTGAGGT GAATGGGTAC CCAACCATTC TCTTCTTTCC TGCTGGCTCG
CAGAAGCCGG AGAAATACAG CGAGGGTCGC GAGGCGAAGG CGTTTGTTTC GTACCTTAAC
AACCAAATAA AGGGCTTGAA CCTTTTTCTT CCCCGTGAAC ACAAGTACGT GATGGCGCTG
GACCAATCCA ACTTTGACAA AGTGGCACTG GACGAAGGGA AAGATGCGTT TGTTTTATTT
TACGCCCCCT GGTGTGGACA CTGCAAGCGC CTCCATCCCT CGTTTGAGAG TCTCGCGAAG
GTGTACCAAA ACGAAAAGGA CCTCATCATT GCTAATGTCG ACGCTGATGA CAAGAGCAAC
TCTGAGGTGA CGAAGCGCTA TAAGGTGGAG GGCTACCCCA CTCTCGTTTT CTTTCCAAAA
GGAAATAAGG GCAACCCGGT GAACTATGAG GAGGGACGAA CCTTGGACGA TATGATCAAG
TTCGTCAATG AGAGAACGGG GAAAAAACGC ACCAGCAGCG GTGATTTTGA CAAAACTGTT
GGTGTCGATG AAACCGTCAC GAACCTCATG AAGGAGATGG CACAGTCAAG CAAGAACAAG
GAGGAGAGGG AACGCCTCCT CGCGCAGATT CAGCAAGCCG TCTCCCCCAA AGCTCTAGGC
GAGGGTGCCA TGCACTACAT CCGTATTGCT ACTAATGTCC TTGAGAATGG TCATGAGTAT
ATTAGCAAGG AACACGAGCG TGTCGGCCGT CTTCTGAAGG GAAGTCTGAC AGGTCCGAAG
AGGGATTCGT TGACGATCCG ATTCAACATT CTTTCGGCAA TCAAAGGTGA GTAA

TbPDI-4 (Tb927.5.1020)

```
ATGCAACGAA CGAAAAGTTT GCTTCATGTC CTGCTTGTGG TAGCACTAGT GCTCTTTTTA
TTGATTACTG AGGTTTCGGT TGCTGGAGGG GTTGTTGACT TGGGAAATGA GGTACGGGAG
GACACCATAC CGCTACTTAA TGACGACACA TTCGACTCGT ATGTTTTTTC AAAGGGATCT
AATGATAAAG AGGGGCGAAA GAGAGGACCT TGGCTCATAT TCTTTTTTGC TCCATGGTGT
GGTCACTGCA AGGCCGCCTT GCCCAAGTAT GCGGATGCTA ATTTGCAGTT GGCGAAACTT
GGAGTTCAGC ACGCGAGGTT TGCAACCGTC AATGCTGTAA AGAGTCCCGA ATTGGCGTTG
CGGTTCCGCG TAAAGGTATA CCCGACGCTC ATATATACCA CTGGCAAGGC AGAAAACTGG
CATGTGTACC GTGGTGCCCT CACTCCCGAG ACTTTAATGA AGTTTGCGAC CCAACTTCAT
TTAGCCGGCA CTGTTGGGTC ATTTGCCGAT TTGACATCAA GCCCCGATGA ATTCGTTAGG
GCGCATCGGC AGGATATGGG AAGGCGCGTG CCCATGTATG TTTATCTTCC TCCAACAGGT
TCTCGAAGTG CCAATGCACC GCACCAGCGA GAAGAACTTG TTCCAGGGCA ACATGGAGTG
GGTGAAGGGG AAGGGCAACA GGGAGTGCAA GCGAATAATA GTAAGTCCAT CCATGAGGAG
AAGCACATGC AGAAACAAGA TCGGCTTTCT GACGCGCGCG AGGCGCGGTG GAATGTTGCC
GTGGACGCTG CTGCTTCGAT GGAAAAGATT CGGTTTGGTG TGATATATGG TGATGACGTG
CCATCTGCTT GGGCAGAGAC GGGGATAACG CCATATATTG CCGTTCTCGA GGCCGCACAG
CGGTGTAAAG GCAGTGGGCC CGATGGCGCA CTTGTGCTTG TTGATTCTGA CGCATACAAA
CAACCACAAT GTTACGAAGG TCCGTGGTTT GAGGAGTCCG TCAATATCCC TCAAAAGAGC
GGGGCGGAGA GCGTGACCCC GCTGCTTCAC CCATCCTTCA TAGCCTTCTT CATGTTAAAC
GGCCTTCATG CAGTGGAACG AGCTTCGTCA TCGCTTCTAT CAGCCGTTAC CACCACGAAC
CATCACTATT TGGGCCTCCT TATCACGAAT GGCCCTATTA CACAAACTGA CACTAATATG
CTTCCAGCTC TGCGTGAGGT GGTGCAAGAA CGAAACGAAG CTCGCCGTGC GGTCGCTGTG
GGGCTTGAAC CTCAGCGGAG GGTCTCGTTG GCGTACGTTG ATGGTCGCAC TCACGAGGCG
TGGCGTACGC AATACAAAGT GAAGCTGGAA GAACTCCCCG CCTTCGTGGT TGTGGACCCG
CGGCGGAACA AGGTATACCG CTTGCGCAGG CACACTCCTC ACTTCGAGGT AATTAAATTC
TTCTTGCCGT GGCCGCGTGG GGGCGAGCAA CAGTCAGTCA TCGTTTCCTT CATTGCTGAT
GTCGAACGTG GGGTCGCGGT TGGTGAAAAG ATGACATTTG TCGGGGAGGT TGCAGAGCAT
CTGCTGCGAC TTCCGGGAGT GGAGTTTCTC TACGAGGTCC TCGAGTTCGA AGACAGTCTA
CTTGTGGTTG TTGTGCTCGC CCTCTTTGTG TTTGCGATCG TACTCCTCAT TGTACTCGTC
GTAGAGCCGG TGTTTGAGCG GCGTGCGGAT GAAAGGATGA CGGGGTACGG GGAAAGGGAC
ACTTCTGCAG TTGAAAAGAA GAGGAACTAA
```

TbPDI-5 (Tb927.7.5790)

```
ATGAGGAACT CATTTGCACT TCTGCTTTTA TCTGTTGCCA TCGCTTTTGT TACTGTTGGC
TCATTTGCTG ATGAGGCGAA GGATTCCGTG GAGTTAACTC CTGATAATTT TGATAAGGTG
GCGTTGGATA CGGAGAAACA TGTGTTTGTA ATGTTTTATG CGCCATGGTG TGGACATTGC
AAACGTCTTA AACCCAAATG GGAGGAACTC GCTAAAGAAA TGAAAGATGA GACATCGGTG
GTTATTGCTC GATTGGATGC TGACAAGCAC CGCAATGTGG CGGAGCGTTT TGATGTTCGC
GGGTACCCGA CACTTCTGCT TTTTGCTCGA AGTAAAAAAG AAGGCCTCCG CTACGAAGGA
GCCCGTGATG TGGCAGCACT GAAGGAATTT GTTAAAAGCA ACATG
```

TbPDI-6 (Tb927.7.5780)

```
ATGATATCAC TACATTTGTC CTCGCACTCA CTAATGCATG TGCTACTTTC ATTCCTGCTG
TTATGTTCTG GAGTTGTTAT GGGATATTAT GCAGATGTGG ATTATGAAGT GCCATCACGT
GTTGTGCAGC TCACTGATGC GACATTTGAC ACAATTGTGA TGGATCCCAC AAAAGATGTA
TTTGTTTTGT ATTGCGTGAC GTGGTCCCGT CATTGCCGCT CTGCTGTAAA ATTATGGGCG
GATCTTTCCA TAAGTCAATT TAAACGACCC ACCGCTGATA CATTTGTGGC TGCATTAATT
GACGCGGAGG CATTTCCCGA AACGGCGAAG CGTATGGGCG TAAAAAGTTA CCCAACAATG
TTATTTTATA CCCGTTTAGA AAAGGCAACA CCACTGGAGT ATAATGGACA GGGCCTCCTT
CCATTACTGG ACAGTTTTGT CTTTCAAAAC TCATAA
```

TbEro 1 (Tb927.8.4890)

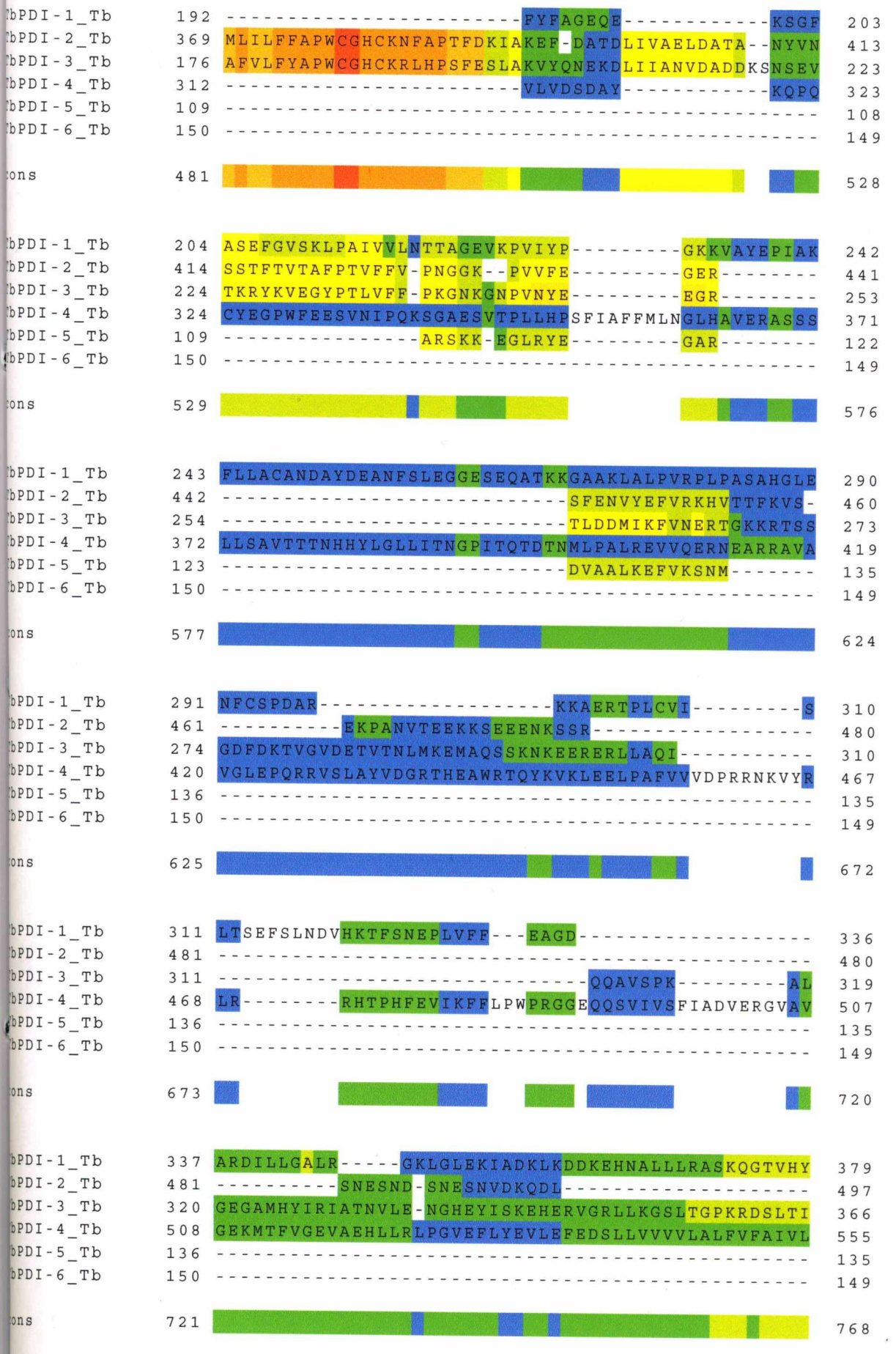
GTTAATCGTA CTATTCACAT CGGGAGATGT TGAAGATGCG ATTATTGATA GTAGTTCCTG
TTCTTCTAGG ACTTGTTTGG CAAATTTTAC TTCGTGCTGA ATTGGATGGT GTTTCTTTTT
TTGGGATGTA CATTTCAGCC AACAATTCCG GGGCTGGCTC ATATGTAAGG ACCAAGAAAG
GAAATGCCCT TAAGGAGGGG TTTTGTTCCC TTACCATGGA TGAGGTCAGT CAGAACACTG
AGGGTATTAC AGGCCTTCTT AACAACATAA CGAGTCACCC ATACTTTCGG TACTTTAAAG
TAAATCTGGA CAGGGAATGT CGTTACTGGG TCGCTGAGGC TAGTTGTACA TGTGATTCAA
ATGGATGCCA GATCTGCACT TGTGATGACA GTGGAATTCC CGAAACACTA AAATACCCAT
ACGATATGAG TGATGTTTCC ACGGTTGAAC GTCGTGCGGC CCCTGACAAA CATGCAGCAA
AAGGTTTCGA GGACGAGATC AAGCCGATCG ACCCAGATAG AGATGCAACA TATGTGGATC
TTCTCCAAAA CCCAGAGGCA AACACCGGAT ATTCGGGACC GAAAGCAGCG CGGGTTTGGC
AGGCTGTATA TAACAACTGC AATATTGACG GACTTCCAAG TAATGATACG GCTGGTGTAG
AAAACAGAGA GAAGGCTTTA CTTCGGCAGC TTCTAAGTGG GCTCCATACT TCTATAACCA
TGCATGTGGC TGCTTTCTTT TATAATGATA CGAAGGGTGA CTCTCCGCTG CGCAGCCTAG
GTGTGCTAAA TAATCCCAAT ATATCTTTCT ATCCTAACTG TGGCATGTTC CGCCGCATCG
TTAAGAATGA CGAATTCATA AGGAACCTTT TTGTTGTCTA CCAGTTTGTC CTTCGCGCTC
TGGCGAAGAC CAAACGTGCA TTTCTTGCTA ACTCAAGTTT GTATAACTCT GGCTTCAACG
GCGCTGCTAC AGATGGAGAT GTGCGTCTCT ACAGCAATAT TGGGGAACTT TTTAGCAGCA
AGCTATTCCG CGTAGCAACT TTCGACGAGC AAAAATTTTT TGAGTCTCCG GGTGCGCACC
TACTGGTTCG GCAAATGAAG CGCGTGGTGC ACAATGTTAC AACTCTCATG GATTGCGTCA
CGTGTGAAAA GTGCCGCGCT TGGGGGAAGT TGGAAACAGC TGCTCTTGCC ACCGCGTTGA
AAATTGTCTT CGGTTCCGCT GAGAATGTGG TGGAACTAAA CAGAGGAGAG AGAGTTGCCC
TTATAAATTT TGCGCGGCAA CTTGCCATTT CTGTTAAAAA TGTTCGTTCA CTAGCTGCTG
TGTGCGAAAA ATTCAATTAC ACGGCTATGA GTTGA

Figure 7.1 Multiple sequence alignment of candidate *trypanosome brucei* PDIs.

A multiple sequence alignment was carried out using TCoffee a multiple sequence alignment (Dereeper et al., 2008, 2010; Notredame et al., 2000). This gave the best alignment of all six protein sequences although as can be seen little sequence similarity exists between the candidate PDIs. The sequences are coloured using the ClustalW colour scheme (Eddy, 1995).

```
-COFFEE, Version_6.85Tue Sep 9 14:03:25 WEST 2008
 Cedric Notredame
 CPU TIME:1 sec.
 SCORE=34
 *
  BAD AVG GOOD
 *
 TbPDI-1_Tb    :   33
 TbPDI-2_Tb    :   34
 TbPDI-3_Tb    :   34
 TbPDI-4_Tb    :   27
 TbPDI-5_Tb    :   46
 TbPDI-6_Tb    :   44
 cons          :   34

 TbPDI-1_Tb        1 ------MKRPSF----LLVSLLFLAIGST-V-----IEAFPFTASSGV    32
 TbPDI-2_Tb        1 ------M-R---AIFLVA--LALAT--------------MRESTAES-    21
 TbPDI-3_Tb        1 -MTPFSCRKTS----LLKAICIFAYFASAA-HAN-----EPDAALEGV    37
 TbPDI-4_Tb        1 MQRTKSLL-HVL----LVVALVLFLLITEVSVAGGVVDLGNEVREDTI    43
 TbPDI-5_Tb        1 -------MRNSFALLLLSVAIAFVTVGS-----------FADEAKDS-    29
 TbPDI-6_Tb        1 -MISLHLSSHSL----MHVLLSFLLLCSGV-VMGYYADVDYEV-PSRV    41

 cons              1                                                   48


 TbPDI-1_Tb       33 VELTPATFNSF-L---------GSHKPVFILFYAPWCGHCKRLHPEWE    70
 TbPDI-2_Tb       22 LKLTKENFNET-I---------AKSEIFLVKFYVDTCGYCQMLAPEWE    59
 TbPDI-3_Tb       38 VDLTSNNFDSS-V---------GKDVAALVEFYAPWCGHCKNLVPEFA    75
 TbPDI-4_Tb       44 PLLNDDTFDSYVFSKGSNDKEGRKRGPWLIFFFAPWCGHCKAALPKYA    91
 TbPDI-5_Tb       30 VELTPDNFDKVAL---------DTEKHVFVMFYAPWCGHCKRLKPKWE    68
 TbPDI-6_Tb       42 VQLTDATFDTIVMDP--------TKDVFVLYCVTWSRHCRSAVKLWA    80

 cons             49                                                   96


 TbPDI-1_Tb       71 KFA-----KSVEGIVRVGAVNADEHQQLGQQFNLRGFPTVKFWGLGEK   113
 TbPDI-2_Tb       60 KAAN-E----TIDNALMGEVDCHSQPELAANFSIRGYPTIILFRNG-K   101
 TbPDI-3_Tb       76 KLGR-AA-AGAKDKVLIAKVDATAQKDLATRFEVNGYPTILFFPAGSQ   121
 TbPDI-4_Tb       92 DANL-QLAKLGVQHARFATVNAVKSPELALRFRVKVYPTLIYTTGKAE   138
 TbPDI-5_Tb       69 ELAK-EM-KD-ETSVVIARLDADKHRNVAERFDVRGYPTLLLF-----   108
 TbPDI-6_Tb       81 DLSISQFKRPTADTFVAALIDAEAFPETAKRMGVKSYPTMLFYTRLEK   128

 cons             97                                                  144


 TbPDI-1_Tb      114 RANAAMDYAGERSAGAIQSQA-----ISLIN-APGIKTVKKAEELREA   155
 TbPDI-2_Tb      102 E---AEHYGGARTKDDII-KYIKA-----NV-GPAVTPASNAEEVTRA   139
 TbPDI-3_Tb      122 K---PEKYSEGREAKAFV-SYLNN-----QIKGLNLFL----------   150
 TbPDI-4_Tb      139 -N--WHVYRGALTPETLM-KFATQLHLAGTV-GSFADLTSSPDEFVRA   181
 TbPDI-5_Tb      109 ------------------------------------------------   108
 TbPDI-6_Tb      129 AT--PLEYNGQGLLPLLD-SFVFQ----------------------   149

 cons            145                                                  192
```

TbPDI-1_Tb 156 AQAAPEKKAVVLFS 169
TbPDI-2_Tb 140 KE---EHDVVCV GL 150
TbPDI-3_Tb 151 150
TbPDI-4_Tb 182 HRQDMGRRVPMYVYLPPTGSRSANAPHQREELVPGQHGVGEGEGQQGV 229
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
cons 193 240
TbPDI-1_Tb 170 169
TbPDI-2_Tb 151 TASNSTSL STTLAEAAQSFRVSLKFF 176
TbPDI-3_Tb 151 PREHKY 156
TbPDI-4_Tb 230 QANNSKSIHEEKHMQKQDRLSDAREARWNVAVDAAASMEKIRFGVIYG 277
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
cons 241 288
TbPDI-1_Tb 170 -------AKSRVPPI 177
TbPDI-2_Tb 177 EAEPKLFPDEKPETIVVYRKGGEKEVYDGPMEVEKLTEFLQISRVAFG 224
TbPDI-3_Tb 157 VM 158
TbPDI-4_Tb 278 DDVPSAWAETGITPY 292
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
cons 289 336
TbPDI-1_Tb 178 177
TbPDI-2_Tb 225 GEITPENYQYYSVIKRPVGWAMVKPNETASIELKESLTEVGKKMRSHM 272
TbPDI-3_Tb 159 -ALDQSNFDKV A 169
TbPDI-4_Tb 293 292
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
cons 337 384
TbPDI-1_Tb 178 177
TbPDI-2_Tb 273 VVLWVNISKHPVWRDFGVPEDAKYPAFLAIHWGANYLHSTAEVVTRES 320
TbPDI-3_Tb 170 169
TbPDI-4_Tb 293 292
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
cons 385 432
TbPDI-1_Tb 178 -------YAVLSLSPRLKSMP 191
TbPDI-2_Tb 321 LEKFILEFAAGRVEPTIKSLPVPEVETVDGKTTIVAKTMQKHLTSGKD 368
TbPDI-3_Tb 170 LDEGKD 175
TbPDI-4_Tb 293 -------IAVLEAAQRCKGSG PDGAL 311
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
cons 433 480


TbPDI-1_Tb 192 203
TbPDI-2_Tb 369 413
TbPDI-3_Tb 176 223
TbPDI-4_Tb 312 323
TbPDI-5_Tb 109 108
TbPDI-6_Tb 150 149
ons 481 528
TbPDI-1_Tb 204 242
TbPDI-2_Tb 414 441
TbPDI-3_Tb 224 253
TbPDI-4_Tb 324 371
TbPDI-5_Tb 109 122
TbPDI-6_Tb 150 149
ons 529 576
TbPDI-1_Tb 243 290
TbPDI-2_Tb 442 460
TbPDI-3_Tb 254 273
TbPDI-4_Tb 372 419
TbPDI-5_Tb 123 135
TbPDI-6_Tb 150 149
ons 577 624
TbPDI-1_Tb 291 310
TbPDI-2_Tb 461 480
TbPDI-3_Tb 274 310
TbPDI-4_Tb 420 467
TbPDI-5_Tb 136 135
TbPDI-6_Tb 150 149
ons 625 672
TbPDI-1_Tb 311 336
TbPDI-2_Tb 481 480
TbPDI-3_Tb 311 319
TbPDI-4_Tb 468 507
TbPDI-5_Tb 136 135
TbPDI-6_Tb 150 149
ons 673 720
TbPDI-1_Tb 337 379
TbPDI-2_Tb 481 497
TbPDI-3_Tb 320 366
TbPDI-4_Tb 508 555
TbPDI-5_Tb 136 135
TbPDI-6_Tb 150 149
ons 721 768

TbPDI-1_Tb 380 RLVGGVNTAETLSSVLQMVVSGEISLTKEQP--K 411
TbPDI-2_Tb 498 ---------- 497
TbPDI-3_Tb 367 RF----------NILSAIKGE---------- 377
TbPDI-4_Tb 556 LIVLVVEPVFERRADERMTGYGERDTSAVEKKRN 589
TbPDI-5_Tb 136 ---------- 135
TbPDI-6_Tb 150 ----------NS 151
cons 769 802

7.2 Analysis of hydrophobic regions of the candidate TbPDIs and TbEro 1 of *Trypanosoma brucei*

7.2.1 TbPDI-1 (Tb927.4.2450)

TMpred predicted 1 strong transmembrane region located between residue 6 to 24 (length of 19 residues) with a significant score of 2234. A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10 residues (Kyte and Doolittle 1982). One strong hit was found in a similar region as before at position 4 to 24 with a score of 1.894. A third program, TMHMM was run to confirm this finding. Again a strong hit was found in a similar region at position 7 to 29. With the position of this hydrophobic region at the *N*-terminus of the protein it was most likely to be an *N*-terminal leader sequence, a peptide signal sequence which allows an mRNA transcript to be directly translated into the endoplasmic reticulum via the SRP (see section 1.2.2). To confirm this finding the program SignalP was employed which is also located on the same website. The program identified the region as a signal peptide with a probability of 0.999 with a maxcleavage site probability of 0.925 between position 23 and 24. With no other hydrophobic regions predicted by the three programs it was decided TbPDI-1 contained no other transmembrane regions or *C*-terminal extension. Graphs of the various plots can be seen in Figure Figure 4-1.

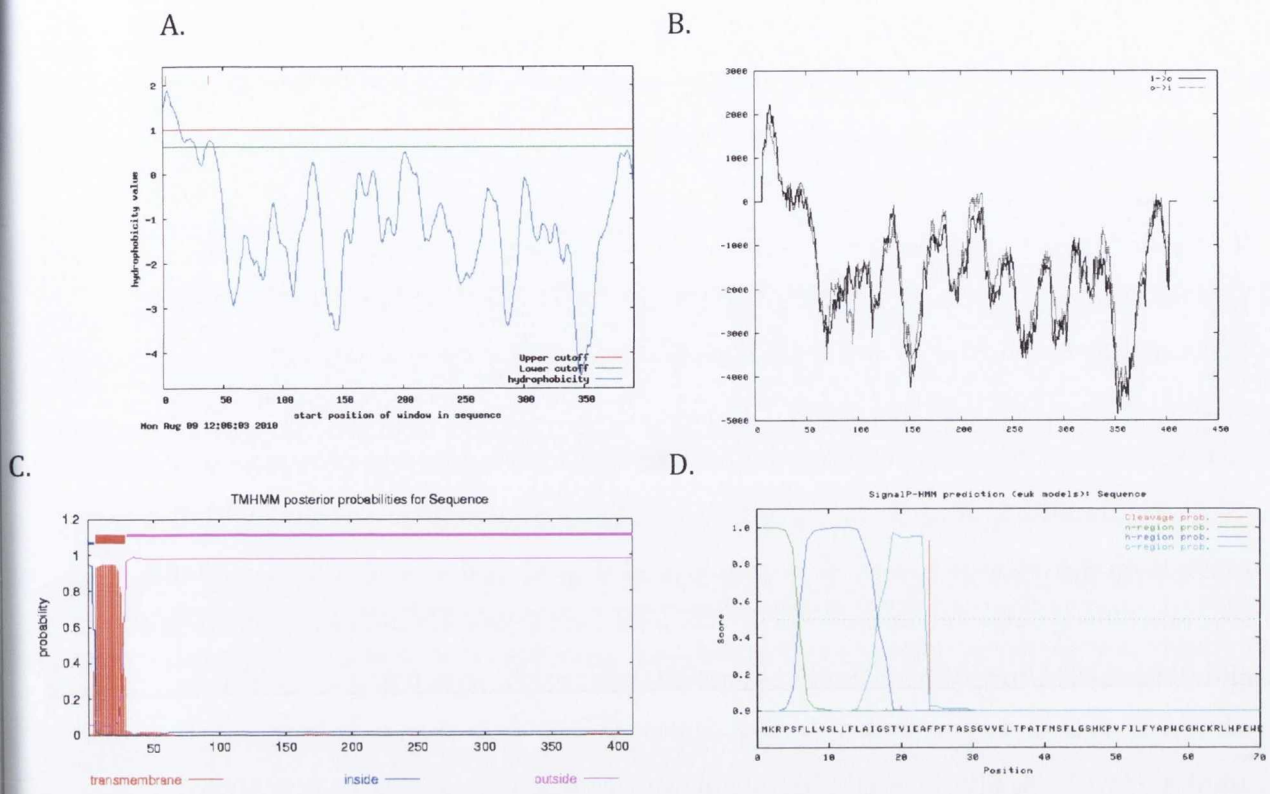


Figure 7.2 Analysis of hydrophobic regions in TbPDI-1.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with an obvious peak at the N-terminus of the protein.

The program TMpred was employed as before and the resulting graph is shown. Again a obvious peak can be seen at the N-terminus of the protein.

The program TMHMM was employed as before and the resulting graph is shown. This third program confirms the other two findings of a hydrophobic region at the N-terminus of the proteins. Due to the position of the hydrophobic region, it was likely this region corresponded to an N-terminal signal peptide.

The program SignalP was employed to predict if the hydrophobic region seen is indeed a possible Signal peptide. The resulting graph is shown. This result confirmed that TbPDI-1 contains a signal peptide between position 6 to 24 with a likelyhood of 0.999 as taken from the program SignalP.

7.2.2 TbPDI-2 (Tb927.10.8230)

TMpred predicted no strong transmembrane regions but only possible hydrophobic regions. The first was located between residue 3 to 19 (length of 16 residues) with a score of 426. Even though this is a very low score it is very likely that this may be an *N*-terminal leader sequence as was found to be the case with TbPDI-1. The other hydrophobic regions found were at positions 145 to 166, 366 to 384 and 411 to 432. Their scores were 186, 521 and 1006 respectively. The last two regions are considered significant as their score was over 500. A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10 residues (Kyte and Doolittle 1982). 3 hits were also found in similar regions as found with the previous program. These were located at position 1 to 21 with a score of 1.107 which was considered certain, position 143 to 163 with a score of 0.6 which was considered putative and the final hit was found at position 412 to 432 with a score of 1.012 which was considered certain. A third program, TMHMM was run to confirm these findings but the results showed no transmembrane regions. As the two previously identified possible transmembrane regions positioned at 143 to 163 and 412 to 432 which corresponded to the PROTEIN DISULFIDE ISOMERASE domain hit previously found, it was no surprise the region did not find a transmembrane region but it was surprising to see no hit for a signal peptide. A search for a signal peptide was carried out in spite of this finding using the program SignalP. A very strong hit was found at position 1 to 19 with a signal peptide probability of 0.999 and with a max cleavage site at position 19 to 20 with a probability of 0.808.

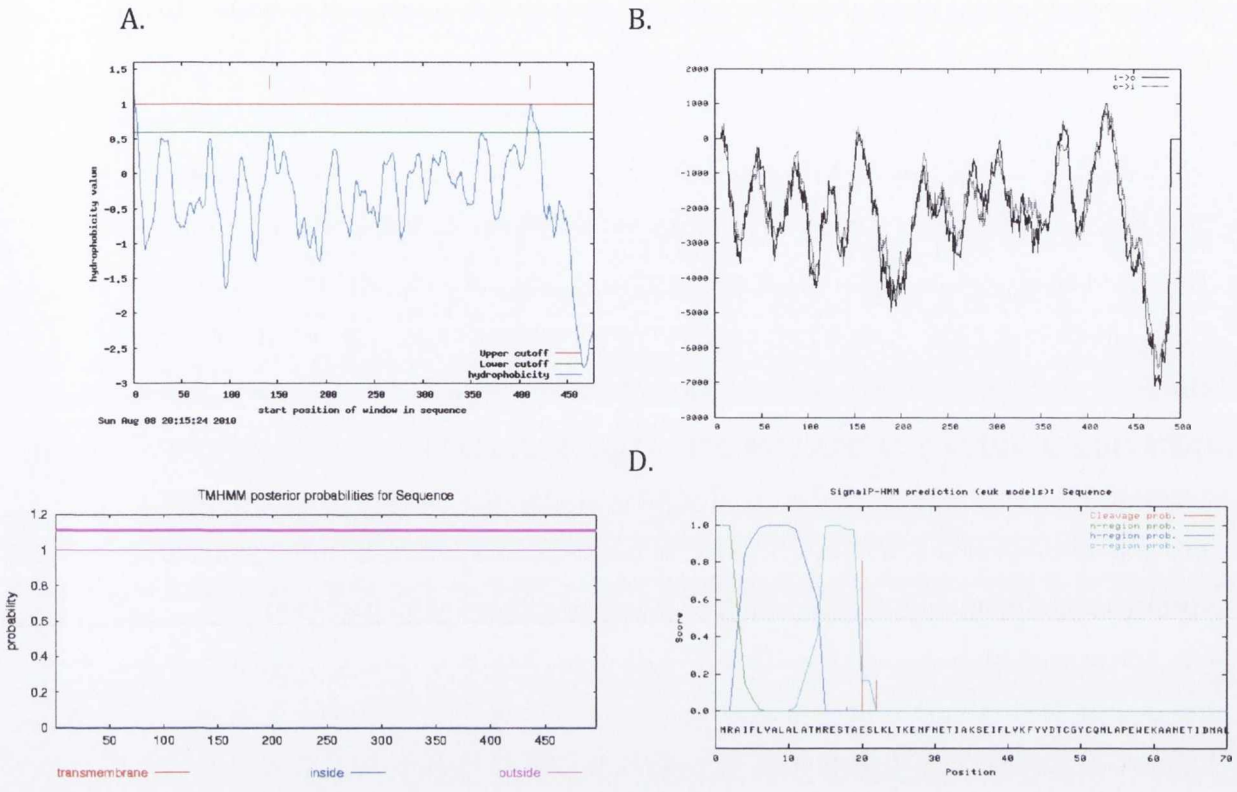


Figure 7.3 Analysis of hydrophobic regions in TbPDI-2.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with three small peaks marked by red vertical lines.
The program TMpred was employed as before and the resulting graph is shown. Again the resulting graph is similar to the previous resulting graph but with less prominent peaks.
The program TMHMM was employed as before and the resulting graph is shown. This third program disagrees with the two previous findings and so it appears the protein contains no transmembrane regions which was the expected outcome due to previous domain analysis.
The program SignalP was employed to predict if the protein contained a signal peptide in spite of the previous findings that the hydrophobic *N*-terminal peak was not significant. The resulting graph is shown. A very strong hit was found at position 1 to 19 with a signal peptide probability of 0.999 and with a max cleavage site at position 19 to 20 with a probability of 0.808.

7.2.3 TbPDI-3 (Tb927.7.1300)

TMpred predicted 1 strong transmembrane region located between residue 9 to 26 (length of 17 residues) with a score of 655 which is considered significant. A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10residues (Kyte and Doolittle, 1982). One strong hit was also found in a similar region as before at position 8 to 28 with a score of 1.252 which is considered certain. A third program, TMHMM was run to confirm this finding but no significant hits were found. Again in spite of this fact the protein was subjected to analysis with the program SignalP. The program identified the *N*-terminal region as a signal peptide with a probability of 0.954 with a maxcleavage site probability of 0.925 between position 27 and 28. With no other hydrophobic regions predicted by the three programs it was decided TbPDI-3 contained no other transmembrane regions or *C*-terminal extension. Graphs of the various plots can be seen in Figure 4-3.

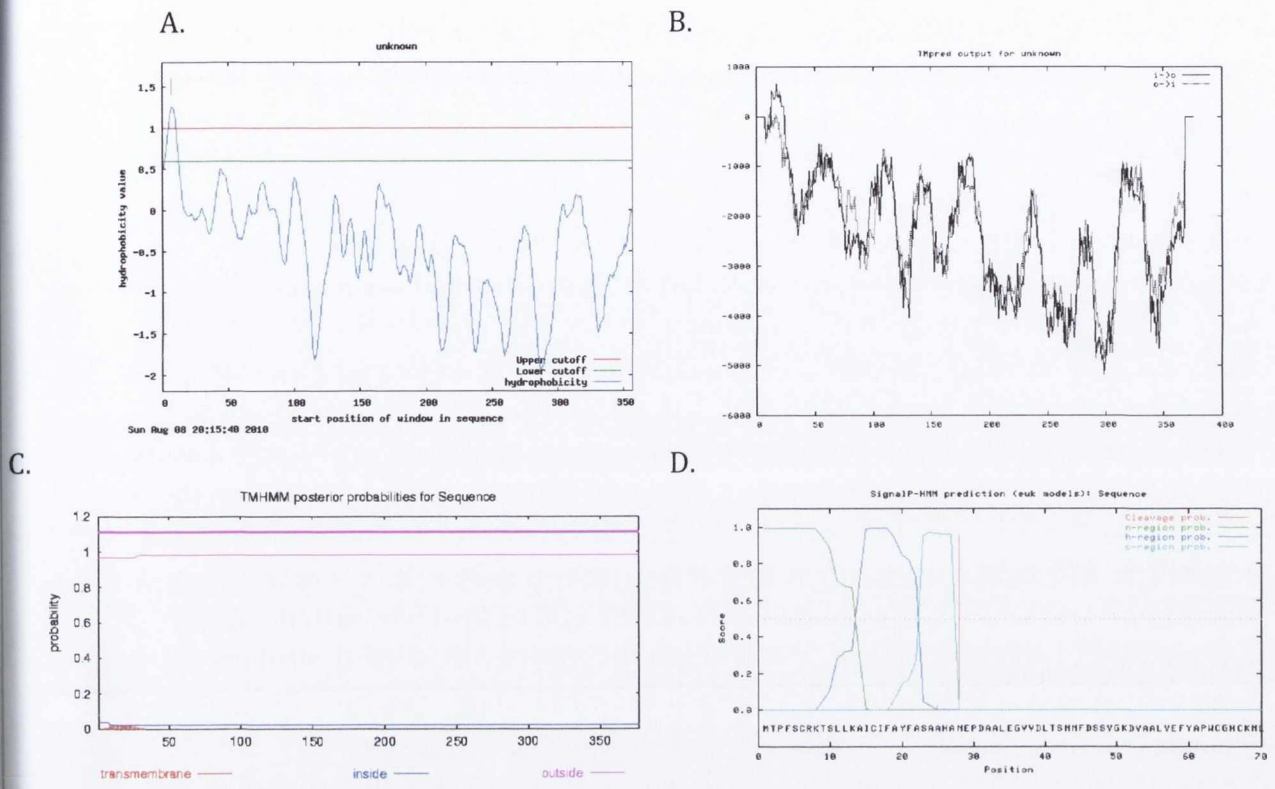


Figure 7.4 Analysis of hydrophobic regions in TbPDI-3.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with an obvious peak at the N-terminus of the protein.

The program TMpred was employed as before and the resulting graph is shown. Again an obvious peak can be seen at the N-terminus of the protein.

The program TMHMM was employed as before and the resulting graph is shown. This third program confirms the other two findings of a hydrophobic region at the N-terminus of the proteins. Due to the position of the hydrophobic region, it was likely this region corresponded to an N-terminal signal peptide.

The program SignalP was employed to predict if the hydrophobic region seen is indeed a possible Signal peptide. The resulting graph is shown. The program predicted that TbPDI-3 contains a signal peptide with a probability of 0.999.

7.2.4 TbPDI-4 (Tb927.5.1020)

TMpred predicted 4 strong hits as possible transmembrane regions located between residue 7 to 25 with a score of 2211, position 68 to 88 with a score of 769, position 347 to 366 with a score of 922 and finally position 543 to 561 with a score of 3243. The first and last hits contain significantly higher scores and as such are likely to be transmembrane regions (a score of over 500 is considered significant). A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10 residues (Kyte and Doolittle, 1982). The results agreed with three of the previously found hydrophobic regions. The following scores were retrieved: position 6 to 26 with a score 2.950 which is considered certain, position 344 to 364 with a score of 1.223 which is considered certain and position 541 to 561 with a score of 3.643 again considered certain. A third program, TMHMM was run to confirm these findings. Two of the previous hydrophobic regions were found during this analysis at position 7 to 29 and 539 to 561. The first of the two regions was most likely to be a signal peptide as was found to be the case with the other PDIs investigated but the second transmembrane region is very unusual. The consensus of the three different programs suggests the second transmembrane region is significant even though it is unexpected. The program SignalP was used to identify if the *N*-terminal transmembrane region was an *N*-terminal signal peptide. The program identified the region as a signal peptide with a probability of 0.946 with a maxcleavage site probability of 0.361 between position 29 and 30. It was concluded that TbPDI-4 contained a signal peptide as well as a transmembrane region. Graphs of the various plots can be seen in Figure 4-4.

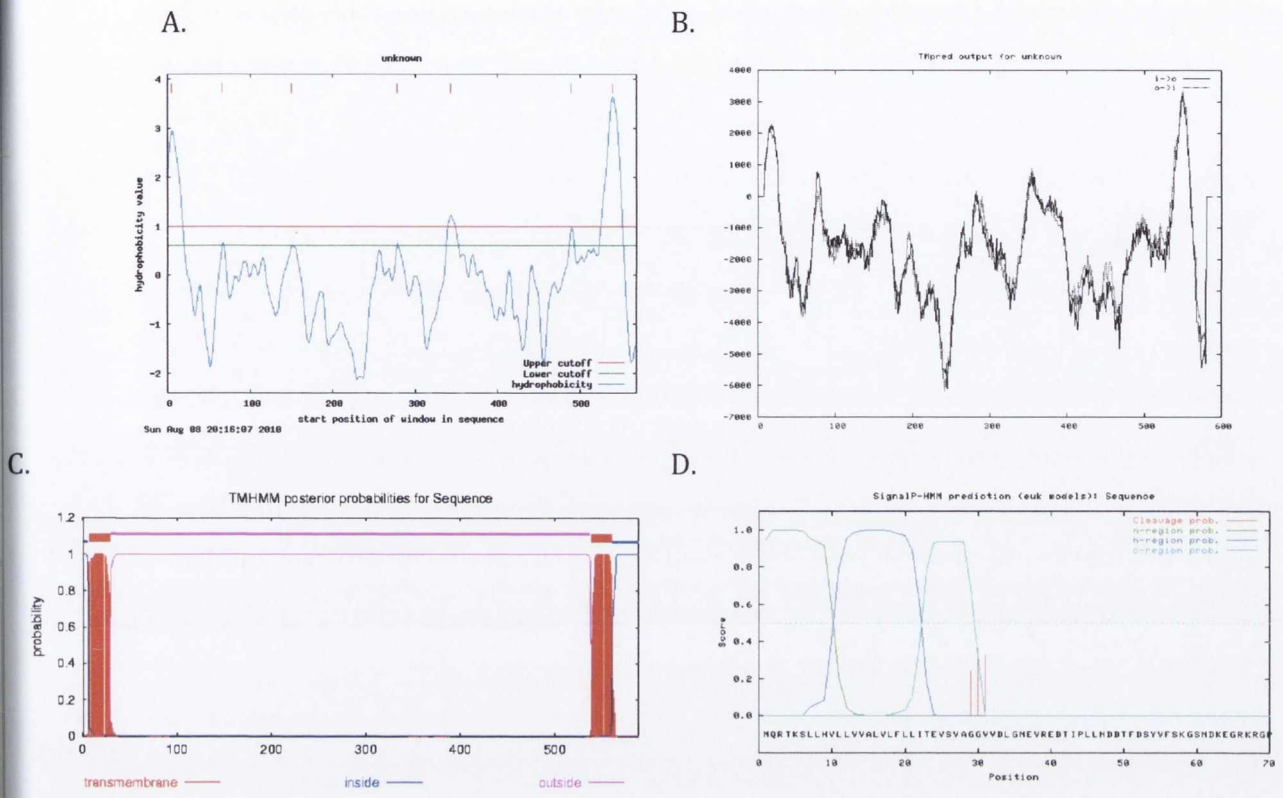


Figure 7.5 Analysis of hydrophobic regions in TbPDI-4.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with two prominent peaks at the *N* and *C*-terminus of the protein.
The program TMpred was employed and the resulting graph is shown. Again two obvious peaks can be seen at the *N* and *C*-terminus of the protein.
The program TMHMM was employed as before and the resulting graph is shown. This third program confirms the other two findings of the first two programs.
The program SignalP was employed to predict if the hydrophobic region seen is indeed a possible Signal peptide. The resulting graph is shown. This result confirmed, and the previous hydrophobic plot results indicate that it is very likely that TbPDI-1 contained a signal peptide between position 6 to 24 with a likelyhood of 0.999 as taken from the program SignalP.

7.2.5 TbPDI-5 (Tb927.7.5790)

TMpred predicted 1 strong transmembrane region located between residue 6 to 24 (length of 19 residues) with a score of 2234. A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10 residues (Kyte and Doolittle, 1982). One strong hit was also found in a similar region as before at position 4 to 24 with a score of 1.894. A third program, TMHMM was run to confirm this finding. Again a strong hit was found in a similar region at position 7 to 29. With the position of this hydrophobic region at the N-terminus of the protein it was most likely to be an *N*-terminal leader sequence. To confirm this finding the program SignalP also located on the same site. The program identified the region as a signal peptide with a probability of 0.999 with a maxcleavage site probability of 0.925 between position 23 and 24. With no other hydrophobic regions predicted by the three programs it was decided TbPDI-1 contained no other transmembrane regions or *C*-terminal extension. Graphs of the various plots can be seen in Figure 4-5.

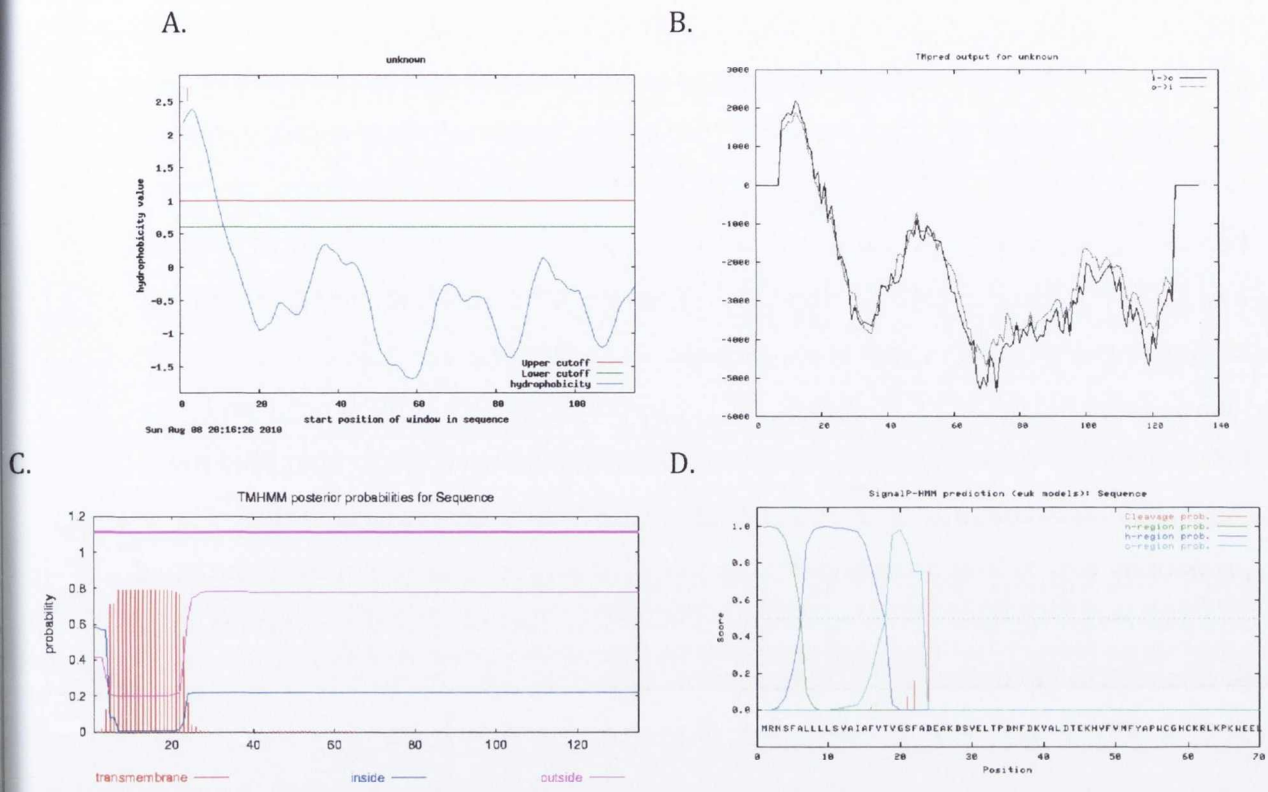


Figure 7.6 Analysis of hydrophobic regions in TbPDI-5.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with an obvious peak at the *N*-terminus of the protein.

The program TMpred was employed as before and the resulting graph is shown. Again an obvious peak can be seen at the *N*-terminus of the protein.

The program TMHMM was employed as before and the resulting graph is shown. This third program confirms the other two findings of a hydrophobic region at the N-terminus of the proteins. Due to the position of the hydrophobic region, it was likely this region corresponded to an N-terminal signal peptide.

The program SignalP was employed to predict if the hydrophobic region seen is indeed a possible Signal peptide. The resulting graph is shown. This result confirmed the previous hydrophobic plot results. The program predicted an *N*-terminal signal peptide with a probability of 0.996.

7.2.6 TbPDI-6 (Tb927.7.5780)

TMpred predicted 1 strong transmembrane region located between residue 6 to 24 (length of 19 residues) with a score of 2234. A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10 residues (Kyte and Doolittle, 1982). One strong hit was also found in a similar region as before at position 4 to 24 with a score of 1.894. TMHMM was run to confirm this finding. Again a strong hit was found in a similar region at position 7 to 29. With the position of this hydrophobic region at the *N*-terminus of the protein it was most likely to be an *N*-terminal leader sequence. The program SignalP was employed which identified the region as a signal peptide with a probability of 0.999 with a maxcleavage site probability of 0.925 between position 23 and 24. It was concluded TbPDI-1 contained no other transmembrane regions or *C*-terminal extension. Graphs of the various plots can be seen in Figure 4-6.

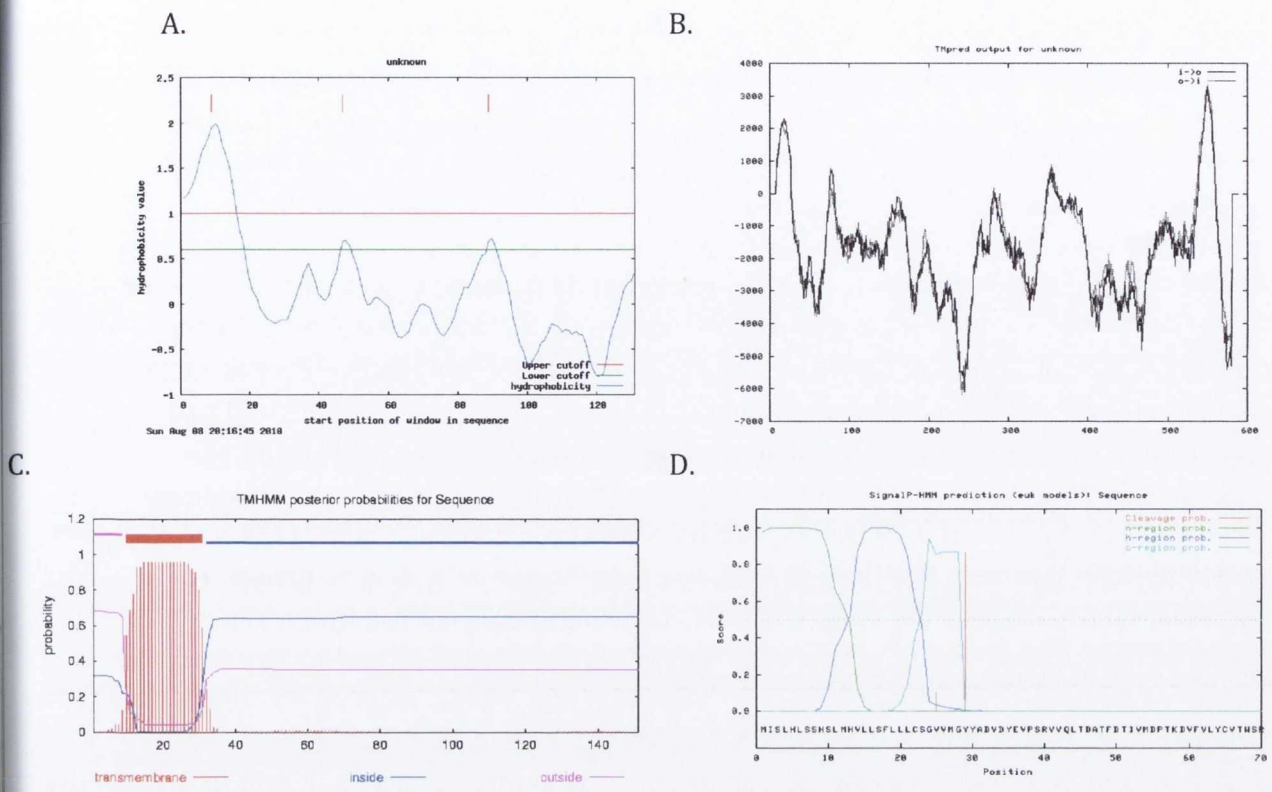


Figure 7.7 Analysis of hydrophobic regions in TbPDI-6.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with an obvious peak at the N-terminus of the protein.

The program TMpred was employed as before and the resulting graph is shown. Again an obvious peak can be seen at the N-terminus of the protein.

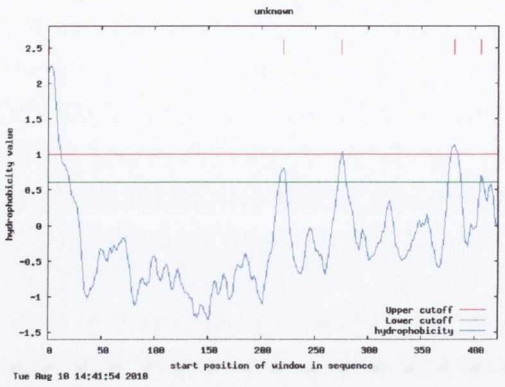
The program TMHMM was employed as before and the resulting graph is shown. This third program confirms the other two findings of a hydrophobic region at the N-terminus of the proteins. Due to the position of the hydrophobic region, it was likely this region corresponded to an N-terminal signal peptide.

The program SignalP was employed to predict if the hydrophobic region seen is indeed a possible Signal peptide. The resulting graph is shown. The program predicted a signal peptide at the N-terminus with a probability of 0.976.

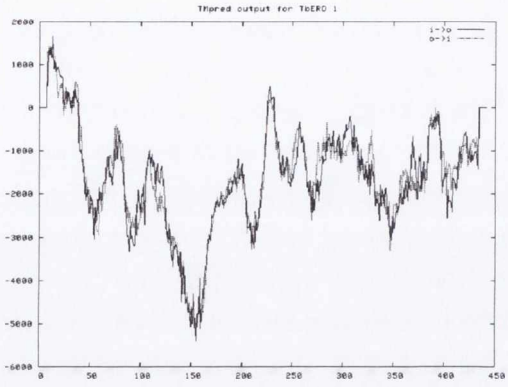
7.2.7 TbEro 1 (Tb927.8.4890)

TMpred predicted 2 strong transmembrane region located between residue 6 to 22 (length of 17 residues) with a score of 1654 and a second transmembrane region located between position 28 to 46 (length of 19 residues) with a score of 601. A hit with a score of greater than 500 is considered significant. A Kyte-Doolittle hydrophobicity plot was carried out using the TopPred program with a window size of 10 residues (Kyte and Doolittle, 1982). The program predicted three transmembrane regions. The first confirmed the previous finding with a hit at position 2 to 22 with a score of 2.251. The second hit was found at position 276 to 296 with a score of 1.032 and finally the third was found at position 382 to 402 with a score of 1.140. A third program, TMHMM was run to confirm the previous results. Again a strong hit was found at position 5 to 22 confirming this transmembrane region as a likely hit but no other regions were found. With the position of this hydrophobic region at the N-terminus of the protein it was most likely to be an *N*-terminal leader sequence. To confirm this hypothesis the program SignalP also located on the same site. The program identified the region as a signal peptide with a probability of 0.976 with a maxcleavage site probability of 0.425 between position 24 and 25. With no other hydrophobic regions predicted by the three programs it was decided TbEro 1 contained no other transmembrane regions or a *C*-terminal extension. Graphs of the various plots can be seen in Figure 4-7.

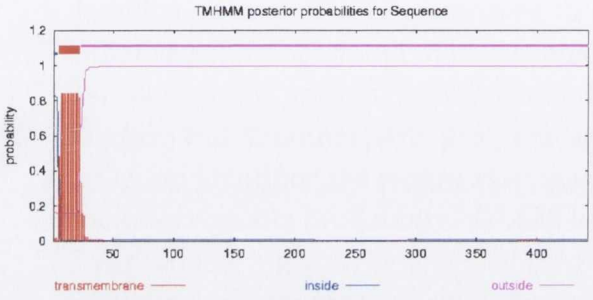
A.



B.



C.



D.

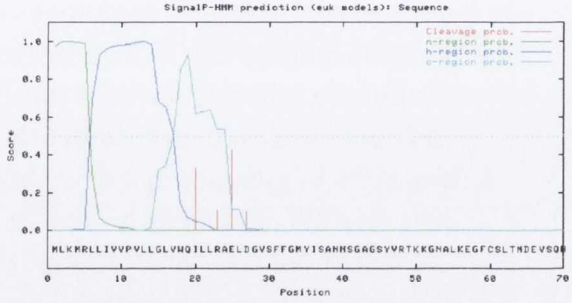


Figure 7.8 Analysis of hydrophobic regions in TbEro 1.

The program TopPred was employed to identify hydrophobic regions using the Kyte-Doolittle method of calculation(Kyte and Doolittle 1982). The resulting graph is shown with 4 possible transmembrane regions indicated with vertical red lines. The program considered two of these peaks to be significant. These can be seen at position 2 to 22 and position 276 to 296.

The program TMpred was employed as before and the resulting graph is shown. The program considered two of these peaks to be significant. These can be seen at position 6 to 22 and position 28 to 46.

The program TMHMM was employed as before and the resulting graph is shown. Again only one significant hydrophobic region was found at position 5 to 22 confirming the other two programs predictions.

To identify if this transmembrane region was a signal peptide, the program SignalP was employed. The resulting graph is shown. The program predicted a signal peptide with a probability of 0.976.

7.3 Primer sequences

Protein	Purpose	Primer Name	Sequence 5'-3'	Note
TbPDI-1 (Tb927.4.2450)	qRT-PCR Forward primer	QRT-PDI1-FP1	ACTTCCAGCC ATTGTTGTC	
	qRT-PCR Reverse primer	QRT-PDI1-RP1	GCCTCATCGT AAGCATCG	
TbPDI-2 (Tb927.10.8230)	qRT-PCR Forward primer	QRT-PDI2-FP1	GCAATAGCACT TCACTTTCTAC	
	qRT-PCR Reverse primer	QRT-PDI2-RP1	TTCTGTCAACT TCTCTACTTCC	
	Cloning into p2T7-177 vector	TbPDI2 RNAiFP1	GGACTAGTCAG CCTCAAATTCT TTGAAGCAG	Contains restriction enzyme cut site for: Spe I
	Cloning into p2T7-177 vector	TbPDI2 RNAiRP1	AATGCAGAAGC ACCTTACGTCG GGATCCGG	Contains restriction enzyme cut site for: BamHI
	Cloning into p2T7-177 vector	TbPDI2 RNAiFP2	CCGCTCGAGTCAGC CTCAAATTCTTTGA AGCAG	Contains restriction enzyme cut site for: XhoI
	Cloning into p2T7-177 vector	TbPDI2 RNAiRP2	AATGCAGAAGCACC TTACGTCGGGAAGC TTCCC	Contains restriction enzyme cut site for: Hind III
	Cloning into pLew100 vector	TbPDI2 FP3	GCGAAGCTTATGCG CGCTATTTTTCTTG TTGC	Contains restriction enzyme cut site for: Hind III
	Cloning into pLew100 vector	TbPDI2 RP3	CGGGATCCCTACAG ATCTTGTTTATCAA CATTGCTC	Contains restriction enzyme cut site for: BamHI

Protein	Purpose	Primer Name	Sequence 5'-3'	Note
TbPDI-2 (Tb927.10.8230)	Amplifying blasticidin resistance gene for knockout of TbPDI-2	BastFP1	GCACGCGTATGCCTTTGTCTCAAGAAGAATCCAC	Contains restriction enzyme cut site for: Mlu I
	Amplifying blasticidin resistance gene for knockout of TbPDI-2	Blast-RP2	CCACTAGTTAGCCCTCCCACACATAACCAG	Contains restriction enzyme cut site for: Spe I
	Amplifying TbPDI-2 to subclone into pUC19	gPDI2FP1	CCTTGTCGTTTACTTAGCATAGGATCCACAACTTTACCC	Contains restriction enzyme site for Bam HI
	Amplifying TbPDI-2 to subclone into pUC19	gPDI2RP1	GCGAGTAGGACTCTAGAGGCGCGCAGC	Contains restriction enzyme site for Xba I
	Amplifying puromycin resistance gene for knockout of TbPDI-2	Puro-RP2	CCACTAGTTCGATCAGGCACCGGGCTTG	Contains restriction enzyme cut site for: Spe I
	Amplifying puromycin resistance gene for knockout of TbPDI-2	PuroFP1	GCACGCGTATGACCGAGTACAAGCCC	Contains restriction enzyme cut site for: Mlu I
	Amplifying flanking regions of TbPDI-2 for knockout construct	gPDI-RP3	GGACTAGTAGCAAGATTAAGAAAACGTTTGGC	Contains restriction enzyme cut site for: Spe I
	Amplifying flanking regions of TbPDI-2 for knockout construct	gPDI-FP3		Contains restriction enzyme cut site for: Mlu I
TbPDI-3 (Tb927.7.1300)	qRT-PCR Forward primer	QRT-PDI3-FP1	GATTCAGCAAGCCGTCTCC	
	qRT-PCR Reverse primer	QRT-PDI3-RP1	ACCATTCTCAAGGACATTAGTAGC	

Protein	Purpose	Primer Name	Sequence 5'-3'	Note
TbPDI-3 (Tb927.7 .1300)	Subcloning into pGEX-5X-1 expression vector	TbPDI3-FP1	GGCGGATCCGTTTTCATGTCGCAAAACC	Contains restriction enzyme cut site for: BamHI
	Subcloning into pGEX-5X-1 expression vector	TbPDI3-RP1	CGCCTCGAGGAGGCGTTCCCTCTCC	Contains restriction enzyme cut site for: XhoI
TbPDI-4 (Tb927.5 .1020)	qRT-PCR Forward primer	QRT-TbPDI4-FP1	GTGCTTGTTGATTCTGAC	
	qRT-PCR Reverse primer	QRT-TbPDI4-RP1	GAAGGCTATGAAGGATGG	
	Subcloning into pGEX-5X-1 expression vector	TbPDI4-FP1	GCGGATCCATACCGCTACTTAATGACGACAC	Contains restriction enzyme cut site for: BamHI
	Subcloning into pGEX-5X-1 expression vector	TBPDI4-RP1	GCGGCCTCGAGAACGGCAATATATGG	Contains restriction enzyme cut site for: XhoI

Protein	Purpose	Primer Name	Sequence 5'-3'	Note
TbPDI-5 (Tb927.7.5790)	qRT-PCR Forward primer	QRT-PDI5-FP1	TTTGATAAGGTGGCGTTG	
	qRT-PCR Reverse primer	QRT-PDI5-RP1	CTTGTCAGCATCCAATCG	
	Subcloning into pGEX-5X-1 expression vector	TbPDI5-FP1	GGCGGATCCTGATAATTTTGATAAGGTGGCG	Contains restriction enzyme cut site for: BamHI
	Subcloning into pGEX-5X-1 expression vector	TBPDI5-RP1	CTCGAAGTAAAAAAGAAGGCCTCGAGGCGG	Contains restriction enzyme cut site for: XhoI
TbEro 1 (Tb927.8.4890)	For use in RNAi vector p2T7-177	TbEro1FP2	GGACTAGTAATGATACGGCTGGTGTAGAAAACAG	Contains restriction enzyme cut site for: SpeI
	For use in RNAi vector p2T7-177	TbEro1RP2	CGGGATCCACCACATTCTCAGCGGAAC	Contains restriction enzyme cut site for: BamHI
	For use in RNAi vector p2T7-177	TbEro1FP3	CCGCTCGAGTAATGATACGGCTGGTGTAGAAAACAG	Contains restriction enzyme cut site for: XhoI
	For use in RNAi vector p2T7-177	TbEro1RP3	CCCAAGCTTCCACCACATTCTCAGCGGACC	Contains restriction enzyme cut site for: HindIII

8 References

Ben Achour, Y., Chenik, M., Louzir, H., and Dellagi, K. (2002). Identification of a disulfide isomerase protein of *Leishmania major* as a putative virulence factor. Infect. Immun *70*, 3576–3585.

Ahlquist, P. (2002). RNA-Dependent RNA Polymerases, Viruses, and RNA Silencing. Science *296*, 1270–1273.

Aitcheson, N., Talbot, S., Shapiro, J., Hughes, K., Adkin, C., Butt, T., Sheader, K., and Rudenko, G. (2005). VSG switching in *Trypanosoma brucei*: antigenic variation analysed using RNAi in the absence of immune selection. Mol Microbiol *57*, 1608–1622.

Alanen, H.I., Salo, K.E.H., Pekkala, M., Siekkinen, H.M., Pirneskoski, A., and Ruddock, L.W. (2003). Defining the domain boundaries of the human protein disulfide isomerases. Antioxid. Redox Signal *5*, 367–374.

Alanen, H.I., Williamson, R.A., Howard, M.J., Hatahet, F.S., Salo, K.E.H., Kauppila, A., Kellokumpu, S., and Ruddock, L.W. (2006). ERp27, a New Non-catalytic Endoplasmic Reticulum-located Human Protein Disulfide Isomerase Family Member, Interacts with ERp57. Journal of Biological Chemistry *281*, 33727 –33738.

Allen, G., and Gurnett, L.P. (1983). Locations of the six disulphide bonds in a variant surface glycoprotein (VSG 117) from *Trypanosoma brucei*. Biochem. J *209*, 481–487.

Andres, D.A., Dickerson, I.M., and Dixon, J.E. (1990). Variants of the carboxyl-terminal KDEL sequence direct intracellular retention. J. Biol. Chem *265*, 5952–5955.

Anelli, T., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., Mezghrani, A., Ruffato, E., Simmen, T., and Sitia, R. (2003). Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44. EMBO J *22*, 5015–5022.

Anelli, T., Ceppi, S., Bergamelli, L., Cortini, M., Masciarelli, S., Valetti, C., and Sitia, R. (2007). Sequential steps and checkpoints in the early exocytic compartment during secretory IgM biogenesis. EMBO J *26*, 4177–4188.

Anelli, T., Bergamelli, L., Margittai, E., Rimessi, A., Fagioli, C., Malgaroli, A., Pinton, P., Ripamonti, M., Rizzuto, R., and Sitia, R. (2012). Ero1α regulates Ca(2+) fluxes at the endoplasmic reticulum-mitochondria interface (MAM). Antioxid. Redox Signal. *16*, 1077–1087.

Anfinsen, C.B., and Haber, E. (1961). Studies on the reduction and re-formation of protein disulfide bonds. J. Biol. Chem *236*, 1361–1363.

Anfinsen, C.B., and Scheraga, H.A. (1975). Experimental and theoretical aspects of protein folding. Adv. Protein Chem *29*, 205–300.

Baker, K.M., Chakravarthi, S., Langton, K.P., Sheppard, A.M., Lu, H., and Bulleid, N.J. (2008). Low reduction potential of Ero1α regulatory disulphides ensures tight control of substrate oxidation. EMBO J *27*, 2988–2997.

Baker, N., Glover, L., Munday, J.C., Aguinaga Andrés, D., Barrett, M.P., de Koning, H.P., and Horn, D. (2012). Aquaglyceroporin 2 controls susceptibility to melarsoprol and pentamidine in African trypanosomes. Proc Natl Acad Sci U S A *109*, 10996–11001.

Bangs, J.D., Andrews, N.W., Hart, G.W., and Englund, P.T. (1986). Posttranslational modification and intracellular transport of a trypanosome variant surface glycoprotein. J Cell Biol *103*, 255–263.

Bangs, J.D., Uyetake, L., Brickman, M.J., Balber, A.E., and Boothroyd, J.C. (1993). Molecular cloning and cellular localization of a BiP homologue in *Trypanosoma brucei*. Divergent ER retention signals in a lower eukaryote. J. Cell. Sci *105 (Pt 4)*, 1101–1113.

Bass, R., Ruddock, L.W., Klappa, P., and Freedman, R.B. (2004). A major fraction of endoplasmic reticulum-located glutathione is present as mixed disulfides with protein. J. Biol. Chem *279*, 5257–5262.

Becker, M., Aitcheson, N., Byles, E., Wickstead, B., Louis, E., and Rudenko, G. (2004). Isolation of the repertoire of VSG expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. Genome Res. *14*, 2319–2329.

Benham, A.M. (2012). The protein disulfide isomerase family: key players in health and disease. Antioxid. Redox Signal. *16*, 781–789.

Benham, A.M., Cabibbo, A., Fassio, A., Bulleid, N., Sitia, R., and Braakman, I. (2000). The CXXCXXC motif determines the folding, structure and stability of human Ero1-Lα. EMBO J *19*, 4493–4502.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Blum, M.L., Down, J.A., Gurnett, A.M., Carrington, M., Turner, M.J., and Wiley, D.C. (1993). A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. Nature *362*, 603–609.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J *17*, 170–180.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol *3*, e85.

Bridges, D.J., Pitt, A.R., Hanrahan, O., Brennan, K., Voorheis, H.P., Herzyk, P., de Koning, H.P., and Burchmore, R.J.S. (2008). Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. Proteomics *8*, 83–99.

Brooks, D.R., McCulloch, R., Coombs, G.H., and Mottram, J.C. (2000). Stable transformation of trypanosomatids through targeted chromosomal integration of the selectable marker gene encoding blasticidin S deaminase. FEMS Microbiology Letters *186*, 287–291.

Bulleid, N.J., and Ellgaard, L. (2011). Multiple ways to make disulfides. Trends in Biochemical Sciences *36*, 485–492.

Cabibbo, A., Pagani, M., Fabbri, M., Rocchi, M., Farmery, M.R., Bulleid, N.J., and Sitia, R. (2000). ERO1-L, a human protein that favors disulfide bond formation in the endoplasmic reticulum. J. Biol. Chem *275*, 4827–4833.

Campbell, D.A., Thomas, S., and Sturm, N.R. (2003). Transcription in kinetoplastid protozoa: why be normal? Microbes Infect *5*, 1231–1240.

Carrington, M., and Boothroyd, J. (1996). Implications of conserved structural motifs in disparate trypanosome surface proteins. Molecular and Biochemical Parasitology *81*, 119–126.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., and Turner, M. (1991). Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. J. Mol. Biol *221*, 823–835.

Castanotto, D., and Rossi, J.J. (2009). The promises and pitfalls of RNA-interference-based therapeutics. Nature *457*, 426–433.

Castillo-Acosta, V.M., Aguilar-Pereyra, F., Vidal, A.E., Navarro, M., Ruiz-Pérez, L.M., and González-Pacanowska, D. (2012). Trypanosomes lacking uracil-DNA glycosylase are hypersensitive to antifolates and present a mutator phenotype. The International Journal of Biochemistry & Cell Biology *44*, 1555–1568.

Cecchi, G., Mattioli, R.C., Slingenbergh, J., and de la Rocque, S. (2008). Land cover and tsetse fly distributions in sub-Saharan Africa. Med. Vet. Entomol. *22*, 364–373.

Chakravarthi, S., Jessop, C.E., Willer, M., Stirling, C.J., and Bulleid, N.J. (2007). Intracellular catalysis of disulfide bond formation by the human sulfhydryl oxidase, QSOX1. Biochem. J. *404*, 403–411.

Chang, T., Milne, K.G., Güther, M.L.S., Smith, T.K., and Ferguson, M.A.J. (2002). Cloning of *Trypanosoma brucei* and *Leishmania major* Genes Encoding the GlcNAc-Phosphatidylinositol De-N-acetylase of Glycosylphosphatidylinositol Biosynthesis That Is Essential to the African Sleeping Sickness Parasite. J. Biol. Chem. *277*, 50176–50182.

Chattopadhyay, A., Jones, N.G., Nietlispach, D., Nielsen, P.R., Voorheis, H.P., Mott, H.R., and Carrington, M. (2005). Structure of the C-terminal Domain from *Trypanosoma brucei* Variant Surface Glycoprotein MITat1.2. Journal of Biological Chemistry *280*, 7228–7235.

Chen, Y., Hung, C., Burderer, T., and Lee, G. (2003). Development of RNA interference revertants in cell lines generated with a double stranded RNA expression construct driven by two opposing promoters. Molecular and Biochemical Parasitology *126*, 275–279.

Chin, K.-T., Kang, G., Qu, J., Gardner, L.B., Coetzee, W.A., Zito, E., Fishman, G.I., and Ron, D. (2011). The sarcoplasmic reticulum luminal thiol oxidase ERO1 regulates cardiomyocyte excitation-coupled calcium release and response to hemodynamic load. FASEB J *25*, 2583–2591.

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. EMBO J *21*, 1881–1888.

Clayton, C., and Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol. Biochem. Parasitol. *156*, 93–101.

Comini, M.A., Krauth-Siegel, R.L., and Flohé, L. (2007). Depletion of the thioredoxin homologue tryparedoxin impairs antioxidative defence in African trypanosomes. Biochem. J *402*, 43–49.

Creighton, T.E. (1984). Disulfide bond formation in proteins. Meth. Enzymol *107*, 305–329.

Cross, G.A. (1978). Antigenic variation in trypanosomes. Proc. R. Soc. Lond., B, Biol. Sci *202*, 55–72.

Danon, A. (2002). Redox reactions of regulatory proteins: do kinetics promote specificity? Trends Biochem. Sci *27*, 197–203.

Darby, N.J., and Creighton, T.E. (1995). Functional properties of the individual thioredoxin-like domains of protein disulfide isomerase. Biochemistry *34*, 11725–11735.

Darby, N.J., Penka, E., and Vincentelli, R. (1998). The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. J. Mol. Biol *276*, 239–247.

Das, A., Gale, M., Carter, V., and Parsons, M. (1994). The protein phosphatase inhibitor okadaic acid induces defects in cytokinesis and organellar genome segregation in *Trypanosoma brucei*. J Cell Sci *107*, 3477–3483.

Delhi, P., Queiroz, R., Inchaustegui, D., Carrington, M., and Clayton, C. (2011). Is there a classical nonsense-mediated decay pathway in trypanosomes? PLoS ONE *6*, e25112.

Van Den Abbeele, J., Claes, Y., Van Bockstaele, D., Le Ray, D., and Coosemans, M. (1999). *Trypanosoma brucei* Spp. Development in the Tsetse Fly: Characterization of the Post-Mesocyclic Stages in the Foregut and Proboscis. Parasitology *118*, 469–478.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. *36*, W465–469.

Dereeper, A., Audic, S., Claverie, J.-M., and Blanc, G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol. Biol. *10*, 8.

Deshaies, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. Nature *349*, 806–808.

Desilva, M.G., Lu, J., Donadel, G., Modi, W.S., Xie, H., Notkins, A.L., and Lan, M.S. (1996). Characterization and chromosomal localization of a new protein disulfide isomerase, PDIp, highly expressed in human pancreas. DNA Cell Biol *15*, 9–16.

Dixon, D.P., Van Lith, M., Edwards, R., and Benham, A. (2003). Cloning and initial characterization of the *Arabidopsis thaliana* endoplasmic reticulum oxidoreductins. Antioxid. Redox Signal *5*, 389–396.

Doyle, J.J. (1977). Antigenic variation in the salivarian trypanosomes. Adv. Exp. Med. Biol *93*, 31–63.

Drozdz, M., and Clayton, C. (1999). Structure of a Regulatory 3′ Untranslated Region from *Trypanosoma brucei*. RNA *5*, 1632–1644.

Durand-Dubief, M., and Bastin, P. (2003). TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. BMC Biol *1*, 2.

Eddy, S.R. (1995). Multiple alignment using hidden Markov models. Proc Int Conf Intell Syst Mol Biol *3*, 114–120.

Engstler, M., and Boshart, M. (2004). Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. Genes & Development *18*, 2798–2811.

Estévez, A.M., Kierszenbaum, F., Wirtz, E., Bringaud, F., Grunstein, J., and Simpson, L. (1999). Knockout of the glutamate dehydrogenase gene in bloodstream *Trypanosoma brucei* in culture has no effect on editing of mitochondrial mRNAs. Molecular and Biochemical Parasitology *100*, 5–17.

Fairlamb, A.H., Blackburn, P., Ulrich, P., Chait, B.T., and Cerami, A. (1985). Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. Science *227*, 1485–1487.

Fass, D. (2008). The Erv family of sulfhydryl oxidases. Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research *1783*, 557–566.

Ferguson, M.A. (1999). The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J. Cell. Sci *112 (Pt 17)*, 2799–2809.

Field, M.C., Sergeenko, T., Wang, Y.-N., Böhm, S., and Carrington, M. (2010). Chaperone requirements for biosynthesis of the trypanosome variant surface glycoprotein. PLoS ONE *5*, e8468.

Fragoso, C.M., Schumann Burkard, G., Oberle, M., Renggli, C.K., Hilzinger, K., and Roditi, I. (2009). PSSA-2, a membrane-spanning phosphoprotein of *Trypanosoma brucei*, is required for efficient maturation of infection. PLoS ONE *4*, e7074.

Fraldi, A., Zito, E., Annunziata, F., Lombardi, A., Cozzolino, M., Monti, M., Spampanato, C., Ballabio, A., Pucci, P., Sitia, R., et al. (2008). Multistep, sequential control of the trafficking and function of the multiple sulfatase deficiency gene product, SUMF1 by PDI, ERGIC-53 and ERp44. Hum. Mol. Genet. *17*, 2610–2621.

Frand, A.R., and Kaiser, C.A. (1998). The ERO1 gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum. Mol. Cell *1*, 161–170.

Frand, A.R., and Kaiser, C.A. (1999). Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. Mol. Cell *4*, 469–477.

Frand, A.R., and Kaiser, C.A. (2000). Two Pairs of Conserved Cysteines Are Required for the Oxidative Activity of Ero1p in Protein Disulfide Bond Formation in the Endoplasmic Reticulum. Mol. Biol. Cell *11*, 2833–2843.

Freedman, R.B., Gane, P.J., Hawkins, H.C., Hlodan, R., McLaughlin, S.H., and Parry, J.W. (1998). Experimental and theoretical analyses of the domain architecture of mammalian protein disulphide-isomerase. Biol. Chem *379*, 321–328.

Freymann, D., Down, J., Carrington, M., Roditi, I., Turner, M., and Wiley, D. (1990). 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. J. Mol. Biol *216*, 141–160.

Frickel, E.-M., Frei, P., Bouvier, M., Stafford, W.F., Helenius, A., Glockshuber, R., and Ellgaard, L. (2004). ERp57 Is a Multifunctional Thiol-Disulfide Oxidoreductase. Journal of Biological Chemistry *279*, 18277–18287.

Fukuda, M., and Kobata, A. (1993). Glycobiology: a practical approach (IRL PRESS).

Gale, M., Carter, V., and Parsons, M. (1994). Translational control mediates the developmental regulation of the *Trypanosoma brucei* Nrk protein kinase. J. Biol. Chem *269*, 31659–31665.

Garbi, N., Tanaka, S., Momburg, F., and Hammerling, G.J. (2006). Impaired assembly of the major histocompatibility complex class I peptide-loading complex in mice deficient in the oxidoreductase ERp57. Nat Immunol *7*, 93–102.

Gibson, W. (2012). The origins of the trypanosome genome strains *Trypanosoma brucei brucei* TREU 927, *T. b. gambiense* DAL 972, *T. vivax* Y486 and *T. congolense* IL3000. Parasit Vectors *5*, 71.

Gillin, F.D., Reiner, D.S., and McCaffery, J.M. (1996). Cell biology of the primitive eukaryote *Giardia lamblia*. Annu. Rev. Microbiol *50*, 679–705.

Glover, L., and Horn, D. (2006). Repression of polymerase I-mediated gene expression at *Trypanosoma brucei* telomeres. EMBO Reports *7*, 93–99.

Goldshmidt, H., Sheiner, L., Bütikofer, P., Roditi, I., Uliel, S., Günzel, M., Engstler, M., and Michaeli, S. (2008). Role of Protein Translocation Pathways across the Endoplasmic Reticulum in *Trypanosoma brucei*. Journal of Biological Chemistry *283*, 32085–32098.

Gottesdiener, K., Garcia-Anoveros, J., Lee, M.G., and Van der Ploeg, L.H. (1990). Chromosome organization of the protozoan *Trypanosoma brucei*. Mol. Cell. Biol. *10*, 6079–6083.

Gross, E., Kastner, D.B., Kaiser, C.A., and Fass, D. (2004). Structure of Ero1p, source of disulfide bonds for oxidative protein folding in the cell. Cell *117*, 601–610.

Gross, E., Sevier, C.S., Heldman, N., Vitu, E., Bentzur, M., Kaiser, C.A., Thorpe, C., and Fass, D. (2006). Generating disulfides enzymatically: reaction products and electron

acceptors of the endoplasmic reticulum thiol oxidase Ero1p. Proc. Natl. Acad. Sci. U.S.A *103*, 299–304.

Grubb, S., Guo, L., Fisher, E.A., and Brodsky, J.L. (2012). Protein disulfide isomerases contribute differentially to the endoplasmic reticulum–associated degradation of apolipoprotein B and other substrates. Mol. Biol. Cell *23*, 520–532.

Gruber, C.W., Cemazar, M., Heras, B., Martin, J.L., and Craik, D.J. (2006). Protein disulfide isomerase: the structure of oxidative folding. Trends in Biochemical Sciences *31*, 455–464.

Gueiros-Filho, F.J., and Beverley, S.M. (1996). Selection against the dihydrofolate reductase-thymidylate synthase (DHFR-TS) locus as a probe of genetic alterations in *Leishmania major*. Mol Cell Biol *16*, 5655–5663.

Gumireddy, K., Sun, F., Klein-Szanto, A.J., Gibbins, J.M., Gimotty, P.A., Saunders, A.J., Schultz, P.G., and Huang, Q. (2007). In vivo selection for metastasis promoting genes in the mouse. Proc. Natl. Acad. Sci. U.S.A *104*, 6696–6701.

Günther, R., Bräuer, C., Janetzky, B., Förster, H.H., Ehbrecht, I.M., Lehle, L., and Küntzel, H. (1991). The *Saccharomyces cerevisiae* TRG1 gene is essential for growth and encodes a lumenal endoplasmic reticulum glycoprotein involved in the maturation of vacuolar carboxypeptidase. J. Biol. Chem *266*, 24557–24563.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611–620.

Güther, M.L.S., Lee, S., Tetley, L., Acosta-Serrano, A., and Ferguson, M.A.J. (2006). GPI-anchored Proteins and Free GPI Glycolipids of Procyclic Form *Trypanosoma brucei* Are Nonessential for Growth, Are Required for Colonization of the Tsetse Fly, and Are Not the Only Components of the Surface Coat. Mol. Biol. Cell *17*, 5265–5274.

Hamm, B., Schindler, A., Mecke, D., and Duszenko, M. (1990). Differentiation of *Trypanosoma brucei* bloodstream trypomastigotes from long slender to short stumpy-like forms in axenic culture. Molecular and Biochemical Parasitology *40*, 13–22.

Hammarton, T.C. (2007). Cell cycle regulation in *Trypanosoma brucei*. Molecular and Biochemical Parasitology *153*, 1–8.

Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., et al. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol. Cell *11*, 619–633.

Hatahet, F., and Ruddock, L.W. (2009). Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. Antioxid. Redox Signal *11*, 2807–2850.

Hirumi, H., and Hirumi, K. (1989). Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. J. Parasitol *75*, 985–989.

Hoare, C.A. (1966). The classification of mammalian trypanosomes. Ergeb Mikrobiol Immunitatsforsch Exp Ther *39*, 43–57.

Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. J. Biol. Chem *264*, 13963–13966.

Horibe, T., Gomi, M., Iguchi, D., Ito, H., Kitamura, Y., Masuoka, T., Tsujimoto, I., Kimura, T., and Kikuchi, M. (2004). Different Contributions of the Three CXXC Motifs of Human Protein-disulfide Isomerase-related Protein to Isomerase Activity and Oxidative Refolding. J. Biol. Chem. *279*, 4604–4611.

Hsu, M.P., Muhich, M.L., and Boothroyd, J.C. (1989). A developmentally regulated gene of trypanosomes encodes a homologue of rat protein-disulfide isomerase and phosphoinositol-phospholipase C. Biochemistry *28*, 6440–6446.

Hwang, C., Sinskey, A.J., and Lodish, H.F. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. Science *257*, 1496–1502.

Janzen, C.J., Fernandez, J.P., Deng, H., Diaz, R., Hake, S.B., and Cross, G.A.M. (2006). Unusual histone modifications in *Trypanosoma brucei*. FEBS Lett. *580*, 2306–2310.

Jasuja, R., Passam, F.H., Kennedy, D.R., Kim, S.H., van Hessem, L., Lin, L., Bowley, S.R., Joshi, S.S., Dilks, J.R., Furie, B., et al. (2012). Protein disulfide isomerase inhibitors constitute a new class of antithrombotic agents. J. Clin. Invest. *122*, 2104–2113.

Jensen, B.C., Kifer, C.T., and Parsons, M. (2011). *Trypanosoma brucei*: Two mitogen activated protein kinase kinases are dispensable for growth and virulence of the bloodstream form. Experimental Parasitology *128*, 250–255.

Jessop, C.E., Chakravarthi, S., Garbi, N., Hämmerling, G.J., Lovell, S., and Bulleid, N.J. (2007). ERp57 is essential for efficient folding of glycoproteins sharing common structural domains. EMBO J *26*, 28–40.

Jessop, C.E., Watkins, R.H., Simmons, J.J., Tasab, M., and Bulleid, N.J. (2009). Protein disulphide isomerase family members show distinct substrate specificity: P5 is targeted to BiP client proteins. J. Cell. Sci *122*, 4287–4295.

Jinushi, M., Vanneman, M., Munshi, N.C., Tai, Y.-T., Prabhala, R.H., Ritz, J., Neuberg, D., Anderson, K.C., Carrasco, D.R., and Dranoff, G. (2008). MHC class I chain-related protein A antibodies and shedding are associated with the progression of multiple myeloma. Proc. Natl. Acad. Sci. U.S.A *105*, 1285–1290.

Jones, N.G., Nietlispach, D., Sharma, R., Burke, D.F., Eyres, I., Mues, M., Mott, H.R., and Carrington, M. (2008). Structure of a Glycosylphosphatidylinositol-anchored Domain from a Trypanosome Variant Surface Glycoprotein. Journal of Biological Chemistry *283*, 3584 –3593.

Jordan, P.A., Stevens, J.M., Hubbard, G.P., Barrett, N.E., Sage, T., Authi, K.S., and Gibbins, J.M. (2005). A role for the thiol isomerase protein ERP5 in platelet function. Blood *105*, 1500–1507.

Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. Curr. Opin. Cell Biol *20*, 253–259.

Kaiser, B.K., Yim, D., Chow, I.-T., Gonzalez, S., Dai, Z., Mann, H.H., Strong, R.K., Groh, V., and Spies, T. (2007). Disulphide-isomerase-enabled shedding of tumour-associated NKG2D ligands. Nature *447*, 482–486.

Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M., and Creighton, T.E. (1996). Structure determination of the *N*-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy. Biochemistry *35*, 7684–7691.

Kettner, K., Blomberg, A., and Rödel, G. (2004). *Schizosaccharomyces pombe* ER oxidoreductin-like proteins SpEro1a p and SpEro1b p. Yeast *21*, 1035–1044.

Kikuchi, M., Doi, E., Tsujimoto, I., Horibe, T., and Tsujimoto, Y. (2002). Functional analysis of human P5, a protein disulfide isomerase homologue. J. Biochem *132*, 451–455.

Kimura, T., Hosoda, Y., Sato, Y., Kitamura, Y., Ikeda, T., Horibe, T., and Kikuchi, M. (2005). Interactions among Yeast Protein-Disulfide Isomerase Proteins and Endoplasmic Reticulum Chaperone Proteins Influence Their Activities. J. Biol. Chem. *280*, 31438–31441.

Knodler, L.A., Noiva, R., Mehta, K., McCaffery, J.M., Aley, S.B., Svärd, S.G., Nystul, T.G., Reiner, D.S., Silberman, J.D., and Gillin, F.D. (1999). Novel Protein-disulfide Isomerases from the Early-diverging Protist *Giardia lamblia*. Journal of Biological Chemistry *274*, 29805 -29811.

Ko, F.C.F., and Chow, K.L. (2002). A novel thioredoxin-like protein encoded by the C. elegans dpy-11 gene is required for body and sensory organ morphogenesis. Development *129*, 1185–1194.

Kodali, V.K., and Thorpe, C. (2010). Quiescin Sulfhydryl Oxidase from *Trypanosoma brucei*: Catalytic Activity and Mechanism of a QSOX Family Member with a Single Thioredoxin Domain. Biochemistry *49*, 2075–2085.

Koumandou, V.L., Natesan, S., Sergeenko, T., and Field, M. (2008). The trypanosome transcriptome is remodelled during differentiation but displays limited responsiveness within life stages. BMC Genomics *9*, 298.

Kozlov, G., Määttänen, P., Thomas, D.Y., and Gehring, K. (2010). A structural overview of the PDI family of proteins. FEBS Journal *277*, 3924–3936.

Krieger, S., Schwarz, W., Ariyanayagam, M.R., Fairlamb, A.H., Krauth-Siegel, R.L., and Clayton, C. (2000). Trypanosomes lacking trypanothione reductase are avirulent and show increased sensitivity to oxidative stress. Molecular Microbiology *35*, 542–552.

Kulp, M.S., Frickel, E.-M., Ellgaard, L., and Weissman, J.S. (2006). Domain Architecture of Protein-disulfide Isomerase Facilitates Its Dual Role as an Oxidase and an Isomerase in Ero1p-mediated Disulfide Formation. Journal of Biological Chemistry *281*, 876–884.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol *157*, 105–132.

LaCount, D.J., Barrett, B., and Donelson, J.E. (2002). *Trypanosoma brucei* FLA1 Is Required for Flagellum Attachment and Cytokinesis. J. Biol. Chem. *277*, 17580–17588.

Lambert, N., and Freedman, R.B. (1983). Structural properties of homogeneous protein disulphide-isomerase from bovine liver purified by a rapid high-yielding procedure. Biochem. J *213*, 225–234.

LeBowitz, J.H., Smith, H.Q., Rusche, L., and Beverley, S.M. (1993). Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes & Development *7*, 996–1007.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet *34*, 77–137.

Leegwater, P.A., Strating, M., Murphy, N.B., Kooy, R.F., van der Vliet, P.C., and Overdulve, J.P. (1991). The *Trypanosoma brucei* DNA polymerase alpha core subunit gene is developmentally regulated and linked to a constitutively expressed open reading frame. Nucleic Acids Res *19*, 6441–6447.

Leifso, K., Cohen-Freue, G., Dogra, N., Murray, A., and McMaster, W.R. (2007). Genomic and proteomic expression analysis of Leishmania promastigote and amastigote life stages: the Leishmania genome is constitutively expressed. Mol. Biochem. Parasitol *152*, 35–46.

Lewandrowski, U., Moebius, J., Walter, U., and Sickmann, A. (2006). Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach. Mol. Cell Proteomics *5*, 226–233.

Lillico, S., Field, M.C., Blundell, P., Coombs, G.H., and Mottram, J.C. (2003). Essential Roles for GPI-anchored Proteins in African Trypanosomes Revealed Using Mutants Deficient in GPI8. Mol Biol Cell *14*, 1182–1194.

Van Lith, M., Karala, A.-R., Bown, D., Gatehouse, J.A., Ruddock, L.W., Saunders, P.T.K., and Benham, A.M. (2007). A developmentally regulated chaperone complex for the endoplasmic reticulum of male haploid germ cells. Mol. Biol. Cell *18*, 2795–2804.

Lopez-Estrano, C., Tschudi, C., and Ullu, E. (1998). Exonic Sequences in the 5' Untranslated Region of alpha -Tubulin mRNA Modulate trans Splicing in *Trypanosoma brucei*. Mol. Cell. Biol. *18*, 4620–4628.

Lustig, Y., Sheiner, L., Vagima, Y., Goldshmidt, H., Das, A., Bellofatto, V., and Michaeli, S. (2007). Spliced-leader RNA silencing: a novel stress-induced mechanism in *Trypanosoma brucei*. EMBO Rep *8*, 408–413.

Mairet-Coello, G., Tury, A., Esnard-Feve, A., Fellmann, D., Risold, P.-Y., and Griffond, B. (2004). FAD-linked sulfhydryl oxidase QSOX: topographic, cellular, and subcellular immunolocalization in adult rat central nervous system. J. Comp. Neurol. *473*, 334–363.

Mandava, V., Fernandez, J.P., Deng, H., Janzen, C.J., Hake, S.B., and Cross, G.A.M. (2007). Histone modifications in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *156*, 41–50.

Marcello, L., and Barry, J.D. (2007). Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Research *17*, 1344 –1352.

Mariappan, M., Radhakrishnan, K., Dierks, T., Schmidt, B., and Figura, K. von (2008). ERp44 Mediates a Thiol-independent Retention of Formylglycine-generating Enzyme in the Endoplasmic Reticulum. J. Biol. Chem. *283*, 6375–6383.

Martin, K.L., and Smith, T.K. (2006a). Phosphatidylinositol synthesis is essential in bloodstream form *Trypanosoma brucei*. Biochem. J *396*, 287–295.

Martin, K.L., and Smith, T.K. (2006b). Phosphatidylinositol synthesis is essential in bloodstream form *Trypanosoma brucei*. Biochem J *396*, 287–295.

Martínez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. (2003). Transcription of *Leishmania major* Friedlin Chromosome 1 Initiates in Both Directions within a Single Region. Molecular Cell *11*, 1291–1299.

Masterson, W.J., Taylor, D., and Turner, M.J. (1988). Topologic analysis of the epitopes of a variant surface glycoprotein of *Trypanosoma brucei*. J. Immunol *140*, 3194–3199.

Mather, J.P., and Roberts, P.E.R. (1998). Introduction to cell and tissure culture Theory and Technique.

Matsuo, Y., Akiyama, N., Nakamura, H., Yodoi, J., Noda, M., and Kizaka-Kondoh, S. (2001). Identification of a Novel Thioredoxin-related Transmembrane Protein. Journal of Biological Chemistry *276*, 10032 –10038.

Matsuo, Y., Nishinaka, Y., Suzuki, S., Kojima, M., Kizaka-Kondoh, S., Kondo, N., Son, A., Sakakura-Nishiyama, J., Yamaguchi, Y., Masutani, H., et al. (2004). TMX, a human transmembrane oxidoreductase of the thioredoxin family: the possible role in disulfide-linked protein folding in the endoplasmic reticulum. Archives of Biochemistry and Biophysics *423*, 81–87.

Matthews, K., and Gull, K. (1997). Commitment to differentiation and cell cycle re-entry are coincident but separable events in the transformation of African trypanosomes from their bloodstream to their insect form. J Cell Sci *110*, 2609–2618.

Matthews, K.R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes & Development *8*, 491–501.

Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S., and Matthews, K. (2006). Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. Nucleic Acids Res *34*, 5312–5324.

McArthur, A.G., Knodler, L.A., Silberman, J.D., Davids, B.J., Gillin, F.D., and Sogin, M.L. (2001). The evolutionary origins of eukaryotic protein disulfide isomerase domains: new evidence from the Amitochondriate protist *Giardia lamblia*. Mol. Biol. Evol *18*, 1455–1463.

McConnell, J., Gurnett, A.M., Cordingley, J.S., Walker, J.E., and Turner, M.J. (1981). Biosynthesis of Trypanosoma brucei variant surface glycoprotein. I. Synthesis, size, and processing of an *N*-terminal signal peptide. Mol. Biochem. Parasitol *4*, 225–242.

McKean, P.G. (2003). Coordination of cell cycle and cytokinesis in *Trypanosoma brucei*. Curr. Opin. Microbiol. *6*, 600–607.

Miller, E.N., Allan, L.M., and Turner, M.J. (1984a). Topological analysis of antigenic determinants on a variant surface glycoprotein of *Trypanosoma brucei*. Mol. Biochem. Parasitol *13*, 67–81.

Miller, E.N., Allan, L.M., and Turner, M.J. (1984b). Mapping of antigenic determinants within peptides of a variant surface glycoprotein of *Trypanosoma brucei*. Mol. Biochem. Parasitol *13*, 309–322.

Milne, K.G., Güther, M.L.S., and Ferguson, M.A.J. (2001). Acyl-CoA binding protein is essential in bloodstream form *Trypanosoma brucei*. Molecular and Biochemical Parasitology *112*, 301–304.

Mizunaga, T., Katakura, Y., Miura, T., and Maruyama, Y. (1990). Purification and characterization of yeast protein disulfide isomerase. J. Biochem *108*, 846–851.

Montgomery, M.K., and Fire, A. (1998). Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. Trends Genet *14*, 255–258.

Munro, S., and Pelham, H.R. (1987). A *C*-terminal signal prevents secretion of luminal ER proteins. Cell *48*, 899–907.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279–289.

Narayan, M. (2012). Disulfide bonds: protein folding and subcellular protein trafficking. FEBS J. *279*, 2272–2282.

Nardelli, S.C., Avila, A.R., Freund, A., Motta, M.C., Manhães, L., de Jesus, T.C.L., Schenkman, S., Fragoso, S.P., Krieger, M.A., Goldenberg, S., et al. (2007). Small-subunit rRNA processome proteins are translationally regulated during differentiation of *Trypanosoma cruzi*. Eukaryotic Cell *6*, 337–345.

Ngô, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc Natl Acad Sci U S A *95*, 14687–14692.

Noiva, R., and Lennarz, W.J. (1992). Protein disulfide isomerase. A multifunctional protein resident in the lumen of the endoplasmic reticulum. J. Biol. Chem *267*, 3553–3556.

Nolan, D.P., and Garcia-Salcedo, J.A. (2008). Loss of actin does not affect export of newly synthesized proteins to the surface of *Trypanosoma brucei*. Molecular and Biochemical Parasitology *157*, 233–235.

Nolan, D.P., Geuskens, M., and Pays, E. (1999). *N*-linked glycans containing linear poly-*N*-acetyllactosamine as sorting signals in endocytosis in *Trypanosoma brucei*. Curr. Biol *9*, 1169–1172.

Nordstrand, K., Åslund, F., Holmgren, A., Otting, G., and Berndt, K.D. (1999). NMR structure of *Escherichia coli* glutaredoxin 3-glutathione mixed disulfide complex: implications for the enzymatic mechanism. Journal of Molecular Biology *286*, 541–552.

Nørgaard, P., and Winther, J.R. (2001). Mutation of yeast Eug1p CXXS active sites to CXXC results in a dramatic increase in protein disulphide isomerase activity. Biochem J *358*, 269–274.

Nørgaard, P., Westphal, V., Tachibana, C., Alsøe, L., Holst, B., and Winther, J.R. (2001). Functional Differences in Yeast Protein Disulfide Isomerases. J Cell Biol *152*, 553–562.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol *302*, 205–217.

Opperdoes, F.R., and Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: The glycosome. FEBS Letters *80*, 360–364.

Ouna, B.A., Stewart, M., Helbig, C., and Clayton, C. (2012). The *Trypanosoma brucei* CCCH zinc finger proteins ZC3H12 and ZC3H13. Mol. Biochem. Parasitol. *183*, 184–188.

Overath, P., Czichos, J., and Haas, C. (1986). The effect of citrate/cis-aconitate on oxidative metabolism during transformation of *Trypanosoma brucei*. Eur. J. Biochem *160*, 175–182.

Padilla, A. (2002). An Atypical Protein Disulfide Isomerase from the Protozoan Parasite Leishmania Containing a Single Thioredoxin-like Domain. Journal of Biological Chemistry *278*, 1872–1878.

Pagani, M., Fabbri, M., Benedetti, C., Fassio, A., Pilati, S., Bulleid, N.J., Cabibbo, A., and Sitia, R. (2000). Endoplasmic reticulum oxidoreductin 1-lbeta (ERO1-Lbeta), a human gene induced in the course of the unfolded protein response. J. Biol. Chem *275*, 23685–23692.

Pagani, M., Pilati, S., Bertoli, G., Valsasina, B., and Sitia, R. (2001). The *C*-terminal domain of yeast Ero1p mediates membrane localization and is essential for function. FEBS Letters *508*, 117–120.

Palenchar, J.B., and Bellofatto, V. (2006). Gene transcription in trypanosomes. Mol. Biochem. Parasitol. *146*, 135–141.

Pan, J.L., and Bardwell, J.C.A. (2006). The origami of thioredoxin-like folds. Protein Sci *15*, 2217–2227.

Park, B., Lee, S., Kim, E., Cho, K., Riddell, S.R., Cho, S., and Ahn, K. (2006). Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing. Cell *127*, 369–382.

Parsons, M., Nelson, R.G., Watkins, K.P., and Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. Cell *38*, 309–316.

Pays, E. (2005). Regulation of antigen gene expression in *Trypanosoma brucei*. Trends in Parasitology *21*, 517–520.

Perelman, D., and Boothroyd, J.C. (1990). Lack of introns in the ribosomal protein gene S14 of trypanosomes. Mol. Cell. Biol. *10*, 3284–3288.

Pollard, M., Travers, K., and Weissman, J. (1998). Ero1p: A Novel and Ubiquitous Protein with an Essential Role in Oxidative Protein Folding in the Endoplasmic Reticulum. Molecular Cell *1*, 171–182.

Ramming, T., and Appenzeller-Herzog, C. (2012). The physiological functions of mammalian endoplasmic oxidoreductin 1: on disulfides and more. Antioxid. Redox Signal. *16*, 1109–1118.

Rancy, P.C., and Thorpe, C. (2008). Oxidative protein folding in vitro: a study of the cooperation between quiescin-sulfhydryl oxidase and protein disulfide isomerase. Biochemistry *47*, 12047–12056.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663–669.

Reckenfelderbäumer, N., Lüdemann, H., Schmidt, H., Steverding, D., and Krauth-Siegel, R.L. (2000). Identification and functional characterization of thioredoxin from *Trypanosoma brucei brucei*. J. Biol. Chem *275*, 7547–7552.

Robinson, N.P. (2002). Inactivation of Mre11 Does Not Affect VSG Gene Duplication Mediated by Homologous Recombination in *Trypanosoma brucei*. Journal of Biological Chemistry *277*, 26185–26193.

Roditi, I., and Pearson, T.W. (1990). The procyclin coat of African trypanosomes (or the not-so-naked trypanosome). Parasitology Today *6*, 79–82.

Roditi, I., Schwarz, H., Pearson, T.W., Beecroft, R.P., Liu, M.K., Richardson, J.P., Bühring, H.J., Pleiss, J., Bülow, R., and Williams, R.O. (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. J. Cell Biol *108*, 737–746.

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. Nat Rev Mol Cell Biol *8*, 519–529.

Roper, J.R., Güther, M.L.S., Milne, K.G., and Ferguson, M.A.J. (2002). Galactose metabolism is essential for the African sleeping sickness parasite *Trypanosoma brucei*. PNAS *99*, 5884–5889.

Rubotham, J., Woods, K., Garcia-Salcedo, J.A., Pays, E., and Nolan, D.P. (2005). Characterization of Two Protein Disulfide Isomerases from the Endocytic Pathway of Bloodstream Forms of *Trypanosoma brucei*. Journal of Biological Chemistry *280*, 10410–10418.

Sahrawy, M., Hecht, V., Lopez-Jaramillo, J., Chueca, A., Chartier, Y., and Meyer, Y. (1996). Intron position as an evolutionary marker of thioredoxins and thioredoxin domains. J. Mol. Evol *42*, 422–431.

Schwaller, M., Wilkinson, B., and Gilbert, H.F. (2003). Reduction-reoxidation cycles contribute to catalysis of disulfide isomerization by protein-disulfide isomerase. J. Biol. Chem *278*, 7154–7159.

Sciandra, F., Angelucci, E., Altieri, F., Ricci, D., Hübner, W., Petrucci, T.C., Giardina, B., Brancaccio, A., and Bozzi, M. (2012). Dystroglycan is associated to the disulfide isomerase ERp57. Experimental Cell Research *318*, 2460–2469.

Sevier, C.S., Qu, H., Heldman, N., Gross, E., Fass, D., and Kaiser, C.A. (2007). Modulation of Cellular Disulfide-Bond Formation and the ER Redox Environment by Feedback Regulation of Ero1. Cell *129*, 333–344.

Shan, S., and Walter, P. (2005). Co-translational protein targeting by the signal recognition particle. FEBS Letters *579*, 921–926.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, K., and Rudenko, G. (2005). Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. Proc Natl Acad Sci U S A *102*, 8716–8721.

Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. Protein Expr. Purif. *15*, 34–39.

Shi, H., Djikeng, A., Tschudi, C., and Ullu, E. (2004). Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. Mol. Cell. Biol *24*, 420–427.

Shi, H., Tschudi, C., and Ullu, E. (2006). An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. RNA *12*, 2063–2072.

Shi, H., Tschudi, C., and Ullu, E. (2007). Depletion of newly synthesized Argonaute1 impairs the RNAi response in *Trypanosoma brucei*. RNA *13*, 1132–1139.

Siegel, T.N., Hekstra, D.R., Wang, X., Dewell, S., and Cross, G.A.M. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucl. Acids Res. gkq237.

Siomi, H., and Siomi, M.C. (2009). On the road to reading the RNA-interference code. Nature *457*, 396–404.

Smith, T.K., Vasileva, N., Gluenz, E., Terry, S., Portman, N., Kramer, S., Carrington, M., Michaeli, S., Gull, K., and Rudenko, G. (2009). Blocking Variant Surface Glycoprotein Synthesis in *Trypanosoma brucei* Triggers a General Arrest in Translation Initiation. PLoS One *4*.

Soldà, T., Garbi, N., Hämmerling, G.J., and Molinari, M. (2006). Consequences of ERp57 deletion on oxidative folding of obligate and facultative clients of the calnexin cycle. J. Biol. Chem *281*, 6219–6226.

Spitznagel, D., Ebikeme, C., Biran, M., Nic a' Bháird, N., Bringaud, F., Henehan, G.T.M., and Nolan, D.P. (2009). Alanine aminotransferase of *Trypanosoma brucei*--a key role in proline metabolism in procyclic life forms. FEBS J *276*, 7187–7199.

Stockdale, C., Swiderski, M.R., Barry, J.D., and McCulloch, R. (2008). Antigenic Variation in *Trypanosoma brucei*: Joining the DOTs. PLoS Biol *6*.

Stockton, J.D., Merkert, M.C., and Kellaris, K.V. (2003). A complex of chaperones and disulfide isomerases occludes the cytosolic face of the translocation protein Sec61p and affects translocation of the prion protein. Biochemistry *42*, 12821–12834.

Stokes, M.J., Güther, M.L.S., Turnock, D.C., Prescott, A.R., Martin, K.L., Alphey, M.S., and Ferguson, M.A.J. (2008). The Synthesis of UDP-N-acetylglucosamine Is Essential for Bloodstream Form *Trypanosoma brucei* in Vitro and in Vivo and UDP-N-acetylglucosamine Starvation Reveals a Hierarchy in Parasite Protein Glycosylation. J. Biol. Chem. *283*, 16147–16161.

Stuart, K. (1983). Kinetoplast DNA, mitochondria DNA with a difference. Molecular and Biochemical Parasitology *9*, 93–104.

Subramanya, S., and Mensa-Wilmot, K. (2006). Regulated cleavage of intracellular glycosylphosphatidylinositol in a trypanosome. Peroxisome-to-endoplasmic reticulum translocation of a phospholipase C. FEBS J *273*, 2110–2126.

Subramanya, S., Hardin, C.F., Steverding, D., and Mensa-Wilmot, K. (2009). Glycosylphosphatidylinositol-specific phospholipase C regulates transferrin endocytosis in the African trypanosome. Biochem. J. *417*, 685.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123–132.

Tachibana, C., and Stevens, T.H. (1992). The yeast EUG1 gene encodes an endoplasmic reticulum protein that is functionally related to protein disulfide isomerase. Mol. Cell. Biol *12*, 4601–4611.

Tachikawa, H., Miura, T., Katakura, Y., and Mizunaga, T. (1991). Molecular structure of a yeast gene, PDI1, encoding protein disulfide isomerase that is essential for cell growth. J. Biochem. *110*, 306–313.

Tachikawa, H., Takeuchi, Y., Funahashi, W., Miura, T., Gao, X.D., Fujimoto, D., Mizunaga, T., and Onodera, K. (1995). Isolation and characterization of a yeast gene, MPD1, the overexpression of which suppresses inviability caused by protein disulfide isomerase depletion. FEBS Lett *369*, 212–216.

Tachikawa, H., Funahashi, W., Takeuchi, Y., Nakanishi, H., Nishihara, R., Katoh, S., Gao, X.D., Mizunaga, T., and Fujimoto, D. (1997). Overproduction of Mpd2p suppresses the lethality of protein disulfide isomerase depletion in a CXXC sequence dependent manner. Biochem. Biophys. Res. Commun *239*, 710–714.

Tian, G., Xiang, S., Noiva, R., Lennarz, W.J., and Schindelin, H. (2006). The Crystal Structure of Yeast Protein Disulfide Isomerase Suggests Cooperativity between Its Active Sites. Cell *124*, 1085–1088.

Tien, A.-C., Rajan, A., Schulze, K.L., Ryoo, H.D., Acar, M., Steller, H., and Bellen, H.J. (2008). Ero1L, a thiol oxidase, is required for Notch signaling through cysteine bridge formation of the Lin12-Notch repeats in *Drosophila melanogaster*. J. Cell Biol. *182*, 1113–1125.

Tokuhiro, K., Ikawa, M., Benham, A.M., and Okabe, M. (2012). Protein disulfide isomerase homolog PDILT is required for quality control of sperm membrane protein ADAM3 and male fertility [corrected]. Proc. Natl. Acad. Sci. U.S.A. *109*, 3850–3855.

Tu, B.P., and Weissman, J.S. (2002). The FAD- and O2-Dependent Reaction Cycle of Ero1-Mediated Oxidative Protein Folding in the Endoplasmic Reticulum. Molecular Cell *10*, 983–994.

Tu, B.P., Ho-Schleyer, S.C., Travers, K.J., and Weissman, J.S. (2000). Biochemical Basis of Oxidative Protein Folding in the Endoplasmic Reticulum. Science *290*, 1571–1574.

Turano, C., Coppari, S., Altieri, F., and Ferraro, A. (2002). Proteins of the PDI family: unpredicted non-ER locations and functions. J. Cell. Physiol *193*, 154–163.

Turnock, D.C., Izquierdo, L., and Ferguson, M.A.J. (2007). The *de Novo* Synthesis of GDP-fucose Is Essential for Flagellar Adhesion and Cell Growth in *Trypanosoma brucei*. J. Biol. Chem. *282*, 28853–28863.

Ullu, E., Tschudi, C., and Chakraborty, T. (2004). RNA interference in protozoan parasites. Cell. Microbiol *6*, 509–519.

VanGuilder, H.D., Vrana, K.E., and Freeman, W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques *44*, 619–626.

Vanhamme, L., Poelvoorde, P., Pays, A., Tebabi, P., Van Xong, H., and Pays, E. (2000). Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. Mol. Microbiol *36*, 328–340.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Abbeele, J.V.D., Pays, A., Tebabi, P., Xong, H.V., Jacquet, A., et al. (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature *422*, 83–87.

Vassella, E., Reuner, B., Yutzy, B., and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J Cell Sci *110*, 2661–2671.

Venetianer, P., and Straub, F.B. (1963). The enzymic reactivation of reduced ribonuclease. Biochim. Biophys. Acta *67*, 166–168.

Vermeulen, A. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. RNA *11*, 674–682.

Vickerman, K. (1962). The mechanism of cyclical development in trypanosomes of the *Trypanosoma brucei* sub-group: An hypothesis based on ultrastructural observations. Transactions of the Royal Society of Tropical Medicine and Hygiene *56*, 487–495.

Vickerman, K. (1970). Functional aspects of the cytology of trypanosomes. Transactions of the Royal Society of Tropical Medicine and Hygiene *64*, 180–181.

Vickerman, K. (1973). The mode of attachment of *Trypanosoma vivax* in the proboscis of the tsetse fly *Glossina fuscipes*: an ultrastructural study of the epimastigote stage of the trypanosome. J. Protozool *20*, 394–404.

Vickerman, K. (1978). Antigenic variation in trypanosomes. Nature *273*, 613–617.

Vickerman, K. (1985). Developemental Cycles and Biology of Pathogenic Trypanosomes. Br Med Bull *41*, 105–114.

Vickerman, K., and Luckins, A.G. (1968). Cyclical transformation in trypanosomes. J. Gen. Microbiol *50*, Suppl:12–13.

Walker, K.W., and Gilbert, H.F. (1997). Scanning and escape during protein-disulfide isomerase-assisted protein folding. J. Biol. Chem *272*, 8845–8848.

Walker, A.K., Soo, K.Y., Levina, V., Talbo, G.H., and Atkin, J.D. (2012). *N*-linked Glycosylation Modulates Dimerization of Protein Disulfide Isomerase Family A Member 2 (PDIA2). FEBS Journal n/a–n/a.

Walker, K.W., Lyles, M.M., and Gilbert, H.F. (1996). Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. Biochemistry *35*, 1972–1980.

Wang, Z.V., Schraw, T.D., Kim, J.-Y., Khan, T., Rajala, M.W., Follenzi, A., and Scherer, P.E. (2007). Secretion of the adipocyte-specific secretory protein adiponectin critically depends on thiol-mediated protein retention. Mol. Cell. Biol. *27*, 3716–3731.

Webb, H., Carnall, N., and Carrington, M. (1994). The role of GPI-PLC in *Trypanosoma brucei*. Braz. J. Med. Biol. Res *27*, 349–356.

Webb, H., Carnall, N., Vanhamme, L., Rolin, S., Abbeele, J.V.D., Welburn, S., Pays, E., and Carrington, M. (1997). The GPI-Phospholipase C of *Trypanosoma brucei* Is Nonessential But Influences Parasitemia in Mice. J Cell Biol *139*, 103–114.

Weiden, M., Osheim, Y.N., Beyer, A.L., and Van der Ploeg, L.H. (1991). Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. Mol. Cell. Biol. *11*, 3823–3834.

Wickstead, B., Ersfeld, K., and Gull, K. (2002). Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. Mol. Biochem. Parasitol *125*, 211–216.

Wirtz, E., and Clayton, C. (1995). Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. Science *268*, 1179–1183.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Mol. Biochem. Parasitol *99*, 89–101.

Woods, K., Nic a'Bhaird, N., Dooley, C., Perez-Morga, D., and Nolan, D.P. (2013). Identification and Characterization of a Stage Specific Membrane Protein Involved in Flagellar Attachment in Trypanosoma brucei. PLoS ONE *8*, e52846.

Wurst, M., Seliger, B., Jha, B.A., Klein, C., Queiroz, R., and Clayton, C. (2012). Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. Mol. Microbiol. *83*, 1048–1063.

Zamore, P., Tuschl, T., Sharp, P., and Bartel, D. (2000). RNAiDouble-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell *101*, 25–33.

Zhang, K., and Kaufman, R.J. (2006). The unfolded protein response: A stress signaling pathway critical for health and disease. Neurology *66*, S102–109.

Zito, E., Chin, K.-T., Blais, J., Harding, H.P., and Ron, D. (2010). ERO1-beta, a pancreas-specific disulfide oxidase, promotes insulin biogenesis and glucose homeostasis. J. Cell Biol *188*, 821–832.